Discourse Structure:

Theory and Practice

Bonnie Webber[1] Markus Egg[2] Valia Kordoni[3,4]

[1] School of Informatics, University of Edinburgh (UK)

[2] Dept. of English and American Studies, Humboldt University Berlin (Germany)

[3] German Research Centre for Artificial Intelligence (DFKI GmbH, Germany)

[4] Dept. of Computational Linguistics, Saarland University (Germany)

July 4, 2011

Outline of Tutorial

1. Introduction
2. Computational approaches to discourse structure
3. Applications
4. Speculating about the future

Outline of Part 1 (Introduction)

- What is discourse?
- What are discourse structures?
- What are the elements of discourse structures?
- What properties of these structures are relevant to LT?

Discourse and sentence sequences

Discourse usually involves a sequence of sentences.

But a discourse can be found even in a single sentence:

(1) If they're drunk
and they're meant to be on parade
and you go to their room
and they're lying in a pool of piss,
then you lock them up for a day.
[The Independent, 17 June 1997]

$$\text{STATE} + \text{STATE} + \text{EVENT} + \text{STATE} \Rightarrow \text{EVENT}$$

Discourse and sentence sequences

The patterns formed by the sentences in a discourse convey more than each does individually.

(2) This parrot is no more.
It has ceased to be.
It's expired and gone to meet its maker.
This is a late parrot.
It's a stiff.
Bereft of life, it rests in peace.
If you hadn't nailed it to the perch, it would be pushing up the daisies.
It's rung down the curtain and joined the choir invisible.
This is an ex-parrot. [Monty Python]

Discourse and language features

Discourse exploits language features that reveal that the speaker is

- still talking about the same thing(s).

Anaphoric expressions

(3) The police are not here to create disorder. **They** are here to preserve **it**. [Attributed to Yogi Berra]

Ellipsis

(4) Pope John XXIII was asked "How many people work in the Vatican?". He is said to have replied, "About **half**".

Discourse and language features

Discourse exploits language features that reveal that the speaker is

- indicating relations that hold between states, events, beliefs, etc.

(5) Men have a tragic genetic flaw. **As a result**, they cannot see dirt **until** there is enough of it to support agriculture.

[Paraphrasing Dave Barry, The Miami Herald - Nov. 23, 2003]

Discourse and language features

Discourse exploits language features that reveal that the speaker is

- changing the topic or resuming an earlier topic.

(6) Man is now able to fly throught the air like a bird
He's able to swim beneath the sea like a fish
He's able to burrow beneath the ground like a mole
Now if only he could walk the Earth like a man,
This would be paradise.

[Lyrics to *This would be Paradise*, Auf der Maur]

These are often called **cue phrases** or **boundary features**.

Discourse and language features

When these same features of language appear in a single sentence,

(7) When Pope John XXIII was asked "How many people work in the Vatican?", **he** is said to have replied, "About **half**".

(8) Everyone who assaults **others**, **does so** for **their** own reasons.

they play the same role as they do across multiple clauses.

So what can we say about discourse?

So it's reasonable to associate discourse with

- a sequence of sentences
- that convey more than its individual sentences through their relationships to one another, and
- that exploit special features of language that enable discourse to be more easily understood.

Discourse structure focuses on the second property.

What are discourse structures?

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

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

Researchers in Language Technology (LT) are beginning to be able to recognize and exploit these patterns for useful ends.

What are the elements of discourse structures?

At a high level,

- topics and their relationships compose structures of **expository** text;
- sentences serving particular roles compose **functional** structures;
- claims and evidence and their relationships compose **argumentation** structures;
- events and circumstances and their relationships compose **narrative** structures;

Feeding into these are low-level structures, variously called **coherence relations**, **discourse relations**, or **rhetorical relations**.

Properties of discourse structure relevant to LT

Properties ascribed to discourse structures have computational implications for LT:

- Are these structures hierarchical or flat?
- Are these structures trees or more complex graphs?
- Are these structures full or partial covers?
- Are the links of these structures symmetric or asymmetric?

Are DStructs hierarchical or flat?

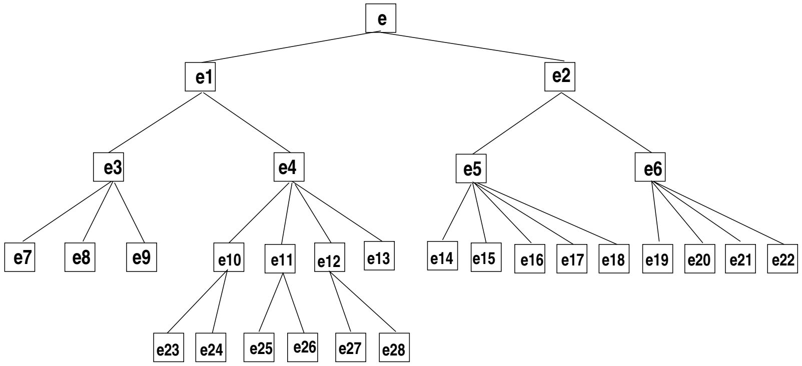
Some discourse structures appear to have a hierarchical structure, like a sentence parse tree [Dal92, GS86, GS90, MT88, WSMP07]

For example, [Dal92] gives this recipe for *Butter bean soup*:

(9) Soak, drain and rinse the butter beans. Peel and chop the onion. Peel and chop the potato. Scrape and chop the carrots. Slice the celery. Melt the butter. Add the vegetables. Saute them. Add the butter beans, the stock and the milk. Simmer. Liquidise the soup. Stir in the cream. Add the seasonings. Reheat.

For [Dal92], this has the structure on the next slide:

[Dale, 1992]



Are DStructs hierarchical or flat?

Other discourse structures appear to have a simpler *linear structure* [BL04, Hea97, MB06, Sib92] – e.g. Wikipedia articles:

	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	...	...	...

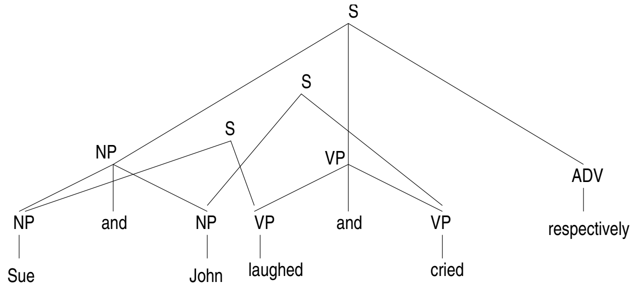
Linear segmentation is a less complex task than recovering an underlying hierarchical structure.

Are DStructs trees or more complex graphs?

The syntactic structure of a sentence is kept a *tree*, in part by understanding more complex dependencies as semantic:

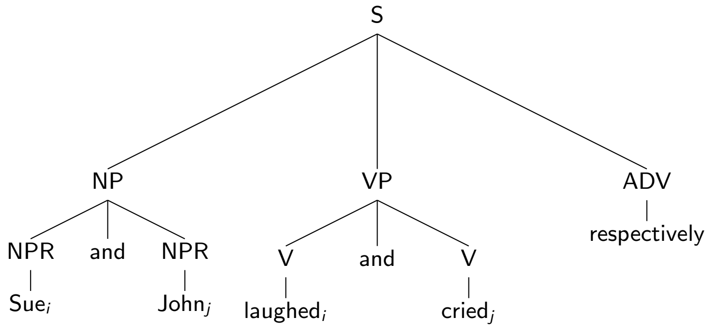
(10) Sue and John laughed and cried, respectively.

Viewed as syntactic structure:



Are DStructs trees or more complex graphs?

Viewed as semantic dependencies:



Are DStructs trees or more complex graphs?

If all discourse dependencies are taken to be syntactic rather than semantic, then discourse will indeed have a complex graph structure.

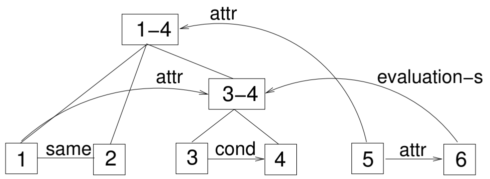
What kind of dependencies lead to a complex graph structure, with crossing edges and multiple directed edges into a node?

They include:

- attribution relations
- coreference relations to entities and events

Discourse Graph Bank [Wolf & Gibson, 2005]

(11) 1. "The administration should now state
2. that
3. if the February election is voided by the Sandinistas
4. they should call for military aid,"
5. said former Assistant Secretary of State Elliot Abrams.
6. "In these circumstances, I think they'd win." [wsj_0655]



A complex graph structure makes the task of discourse parsing even more complex.

Are DStructs full or partial covers?

The parse of a sentence **fully covers** its elements (words).

In contrast, the **coherence relations** in a discourse (its low-level structure) need not do so.

(12) "**I'm sympathetic with workers who feel under the gun**," says Richard Barton of the Direct Marketing Association of America, which is lobbying strenuously against the Edwards beeper bill. "**But the only way you can find out how your people are doing is by listening**." [wsj_1058]

The coherence relation (CONCESSION) holds only between the highlighted elements.

So recovering coherence relations (*discourse chunking*, Part 2.3) may be a less complex task than discourse parsing.

Are the links in DStructs symmetric or asymmetric?

Rhetorical Structure Theory [MT88] takes certain discourse relations to be **asymmetric**, with one argument (the *nucleus*) more essential to the purpose of the communication than the other (the *satellite*).

So, Part 3.1 shows how RST-based *extractive summarization* takes satellites to be removable without harm [DM02].

Stede [Ste08b] argues that asymmetry in discourse has several sources that should not be conflated: Other discourse relations are simply **symmetric**.

This may be the source of problems noticed with RST-based extractive summarization [Mar98].

Outline of Part 2

- Recovering topic structure
- Recovering functional structure
- Recovering relational structure (Discourse Chunking)
- Recovering hierarchical structure (Discourse Parsing)
- Classifying unmarked relations
- Identifying entity structure
- Useful resources

Topic Structure

Expository text can be viewed as a sequence of **topically coherent** segments, whose order may become conventionalized over time:

	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	Education	Sports	Law and government
9	Culture	Culture	Public Health
10	...	...	...

Wikipedia articles about US states

Topic Structure

Being able to recognize topic structure was originally seen as benefitting **information retrieval** [Hea97]

Recent interest comes from the potential use of topic structure in **segmenting lectures, meetings or other speech events**, making them more amenable to search [GMFLJ03, MB06].

Topic Structure

Computational work on topic structure and segmentation takes:

- the topic of each segment to relate only to the topic of the discourse as a whole (eg, History of Vermont → Vermont).
- sequence to be the only relation holding between sister segments, although certain sequences may be more common than others (cf. Wikipedia articles).
- the topic of each segment to differ from those of its adjacent sisters. (Adjacent spans that share a topic are taken to belong to the same segment.)
- topic to predict lexical choice, either of all words of a segment or just its content words (ie, excluding "stop-words").

Topic Structure

Making topic structure explicit (ie, topic segmentation) uses either

- **semantic-relatedness**, where words within a segment are taken to relate to each other more than to words outside the segment [Hea97, CWHM01, Bes06, GMFLJ03, MB06]
- **topic models**, where each segment is taken to be produced by a distinct, compact lexical distribution [CBBK09, EB08a, PGKT06]

See [Pur11] for an excellent overview and survey of this work.

Topic Segmentation through Semantic-relatedness

All computational models that use semantic-relatedness for topic segmentation have:

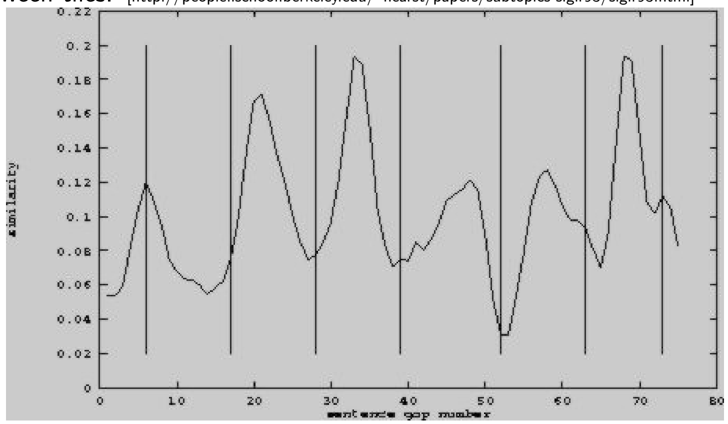
1. a **metric** for assessing the semantic relatedness of terms within a proposed segment;
2. a **locality** that specifies what units within the text are assessed for semantic relatedness;
3. a **threshold** for deciding how low relatedness can drop before it signals a shift to another topic.

TextTiling [Hearst 1997]

1. **Metric**: *Cosine similarity*, using a vector representation of fixed-length spans (pseudo-sentences) in terms of frequency of word stems (ie, words from which inflection has been removed)
2. **Locality**: Cosine similarity is computed between adjacent spans (and only adjacent spans)
3. **Threshold**: Empirically-determined in order to select where to place segment boundaries.

TextTiling – Computed similarity of adjacent blocks

TextTiling a popular science article. Vertical lines show manually-assigned topic boundaries. Peaks indicate coherency, and valleys, potential breaks between tiles. [http://people.ischool.berkeley.edu/˜hearst/papers/subtopics-sigir93/sigir93.html]



Topic Segmentation through Topic Models

Topic segmentation using topic models can take advantage of both

- Features internal to a segment (*Segmental Features*), including words (all words or just content words) and syntax
- Features occurring at segmental boundaries (*Boundary Features* — cf. Part 1.1), including **cue words** (eg, "now", "so", "anyway"), syntax and (in speech) pauses and intonation.

N.B. These cue words don't indicate anything about the content of the current topic or the next one — only a particular kind of change from one to the other.

Outline of Part 2

- Recovering topic structure
- Recovering functional structure
- Recovering relational structure (Discourse Chunking)
- Recovering hierarchical structure (Discourse Parsing)
- Classifying unmarked relations
- Identifying the entity structure
- Useful resources

Functional discourse structure

Texts within a given genre – eg,

- news reports
- errata
- scientific papers
- letters to the editor of the *New York Times*
- . . .

generally share a similar structure, that is independent of topic (eg, sports, politics, disasters; or molecular genetics, radio astronomy, SMT), instead reflecting the **function** played by their parts.

Example: News Reports

Best known is the **inverted pyramid** structure of news reports:

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

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

Example: Errata

Also recognizable are **errata** – declarations of errors made in previous issue of a periodical and correct versions:

- Correct statement
- Description of error

(13) EMPIRE PENCIL, later called Empire-Berol, developed the plastic pencil in 1973. Yesterday's Centennial Journal misstated the company's name. [wsj_1751]

(14) PRINCE HENRI is the crown prince and hereditary grand duke of Luxembourg. An article in the World Business Report of Sept. 22 editions incorrectly referred to his father, Grand Duke Jean, as the crown prince. [wsj_1871]

Example: Scientific articles/abstracts

Well-known in academia is the multi-part structure of scientific papers (and, more recently, their abstracts):

- **Objective** (aka *Introduction*, *Background*, *Aim*, *Hypothesis*)
- **Methods** (aka *Method*, *Study Design*, *Methodology*, etc.)
- **Results** or *Outcomes*
- **Discussion**
- Optionally, **Conclusions**

N.B. Not every sentence within a section need realise the same function: Fine-grained functional characterizations of scientific papers show them serving a range of functions [LTSB10].

Functional Structure

Automatic annotation of functional structure is seen as benefitting:

- **Information extraction**: Certain types of information are likely to be found in certain sections [MUD99, MUD00, Moe01]
- **Extractive summarization**: More “important” sentences are more likely to be found in certain sections.
- **Sentiment analysis**: Words that have an objective sense in one section may have a subjective sense in another [TBS09]
- **Citation analysis**: A citation may serve different functions in different sections [Teu10]

Functional structure

Computational work on functional structure and segmentation assumes that:

- The function of a segment relates to the discourse as a whole.
- While relations may hold between sisters (eg, *Methods* constrain *Results*), only sequence has been used in modelling.
- Function predicts more than lexical choice:
  - indicative phrases such as “results show” (→ *Results*)
  - indicative stop-words such as “then” (→ *Methods*).
- Functional segments usually appear in a specific order, so either sentence position is a feature in the models or sequential models are used.

Functional structure

Much computational work has ignored the internal structure of segments [Chu09, MS03, LKDFK06, RBC$^+$07].

However, Hirohata et al [HOAI08] found that, within a segment:

- Properties of the first sentence differ from those of the other sentences (as in ’BIO’ approaches to Named Entity Recognition).
- Modelling this leads to improved performance in high-level functional segmentation (ie, 94.3% per sentence accuracy vs. 93.3%).

This accords with work in low-level (fine-grained) modelling of functional structure [LTSB10].

Labelled biomedical abstracts

Functional segmentation takes biomedical abstracts with labelled sections as **training data** for segmenting unlabelled abstracts [Chu09, GKL$^+$10, HOAI08, LTSB10, LKDFK06, MS03, RBC$^+$07],

(15) **BACKGROUND**: Mutation impact extraction is a hitherto unaccomplished task in state of the art mutation extraction systems. . . . **RESULTS**: We present the first rule-based approach for the extraction of mutation impacts on protein properties, categorizing their directionality as positive, negative or neutral. . . . **CONCLUSION**: . . . Our approaches show state of the art levels of precision and recall for Mutation Grounding and respectable level of precision but lower recall for the task of Mutant-Impact relation extraction. . . . [**PMID 21143808**]

Part 3.1 discusses segmentation of legal texts [MUD99, MUD00, Moe01] and student essays [BMAC01, BMK03] for Info Extraction.

Outline of Part 2

- Recovering topic structure
- Recovering genre-specific structure
- Recovering relational structure (Discourse Chunking)
- Recovering hierarchical structure (Discourse Parsing)
- Classifying unmarked relations
- Identifying the entity structure
- Useful resources

Elements of Discourse Chunking

Like “NP chunking”, **discourse chunking** is a lightweight approximation to full discourse parsing.

It produces a flat structure of (possibly overlapping) coherence relations relations”) by identifying

1. what is relating elements in a discourse;
2. what elements are being related;
3. what type(s) of relation hold(s) between them.

From discourse connectives to their arguments

The general problems are:

1. Given a language, what affixes, words, terms and/or constructions can serve to relate elements in a discourse (ie, as its discourse connectives)?
2. Given a particular token in context, does it serve to relate discourse elements?
3. Given such a token, what elements does it relate (ie, its arguments)?
4. Given such a token and its arguments, what sense relation(s) hold between the arguments?

Examples are from the *Penn Discourse TreeBank* (PDTB) [PDL$^+$08]. Part 2.7 will highlight other corpora.

From DConns to their Args: Problem 1

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

- Coordinating conjunctions (on clauses or sentences)

(16) Finches eat seeds, and/but/or robins eat worms.

(17) Finches eat seeds. But today, I saw them eating grapes.

- Subordinating conjunctions

(18) While finches eat seeds, robins eat worms.

(19) Robins eat worms, just as finches eat seeds.

From DConns to their Args: Problem 1

In other cases, only a subset of a given part-of-speech (PoS) indicates a relation between discourse elements

- eg, not all adverbials, just *discourse adverbials*

(20) Robins eat both worms and seeds. Consequently they are omnivores. *(discourse adverbial)*

(21) Robins eat both worms and seeds. Fortunately they prefer worms. *(sentential adverbial)*

From DConns to their Args: Problem 1

Constructions other than particular parts-of-speech regularly serve to indicate a discourse relation [PDL$^+$08, PJW10b]:

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

Both **bootstrapping** and **back translation** have been used to discover new instances of named entities and paraphrases.

Can these same techniques be used to identify other constructions that can indicate discourse relations?

From DConns to their Args: Problem 1

Syntactically constrained **back translation** [CB08] on EuroParl translation pairs yields many phrases that were either not annotated as discourse connectives in the PDTB or don't appear there – eg,

Not annotated	Doesn't appear
above all	*as a consequence*
after all	*as an example*
despite that	*by the same token*
...	...

But this doesn't reveal instances with different syntactic structure than their source phrase, without introducing extensive noise.

From DConns to their Args: Problem 2

It is hard to decide whether an individual token signals a coherence relation because such tokens are often *syntactically ambiguous* [PN09]:

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

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

PoS-tagging can often distinguish discourse from non-discourse use.

Even without PoS-tagging, surface cues allow discourse and non-discourse use to be distinguished with at least 94% accuracy [PN09].

From DConns to their Args: Problem 3

Discourse relations in all languages analyzed so far relate two arguments:

- **Arg2** – argument syntactially bound to the relation
- **Arg1** – the other argument

With **Arg2**, the challenge is whether *attribution* is included.

(24) **We pretty much have a policy of not commenting on rumors**, and I think **that falls in that category**. [wsj_2314]

(25) **Advocates said the 90-cent-an-hour rise, to $4.25 an hour by April 1991, is too small for the working poor**, while **opponents argued that the increase will still hurt small business and cost many thousands of jobs**. [wsj_0098]

From DConns to their Args: Problem 3

Identifying **Arg1** is harder because it need not be adjacent to **Arg2**:

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

(26) On a level site you can provide a cross pitch to the entire slab by **raising one side of the form** (step 5, p. 153), but for a 20-foot-wide drive this results in an awkward 5-inch (20 x 1/4 inch) slant across the drive's width. Instead, **make the drive higher at the center**.

From DConns to their Args: Problem 3

2. All parts of a text are not equally essential to an argument (cf. Part 1.4):

(27) **Big buyers like Procter & Gamble say there are other spots on the globe and in India, where the seed could be grown**. "It's not a crop that can't be doubled or tripled," says Mr. Krishnamurthy. But **no one has made a serious effort to transplant the crop**. [wsj_0515]

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

From Discourse Connectives to their Arguments

There is a growing amount of work on automatically identifying discourse connectives and locating their arguments:

- Initial results [WP07] suggest that **connective specific** models might perform better than models that just consider the type of connective (coordinating, subordination, adverbial).
- [EB08b] show that connective specific models significantly improve results for discourse adverbials: (67.5% vs. 49.0%).
- [PJW10a] show that for inter-sentential 'And', 'But' and discourse adverbials, performance is significantly higher for **within-paragraph** tokens, since 4301/4373 = 98% of the time, Arg1 is in same paragraph, simplifying the problem.

From Discourse Connectives to their Arguments

[LNK10] demonstrate the first end-to-end processor for discourse chunking, identifying

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

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

- Similar to previous results for each type of explicit connective;
- 40% for implicit connects, dropping to around 25-26% with error propagation.

Outline of Part 2

- Recovering topic structure
- Recovering genre-specific structure
- Recovering relational structure (Discourse Chunking)
- **Recovering hierarchical structure** (Discourse Parsing)
- Classifying unmarked relations
- Identifying entity structure
- Useful resources

Discouse Parsing

Discourse parsing automatically constructs a discourse structure (usually, but not always, a tree) covering the entire input text.

Discourse Parsing is useful for:

- QA
- IE
- Text-to-Text Applications (Summarisation, Paraphrasing)
- Recognising Textual Entailment
- Modeling and Evaluating Text Coherence, etc.

History of discourse parsing:

- rule-based systems – either knowledge-rich or knowledge-poor;
- more recently, systems based on ML from corpora

Discourse Parsing: Rule-based Approaches

Knowledge-rich models [HSAM93, KR93, AL03]

- logic-based
- explicit representation of world knowledge in knowledge base
- discourse meaning as an extension of sentence meaning (i.e., the aim is to find the best logical form)

Discourse Parsing: Rule-based Approaches

Knowledge-poor models [Mar97, COCO98]

- input: syntactically analysed texts
- heuristics to compute discourse structure
- no extensive semantic knowledge (no knowledge base)
- surface form (syntactic structure, deixis, anaphora, cue words, etc.) provides cues for discourse structure

Discourse Parsing: Corpus-based Approaches

Corpus-based approaches [Mar99, BL05]

- supervised machine learning
- training data: e.g., RST Discourse Treebank
- discourse parsing analogous to syntactic parsing

Discourse Parsing: Marcu (1999)

Marcu automatically derives the discourse structure of texts. Two subtasks:

1. Find boundaries between elementary discourse units (EDUs);
2. Find rhetorical relations that connect EDUs, building discourse trees.

Marcu's approach:

- relies on manual annotation;
- based on Rhetorical Structure Theory (RST);
- uses decision-tree learning.

Discourse Parsing: Marcu (1999) – Annotation

The data is extracted from the following corpora:

- MUC7 corpus (30 stories);
- Brown corpus (30 scientific texts);
- Wall Street (30 editorials).

The corpora are marked up:

- *elementary discourse units* (EDUs);
- discourse trees in the style of RST.

Discourse Parsing: Marcu (1999)

Task: process each lexeme (word or punctuation mark) and recognize sentence and *EDU* boundaries and parenthetical units.
To generate the learning cases:

- use the leaves of the manually built discourse trees;
- associate each lexeme in the text with one learning case;
- associate with each lexeme one of the following classes to be learnt: *sentence-break*, *EDU-break*, *start-paren*, *end-paren*, *none*.

Approach: determine a set of features that will predict these classes, then:

- extract features from annotated text;
- use decision-tree learning to combine features and perform segmentation.

Discourse Parsing: Marcu (1999) – Features

Local context features:

- POS tags preceding and following the lexeme (2 before, 2 after);
- discourse connectives (*because, and*);
- abbreviations.

Global context features, pertaining to the boundary identification process:

- discourse connectives that introduce expectations (e.g., *on the one hand, although* [CW97];
- commas or dashes before the estimated end of the sentence;
- verbs in unit of consideration.

Discourse Parsing: Marcu (1999) – Results

Corpus	B1 (%)	B2 (%)	DT (%)
MUC	91.28	93.1	96.24
WSJ	92.39	94.6	97.14
Brown	93.84	96.8	97.87

- B1: defaults to *none*;
- B2: defaults to *sentence-break* for every full-stop and *none* otherwise;
- DT: decision tree classifier.

Discourse Parsing: Marcu (1999) – Discourse Structure

Task: determine rhetorical relations and construct discourse trees as defined by RST.

Approach:

- exploit RST trees created by annotators;
- map tree structure onto SHIFT/REDUCE operations;
- extract features for these operations;
- distinguish nuclei and satellites (following RST).

Discourse Parsing: Marcu (1999) – Discourse Structure

Operations:

- 1 SHIFT operation;
- 3 REDUCE operations: RELATION-NS, RELATION-SN, RELATION-NN.

Rhetorical relations:

- taken from RST;
- 17 in total: CONTRAST, PURPOSE, EVIDENCE, EXAMPLE, ELABORATION, etc.

Discourse Parsing: Marcu (1999) – Discourse Structure Features

Features used for decision tree classifier for operations:

- *structural*: rhetorical relations that link the immediate children of the link nodes;
- *lexico-syntactic*: discourse markers and their position;
- *operational*: last five operations;
- *semantic*: similarity between trees (bags of words).

Discourse Parsing: Marcu (1999) – Discourse Structure Results

Corpus	B3 (%)	B4 (%)	DT (%)
MUC	50.75	26.9	61.12
WSJ	50.34	27.3	61.65
Brown	50.18	28.1	61.81

- B3: defaults to SHIFT;
- B4: chooses SHIFT and REDUCE operations randomly;
- DT: decision tree classifier.

Discourse Parsing: Marcu (1999) – Discourse Structure Results

Strengths:

- First fully automated system for parsing discourse structure.

Weaknesses:

- relies on manual annotation, which is time-consuming and difficult.

Getting around manual annotation requires ability to automatically annotate both marked and unmarked discourse relations.

Discourse Parsing: Summary

- knowledge-based systems:
  - few real implementations for small and well-defined domains
- heuristics-based systems:
  - relatively good for easy cases, with bad coverage, though, for unmarked relations
- corpus-trained systems:
  - few annotated data available
  - accuracy approx. 60%

There may be many applications for which full discourse parsing may not necessary.

Outline of Part 2

- Recovering topic structure
- Recovering genre-specific structure
- Recovering relational structure (Discourse Chunking)
- Recovering hierarchical structure (Discourse Parsing)
- **Classifying unmarked relations**
- Identifying the entity structure
- Useful resources

Classifying Unmarked Relations

In both discourse chunking and discourse parsing, there may be no specific cue (e.g., a discourse connective) of the relation that holds between elements.

For otherwise unmarked relations, evidence for the relation may be derivable from other features.

(28) [A car had broken down on an unmanned level crossing and was hit by a high speed train.]
[The train derailed.]
→ Result

(29) [The damage to the train was substantial,]
[fortunately nobody was injured]
→ Contrast

Classifying unmarked relations

To derive a classifier, considerable training data is needed (here, pairs of discourse elements annotated with the discourse relations holding between them).

- start with manually annotated texts;
- perform active learning [NM01];
- automatically label training data [ME02].

Automatic Labeling of Data: Marcu and Echihabi (2002)

Data:

- 4 relations from RST [MT87]: contrast, cause-explanation-evidence, condition, elaboration;
- 2 non-relations: no-relation-same-text, no-relation-different-text;
- 900,000 to 4 million automatically labelled examples per relation, derived from clauses connected by relatively unambiguous subordinating or coordinating conjunctions.

Model:

- Naive Bayes
- Word co-occurence features taken to predict the relation indicated by the explicit conjunction found between the clauses.

Automatic Labeling of Data: Marcu and Echihabi (2002)

Results:

- test on automatically labelled data: 49.7% accuracy for 6-way classifier
- test on manually labelled examples from RST-DT (marked and unmarked) without removing discourse connectives from training data and by using binary classifiers: 63% to 87% accuracy
- test on manually labelled, unmarked examples using binary classifiers (contrast vs. elaboration, and cause-explanation-evidence vs. elaboration): 69.5% recall for contrast, 44.7% recall for cause-explanation-evidence

Subsequent work has used manually-labelled unmarked relations from the PDTB and other corpora (cf. Part 2.7).

Outline of Part 2

- Recovering topic structure
- Recovering genre-specific structure
- Recovering relational structure (Discourse Chunking)
- Recovering hierarchical structure (Discourse Parsing)
- Classifying unmarked relations
- **Identifying entity structure**
- Useful resources

Coherence and entity structure

Coherence

- is a property of well-written texts;
- makes them easier to read and understand;
- ensures that sentences are meaningfully related.

The way entities are introduced and discussed influences coherence [GJW95].

By ignoring this, extractive summaries can appear incoherent.

Coherence and entity structure

Summary A

Britain said he did not have diplomatic immunity. The Spanish authorities contend that Pinochet may have committed crimes against Spanish citizens in Chile. Baltasar Garzon filed a request on Wednesday. Chile said, President Fidel Castro said Sunday he disagreed with the arrest in London.

Summary B

Former Chilean dictator Augusto Pinochet, was arrested in London on 14 October 1998. Pinochet, 82, was recovering from surgery. The arrest was in response to an extradition warrant served by a Spanish judge. Pinochet was charged with murdering thousands, including many Spaniards. Pinochet is awaiting a hearing, his fate in the balance. American scholars applauded the arrest.

Centering Theory

- Salience is associated (*inter alia*) with referring forms (headed NP or pronoun) and syntactic position (eg, subject, object, indirect object).
- *Entities* referred to in an *utterance* are ranked by salience.
- Each utterance has one *center* (topic or focus).
- Coherent discourse have utterances with common centers.
- Entity transitions capture degrees of coherence (e.g., in Centering theory CONTINUE > SHIFT.)

Entity-based Local Coherence

John went to his favourite music store to buy a piano.

He had frequented the store for many years.

He was excited that he could finally buy a piano.

He arrived just as the store was closing for the day.

John went to his favourite music store to buy a piano.

It was a store John had frequented for many years.

He was excited that he could finally buy a piano.

It was closing just as John arrived.

The Entity Grid

Can we compute entity structure automatically?

- Does it capture coherence characteristics?
- What linguistic information matters for coherence?
- Is it robust across domains and genres?

What is an appropriate coherence model?

- View coherence rating as a machine learning problem.
- Learn a ranking function without manual involvement.
- Apply to text-to-text generation tasks.

Inspired by Centering Theory, rather than a direct implementation.

The Entity Grid

1 Former Chilean dictator Augusto Pinochet, was arrested in London on 14 October 1998.
2 Pinochet, 82, was recovering from surgery.
3 The arrest was in response to an extradition warrant served by a Spanish judge.
4 Pinochet was charged with murdering thousands, including many Spaniards.
5 He is awaiting a hearing, his fate in the balance.
6 American scholars applauded the arrest.

The Entity Grid

1 Former Chilean dictator Augusto Pinochet$_S$, was arrested in London$_X$ on 14 October$_X$ 1998.
2 Pinochet$_S$, 82, was recovering from surgery$_X$.
3 The arrest$_S$ was in response$_X$ to an extradition warrant$_X$ served by a Spanish judge$_S$.
4 Pinochet$_S$ was charged with murdering thousands$_O$, including many Spaniards$_O$.
5 Pinochet$_S$ is awaiting a hearing$_O$, his fate$_X$ in the balance$_X$.
6 American scholars$_S$ applauded the arrest$_O$.

The Entity Grid

1 Pinochet$_S$ London$_X$ October$_X$
2 Pinochet$_S$ surgery$_X$
3 arrest$_S$ response $_X$ warrant$_X$ judge$_O$
4 Pinochet$_S$ thousands$_O$ Spaniards$_O$
5 Pinochet$_S$ hearing$_O$ Pinochet$_X$ fate$_X$ balance$_X$
6 scholars$_S$ arrest$_O$

The Entity Grid

	Pinochet	London	October	Surgery	Arrest	Warrant	Judge	Thousands	Spaniards	Hearing	Fate	Balance	Scholars
1													
2													
3													
4													
5													
6													

The Entity Grid

S	X	X	–	–	–	–	–	–	–	–	–	–	–
S	–	–	X	–	–	–	–	–	–	–	–	–	–
–	–	–	–	S	X	X	O	–	–	–	–	–	–
S	–	–	–	–	–	–	–	O	O	–	–	–	–
S	–	–	–	–	–	–	–	–	–	O	X	X	–
–	–	–	–	O	–	–	–	–	–	–	–	–	S

Entity Transitions

Definition A local entity transition is sequence $\{\mathbf{s}, \mathbf{o}, \mathbf{x}, -\}^n$ that represents entity occurrences and their syntactic roles in n adjacent sentences.

Feature Vector Notation Each grid x_{ij} for document d_i is represented by a feature vector:

$$\Phi(x_{ij}) = (p_1(x_{ij}), p_2(x_{ij}), \ldots, p_m(x_{ij}))$$

m is the number of predefined entity transitions
$p_t(x_{ij})$ the probability of transition t in grid x_{ij}

Entity Transitions

Example (transitions of length

		S S	S O	S X	S –	O S	O O	O X	O –	X S	X O	X X	X –	– S	– O	– X	– –
2)	d_1	0	0	0	.03	0	0	0	.02	.07	0	0	.12	.02	.02	.05	.25
	d_2	0	0	0	.02	0	.07	0	.02	0	0	.06	.04	0	0	0	.36
	d_3	.02	0	0	.03	0	0	0	.06	0	0	0	.05	.03	.07	.07	.29

Linguistic Dimensions

Salience: Are some entities more important than others?

- Discriminate between salient (frequent) entities and the rest.
- Collect statistics separately for each group.

Coreference: What is its contribution?

- Entities are coreferent if they have the same surface form.
- Coreference resolution tool [NC02].

Syntax: Does syntactic knowledge matter?

- Use four categories { **S**, **O**, **X**,– }.
- Reduce categories to { **X**,– }.

Learning a Ranking Function

Training Set:
Ordered pairs (x_{ij}, x_{ik}), where x_{ij} and x_{ik} represent the same document d_i, and x_{ij} is more coherent than x_{ik} (assume $j > k$).

Goal:
Find a parameter vector $\vec{w}$ such that:

$$\vec{w} \cdot (\Phi(x_{ij}) - \Phi(x_{ik})) > 0 \ \forall j, i, k \text{ such that } j > k$$

Support Vector Machines:
Constraint optimization problem can be solved using the search technique described in [Joa02]; see also [TMM04] for an application to parse selection.

Text Ordering

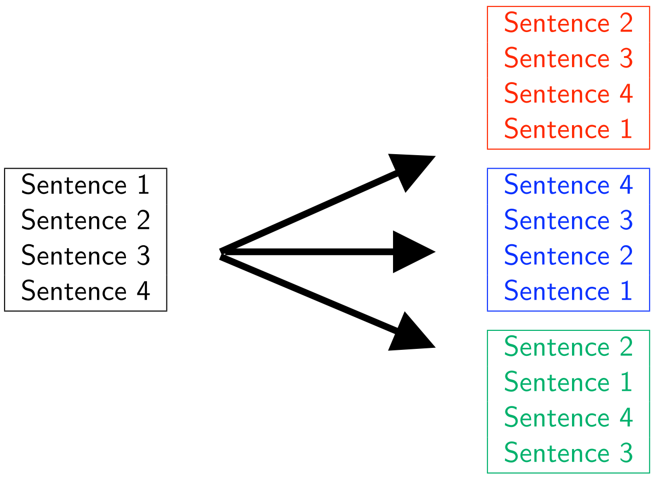
Motivation:

- Determine a sequence in which to present a set of information-bearing items.
- Information-ordering is used to evaluate text structuring algorithms.
- Essential step in generation applications.

Data:

- Source document and permutations of its sentences.
- Original order *assumed coherent*.
- Given k documents, with n permutations, we obtain $k \cdot n$ pairwise rankings for training and testing.
- Two corpora, Earthquakes and Accidents, 100 texts each.

Text Ordering



Sentence 1
Sentence 2
Sentence 3
Sentence 4

Sentence 2
Sentence 3
Sentence 4
Sentence 1

Sentence 4
Sentence 3
Sentence 2
Sentence 1

Sentence 2
Sentence 1
Sentence 4
Sentence 3

Comparison with LSA Vector Model [FKL98]

- Meaning of individual words is represented in vector space.
- Sentence meaning is the mean of the vectors of its words.
- Average distance of adjacent sentences.
- *Unsupervised, local, unlexicalised, domain independent.*

Comparison with HMM-based Model [BL04]

- Model topics and their order in text.
- Model is an HMM: states correspond to topics (sentences).
- Model selects sentence order with highest probability.
- *Supervised, global, lexicalized, domain dependent.*

Results

Model	Earthquakes	Accidents
Coreference+Syntax+Salience+	**87.2**	**90.4**
Coreference+Syntax+Salience−	88.3	90.1
Coreference+Syntax−Salience+	86.6	88.4**
Coreference−Syntax+Salience+	83.0**	89.9
Coreference+Syntax−Salience−	86.1	89.2
Coreference−Syntax+Salience−	82.3**	88.6*
Coreference−Syntax−Salience+	83.0**	86.5**
Coreference−Syntax−Salience−	81.4**	86.0**
HMM-based Content Models	88.0	75.8**
Latent Semantic Analysis	81.0**	87.3**

Evaluation metric: % correct ranks in test set.
**: significant different from Coreference+Syntax+Salience+

Results

- Omission of coreference causes performance drop.
- Linguistically poor model generally worse.
- Entity model is better than LSA.
- HMM-based content models exhibit high variability.
- Models seem to be complementary.

This appears a fruitful area, in which work is continuing.

Outline of Part 2

- Recovering topic structure
- Recovering genre-specific structure
- Recovering relational structure (Discourse Chunking)
- Recovering hierarchical structure (Discourse Parsing)
- Classifying unmarked relations
- Identifying the entity structure
- **Useful resources**

Useful resources

English

- RST Discourse TreeBank [CMO03]
  http://www.ldc.upenn.edu, CatalogEntry=LDC2002T07
- Discourse Graph Bank [WG05]
  http://www.ldc.upenn.edu, CatalogEntry=LDC2005T08
- Penn Discourse TreeBank [PDL^{+}07, PDL^{+}08]
  http://www.ldc.upenn.edu, CatalogEntry=LDC2008T05

Dutch

- Discourse-annotated Dutch corpus [vdVBB^{+}11]

German

- Potsdam Commentary Corpus [Ste04]

Useful resources

Danish, English, German, Italian and Spanish

- Copenhagen Dependency TreeBank [BKrKeM09]
  Parallel treebanks (˜40K words per language)
  http://code.google.com/p/copenhagen-dependency-treebank/wiki/CDT

Turkish

- METU Turkish Discourse Resource
  [ZW08, ZTB^{+}09, ZDSc^{+}10]
  http://www.ii.metu.edu.tr/research_group/metu-turkish-discourse-resource-project

Useful resources

Hindi

- Hindi Discourse Relation Bank [OPK^{+}09] 200K words corpus from the newspaper *Amar Ujala*

Arabic

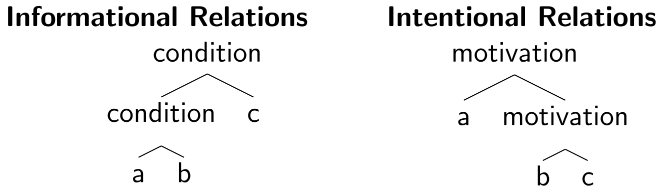
- Leeds Arabic Discourse TreeBank [ASM10, ASM11]

Multi-layer resources 1

Resources mentioned so far have one layer of discourse annotation.

- But annotating multiple discourse structures is revealing.
- e.g., the structure of CONDITION coherence relations in (30) differs from its intentional structure in terms of MOTIVATION

(30) (a) Come home by five o'clock. (b) Then we can go to the hardware store. (c) That way we can finish the bookshelves tonight.



Informational Relations

Intentional Relations

Multi-layer resources 2

Multi-level annotation [KOOM01, PSEH04] can handle this.

Multi-level annotation of discourse is not new.

- the Potsdam Commentary Corpus (PCC) [Ste04] annotates both coreference and discourse connectives and their arguments
- Complementing the Penn Discourse Treebank (PDTB) [PDL^{+}07, PDL^{+}08] annotation of discourse relations is entity coreference annotation in the OntoNotes project [HMP^{+}06]

In multi-level discourse annotation, discourse structures themselves form multiple layers.

Multi-layer resources 3

The new PCC resource [Ste08a] distinguishes four layers

- conjunctive (aka *coherence*) relations: read off the text surface
- intentional structure: the speaker's intention
- thematic structure: what the discourse and its parts are about
- referential structure: coreference relations

Typically, conjunctive relations link only parts of a discourse

- the majority of discourse relations is not signalled explicitly

Multi-layer resources 4

The atomic segments of intentional structures are classified as 'speech acts'

- examples: 'stating an option', 'making a suggestion'

These speech acts combine into larger units though relations like

- 'encourage acting' (≈ MOTIVATION in RST)
- 'ease understanding' (≈ BACKGROUND)

Segments can simultaneously enter conjunctive and intentional relations on different levels of the annotation.

Multi-layer resources 5

The AMI meeting corpus (http://corpus.amiproject.org/) — richly annotated with dialogue acts — has recently been annotated with rhetorical relations [PPB11], relating both semantic content and communicative function.

(31) B1: I'm afraid we have no time for that.
B2: We're supposed to finish this before twelve. [AMI meeting corpus ES2002a]

B2 has an EXPLANATION relation to the DECLINE REQUEST act in B1.

Outline of Part 3

- **Summarization**
- Information Extraction
- Essay analysis and scoring
- Sentiment analysis and opinion mining

Summarization 1

Summarization is one of the earliest applications of discourse structure analysis.

It motivated much research on theories of hierarchical discourse structure (e.g., RST) [OSM94, Mar00a, TvdBPC04]

Discourse segmentation is also applied to summarization.

There are different types of summarizers:

- based on hierarchical or sequential discourse structure,
- designed for different kinds of documents,
- using different ways of identifying discourse structure.

Summarization 2

Summarization based on hierarchical discourse structure

- This kind of summarization exploits the asymmetry of many discourse relations (slide 22).
- Information in nuclei is more central than the one in a satellite.
- Most satellites can be left out without diminishing the readability of a text.
- From a hierarchical structure, one can derive a partial ordering of units according to importance.
- Advantage: cutoff points can be chosen freely, which makes summarisation scalable.
- Different ways of weighing units yield similar results [UPN10].

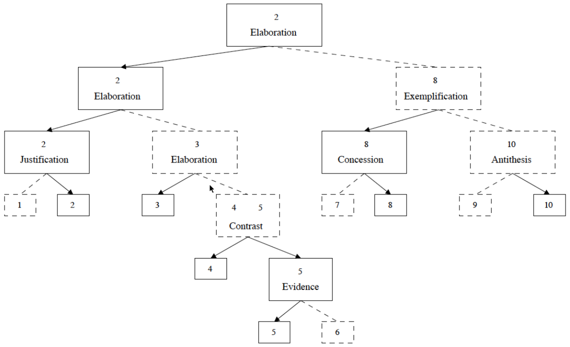
Summarization 3

With its distant orbit (...), (C_1) Mars experiences frigid weather conditions. (C_2) Surface temperatures typically average about -60 degrees Celsius (-76 degrees Fahrenheit) at the equator (...). (C_3) Only the midday sun at tropical latitudes is warm enough to thaw ice on occasion, (C_4) but any liquid water formed in this way would evaporate almost instantly (C_5) because of the low atmospheric pressure. (C_6) Although the atmosphere holds a small amount of water (...), (C_7) most Martian weather involves blowing dust or carbon dioxide. (C_8) Each winter, for example, a blizzard of frozen carbon dioxide rages over one pole (...). (C_9) Yet even on the summer pole (...) temperatures never warm enough to melt frozen water. (C_{10})

(example from [Mar00b])

Summarization 4

- The discourse structure tree of [Mar00b] for this example:



- The resulting equivalence classes for the segments: $2 > 8 > 3, 10 > 1, 4, 5, 7, 9 > 6$

Summarization 5

Summarization based on hierarchical discourse structure (ctd)

- This approach instantiates extractive summarization, which selects the most important sentences of a text.
- This is in contrast with sentence simplification systems, which shorten ('compress') the individual sentences.
- It is used for summaries representing textual content [DM02].
- There are other goals for summarization:
  - Indicative summary: Is a text worth while reading? [BE97]
  - For scientific articles: highlight their contribution and relate it to previous work [TM02].
- Bottleneck: automatic parsing of unrestricted discourse.
- Alleviation: underspecification of discourse structure [Sch02].

Summarization 6

Summarization based on flat genre-specific discourse structure

- [TM02] use discourse segmentation ('argumentative zoning') for the summarization of scientific articles.
- They assume a structure of scientific papers comprising:
  - Aim (research goal),
  - Textual (outline of paper),
  - Own (own contribution; methods, results, discussion),
  - Other (presentation of other work).
- They classify sentences for membership in these classes.
- Summarization can then focus on specific parts of the paper.

Summarization 7

Mostly, the types of documents to be summarized are news articles or scientific articles.

Their structure is radically different - and so are ways of approaching their summarisation.

- The first sentences of news articles are often good summaries (due to their 'inverted pyramid' structure, see slide 34).
- For scientific articles, core sentences are more evenly distributed.

Summarizers are optimized for one class of documents, e.g., the one of [Mar00a] targets essays and argumentative texts.

Summarization 8

Discourse structure can be identified by cue phrases (e.g., discourse markers) or punctuation [Mar00b].

[TM02]'s features include location in the document, length, and lexical and phrasal cue elements (e.g., *along the lines of*).

[BE97, CL10] use lexical chains:

- They are useful for extraction and compression: identification of summary-worthy sentences or key expressions.
- Strength of lexical chains is calculated in terms of chain length/homogeneity or amount of units covered.

Outline of Part 3

- Summarization
- **Information Extraction**
- Essay analysis and scoring
- Sentiment analysis and opinion mining

Information Extraction 1

Information Extraction (IE) extracts from texts named entities and their roles in event descriptions.

- Named entities comprise persons, organizations, or locations.

IE systems focus on specific domains (e.g., terrorist incidents), searching only for information relevant to the domain.

Often, requests for information are described by templates:

```
Name: %MURDERED%
Event Type: MURDER
TriggerWord: murdered
Activating Conditions: passive-verb
Slots: VICTIM <subject>(human)
       PERPETRATOR<prep-phrase, by>(human)
       INSTRUMENT<prep-phrase, with>(weapon)
```

Information Extraction 2

Discourse structure is used to guide the selection of parts of a document which are relevant for IE

This is part of a larger tendency towards a two-step IE

- identify relevant regions for a specific piece of information first
- then try to extract this piece of information from these regions

This boosts the overall performance of IE systems [PR07]:

- many fewer false multiple retrievals of fillers for the same slot,
- fewer false hits (often in irrelevant parts) ,
- a more confident search in the relevant parts.

Information Extraction 3

Discourse structure information helps identify relevant parts of documents with a strongly conventionalised structure.

Different kinds of discourse structure can be used for IE:

- a flat discourse structure based on [TM02]'s argumentative zoning (see slide 111) for biology articles [MKMC06],
- the top levels of a hierarchical discourse structure [MUD99, MUD00],
- the lower levels of a hierarchical discourse structure [MC07].

Information Extraction 4

Example 1: extracting the novel contribution of a scientific paper.

- Discourse parts expressing results might report earlier work.
- Argumentative zoning identifies parts with novel contributions.
- [MKMC06] refine the *Own* class of [TM02] into *Method*, *Result*, *Insight*, and *Implication*.
- Then they investigate the distribution of these subclasses across the common fourfold division of scientific articles in: *Introduction*, *Materials and methods*, *Results*, and *Discussion*.
- Subclasses and divisions do not correlate perfectly:
  - high for *Materials and methods* vs. *Method*,
  - low for *Results* vs. *Result*.

Information Extraction 5

Example 2: extraction of offences and verdicts from criminal cases.

- The structure of this genre is highly conventionalized.
- Discourse parsing is used to identify the parts that convey this information [MUD99, MUD00].
- The top part of the hierarchical discourse structure for legal texts is follows a fixed order.
- This information is then the basis for short indicative summaries.

Outline of Part 3

- Summarization
- Information Extraction
- **Essay analysis and scoring**
- Sentiment analysis and opinion mining

Essay analysis and scoring 1

Here, the overall goal is improving essay quality by giving feedback on its organizational structure.

For this, specific discourse elements in an essay are identified.

The elements are part of a non-hierarchical genre-specific conventional discourse structure (slide 34).

First, thesis statements are automatically identifed [BMAC01]:

- They explicate purpose and/or main ideas of the essay.
- This can itself serve as feedback to the authors.
- Assessing essay structure centers around the thesis statement.

Essay analysis and scoring 2

For the automatic identification of thesis statements, probabilistic classifiers are trained on corpora of manually annotated essays.

Features include position in the essay and specific lexical items.

RST-based features (relation, nuclearity) are obtained from discourse parsing [SM03].

This approach generalizes across essay topics.

Essay analysis and scoring 3

This approach was extended to identify the main parts of an argumentative essay:

- introductory material,
- thesis,
- main idea [thesis + main idea(s) = thesis statement],
- supporting ideas,
- conclusion [BMK03].

(32) <**Introductory material**> I've seen many successful people who are doctors, artists, teachers, designers, etc. </**Introductory material**> <**Main point**> In my opinion they were considered successful people because they were able to find what they enjoy doing and worked hard for it. </**Main point**>

Essay analysis and scoring 4

Two kinds of discourse analysers are used to identify the main parts of an argumentative essay:

- decision-based, with the features discourse structure markers (not parsing), syntactic structure, and position in the essay,
- stochastic, targeting sequences of segments (e.g., no conclusion at the beginning).

Combining the best analysers in a voting system optimizes results.

Essay analysis and scoring 5

The next level is to assess the internal coherence of the essay.

This presupposes identification of the discourse units in an essay.

[HBMG04] define coherence in terms of relatedness of units:

- to the essay topic (in particular, for thesis statement, background, and conclusion),
- to thesis (especially for main ideas, background material, and conclusion),
- within units.

Relatedness is modelled as semantic similarity, i.e., the amount of terms in the same semantic domain.

Outline of Part 3

- Summarization
- Information Extraction
- Essay analysis and scoring
- **Sentiment analysis and opinion mining**

Sentiment analysis and opinion mining 1

Its goal is to assess the overall opinion expressed in a review.

The impact of evaluative words in a text is rated and a score for the text is calculated.

But this impact depends on their position in the discourse.

As a first approximation, evaluative words can get more weight at the beginning and the end [PLV02] or only at the end [VT07].

As a second approximation, discourse markers can be used to weigh evaluative words [PZ04].

(33) Though Al is brilliant at math, he is a horrible teacher.

Sentiment analysis and opinion mining 2

But such approaches will encounter problems in cases like (34):

(34) Aside from a couple of unnecessary scenes, The Sixth Sense is a low-key triumph of mood and menace; the most shocking thing about it is how hushed and intimate it is, how softly and quietly it goes about its business of creeping us out. The movie is all of a piece, which is probably why the scenes in the trailer, ripped out of context, feel a bit cheesy.

Sentiment analysis and opinion mining 3

The central statement in (34) is the second clause.

Its evaluative word *triumph* outweighs the majority of negative evaluative words.

A related observation is that evaluative words in highly topical sentences get higher weight [PL05, Tur02].

Appraisal Analysis [Mar00c] refines the notion of opinion by distinguishing three components:

- affect (emotional),
- judgement (ethical),
- appreciation (aesthetic).

Sentiment analysis and opinion mining 4

Movie reviews are also complicated because they are a mixture of descriptive and evaluative segments.

Evaluative words in descriptive segments do not count [PLV02].

(35) I love this movie

(36) The colonel's wife (played by Deborah Kerr) loves the colonel's staff sergeant (played by Burt Lancaster).

This calls for a discourse analysis of evaluative texts.

- [TBS09] successfully include discourse-structure information in a system that classifies reviews as either positive or negative.
- Argumentative zoning works better here than discourse parsing [VT07, TBS09].

Outline of Part 4

In the next 5-10 years, we expect to see:

- Improved recognition of discourse structures
- New applications of discourse structures – in particular, Machine Translation

Improved recognition of discourse structures

Theory: Better understanding of

- each type of discourse structure;
- relations between different types/layers of structure.

Practice: More training data through

- Easier, cheaper ways acquisition of manual annotation;
- More effective use of unlabelled data.

New Applications

(Statistical) Machine Translation could draw benefits from three aspects of discourse structure:

- Segments of topic structure and functional structure vary in their syntactic and lexical features;
- Relational and hierarchical structure convey meaning through structure;
- With entity structure, reference is constrained through structure.

Topic Structure and Machine Translation (MT)

○ Text heterogeneity across topic and functional structures can be exploited to improve translation.

⇒ Tailor sub-language models to sub-structure. Overcome lack of a natural back-off strategy for gaps in data [FIK10].

○ Propagation of corrections made in post-editting a document can already improve translation to the rest [HE10].

○ For highly structured documents such as patents, corrections made to near-by sentences provide more value than corrections further away [HE10].
⇒ Given a source text annotated with topic structure breaks, one could focus correction propagation to all/only sentences within the same segment of topic or functional structure.

Entity Structure and Statistical MT

Anaphors (pronouns and 0-anaphors) are constrained by their **antecedents** in all languages, but in different ways.

- English: Pronoun gender reflects the **referent** of the antecedent.
- French, German, Czech: Pronoun gender reflects the **form** of the antecedent.

(37) a. Here's a book. I wonder if **it** is new. (inanimate, neuter referent)
b. Voici un livre. Je me demande si **il** est nouveau. (masculine form)

Entity Structure and SMT

Phrase-based and syntax-based SMT just consider the local context - cf. Google translate

(38) I wonder if it is new.
Google translate: Je me demande si **elle** est nouvelle.

(39) I wondered if it was new.
Google translate: Je me demandais si **il** était neuf.

Entity Structure and SMT

Recognizing **entity structures** in a source text links anaphors to their antecedents, potentially allowing appropriate anaphoric forms to be projected into the target.

Preliminary work [NK10, HF10] is based on annotating source text:

- Identify the antecedent(s) of a source language pronoun through anaphor resolution;
- Identify the gender of the target text aligned with that antecedent;
- Annotate the source text pronoun with this gender, and use the annotated text to produce a translation model;
- Annotate source text pronouns with their antecedents in test data, to make use of this enriched translation model.

Relational structure and SMT

○ Aspects of meaning are conveyed through relational structures, that can be marked by discourse connectives.

○ Discourse connectives cover different senses in different languages.

- *Since* in English can express either an explanation (like *because*) or a temporal relation (like *after*).
- *Puisque* in French expresses only the former sense, while *depuis* expresses only the latter.

⇒ Preliminary work [Mey11] suggests that recognizing and annotating **relational structures** in the source can allow appropriate discourse connectives to be selected in the target.

Relational structure and SMT

○ Translators often make discourse connectives **explicit** in their target translation that were implicit in the source [KO11]:

Connective	Orig Frequency	Trans Frequency
therefore	0.153%	0.287%
nevertheless	0.019%	0.045%
thus	0.015%	0.041%
moreover	0.008%	0.035%

○ This can produce source-target mis-alignments that produce bad entries in the translation model.

Relational structure and SMT

○ Using **relational structure** to explicitate **implicit connectives** in source texts [PDL$^+$08] should improve alignment and thus SMT.

E.g. Implicit THEREFORE (114 tokens) and THUS (179 tokens) in the PDTB.

(40) **Its valuation methodologies**, she said, "**are recognized as some of the best on the Street**. Implicit = THEREFORE **Not a lot was needed to be done**." [wsj_0304]

(41) "**In Asia, as in Europe, a new order is taking shape**," Mr. Baker said. Implicit = THUS "**The U.S., with its regional friends, must play a crucial role in designing its architecture**." [wsj_0043]

Hierarchical structure and SMT

○ Languages may differ in their common ways of expressing relational and hierarchical structure [MCW00].

○ Syntactic/dependency structure is beginning to be used as an **inter-lingua** so that features of the source conveyed through syntactic and/or dependency structure can be preserved in translating to the target.

○ In the same way, a hierarchical structure such as RST could be used as an **inter-lingua** for a larger unit of text.

○ Here it would be features of the source expressed through hierarchical structure that would be preserved, even if sentence order were violated [GBC01].

Conclusion

This tutorial has tried to introduce you to:

- Ways in which discourse is structured;
- Ways of recovering these structures from text;
- Ways that discourse structures can support LT applications;
- Discourse resources that will support new discoveries and applications;
- Opportunities for improving and exploiting discourse structures in the future.

The future is in your hands.

Thank you!