Twitter is a service for friends, family, and co–workers to communicate and stay connected through the exchange of quick, frequent answers to one simple question: **What are you doing?**

The Daily W

Sunday, August 30, 2006

Martians invade earth

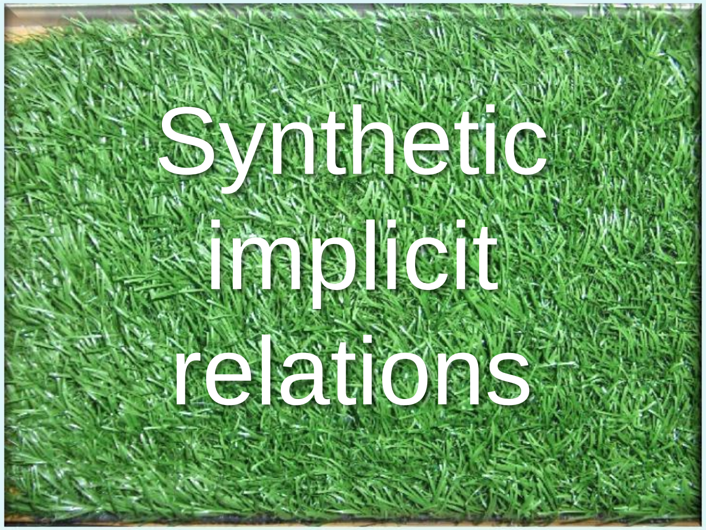
Incredible as it may seem, it has been confirmed that a large martian invasion fleet has landed on earth tonight.

First vessels were sighted over Great Britain, Denmark and Norway already in the late evening from where, as further reports indicate, the fleet headed towards the North Pole and Santa Claus was taken hostage by the invaders.

Afterwards they split apart in order to approach most major cities around the earth. The streets filled as thousands fled their homes, many only wearing their pajamas...

Empirical approaches to discourse

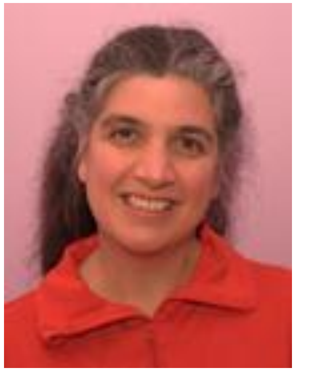
ESSLLI 2012
Jennifer Spenader



- Marcu & Echihabi (2002).
- Sporleder & Lascarides (2008)

- Create **synthetic** examples of implicit relations by taking unambiguously marked relations and removing the connective.

Sporleder & Lascarides (2007)



Using Automatically Labelled Examples to Classify Rhetorical Relations: An Assessment

- Automatic rhetorical relation identification is a goal
  - To be able to use supervised machine learning to create such an application, you need manually annotated data
  - creating manually annotated data is time-consuming
- Some rhetorical relations are unambiguously marked
  - these examples can be used to create models that can then be applied to unmarked (implicit) examples

- Extracted a set 8.3 million unambiguously marked examples for training
  - Used 55 unambiguous markers for extraction, based on SDRT
  - Remove the connective and they resemble Implicit relations
- Synthetic Examples taken from:
  - the British National Corpus (BNC,100 million words),
  - the North American News Text Corpus (350 million words)
  - the English Gigaword Corpus (1.7 billion)

Table 1. *Number of Automatically Extracted Examples per Relation*

	CONTRAST	EXPLANATION	RESULT	SUMMARY	CONTINUATION
examples	6,753,104	1,490,274	14,978	16,718	8,495

- Used the RST Discourse Treebank to extract implicit relations (Carlson et al., 2002)
  - Potential implicit relations of the right type were extracted from the corpus
    - only relations that did NOT include any of the 55 unambiguous markers used to extract the synthetic examples were used
  - The were then manually checked and categorized to create a set of implicit relations of the same types that were extracted for training.
  - 1,050 relations in total

	Contrast	Explanation	Result	Summary	Continuation
# of manual examples	213	268	266	44	260

	Accuracy	Kappa
intra-annotator agreement	79.47%	.679
inter-annotator agreement	71.86%	.592

Table 2. *Intra- and Inter-Annotator Agreement for Manual Labelling of Relations*

Selection of 200 of 1,050 relations

- Intra-annotator agreement= same annotator 6 mths later
- Inter-annotator agreement = second annotator

Sporleder & Lascarides

- Two Language Models
  - LM1 : Naïve Bayes Word frequency model
    - Almost identical to model used by Marcu & Echihabi
    - **‘knowledge lean’**
  - LM2: Model with 41 Linguistically motivated features
    - POS information
    - Positional features
      - E.g. Beginning or end of a paragraph
    - Length features
      - E.g. EXPLANATION often longer than e.g. SUMMARY
    - Temporal features
      - About verbs
    - Cohesion features
      - Ellipsis? Number of pronouns, etc…
    - **‘knowledge rich’**

LM1: Naïve Bayes, Unambiguously marked data

Table 3. *Applying the Naive Bayes Word Pair Model to unambiguously marked data, 10-fold cross-validation*

Relation	Avg. Acc	Avg. Prec	Avg. Rec	Avg. F-Score
continuation	n/a	23.54	62.36	34.17
result	n/a	52.07	27.41	35.90
summary	n/a	56.49	32.79	41.46
explanation	n/a	47.56	71.32	57.05
contrast	n/a	50.31	26.06	34.29
all	42.34	45.99	43.99	40.57

LM2: BoosTexter,
Unambiguously marked data

Table 4. *Applying the BoosTexter model to unambiguously marked data, 10-fold cross-validation*

Relation	Avg. Acc	Avg. Prec	Avg. Rec	Avg. F-Score
continuation	n/a	53.37	54.90	54.11
result	n/a	56.33	47.08	51.26
summary	n/a	61.41	60.98	61.16
explanation	n/a	67.75	79.35	73.05
contrast	n/a	59.20	57.85	58.42
all	60.88	59.61	60.03	59.60

When LM is trained and tested on the same type of synthetic examples, it works better than the simple Word Pair LM1.

LM1: Naïve Bayes,
Manually annotated data, trained on
Unambiguous data

Table 5. *Applying the Naive Bayes Word Pair Model to data that is not unambiguously marked, averaged over 10 training runs*

Relation	Avg. Acc	Avg. Prec	Avg. Rec	Avg. F-Score
continuation	n/a	26.62	62.85	37.40
result	n/a	24.87	8.12	12.24
summary	n/a	5.47	8.41	6.63
explanation	n/a	31.55	25.15	27.97
contrast	n/a	23.40	7.65	11.53
all	25.92	22.38	22.44	19.15

Simple Word Pair LM trained on extracted relations, tested on manually identified implicit relations doesn't work very well.

LM2: BoosTexter,
Training: Unambiguous data
Testing: Manually annotated data

Table 6. *Applying the BoosTexter Model to unmarked data, averaged over 10 training runs*

Relation	Avg. Acc	Avg. Prec	Avg. Rec	Avg. F-Score
continuation	n/a	36.70	20.35	26.17
result	n/a	25.08	19.74	22.08
summary	n/a	9.32	45.91	15.49
explanation	n/a	37.51	37.13	37.30
contrast	n/a	21.38	21.60	21.47
all	25.80	26.00	28.94	24.50

More complex LM trained on synthetic examples leads to better performance on implicit relations than simple Word Pair LM, but still not very good.

LM1: Naïve Bayes, Training and Testing: Unambiguously marked data

Table 7. *Training and Testing on Manually Labelled Data, Naive Bayes Word Pair Model, 5 times 2-fold cross-validation*

Relation	Avg. Acc	Avg. Prec	Avg. Rec	Avg. F-Score
continuation	n/a	27.27	12.00	16.48
result	n/a	27.65	9.70	13.41
summary	n/a	2.44	29.09	4.50
explanation	n/a	29.85	5.97	9.89
contrast	n/a	19.43	23.28	20.54
all	12.88	21.33	16.01	12.96

Training even on a small data set of “good” Implicit relations with a Word Pair model leads to performances worse than a simple baseline!

Table 8. *Training and Testing on Manually Labelled Data, BoosTexter Model, 5 times 2-fold cross-validation*

Relation	Avg. Acc	Avg. Prec	Avg. Rec	Avg. F-Score
continuation	n/a	36.78	36.85	36.77
result	n/a	38.53	46.32	41.99
summary	n/a	13.75	3.64	5.63
explanation	n/a	49.80	50.15	49.85
contrast	n/a	36.70	32.21	34.19
all	40.30	35.11	33.83	33.69

Training even on a small data set of “good” Implicit relations leads to better classification with more sophisticated LM

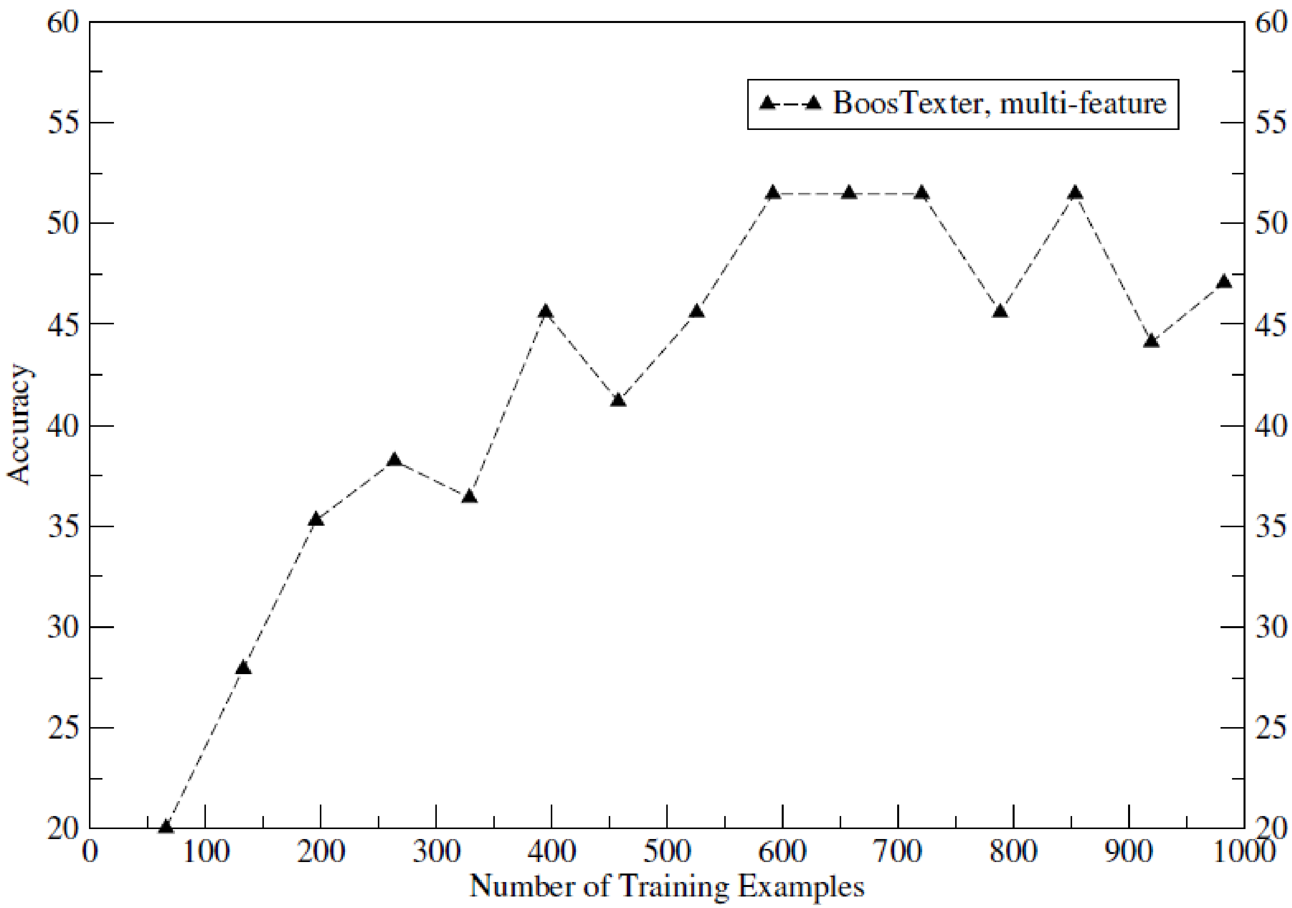


Fig. 5. Learning curve for training and testing on manually labelled, unmarked data

How much data is needed?

> Our results suggest that training on this type of data may not be such a good strategy, as models trained in this way do not seem to generalize very well to unmarked data. Furthermore, we found some evidence that this behavior is largely independent of the classifiers used and seems to lie in the data itself (e.g., marked and unmarked examples may be too dissimilar linguistically and removing unambiguous markers in the automatic labeling process may lead to a meaning shift in the examples)

Spoorleder & Lascarides

Recognizing Implicit Discourse Relations in the Penn Discourse Treebank
Lin, Kan and Ng
(*EMNLP 2009)*



Four sets of features

- **Production rules**
  - =Constituent Parse Tree information extracted from Gold Standard PTB annotation
- **Dependency rules**
  - (dependency parse derived from constituent parse tree, encodes additional word level dependencies not explicit in the constituent parse tree
- **Word pairs** (same as Marcu & Echihabi)
- **Context**
  - the connectives of **Prev** and **Next** when they are explicit relations, etc.

- Used the Implicit Relations from the PDTB
- Lin et al. used MaxEnt learner
  - recall Marcu & Echihabi used Naïve Bayes
- Test set accuracy for baselines.
  - Majority class baseline (Cause):
    - 26% accuracy
  - Random baseline:
    - 9.1% accuracy

From Lin et al. (2009). Recognizing Implicit discourse relations in the Penn Discourse Treebank

Adjusted total: removed Cases where there were too few training instances

Level 1 Class	Level 2 Type	Training instances	%	Adjusted %
Temporal	Asynchronous	583	4.36	4.36
	Synchrony	213	1.59	1.59
Contingency	Cause	3426	25.61	25.63
	Pragmatic Cause	69	0.52	0.52
	Condition	1	0.01	–
	Pragmatic Condition	1	0.01	–
Comparison	Contrast	1656	12.38	12.39
	Pragmatic Contrast	4	0.03	–
	Concession	196	1.47	1.47
	Pragmatic Concession	1	0.01	–
Expansion	Conjunction	2974	22.24	22.25
	Instantiation	1176	8.79	8.80
	Restatement	2570	19.21	19.23
	Alternative	158	1.18	1.18
	Exception	2	0.01	–
	List	345	2.58	2.58
Total		13375		
Adjusted total		13366		

Lin et al. : word pairs work well, even with a small corpus

MaxEnt vs Naive Bayes (Marcu & Echihabi)

	# Production rules	# Dependency rules	# Word pairs	Context	Acc.
R1	11,113	–	–	No	36.7%
R2	–	5,031	–	No	26.0%
R3	–	–	105,783	No	30.3%
R4	–	–	–	Yes	28.5%
R5	11,113	5,031	105,783	Yes	35.0%

Table 3: Classification accuracy with all features from each feature class. Rows 1 to 4: individual feature class; Row 5: all feature classes.

Results are pretty good, task much harder
Marcu & Echihabi

Level 2 Type	Precision	Recall	F_1	Count in test set
Asynchronous	0.50	0.08	0.13	13
Synchrony	–	–	–	5
Cause	0.39	0.76	0.51	200
Pragmatic Cause	–	–	–	5
Contrast	0.61	0.09	0.15	127
Concession	–	–	–	5
Conjunction	0.30	0.51	0.38	118
Instantiation	0.67	0.39	0.49	72
Restatement	0.48	0.27	0.35	190
Alternative	–	–	–	15
List	0.80	0.13	0.23	30
All (Micro Avg.)	0.40	0.40	0.40	780

Table 6: Recall, precision, F_1, and counts for 11 Level 2 relation types. "–" indicates 0.00.

Conclusion: Lin et al.

- **Production rules** (Syntactic constituency information) contribute the most to the performance, followed by word pairs

- But why is it still so difficult?
  – Lin et al. looked manually at their results and identified four major **challenges**

1. Ambiguity

In the third quarter, AMR said, net **fell** to $137 million, or $2.16 a share, from $150.3 million, or $2.50 a share.

[while]

Revenue rose 17% to $2.73 billion from $2.33 billion a year earlier.

(Contrast - wsj 1812)

Dow's third-quarter net **fell** to $589 million, or $3.29 a share, from $632 million, or $3.36 a share, a year ago.

[while]

Sales in the latest quarter rose 2% to $4.25 billion from $4.15 billion a year earlier.

(Conjunction - wsj 1926)

1. Ambiguity

In the third quarter, AMR said, **net** **<u>fell</u>** to $137 million, or $2.16 a share, from $150.3 million, or $2.50 a share.

[while]

Revenue **<u>rose</u>** **17% to $2.73 billion from $2.33 billion a year earlier.**

(Contrast - wsj 1812)

Dow's third-quarter **net** **<u>fell</u>** to $589 million, or $3.29 a share, from $632 million, or $3.36 a share, a year ago.

[while]

Sales in the latest quarter **<u>rose</u>** **2% to $4.25 billion from $4.15 billion a year earlier.**

(Conjunction - wsj 1926)

1. Ambiguity

In the third quarter, AMR said, **net** **fell** to $137 million, or $2.16 a share, from $150.3 million, or $2.50 a share.

[while]

Revenue **rose** **17% to $2.73 billion from $2.33 billion a year earlier.**

(Contrast - wsj 1812)

Dow's third-quarter **net** **fell** to $589 million, or $3.29 a share, from $632 million, or $3.36 a share, a year ago.

[while]

Sales in the latest quarter **rose** **2% to $4.25 billion from $4.15 billion a year earlier.**

(Conjunction - wsj 1926)

1. Ambiguity

In the third quarter, AMR said, **net** **fell** to $137 million, or $2.16 a share, from $150.3 million, or $2.50 a share.

[while]

Revenue **rose** **17% to $2.73 billion from $2.33 billion** **a year earlier.**

(Contrast - wsj 1812)

Dow's third-quarter **net** **fell** to $589 million, or $3.29 a share, from $632 million, or $3.36 a share, a year ago.

[while]

Sales **in the latest quarter** **rose** **2% to $4.25 billion from $4.15 billion** **a year earlier.**

(Conjunction - wsj 1926)

1. Ambiguity

In the third quarter, AMR said, **net** **fell** to $137 million, or $2.16 a share, from $150.3 million, or $2.50 a share.

[while]

Revenue **rose** **17% to $2.73 billion from $2.33 billion** **a year earlier.**

(Contrast - wsj 1812)

Dow's third-quarter **net** **fell** to $589 million, or $3.29 a share, from $632 million, or $3.36 a share, a year ago.

[while]

Sales **in the latest quarter** **rose** **2% to $4.25 billion from $4.15 billion** **a year earlier.**

(Conjunction - wsj 1926)

2. Inference

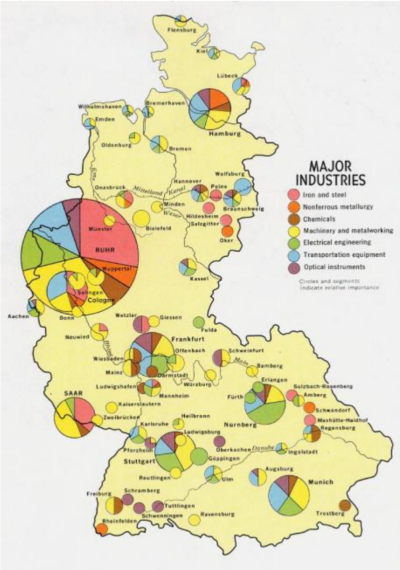
"I had calls all night long from the States," he said. **[in fact]** **I was woken up every hour – 1:30, 2:30, 3:30, 4:30."**

(Restatement - wsj 2205)

3. Context

- **the Minimality Principle** in PDTB argument selection:
  - only include in the argument the minimal span of text that is sufficient for the interpretation of the relation.

3. Context



West German Economics Minister Helmut Haussmann said, “In my view, the stock market will stabilize relatively quickly. There may be one or other psychological or technical reactions,

but they aren’t based on fundamentals.

[in short]

The economy of West Germany and the EC European Community is highly stable.”

(Conjunction – wsj 2210)

4. World knowledge

Senator Pete Domenici calls this effort “the first gift of democracy”.

[but]

The Poles might do better to view it as a Trojan Horse.

(Contrast - wsj 2237)

Lin et al.’s conclusions

- show that implicit discourse relation classification needs deeper semantic representations, more robust system design, and access to more external knowledge

- Language Models could be more sophisticated
  - Can use additional semantic information
    - E.g. Levin verb classes taken from VerbNet, etc.
    - lexical relation information (is word-x in Arg1 an antonym of word-y in Arg2?)
    - Meronymy information, e.g. a brake is part of a car…
  - Could use information about syntactic structure of the sentence
  - Hope that the content of the arguments is rich enough that the connective information is actually redundant

How difficult is Discourse Parsing?

- Depends on how you define the task.
- For explicit relations, with PDTB style annotation: not so difficult
- For implicit relations:
  - Much harder
  - Linguistically informed models work better than bag-of-word methods
  - Manually annotated training data works better than synthetically created training data
    - Suggests that implicit and explicit discourse relations **are** qualitatively different

Entity-based coherence structure

Halliday & Hasan (1976). Cohesion in English.

Cohesion

how textual units are linked
or related via words or referents

you can identify and quantify the cohesive
relationships and use this to measure cohesion
in different parts of a text.

Lexical and entity-base cohesion

Coherence

how events are linked

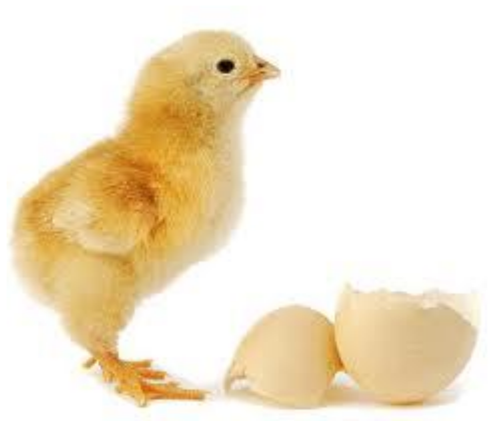
often this link is left implicit

requires world knowledge

requires inferencing

For the speaker:
Coherence comes before cohesion

For the hearer:
Cohesion helps us figure out coherence



For the speaker:

Coherence comes before cohesion

(the speaker has a message. The parts of the message fit together rhetorically. Cohesive lexical relations are just a by-product)

For the hearer:

Cohesion helps us figure out coherence

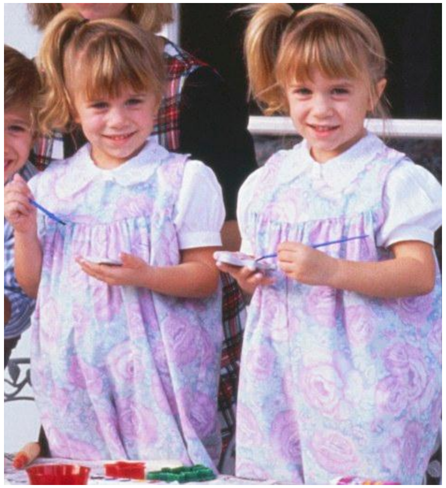
(rhetorical connections are sometimes implicit. Paying attention to cohesive relations lets the hearer reconstruct the discourse structure)

In a biography of Churchill:

"one would expect frequent mention of words like Churchill, he, him, his, and so on. The source of coherence would lie in the content, and the repeated occurrences of certain words would be the consequence of content coherence, not something that was a source of coherence."

(Morgan & Seller, 1980)

- Lexical cohesion alone is not sufficient for coherence



But it seems a bit abstract until you see some minimal pairs

- **Ferstl and von Cramen (2001):**
  The role of coherence and cohesion in text comprehension: an event-related fMRI study

- **Coherent/Cohesive**
- Mary’s exam was about to start. *Therefore*, *her* palms were sweaty.
- Laura got a lot of mail today. *Her* friends had remembered *her* birthday.

- **Coherent/ Incohesive**
- Mary’s exam was about to start. The palms were sweaty.
- Laura got a lot of mail today. Some friends had remembered the birthday.

- **Coherent/Cohesive**
- Mary’s exam was about to start. *Therefore*, ***her* palms** were sweaty.
- Laura got a lot of mail today. ***Her* friends** had remembered ***her* birthday**.

- **Coherent/ Incohesive**
- Mary’s exam was about to start. **The palms** were sweaty.
- Laura got a lot of mail today. **Some friends** had remembered **the birthday**.

- **Incoherent /Cohesive**
- Laura got a lot of mail today. *Therefore*, *her* palms were sweaty.
- Mary’s exam was about to start. *Her* friends had remembered *her* birthday.

- **Incoherent / Incohesive**
- Laura got a lot of mail today. The palms were sweaty.
- Mary’s exam was about to start. Some friends had
- remembered the birthday.

Ferstl and von Cramon (2001).

- Tested reading times and reaction times during an fMRI experiment, confirmed

- **Results:**

- lexical cohesion facilitates inference processes
- lexical cohesion makes the detection of incoherence more difficult

Cohesive devices

Grammatical or lexical

Halliday & Hasan identified five general categories of cohesive devices:

- Reference
- Substitution
- Ellipsis
- Lexical cohesion
- Conjunction

Type	Examples
Reference	Wash and core six cooking apples. Put **them** into a fireproof dish.
Substitution	My axe is blunt. I have to get a sharper **one.**
Ellipsis	Did you see John? - Yes **Ø**.
Lexical Cohesion	There is **a boy** climbing the tree. **The child's** going to fall if he does not take care.
Conjunctions	They fought a battle. **Afterwards**, it snowed.
Four types	Additive, adversative, causal and temporal

All devices related to referential form except for "Conjunction"

Halliday & Hasan (1980): extremely influential

- Google scholar: 8890 citations

- linguistic form reflects and molds discourse structure
- Separation of world knowledge and intention from the form used, which reflects it (and is our clue to it)

- Not a very practical theory: what can we use these ideas for, what claims made are specific enough to be testable?

Slide from: Nynke van der Vliet

Introduction

- **M**odeling **T**extual **O**rganization (MTO) Program

- Build a **Dutch text corpus**, annotated for discourse structure, genre structure, lexical cohesion, coreference, and discourse connectives

- Project Goals:
- Investigate the genre-dependent interaction between discourse structure and lexical cohesion (Project 1, Ildikó Berzlánovich)
- Investigate the mechanisms that establish coherence in text and develop algorithms for discourse parsing (Project 2, Nynke van der Vliet)
    - http://www.let.rug.nl/mto/

Slide from: Nynke van der Vliet

Lexical cohesion (1)

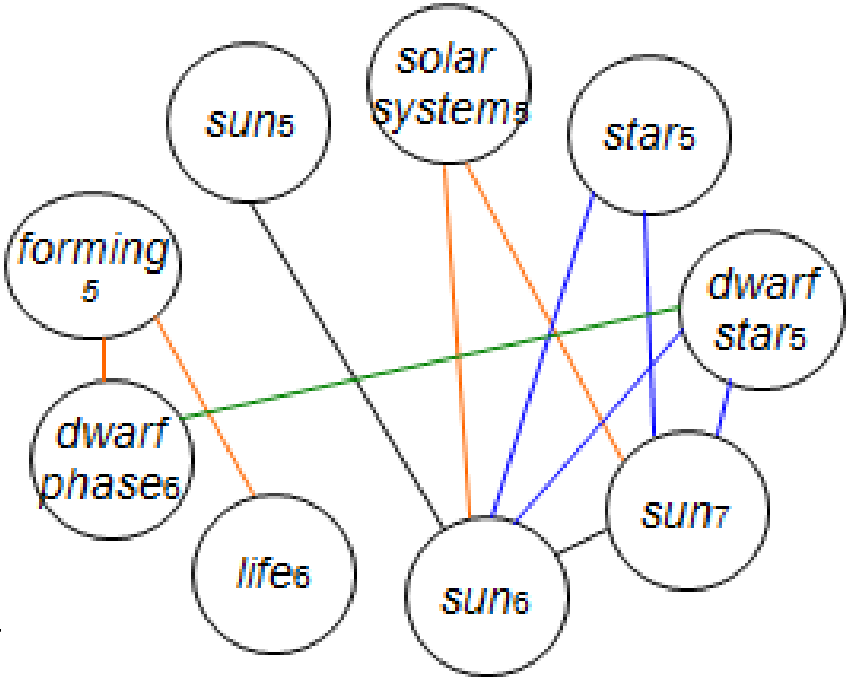
- Lexical cohesive items build up graph structures in the text
- For each lexical item, lexical links to items in preceding and following EDUs are identified

Category		Example
Repetition	Full repetition	*planet - planet*
	Partial repetition	*planet - planetary*
Systematic semantic relations	Hyponymy	*sun - star*
	Hyperonymy	*gas - hydrogen*
	Co-hyponymy	*Venus - Mercury*
	Meronymy	*planet - solar system*
	Holonymy	*solar system - sun*
	Co-meronymy	*Earth - sun*
	Synonymy	*life -existence*
	Antonymy	*light - heavy*
Collocation		*light - star*

Slide from: Nynke van der Vliet

Lexical cohesion (2)

EDU5 [*After the* ***forming*** *of the* ***sun*** *and the* ***solar system****, our* ***star*** *began its long existence as a so-called* ***dwarf star***] EDU6 [*In the* ***dwarf phase*** *of its* ***life****, the energy that the* ***sun*** *gives off is generated in its core through the fusion of hydrogen into helium.*] EDU7[*The* ***sun*** *is about five billon years*]



Lexical Cohesion

- Could be done automatically
  - use WordNet, automatical extracted lexical relations, etc.
- Useful for telling use
  - can use to study difference between genres
  - or, e.g. automatic essay grading
    - assumption: the more lexically cohesive a text is, the more coherent it is
    - Recall: `Maximize Discourse Coherence´ from SDRT
      - the more links you can identify, the better
      - also includes anaphoric links
      - but anaphoric linking is just one type of link
        - » has been interesting because it´s an obvious difficulty for automatically interpreting a text

Coreference tracking

- Simply keeping track of what referents were referred to when, is also important aspect of determining how coherent a text is (e.g. Churchill example).
  - or e.g. topic recognition,

- “Coreference resolution”

Op **9 december 1983** werd **Alfred Heineken** samen met **zijn chauffeur** ontvoerd.

On **the 9th of december 1983 Alfred Heineken** was kidnapped together with **his driver.**

De kidnappers vroegen **43 mijoen gulden losgeld. Een bescheiden bedrag**, vonden **ze** zelf.

The kidnappers demanded **43 million guilders** in **ransom. A modest amount, they** thought.

- Coreference resolution:
  - Key task
    - Machine translation, automatic summarization, information extraction, essay rating, topic segmentation
  - Complex
    - Requires many different kinds of knowledge
      - Morphological, lexical information
      - Syntactic function of bothe the anaphor and antecedent
      - Semantic information about hyponyms and synonyms
      - Semantic information abotu different named entities

Hoste & Daelemans

- **Steps**

1. Created an annotated corpus of coreference chains
2. Preprocessing steps
3. Created positive and negative instances for training and test data
4. Experiments with three seperate data sets for different NP types
5. Selection of features for the machine learning
6. Compared two machine learning approaches
7. Error analysis to determine how to improve results

Hoste & Daelemans

Op den Akker (2002)	802 pronouns
Bouma (2003)	222 pronouns
KNACK 2002	12,546 noun phrases (267 documents)

Hoste & Daelemans

Ongeveer een maand geleden stuurde
<COREF ID = ”1”> **American Airlines**</COREF>
<COREF ID = ”2” MIN = ”toplui”> **enkele toplui**</COREF>
naar Brussel.
<COREF ID = ”3” TYPE = ”IDENT” REF = ”1”
MIN=”vliegtuigmaatschappij”> **De grote**
vliegtuigmaatschappij </COREF>
had interesse voor DAT en wou daarover
<COREF ID = ”5”> **de eerste minister**</COREF>
spreken. Maar
<COREF ID = ”6” TYPE = ”IDENT” REF
= ”5”> **Guy Verhofstadt** </COREF>
(VLD) weigerde
<COREF ID = ”7” TYPE = ”BOUND” REF = ”2”> **de delegatie**
</COREF>
te ontvangen.

Hoste & Daelemans

Ongeveer een maand geleden stuurde
<COREF ID = ″1″> American Airlines</COREF>
<COREF ID = ″2″ MIN = ″toplui″> enkele toplui</COREF>
naar Brussel.
<COREF ID = ″3″ TYPE = ″IDENT″ REF = ″1″
MIN=″vliegtuigmaatschappij″> De grote
vliegtuigmaatschappij </COREF>
had interesse voor DAT en wou daarover
<COREF ID = ″5″> de eerste minister</COREF>
spreken. Maar
<COREF ID = ″6″ TYPE = ″IDENT″ REF
= ″5″> Guy Verhofstadt </COREF>
(VLD) weigerde
<COREF ID = ″7″ TYPE = ″BOUND″ REF = ″2″> de delegatie
</COREF>
te ontvangen.

Hoste & Daelemans

Ongeveer een maand geleden stuurde
<COREF ID = ”1”> American Airlines</COREF>
<COREF ID = ”2” MIN = ”toplui”> enkele toplui</COREF>
naar Brussel.
<COREF ID = ”3” TYPE = ”IDENT” REF = ”1”
MIN=”vliegtuigmaatschappij”> De grote
vliegtuigmaatschappij </COREF>
had interesse voor DAT en wou daarover
<COREF ID = ”5”> de eerste minister</COREF>
spreken. Maar
<COREF ID = ”6” TYPE = ”IDENT” REF
= ”5”> Guy Verhofstadt </COREF>
(VLD) weigerde
<COREF ID = ”7” TYPE = ”BOUND” REF = ”2”> de delegatie
</COREF>
te ontvangen.

Hoste & Daelemans

Ongeveer een maand geleden stuurde
<COREF ID = ″1″> American Airlines</COREF>
<COREF ID = ″2″ MIN = ″toplui″> enkele toplui</COREF>
naar Brussel.
<COREF ID = ″3″ TYPE = ″IDENT″ REF = ″1″
MIN=″vliegtuigmaatschappij″> De grote
vliegtuigmaatschappij </COREF>
had interesse voor DAT en wou daarover
<COREF ID = ″5″> de eerste minister</COREF>
spreken. Maar
<COREF ID = ″6″ TYPE = ″IDENT″ REF
= ″5″> Guy Verhofstadt </COREF>
(VLD) weigerde
<COREF ID = ″7″ TYPE = ″BOUND″ REF = ″2″> de delegatie
</COREF>
te ontvangen.

Hoste & Daelemans

Ongeveer een maand geleden stuurde
<COREF ID = ″1″> American Airlines</COREF>
<COREF ID = ″2″ MIN = ″toplui″> enkele toplui</COREF>
naar Brussel.
<COREF ID = ″3″ TYPE = ″IDENT″ REF = ″1″
MIN=″vliegtuigmaatschappij″> De grote
vliegtuigmaatschappij </COREF>
had interesse voor DAT en wou daarover
<COREF ID = ″5″> de eerste minister</COREF>
spreken. Maar
<COREF ID = ″6″ TYPE = ″IDENT″ REF
= ″5″> Guy Verhofstadt </COREF>
(VLD) weigerde
<COREF ID = ″7″ TYPE = ″BOUND″ REF = ″2″> de delegatie
</COREF>
te ontvangen.

Hoste & Daelemans

Ongeveer een maand geleden stuurde
<COREF ID = ”1”> American Airlines</COREF>
<COREF ID = ”2” MIN = ”toplui”> enkele toplui</COREF>
naar Brussel.
<COREF ID = ”3” TYPE = ”IDENT” REF = ”1”
MIN=”vliegtuigmaatschappij”> De grote
vliegtuigmaatschappij </COREF>
had interesse voor DAT en wou daarover
<COREF ID = ”5”> de eerste minister</COREF>
spreken. Maar
<COREF ID = ”6” TYPE = ”IDENT” REF
= ”5”> Guy Verhofstadt </COREF>
(VLD) weigerde
<COREF ID = ”7” TYPE = ”BOUND” REF = ”2”> de delegatie
</COREF>
te ontvangen.

Hoste & Daelemans

Ongeveer een maand geleden stuurde
<COREF ID = ”1”> American Airlines</COREF>
<COREF ID = ”2” MIN = ”toplui”> enkele toplui</COREF>
naar Brussel.
<COREF ID = ”3” TYPE = ”IDENT” REF = ”1”
MIN=”vliegtuigmaatschappij”> De grote
vliegtuigmaatschappij </COREF>
had interesse voor DAT en wou daarover
<COREF ID = ”5”> de eerste minister</COREF>
spreken. Maar
<COREF ID = ”6” TYPE = ”IDENT” REF
= ”5”> Guy Verhofstadt </COREF>
(VLD) weigerde
<COREF ID = ”7” TYPE = ”BOUND” REF = ”2”> de delegatie
</COREF>
te ontvangen.

Hoste & Daelemans

Ongeveer een maand geleden stuurde
<COREF ID = ”1”> American Airlines</COREF>
<COREF ID = ”2” MIN = ”toplui”> enkele toplui</COREF>
naar Brussel.
<COREF ID = ”3” TYPE = ”IDENT” REF = ”1”
MIN=”vliegtuigmaatschappij”> De grote
vliegtuigmaatschappij </COREF>
had interesse voor DAT en wou daarover
<COREF **ID = ”5”>** **de eerste minister**</COREF>
spreken. Maar
<COREF ID = ”6” TYPE = ”IDENT” REF
= ”5”> Guy Verhofstadt </COREF>
(VLD) weigerde
<COREF ID = ”7” TYPE = ”BOUND” REF = ”2”> de delegatie
</COREF>
te ontvangen.

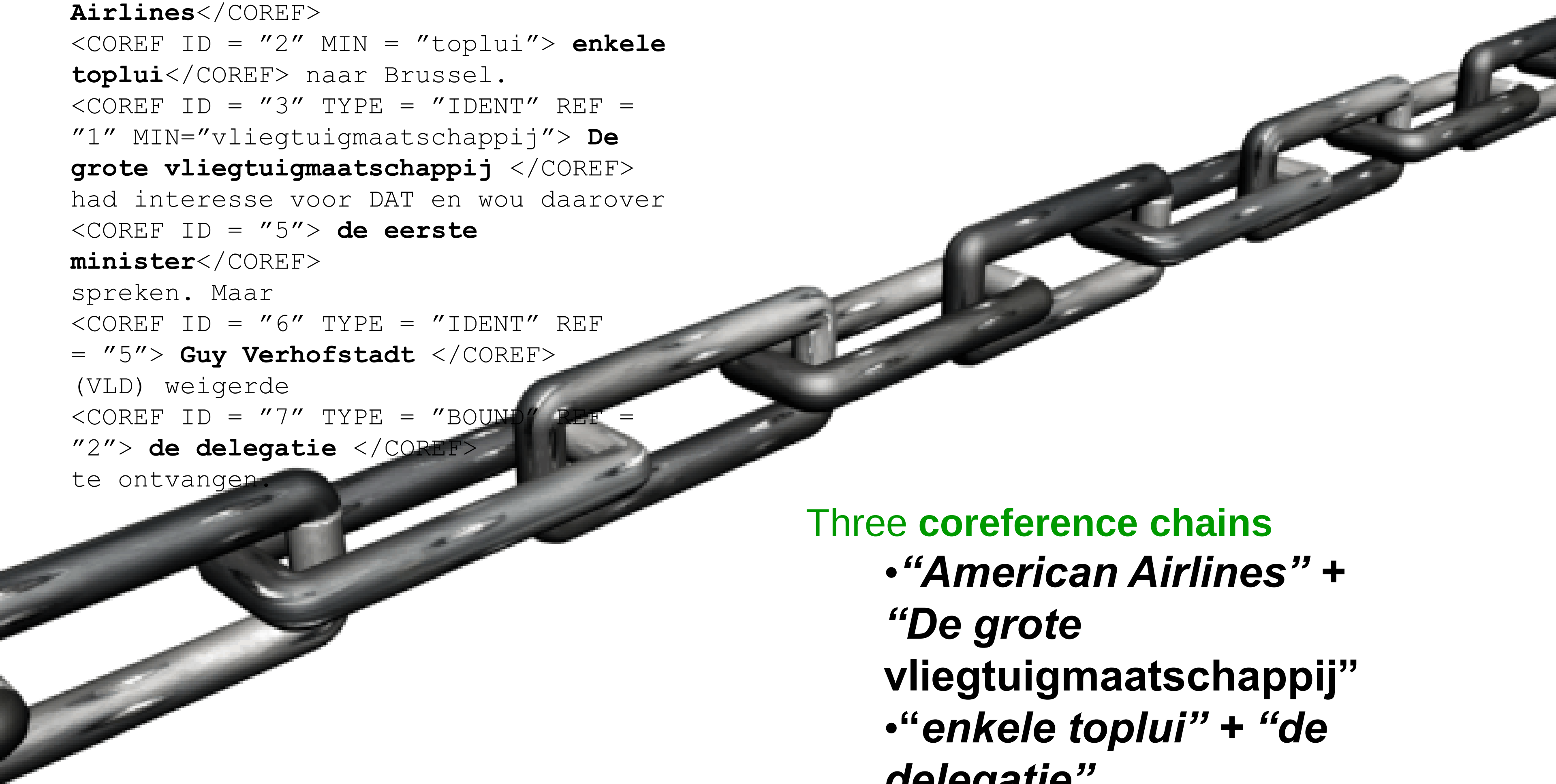
Hoste & Daelemans

Ongeveer een maand geleden stuurde
<COREF ID = ”1”> American Airlines</COREF>
<COREF ID = ”2” MIN = ”toplui”> enkele toplui</COREF>
naar Brussel.
<COREF ID = ”3” TYPE = ”IDENT” REF = ”1”
MIN=”vliegtuigmaatschappij”> De grote
vliegtuigmaatschappij </COREF>
had interesse voor DAT en wou daarover
<COREF ID = ”5”> de eerste minister</COREF>
spreken. Maar
<COREF ID = ”6” TYPE = ”IDENT” REF
= ”5”> Guy Verhofstadt </COREF>
(VLD) weigerde
<COREF ID = ”7” TYPE = ”BOUND” REF = ”2”> de delegatie
</COREF>
te ontvangen.

Hoste & Daelemans

Ongeveer een maand geleden stuurde
<COREF ID = ”1”> American Airlines</COREF>
<COREF ID = ”2” MIN = ”toplui”> **enkele toplui**</COREF>
naar Brussel.
<COREF ID = ”3” TYPE = ”IDENT” REF = ”1”
MIN=”vliegtuigmaatschappij”> De grote
vliegtuigmaatschappij </COREF>
had interesse voor DAT en wou daarover
<COREF ID = ”5”> de eerste minister</COREF>
spreken. Maar
<COREF ID = ”6” TYPE = ”IDENT” REF
= ”5”> Guy Verhofstadt </COREF>
(VLD) weigerde
<COREF ID = ”7” TYPE = ”BOUND” REF = ”2”> de delegatie
</COREF>
te ontvangen.

Ongeveer een maand geleden stuurde
<COREF ID = ″1″> **American Airlines**</COREF>
<COREF ID = ″2″ MIN = ″toplui″> **enkele toplui**</COREF> naar Brussel.
<COREF ID = ″3″ TYPE = ″IDENT″ REF = ″1″ MIN=″vliegtuigmaatschappij″> **De grote vliegtuigmaatschappij** </COREF>
had interesse voor DAT en wou daarover
<COREF ID = ″5″> **de eerste minister**</COREF>
spreken. Maar
<COREF ID = ″6″ TYPE = ″IDENT″ REF = ″5″> **Guy Verhofstadt** </COREF>
(VLD) weigerde
<COREF ID = ″7″ TYPE = ″BOUND″ REF = ″2″> **de delegatie** </COREF>
te ontvangen.

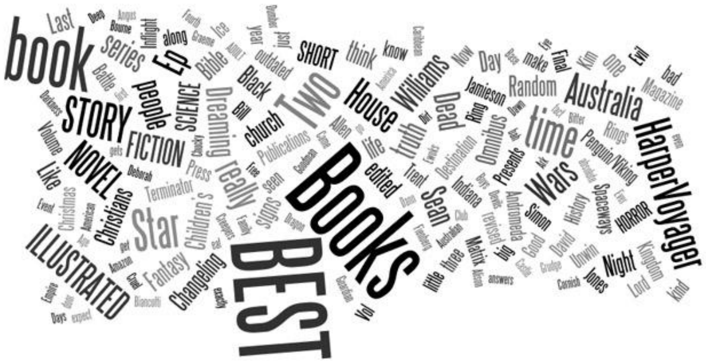


Three **coreference chains**

- ***“American Airlines” + “De grote vliegtuigmaatschappij”***
- ***“enkele toplui” + “de delegatie”***
- ***“de eerste minister” + “Guy Verhofstadt”***

Hoste & Daelemans

- Experiments



25,994 words

50 documents

3,014 coreferential tags

Hoste & Daelemans

- Preprocessing
  - Tokenization
  - Named Entity recognition
  - Part-of-speech tagging
  - Text chunking
  - Relation finding
  - Morphological analysis

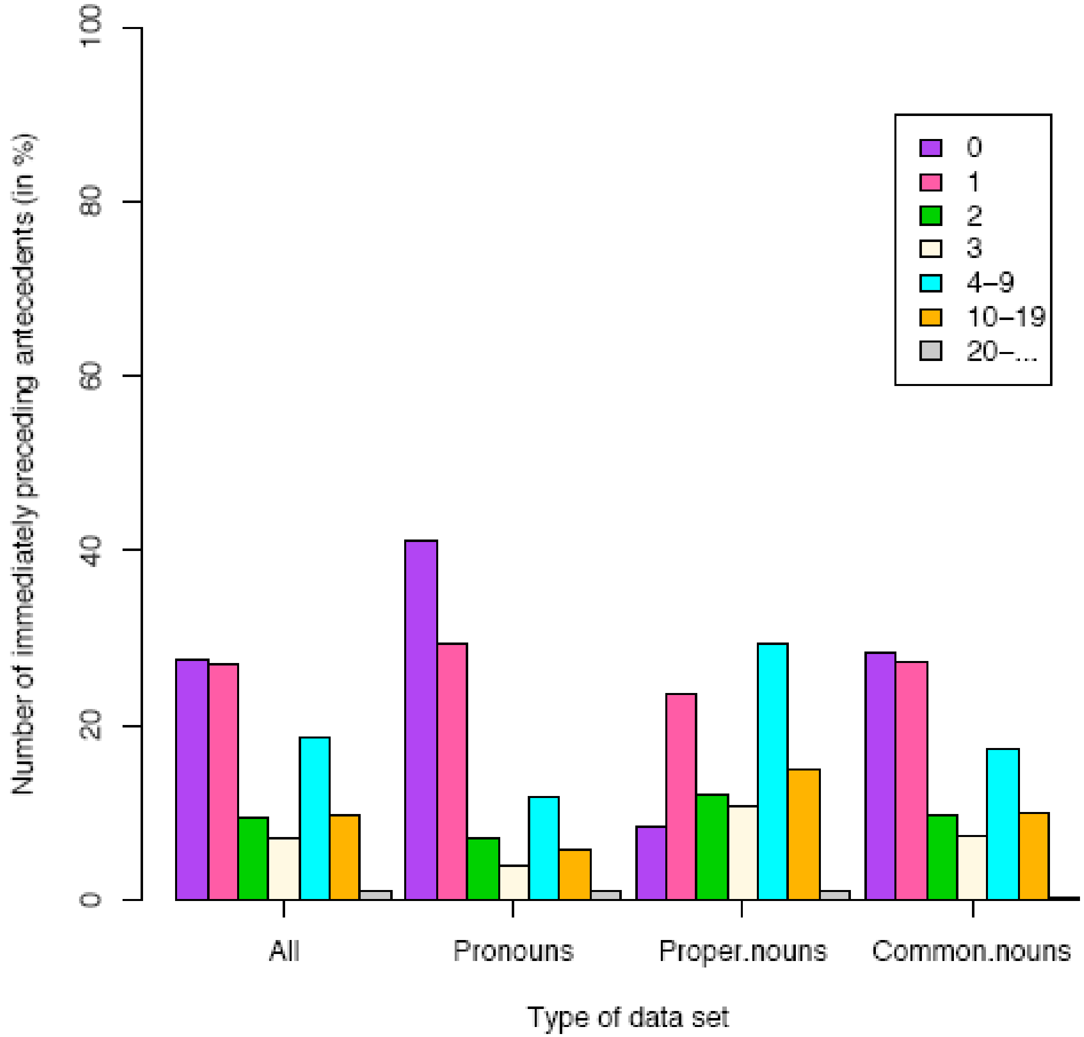
- Creation of positive and negative instances for machine learning

Hoste & Daelemans

Op **9 december 1983** werd **Alfred Heineken** samen met **zijn chauffeur** ontvoerd.
De kidnappers vroegen **43 mijoen gulden losgeld. Een bescheiden bedrag**, vonden **ze** zelf.

ze	een beschieden bedrag	neg
ze	43 mijoen gulden losgeld	neg
ze	**de kidnappers**	**pos**
ze	zijn chauffeur	neg
ze	zijn	neg
ze	Alfred Heineken	neg
ze	9 december 1983	neg

Figure 3: Distance in number of sentences between a given referring expression and its immediately preceding antecedent in the KNACK-2002 training set.



Hoste & Daelemans

- **Pronouns:**
  - all NPs in a context of 2 sentences before pronouns in test set
- **Proper and common nouns:**
  - all partially matching NPs included for non-matching NPs, only two sentences included

Hoste & Daelemans

- Train separate classifiers for each type of NP

(3) Vlaams minister van Mobiliteit Steve Stevaert dreigt met een regeringscrisis als de federale regering blijft weigeren mee te werken aan het verbeteren van de verkeersveiligheid. (...) Stevaert ergert zich aan de manier waarop de verschillende ministeries het dossier naar elkaar toeschuiven.

(4) De beklaagde, die de doodstraf riskeert, wil dat zijn proces op televisie uitgezonden wordt.

Hoste & Daelemans

Table 2: Number of instances per NP type in the KNACK-2002 corpus.

	TRAIN		TEST
NP type	positive	negative	
Pronouns	3,111	33,155	5,897
Proper nouns	2,065	31,370	10,954
Common nouns	1,281	31,394	24,677
Complete	6,457	95,919	41,528

- **Features**

Positional features	DIST_SENT DIST_NP (# NPs inbetween)
Local context features	3 words before and after POS-tag
Morpholoigical features	DEMON, PRON, PROP NUM_AGREE
Syntactic features	ANA_SYNT, ANT_SYNT (subject, object, predicate) APPOSITIVE
String-matching features	COMP-MATCH, PART_MATCH
Semantic features	SYNONYM, HYPONYM, SAME_NE

Hoste & Daelemans

		Prec.	Rec.	$F_{\beta=1}$
Timbl	PPC	65.9	42.2	51.4
	Pronouns	64.9	—	—
	Proper nouns	79.4	—	—
	Common nouns	47.6	—	—
Ripper	PPC	66.3	40.9	50.6
	Pronouns	66.7	—	—
	Proper nouns	79.0	—	—
	Common nouns	47.5	—	—

More than just chains

- Coreference chain identification is important for NLU tasks

- For NLG we have to pay attention to the form of the references
  - Certain referential forms are ruled out in certain contexts

- Referential form also tells us something important about the **salience** of the referent at a particular point in a discourse

Information structure

Referential form choice makes or breaks cohesion

(modified from Gordon 1993)

***1.Susan** gave **Betsy a hamster**.*

***2.She** told **her** to feed **the hamster** well.*

3a**. **Betsy asked her what to feed him.

3b. ???She** asked **Susan** what to feed **him.

•Complex rules govern when you should use a pronoun and when you shouldn´t

•When the dialogue doesn´t follow these rules it creates confusion

Centering Theory

- Centering Theory (Grosz, Joshi, and Weinstein 1995)
- salience concerns how entities are realized in an utterance
  - salience status often reflected in a referent´s grammatical function **and** the linguistic form of its subsequent mentions
  - Salient entities are more likely to be subjects, to appear in the main clause, etc.
  - Pronominalization—is linked to salience
  - the more `underspecified ´ your referring expression is, the more salient the referent of that expression is

Transition can be smooth or rough

- Texts about the same discourse entity more coherent than texts that frequently switch
- CT formalizes fluctuations in topic continuity with **transitions**
- Transitions are ranked,
  - texts with many smooth transitions are deemed more coherent than texts where such transitions are absent or infrequent.

- Forward looking centers
  - An ordered set of entities
  - What could we expect to hear about next
  - Ordered by salience as determined by grammatical function
  - Subject > Indirect object > Object > Others
- John gave the textbook to Mary.
  - C_f = {John, Mary, textbook}
- Preferred center C_p
  - The highest ranked forward looking center
  - High expectation that the next utterance in the segment will be about C_p

- Single backward looking center, C_b (U)
  - For each utterance other than the segment-initial one
- The backward looking center of utterance U_{n+1} connects with one of the forward looking centers of U_n
- C_b (U+1) is the most highly ranked element from C_f (Un) that is also realized in U+1

Centering transitions ordering

	Cb(Un+1)=Cb(Un) OR Cb(Un)=[?]	Cb(Un+1) != Cb(Un)
Cb(Un+1) = Cp (Un+1)	continue	smooth-shift
Cb(Un+1) != Cp (Un+1)	retain	rough-shift

a. Terry really goofs sometimes.
b. Yesterday was a beautiful day and he was excited about trying out his new sailboat.
c. He wanted Tony to join him on a sailing expedition.
d. He called him at 6am.
e. He was sick and furious at being woken up so early.

Centering analysis

- Terry really goofs sometimes.
  - Cf={Terry}, Cb=?, undef
- Yesterday was a beautiful day and he was excited about trying out his new sailboat.
  - Cf={Terry,sailboat}, Cb=Terry, continue
- He wanted Tony to join him in a sailing expedition.
  - Cf={Terry, Tony, expedition}, Cb=Terry, continue
- He called him at 6am.
  - Cf={Terry,Tony}, Cb=Terry, continue

- He called him at 6am.
  - Cf={Terry,Tony}, Cb=Terry, continue
- Tony was sick and furious at being woken up so early.
  - Cf={Tony}, Cb=Tony, smooth shift
- He told Terry to get lost and hung up.
  - Cf={Tony,Terry}, Cb=Tony, continue
- Of course, Terry hadn’t intended to upset Tony.
  - Cf={Terry,Tony}, Cb = Tony, retain

Ranking forward looking centers

This is being empirically investigated

Subject > Indirect object > Object > Others > Quantified indefinite subjects (people, everyone) > Arbitrary plural pronominals

- STRUBE and Hahn: rank by function. argue that that makes more sense for German…
- Poesio

a. Terry really goofs sometimes.
b. Yesterday was a beautiful day and he was excited about trying out his new sailboat.
c. He wanted Tony to join him on a sailing expedition.
d. He called him at 6am.
e. He was sick and furious at being woken up so early.

Centering analysis

- Terry really goofs sometimes.
  - Cf={Terry}, Cb=?, undef
- Yesterday was a beautiful day and he was excited about trying out his new sailboat.
  - Cf={Terry,sailboat}, Cb=Terry, continue
- He wanted Tony to join him in a sailing expedition.
  - Cf={Terry, Tony, expedition}, Cb=Terry, continue
- He called him at 6am.
  - Cf={Terry,Tony}, Cb=Terry, continue

- He called him at 6am.
  - Cf={Terry,Tony}, Cb=Terry, continue
- Tony was sick and furious at being woken up so early.
  - Cf={Tony}, Cb=Tony, smooth shift
- He told Terry to get lost and hung up.
  - Cf={Tony,Terry}, Cb=Tony, continue
- Of course, Terry hadn't intended to upset Tony.
  - Cf={Terry,Tony}, Cb = Tony, retain

Rough shifts in evaluation of writing skills

- One of the graders of student essays in standardized tests is an automatic program

- ETS researchers have developed a number of applications that use natural language processing technologies to evaluate and score the writing abilities of test takers:
  - The *CriterionSM* Online Essay Evaluation Service automatically evaluates essay responses using *e-rater* and the Critique writing analysis tools.

  - *E-rater*® gives holistic scores for essays.

  - *CritiqueTM* provides real-time feedback about grammar, usage, mechanics and style, and organization and development.

  - *C-raterTM* offers automated analysis of conceptual information in short-answer, free responses.

Ranking forward looking centers

- Subject >
- Indirect object >
- Object >
- Others >
- Quantified indefinite subjects (people, everyone) >
- Arbitrary plural pronominals

- STRUBE and Hahn: rank by function. argue that that makes more sense for German…

Entity Grid from Barzilay & Lapata

	Department	Trial	Microsoft	Evidence	Competitors	Markets	Products	Brands	Case	Netscape	Software	Tactics	Government	Suit	Earnings	
1	S	O	S	X	O	–	–	–	–	–	–	–	–	–	–	1
2	–	–	O	–	–	X	S	O	–	–	–	–	–	–	–	2
3	–	–	S	O	–	–	–	–	S	O	O	–	–	–	–	3
4	–	–	S	–	–	–	–	–	–	–	–	S	–	–	–	4
5	–	–	–	–	–	–	–	–	–	–	–	–	S	O	–	5
6	–	X	S	–	–	–	–	–	–	–	–	–	–	–	O	6

Summary

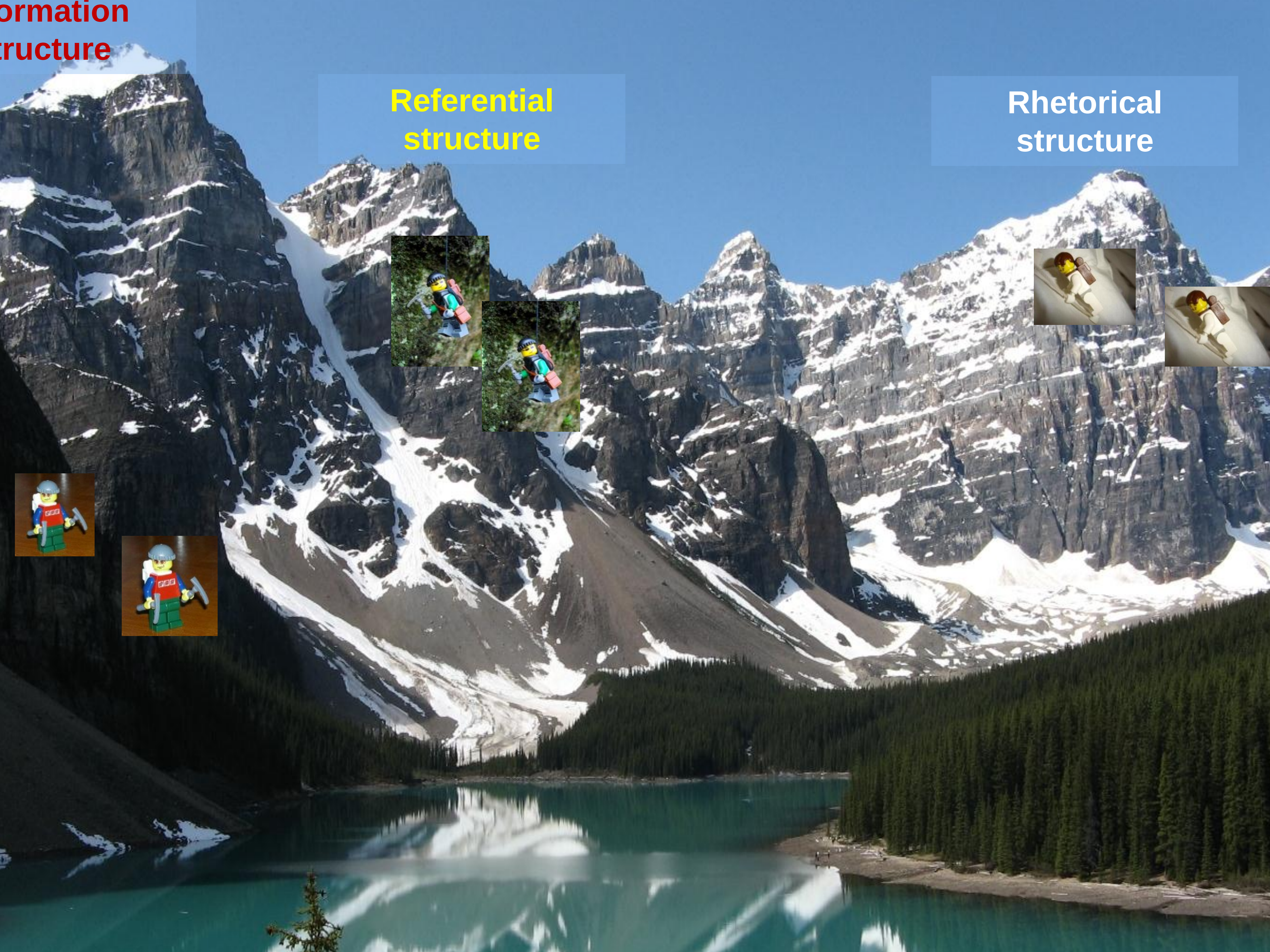
- What should a theories of discourse coherence deal with?
  - coherence relations
  - entity-based coherence
  - information structure
- Coherence relations
  - Hobbs
  - Grosz & Sidner
  - Mann & Thompson and Rhetorical Structure Theory (RST)
  - SDRT
  - PDTB

- What problem are there with coherence theories
  - inventory of relations may be unprinciples
  - very different types of information may be conflated into one format in a framework
  - implicit discourse relations seem to be qualitatively different than explicitly marked ones, yet these are the ones we need to recognize
  - annotation is very difficult
- What is entity-based coherence and how are computational linguistics approaching it?
  - lexical cohesion chains
  - coreference cains
- What about information structure and topics
  - centering theory?

Clearly research on discourse structure is very important, useful work!

Discourse

is a very important topic that more people should be interested in!


ormation
ructure
Referential
structure
Rhetorical
structure