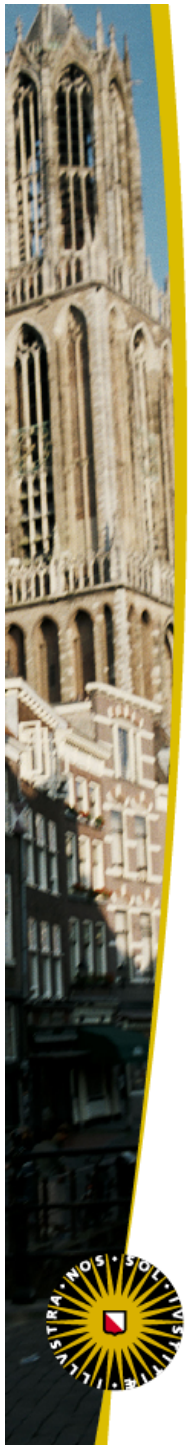
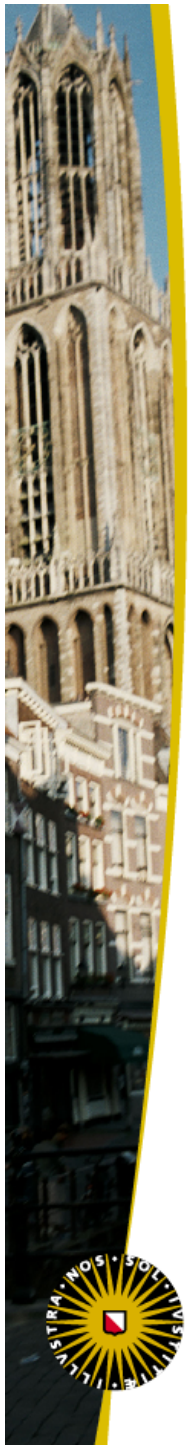


Let's try to make annotation systems communicate – towards a systematic approach of coherence relations

Ted Sanders

Dept of Languages, Literature and Communication
Utrecht institute of Linguistics OTS
Universiteit Utrecht
The Netherlands

Universiteit Utrecht

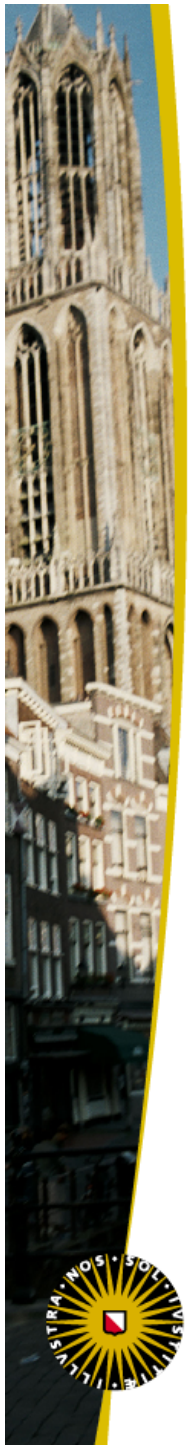


Textlink: goals

- To use and develop viable annotation systems of relations, which are empirically and cognitively sound.

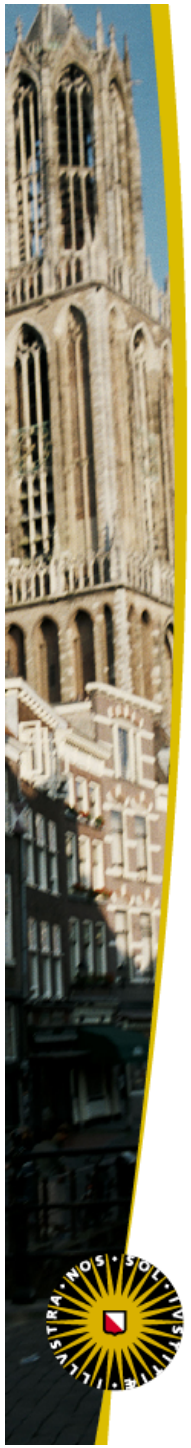
- Goal here today:
- Investigate similarities between different systems
- So that we can converge; make existing systems communicate

- A minimal set, that can be extended, specified
- Is useful in discourse annotation
- Start from abstract categories, then down to specific ones found in corpora



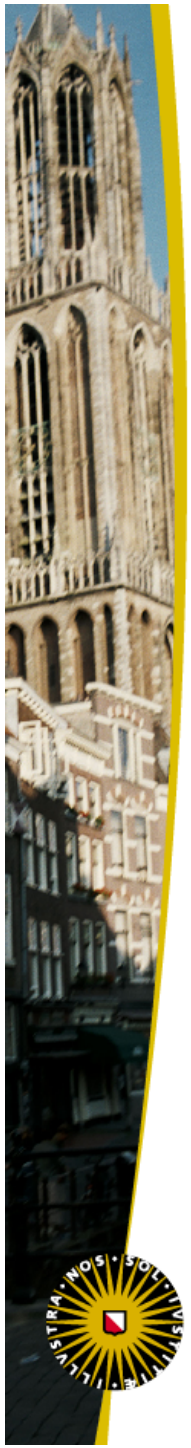
Structure of the talk

- Look at PDTB and RST
- Show some similarities in terms of underlying dimensions
- Illustrate such a minimal annotation scheme
- As it has been used in previous and ongoing research
- Analyze the examples we studied for the workshop



An outstanding example of discourse annotation: PDTB

- Penn Discourse Treebank (Prasad, Webber, Joshi)
- Often used in real corpora and applied to various corpora in many languages
- Theory-neutral approach: does not predict what kind of high-level structures can be created from the low-level annotations of relations.
- Tagset consists of three levels: class, type and subtype



Relations in Penn Discourse Treebank

TEMPORAL
- → Synchronous
- → Asynchronous
  - → precedence
  - → succession

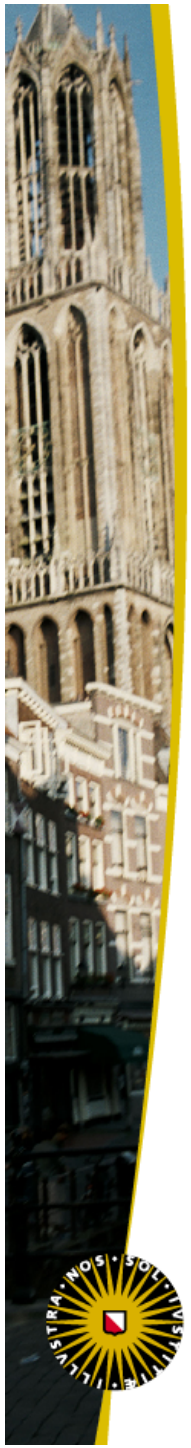
COMPARISON
- → Contrast
  - → juxtaposition
  - → opposition
- → *Pragmatic Contrast*
- → Concession
  - → expectation
  - → contra-expectation
- → *Pragmatic Concession*

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

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

Universiteit Utrecht

Figure 1: Hierarchy of sense tags in Penn Discourse Tree Bank



Discourse annotation of corpora

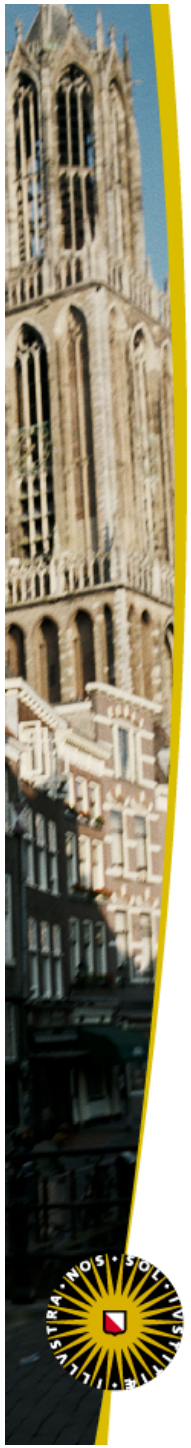
- Conceptually related relations fall in different categories in the scheme.

- For example: contrastive relations that are expressed with ***but*** fall in two totally different classes: *comparison* and *expansion.*
- Issues:

1. This maybe something to avoid for theory/ internal reasons;
2. Such counter-intuitive aspects can be confusing for annotators

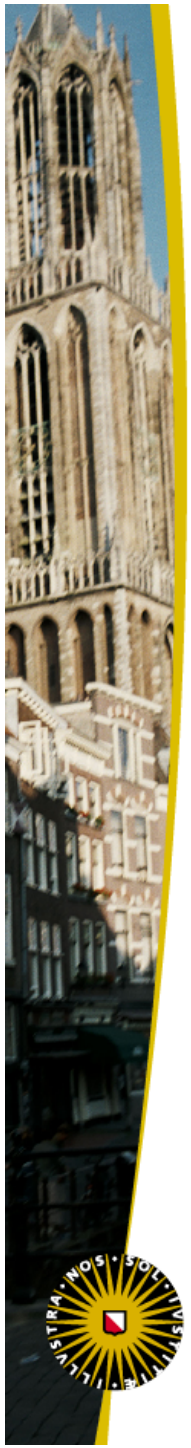
- A more systematically organized set of relations might be theoretically attractive, and useful in discourse annotation.

Universiteit Utrecht



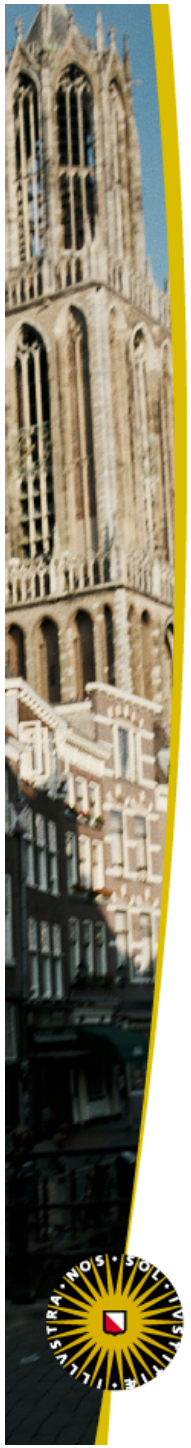
Another outstanding example of discourse annotation: RST

- Rhetorical Structure Theory (Mann & Thompson 1988; Taboada & Mann; Taboada et al.)
- Often used in real corpora and applied to various corpora in many languages
- Top-down approach: texts are ordered hierachically; one span at top level; then further down the tree until adjacent segments.
- Questions can be asked about the exact set (Marcu's is different, etc.) and about its organization



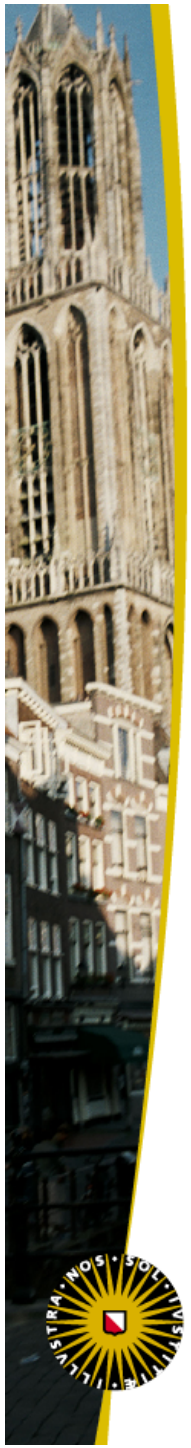
RST relations

Relation name	Nucleus	Satellite
Contrast	One alternate	The other alternate
Antithesis	Ideas favored by the author	Ideas disfavored by the author
Background	Text whose understanding is being facilitated	Text for facilitating understanding
Circumstance	Text expressing the events or ideas occurring in the interpretative context	An interpretive context of situation or time
Concession	Situation affirmed by author	Situation which is apparently inconsistent but also affirmed by the author
Condition	Action or situation whose occurrence results from the occurrence of the conditioning situation	Conditioning situation
Elaboration	Basic information	Additional information
Enablement	An action	Information intended to aid the reader in performing an action
Evaluation	A situation	An evaluative comment about the situation
Evidence	A claim	Information intended to increase the reader's belief in the claim
Interpretation	A situation	An interpretation of the situation
Joint	Unconstrained	Unconstrained
Justify	Text	Information supporting the writer's right to express the text
List	An item	The next item
Motivation	An action	Information intended to increase the reader's desire to perform the action
Non-volitional cause	A situation	Another situation which causes that one, but not by anyone's deliberate action
Non-volitional result	A situation	Another situation which is caused by that one, but not by anyone's deliberate action
Otherwise (anti-conditional)	Action or situation whose occurrence results from the lack of the occurrence of the conditioning situation	Conditioning situation
Preparation	Text to be presented	Text which prepares the reader to expect and interpret the text to be presented
Purpose	An intended situation	The intent behind the situation
Restatement	A situation	A reexpression of the situation
Sequence	An item	A next item
Solutionhood	A situation or method supporting full or partial satisfaction of the need	A question, request, problem or other expressed need
Summary	Text	A short summary of that text
Volitional cause	A situation	Another situation which causes that one, by someone's deliberate action
Volitional result	A situation	Another situation which is caused by that one, by someone's deliberate action



RST relations; a first grouping

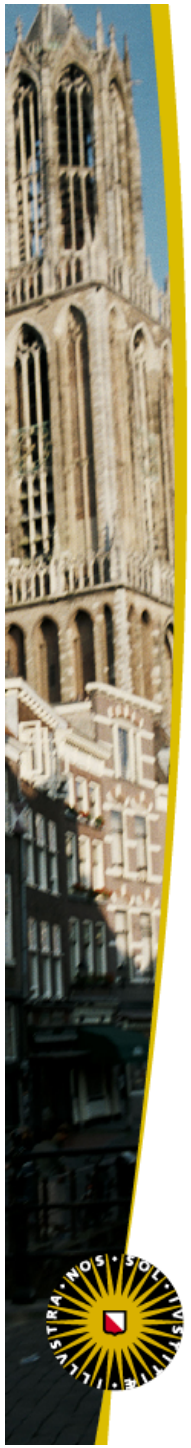
Causal-Conditional	***Contrastive***	***Additive***
Condition	Contrast	Background
Enablement	Antithesis	Circumstance
Evaluation	Concession	Elaboration
Evidence		List
Interpretation		Joint
Justify		Preparation
Motivation		Restatement
Non-volitional cause		Sequence
Non-volitional result		Summary
Otherwise		
Purpose		
Solutionhood		
Volitional cause		
Volitional result		



Can we identify dimensions common to such relation sets ?

- Characteristics shared by all relations:
  - Positive – Negative (Polarity)
  - Additive – Temporal – Causal / Conditional (Basic Operation)
  - Subjective – Objective (Source of Coherence)
  - Basic - Non-basic Order (Order)

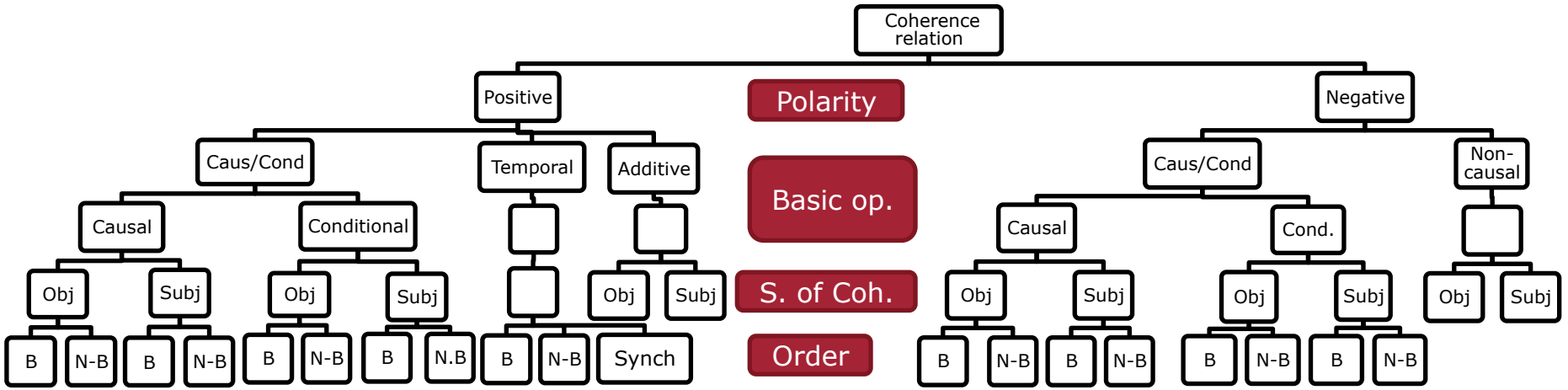
- **These are not all criteria, just shared ones**

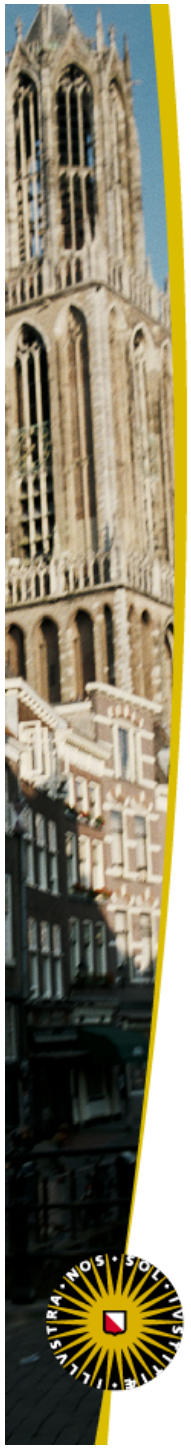


Four dimensions common to all relations

- Subjective – Objective Source of Coherence:
  - Pragmatic vs Semantic
  - Presentational vs content
  - Speech act – epistemic - content
- Objective: events, facts in the world, versus Subjective: speaker / Subject of Consciousness is arguing, reasoning or explaining a speech act
- Basic - Non-basic Order (Order)
  - Antecedens – consequens (P, Q) or reverse

Taxonomy, organized by four categories of relational characteristics



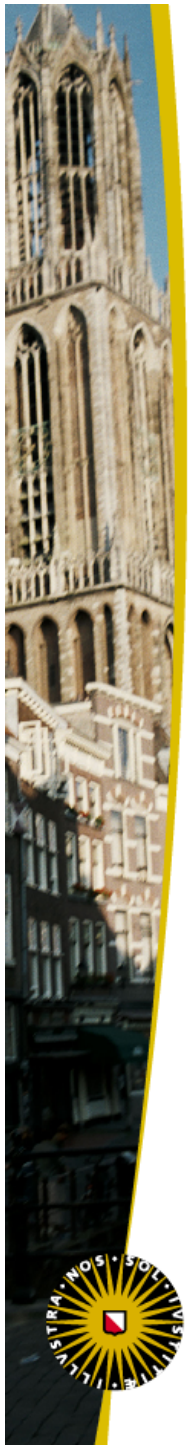


Some examples

- There was a lot of rain. Later, storms came in.
- S1. *Later* S2: positive temporal objective

- She got wet because it rained
- S1 *because* S2: positive, causal, objective, non-basic (Q, P)

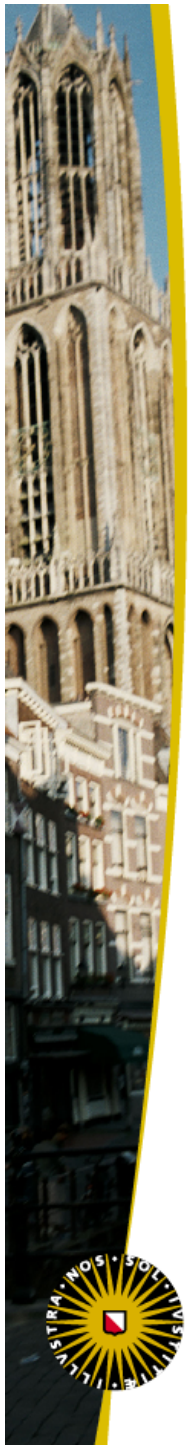
- Although he worked hard, he failed the exam
- *Although* s1, s2: negative, causal, objective, basic (P,Q)



Some more examples

- Something must have come up, because he is never late
- S1 *because* S2: positive, causal, SUBJECTIVE (epistemic), non-basic (Q, P)
- My claim / conclusion is, the argument is..

- Does anyone need to go to bathroom? We're leaving in a minute.
- S1 (*because)* S2: positive, causal, SUBJECTIVE (speech act), non-basic (Q, P)
- I am asking / inviting you to, and the reason for this that…



A paraphrase test for Source of Coherence

- Subjective-Objective distinction

She got wet because it rained

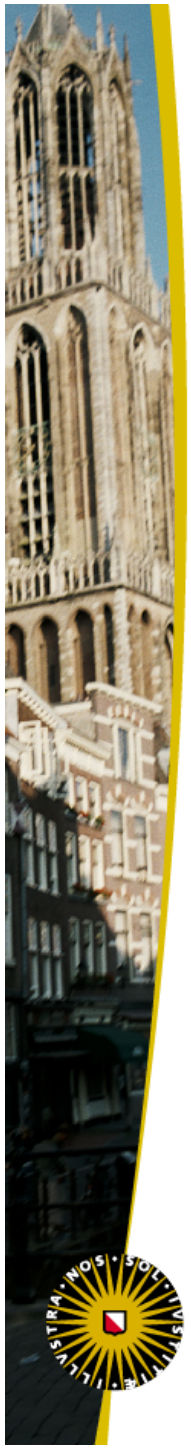
S1 *because* S2: positive, causal, objective, non-basic (Q, P)

- The fact that P causes / leads to the fact / situation that Q
- The fact that it rained leads to the situation she got wet

Something must have come up, because he is never late

S1 *because* S2: positive, causal, SUBJECTIVE (epistemic), non-basic (Q, P)

- The fact that P leads to my conclusion that Q
- ## The fact that P causes / leads to the fact / situation that Q



A paraphrase test

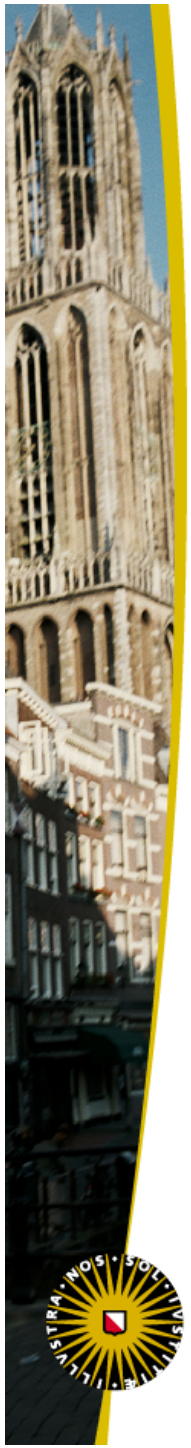
- Subjective-Objective distinction

Does anyone need to go to bathroom? We're leaving in a minute.

S1 (*because)* S2: positive, causal, SUBJECTIVE (speech act), non-basic (Q, P)

The fact that P causes / leads to me saying Q

The fact that P causes / leads to the fact / situation that Q



What is this proposal based on ?

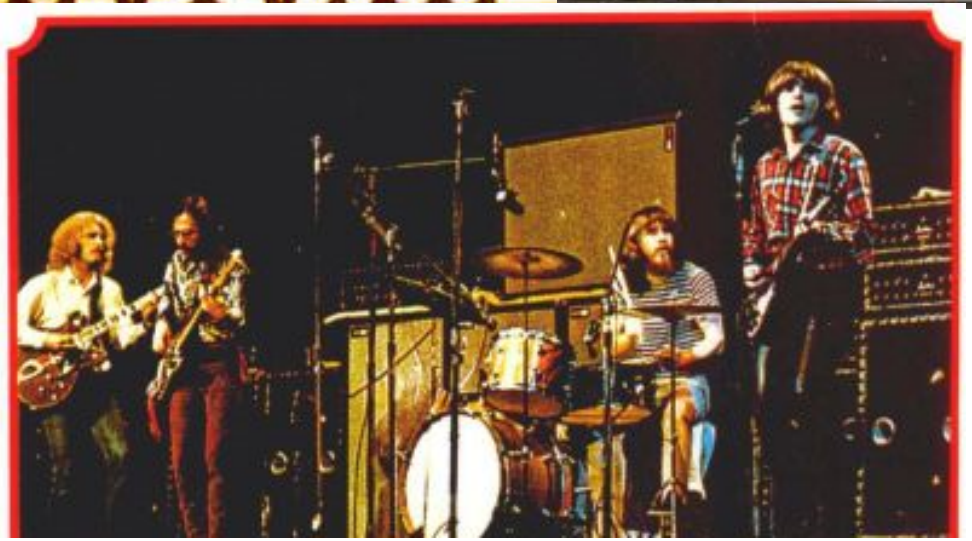
- An analysis of what systems have in common:
- RST- and PDTB-relations map onto these dimensions

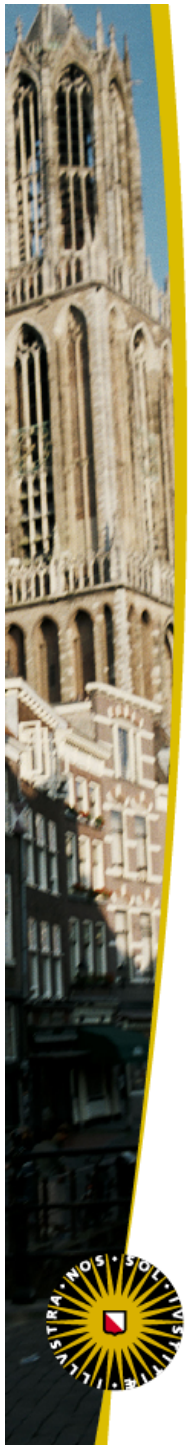
- There is evidence for the relevance of the basic categories from empirical research:

  - Cross-linguistic comparison
  - Acquisition
  - Processing

- Goes back on Sanders, Spooren & Noordman (1992, 93), and elsewhere up to Sanders & Spooren 2015
- ….

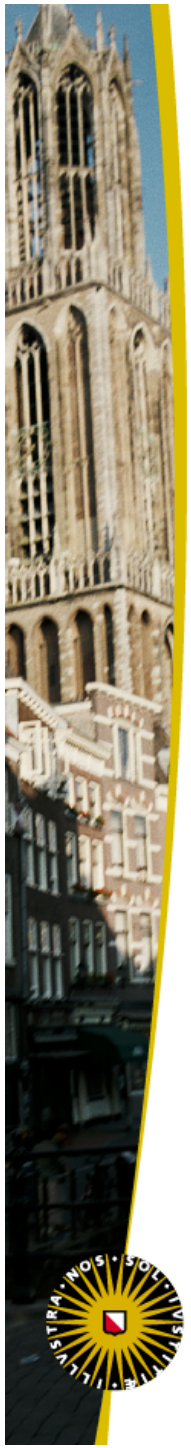
Based on a Cognitive approach to Coherence relations (CCR)





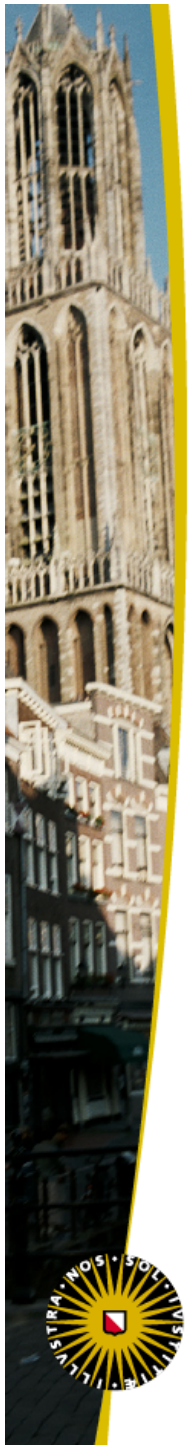
Implications for discourse annotation

1. Systematic: cross-classification defines relations. Conceptually related relations fall in the same categories
2. Claim: all possible relations can be described in these terms. We did that for RST and PDTB.
3. A systematically organized set of relations is useful in discourse annotation: similar steps for each coherence relation might be easier to annotate (Scholman, Evers-Vermeul & Sanders, submitted)



CCR – RST mapping

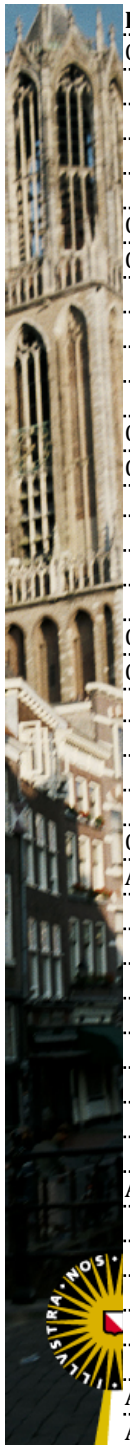
Basic op.	Source of coh.	Order	Polarity	CCR Relation	Additional criteria	RST Relation
Causal	Objective	Basic	Positive	Cause-consequence	+volitional	Volitional cause/result
					-volitional	Non-volitional cause/result
				Condition-consequence		Condition
Causal	Objective	Basic	Negative	Contrastive cause-consequence		Contrast
Causal	Objective	Non-basic	Positive	Consequence-cause	+volitional	Volitional cause/result
					-volitional	Non-volitional cause/result
				Consequence-condition		Condition
Causal	Objective	Non-basic	Negative	Contrastive consequence-cause		Contrast
Causal	Subjective	Basic	Positive	Argument-claim	+evaluation	Evaluation
					-evaluation	Interpretation
				Condition-claim		Condition
Causal	Subjective	Basic	Negative	Contrastive argument-claim		Anti-thesis
Causal	Subjective	Non-basic	Positive	Claim-argument	Content claim	Evidence
						Justify
						Motivation
				Claim-condition		Condition
Causal	Subjective	Non-basic	Negative	Contrastive claim-argument		Anti-thesis
Additive	Objective	-	Positive	List	+temp order	Sequence
					-temp order -specification	Joint
					-temp order +specification	Elaboration
						Restatement
						Summary
						Circumstance
						Background
Additive	Objective	-	Negative	Opposition		Contrast
				Exception		Contrast
Additive	Subjective	-	Positive	Enumeration		Presentational sequence
Additive	Subjective	-	Negative	Concession		Concession



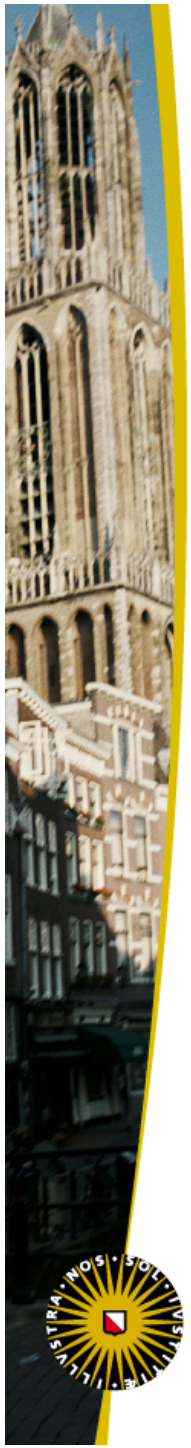
CCR-RST mapping; some highlights

- Causes and results together: Objective causals
- Systematically different from Subjective causals: Claim-Argument:
- Evidence, Justification, Motivation

- Additional criteria needed to distinguish between these three.
- Or between volitional and non-volitional result and cause
- Additives like specification and restatement



Basic op.	S. of coh.	Order	Polarity	CCR Relation	Additional criteria	PDTB Type and Subtype
Causal	objective	Basic	Positive	Cause-consequence		Cause - result
				Condition-consequence	+one time event	Condition - hypothetical
					-one time event	Condition - general
					+present	Condition - factual present
					-present	Condition - factual past
Causal	objective	Basic	Negative	Contrastive cause-consequence		Concession - expectation
Causal	objective	Non-basic	Positive	Consequence-cause		Cause - reason
				Consequence-condition	+one time event	Condition - hypothetical
					-one time event	Condition - general
					+present	Condition - factual present
					-present	Condition - factual past
Causal	objective	Non-basic	Negative	Contrastive consequence-cause		Concession - contra-expectation
Causal	subjective	Basic	Positive	Argument-claim		subjective cause - justification
				Condition-claim	+implicit assertion	Pragmatic condition - implicit assertion
					-implicit assertion	Pragmatic condition - relevance
					+present	Condition - unreal present
					-present	Condition - unreal past
Causal	subjective	Basic	Negative	Contrastive argument-claim		Pragmatic contrast
Causal	subjective	Non-basic	Positive	Claim-argument		Pragmatic cause - justification
				Claim-condition	+implicit assertion	Pragmatic condition - implicit assertion
					-implicit assertion	Pragmatic condition - relevance
					+present	Condition - unreal present
					-present	Condition - unreal past
Causal	subjective	Non-basic	Negative	Contrastive claim-argument		Pragmatic contrast
Additive	objective	-	Positive	List		List
					temp basic order	Asynchronous - succession
					temp nonbasic order	Asynchronous - precedence
						Synchronous
						Instantiation
					-different perspectives	Restatement - specification
					+different perspectives	Restatement - equivalence
						Restatement - generalization
						Instatiation
Additive	objective	-	Negative	Opposition	-gradable scale	Contrast - opposition
					+gradable scale	Contrast - juxtaposition
				Exception		Exception
					+both hold	Alternative - conjunctive
					-both hold	Alternative - disjunctive
					-both hold	Chosen alternative
Additive	subjective	-	Positive	Enumeration		Conjunction
Additive	subjective	-	Negative	Concession		Pragmatic concession



CCR-PDTB mapping; some highlights

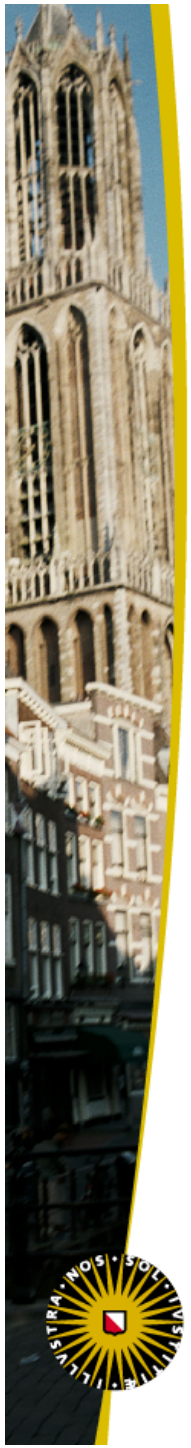
- Contrastives (but) together

- Makes sense for existing distinctions:
  - Cause – Pragmatic cause
  - Contrast – Pragmatic contrast

- Temporals,
- Causal-Conditional remains together

- Additional criteria needed for
- Further distinctions in conditionals (hypothetical-factual-general)
- Additives like specification and restatement

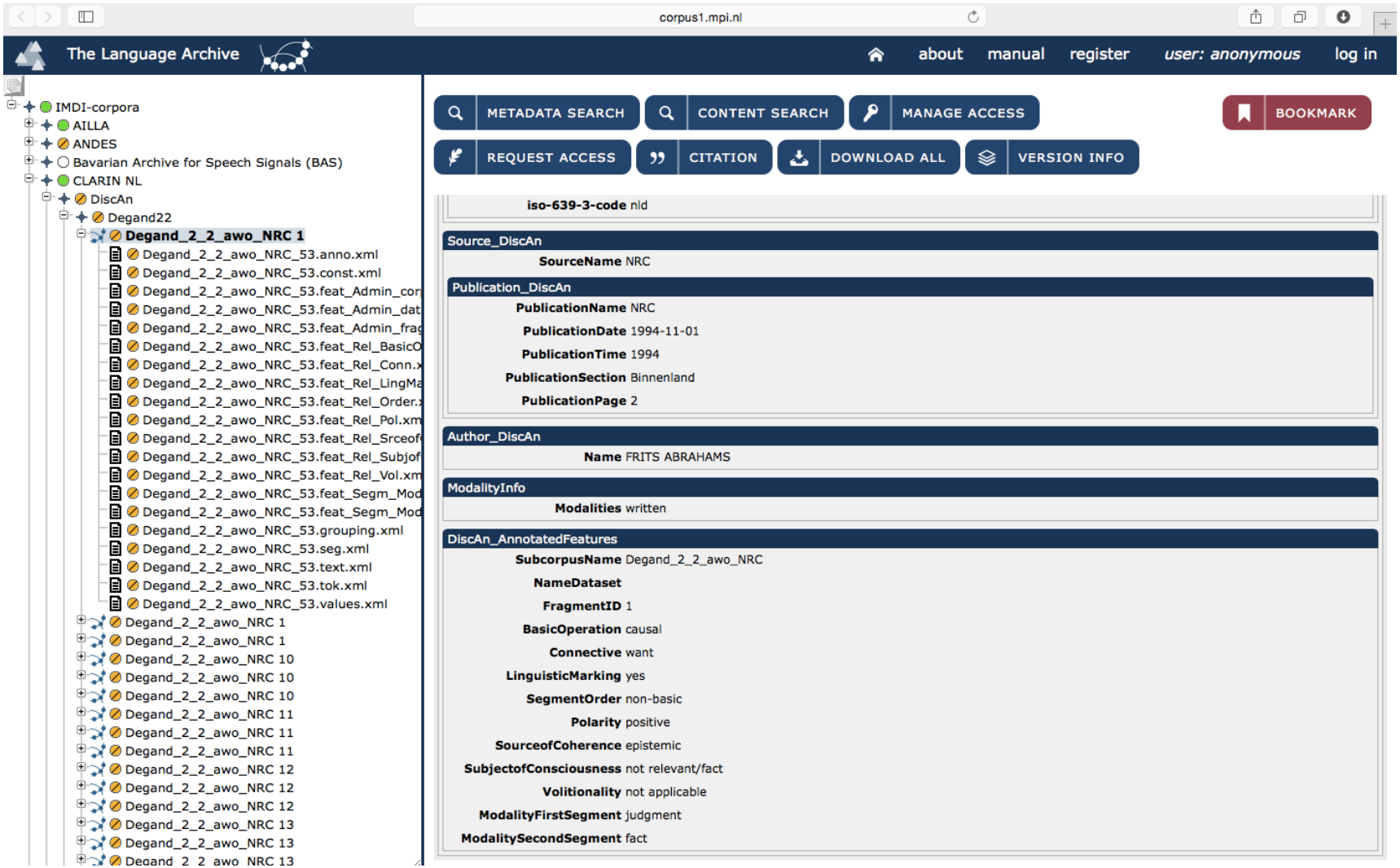
Universiteit Utrecht



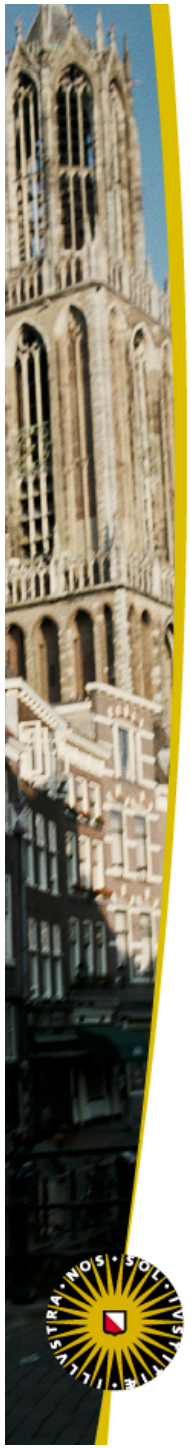
Applications of these ideas in concrete discourse annotation

- 1.DiscAn
- 2. Hoek & Zufferey
- 3. examples for this workshop

DiscAn corpus – example



Universiteit Utrecht



DiscAn annotation – example fragment

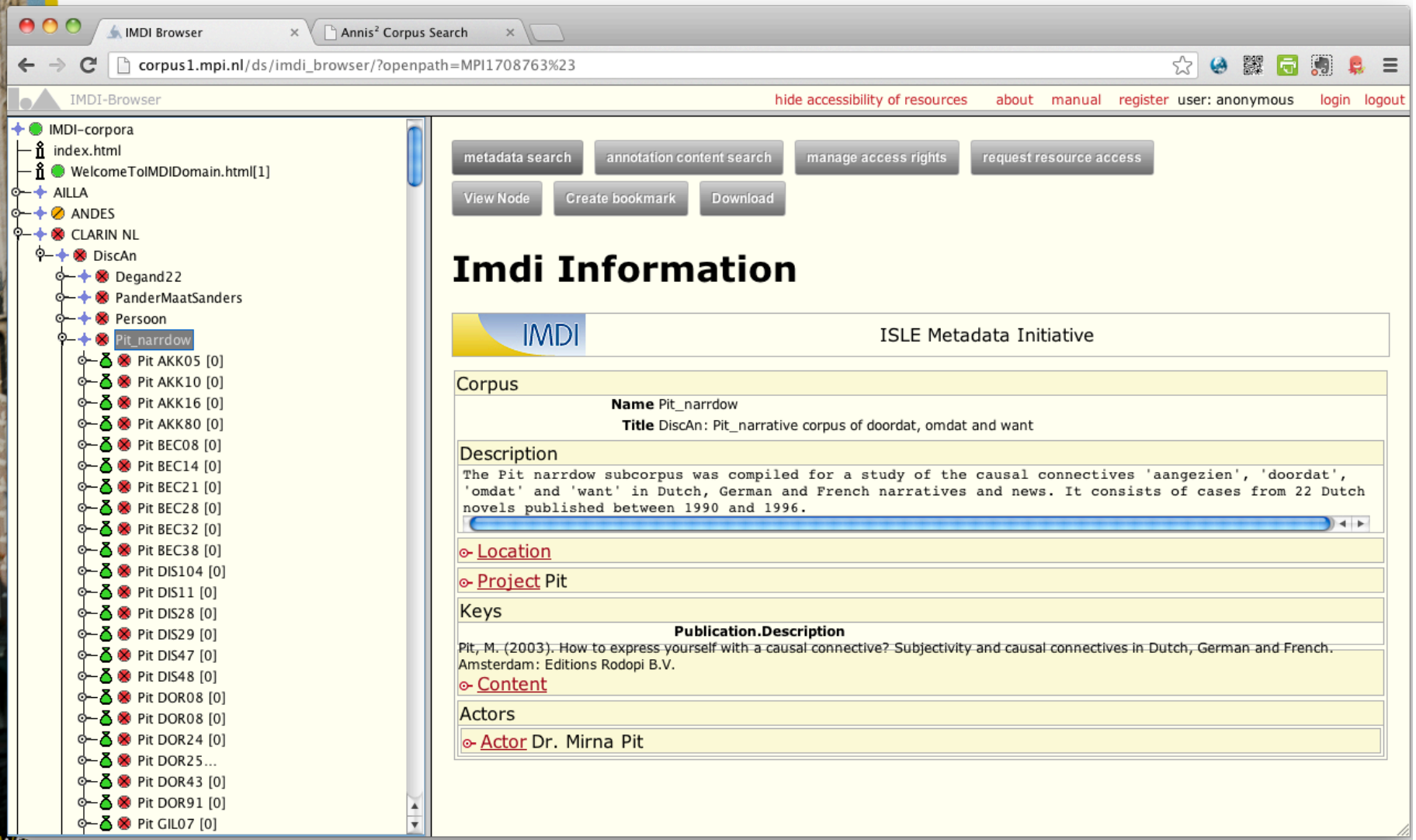
- De atletiekunie was gedwongen om uit te wijken naar België, omdat er geen accommodatie beschikbaar was in Nederland.

(The athletics union was forced to emigrate to Belgium, because there was no accommodation available in the Netherlands.)

Annotation:

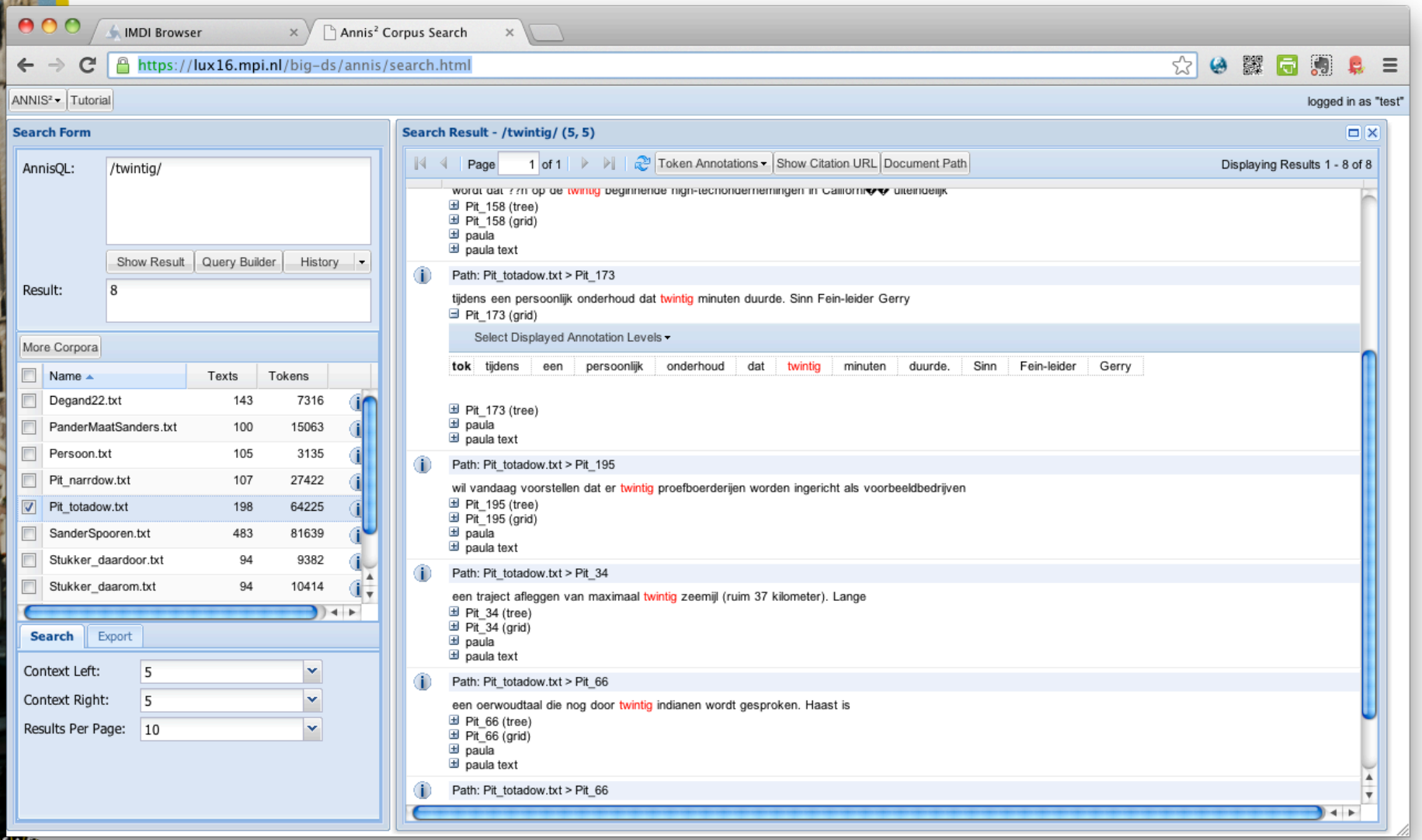
Polarity	Positive
Basic operation	Causal
S. of Coh.	Objective
Order	Non-basic
Volitionality	Non-volitional
Subj. of Consciousness	Not relevant
Linguistic marker	omdat (because)

DiscAn corpus - (C)IMDI view

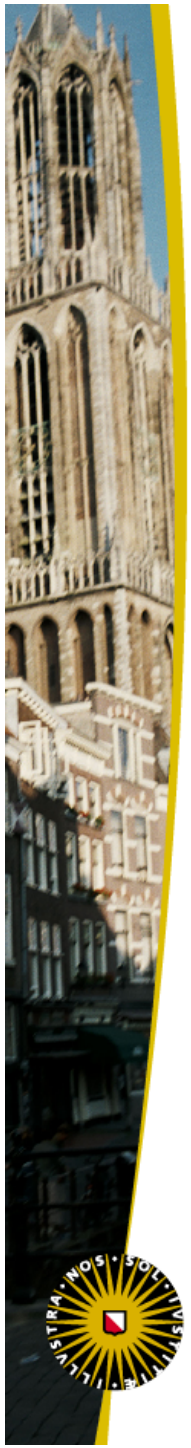


Universiteit Utrecht

DiscAn corpus – ANNIS view



Universiteit Utrecht



Hoek & Zufferey (2015): Parallel corpus study on translation

- Implicit relations
  - Explicit: I went to the party *because* it seemed fun.
  - Implicit: I went to the party. It seemed fun.

1 Source L: English

4 Target Ls: Dutch, German, French, Spanish

- RQs:
- Which relations can be (easily) expressed without a connective, which ones cannot?
- What factors influence the implicitness vs. explicitness of a relation?

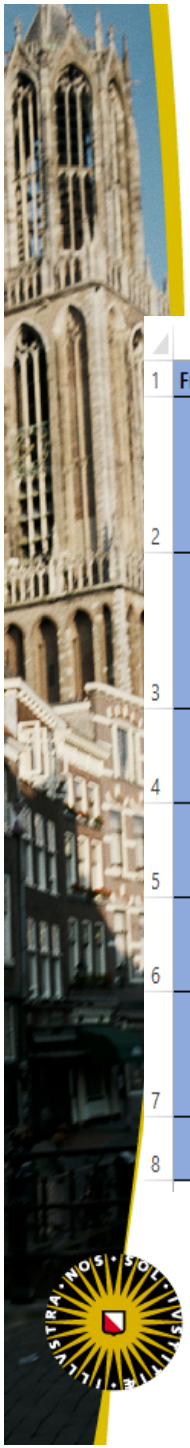
→ *When do you 'need' a connective?*

Manual annotation of relations in SL, **translation spotting** in TLs.

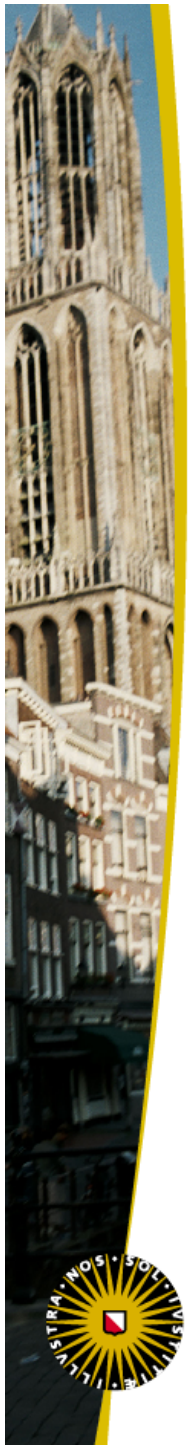
Universiteit Utrecht

Hoek & Zufferey (2015), example of annotation

	A	B	C	D	E	F	G	H	I
1	Fragment nr	EP nr	Fragment	Polarity	Basic operation	Source of Coherence	Order		
2	1	ep-00-01-18 65	[(This poses enormous challenges for competition policy,) which I hope it will be able to meet] [[because]], [for certain, many of those mergers are going to be designed to protect profit margins of the businesses from competition rather than merely to enhance productivity and make those businesses able to operate on a bigger scale.]	positive	causal	subjective	basic		
3	2	ep-00-01-18 789	Can I first thank Mr van Hulten for this report. It is an excellent report. [It would have been wrong for Parliament to have put in willy nilly every single recommendation that came from an external body] [[because]] Parliament should have its own opinion on these issues.] It is right for us to have a focused report which is what Mr van Hulten has produced.	positive	causal	subjective	basic		
4	3	ep-00-02-02 138	[Only the Amsterdam Treaty lasted a year and a half] and that was [[because]] [everyone knew you had to wait for the results of the British election if you were going to have any outcome from that IGC so that was a different reason.]	positive	causal	objective	basic		
5	4	ep-00-02-02 265	[If there is more symbolism than reality in what we can achieve at Community level, this is a pity] [[because]] [there is a lot of enthusiasm at local level for action, including Community actions, on energy saving.]	positive	causal	subjective	basic		
6	5	ep-00-02-14 35	Madam President, it is in itself an achievement that we are having this debate on the new URBAN Community initiative and [it is an achievement that I am here tonight] [[because]] [Air France cancelled my flight at 2.10 p.m.] but I am here!	positive	causal	subjective	basic		
7	6	ep-00-02-15 18	The challenge of a common foreign security and defence policy is very political. [[Because]] [it is political] [it is more complex.] Because it is political it is more sovereignty-sensitive. Because it is more sovereignty-sensitive it is more voter-and citizen-sensitive.	positive	causal	objective	non-basic		
8	7	ep-00-02-15 45	However, [we in the Liberal Group welcome this] [[because]] [it means we are now settling down to business.]	objective	causal	objective	basic		



Universiteit Utrecht

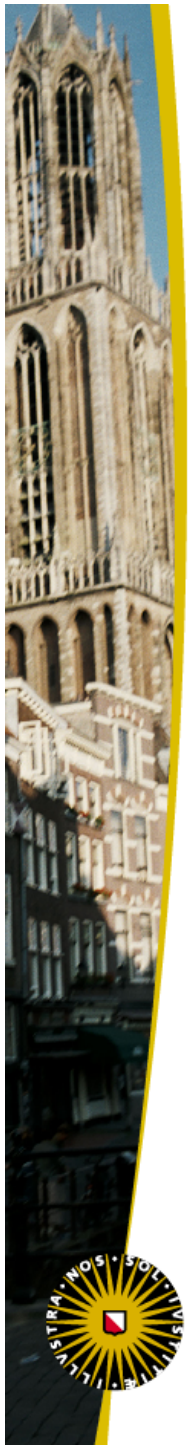


Interannotator agreement,example

100 causal relations

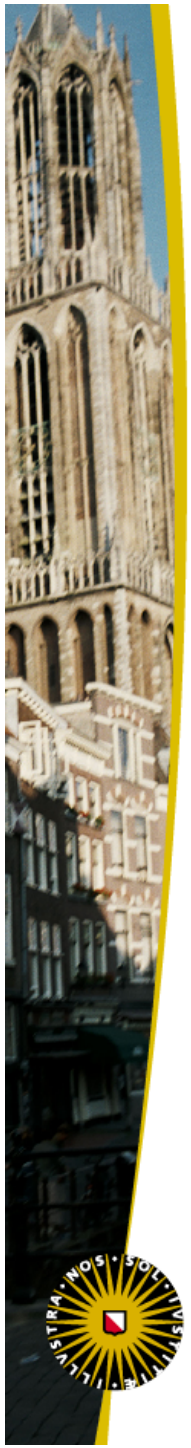
κ = 0.66 **before** discussion

All disagreements on *source of coherence* (objective/ subjective); agreed after discussion, sometimes including a third judge.



Type of results

- Positive additive & positive causal more often implicit than negative and conditional
- Negative more often implicit than conditional
- Positive additive more often implicit than positive causal
- Cf also Asr & Demberg (2012)



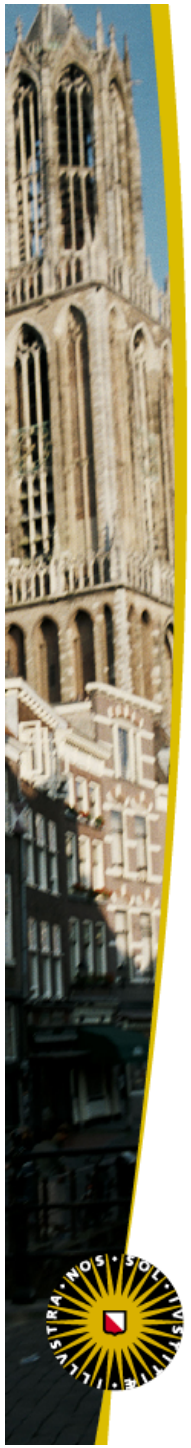
Finally: Examples for this workshop

Example 1:

The door slammed because there is strong wind outside.

- RST: Non-Volitional cause / Explanation ?
- PDTB: (CONTINGENCY.Cause.) reason ?

- CCR: positive, causal, objective, non-basic (Consequence-cause)



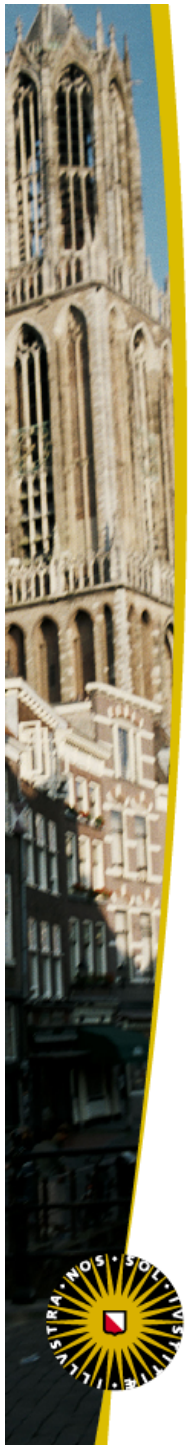
Examples for this workshop

Example 2:

Max is a very good skier, because he won the competition twice last year.

- RST: Evidence ?
- PDTB:(CONTINGENCY.) Pragmatic cause / justification ?

- CCR: Positive, causal, subjective, non-basic (Claim-argument)

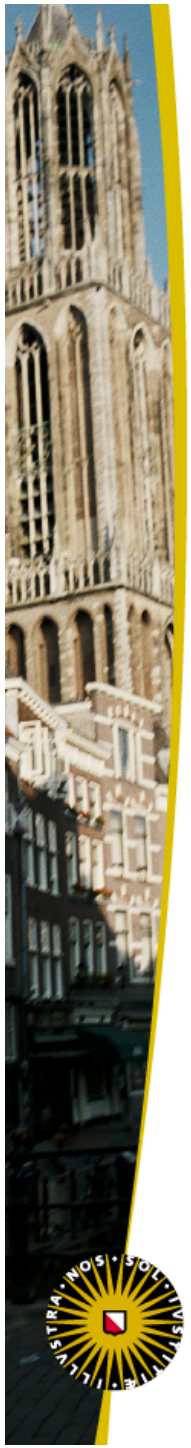


Example sentences

Example 3:

John is tall but Fred is small.

- RST: Contrast ?
- PDTB: (COMPARISON.Contrast.)opposition ?
- CCR: Negative, additive, objective (opposition)



Example sentences

Example 4:

Jane married Mark even though she does not love him.

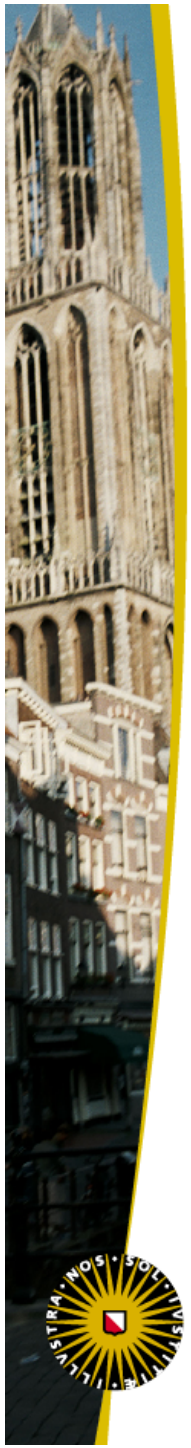
- RST: Antithesis ?
- PDTB: (COMPARISON.Concession.) contra-expectation **?**
- CCR: negative, causal, objective, non-basic (Contrastive consequence-cause)

NOT a concession (Lakoff, 1971, Spooren, 1989):

Should we buy the house? It has a great view, but it is expensive

One argument in favor, one argument against.

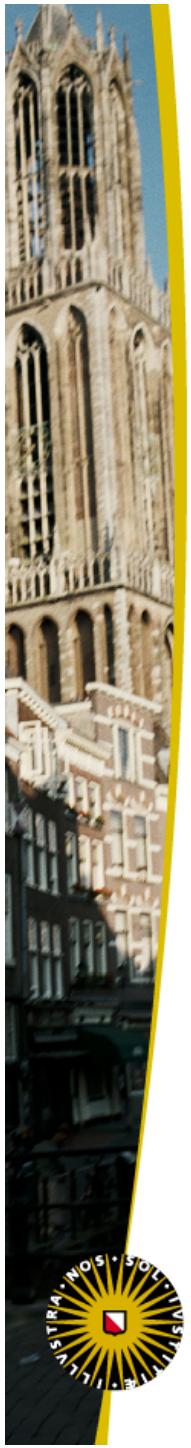
PDTB: pragmatic contrast?



Finally, during this workshop

- It is worthwhile to find out
- Whether we can agree on analyses of examples
- Whether we can see that systems indeed communicate
- For instance via CCR-like dimensions
- See which additional criteria are needed

- Challenges: contrastives
- Further and more precise definitions



Joint work

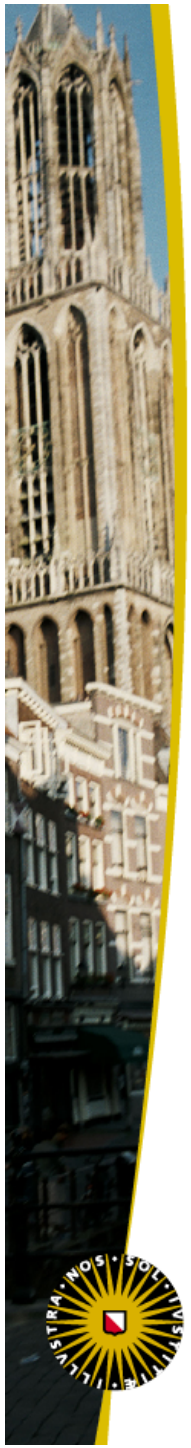
Utrecht team

- Jacqueline Evers-Vermeul
- Merel Scholman
- Jet Hoek
- Martin Groen

- Sandrine Zufferey (Fribourg)
- José Sanders (Nijmegen)
- Wilbert Spooren (Nijmegen)
- Eve Sweetser (Berkeley)

- Discussions with Fatemeh Asr, Vera Demberg

Universiteit Utrecht

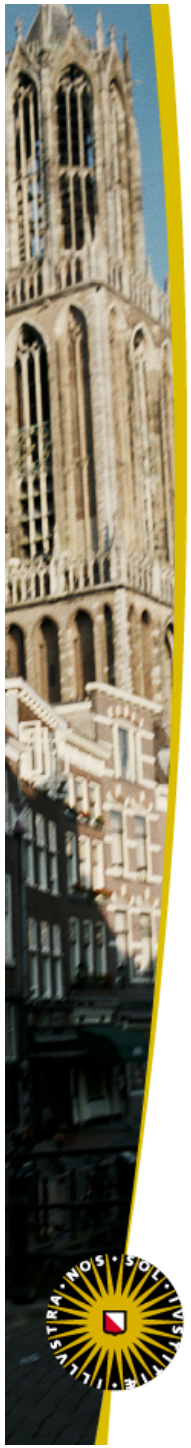


Let's try to make annotation systems communicate – towards a systematic approach of coherence relations

Ted Sanders

Dept of Languages, Literature and Communication
Utrecht institute of Linguistics OTS
Universiteit Utrecht
The Netherlands

Universiteit Utrecht



Possible advantages in annotation

- Allows for substitution and paraphrase tests to be used (Knott & Dale, 1994; Knott & Sanders, 1998; Pander Maat 1994, 1998; Pander Maat & Sanders, 1994)

- Substitution tests:
  - Connectives signal certain types of relations
  - E.g.: *because* signals a causal relation, *meanwhile* a temporal relation and *but* a negative relation.
  - Substitution tests can test the semantic intuitions and thus guide an annotator
  - "Can you connect the two segments with a *but* ?"

- Paraphrase tests:
  - Restate the meaning of the segments in a simpler form
  - E.g.: 'segment 1 presents the cause; segment 2 presents the consequence' OR 'segment 1 presents the consequence, segment 2 presents the cause'
  - Subjective-Objective distinction