Categories of coherence relations in discourse annotation

Ted Sanders & Merel Scholman

Utrecht institute of Linguistics
Universiteit Utrecht

Supported by
Clarin-NL *DiscAn*-project

Universiteit Utrecht

Discourse annotation of corpora

- In search of a system for annotating discourse relations
- Inspired by seminal work on RST, Penn Discourse Treebank (Prasad, et al.), Potsdam Corpus (Stede et al.)

- Our own small context
- DiscAn: Clarin-NL project, with two main goals:
  1. develop a system that is useful for future annotations
  2. Standardize existing data on connectives in Dutch

- In this talk, a proposal: MINIMAL SET of characteristics
  - Today: focus on relational characteristics
  - Based on natural categories of relations
  - A first test of usefulness in annotation

Discourse annotation of corpora

- An inspiring example
- Which is often used in real corpora
- Across languages:
- PDTB

Relations in Penn Discourse Treebank (Prasad, Joshi, Webber)

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

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

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

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

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

Discourse annotation of corpora

RQ:

- Can we come up with a similar, but systematically organized set?
- Is such a set useful in annotation?

A minimal set:
Four fundamental characteristics

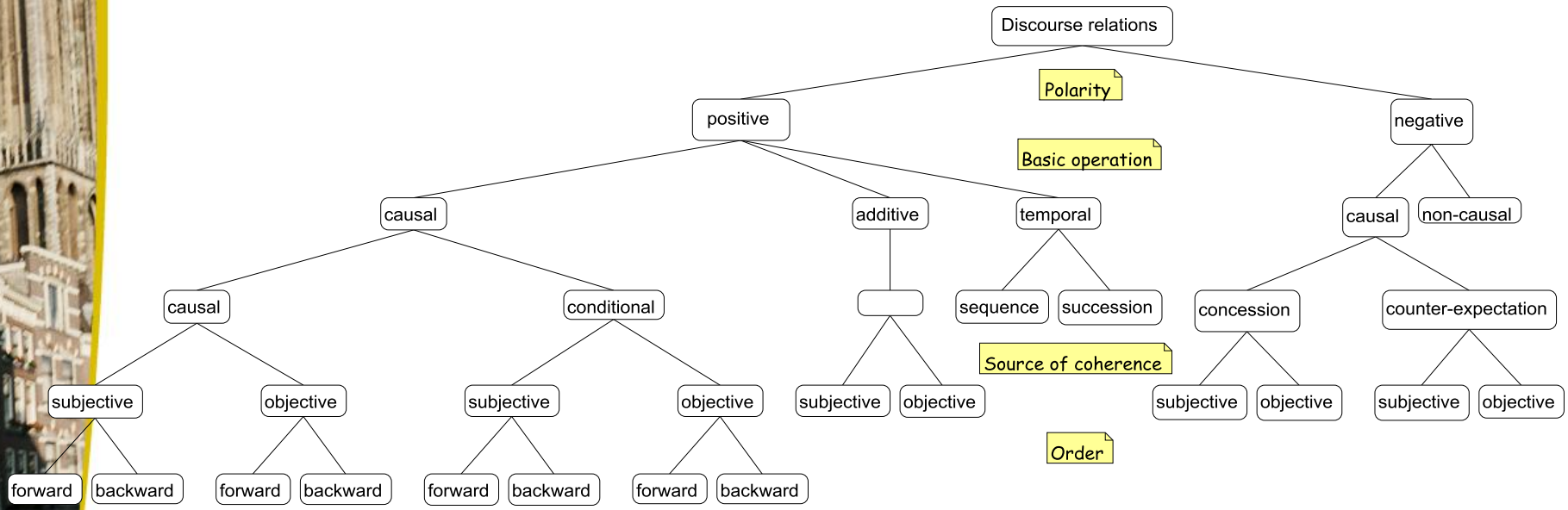
- Based on a Cognitive approach to Coherence relations
  (Sanders, Spooren, Noordman 92, 93 and elsewhere)

1. Polarity positive - negative
2. Basic operation additive – temporal - causal
3. Source of Coherence objective (content/semantic) – subjective (epistemic – speech act / pragmatic)
4. Order forward, backward (p, q or q, p)

Categories of relations

- The combination of these 4 categories produces a classification scheme, in which all coherence relations 'fit'.
- A taxonomy, showing relations among relations:
  - Conclusion (*so*) is subjective positive causal in forward direction
  - Claim-argument (*since*) is subjective positive causal backward
  - Concession is negative causal: *although*
  - Temporal sequence (*and then*) is positive additive, but ordered in time

Taxonomy, organized by a minimal set of relational characteristics

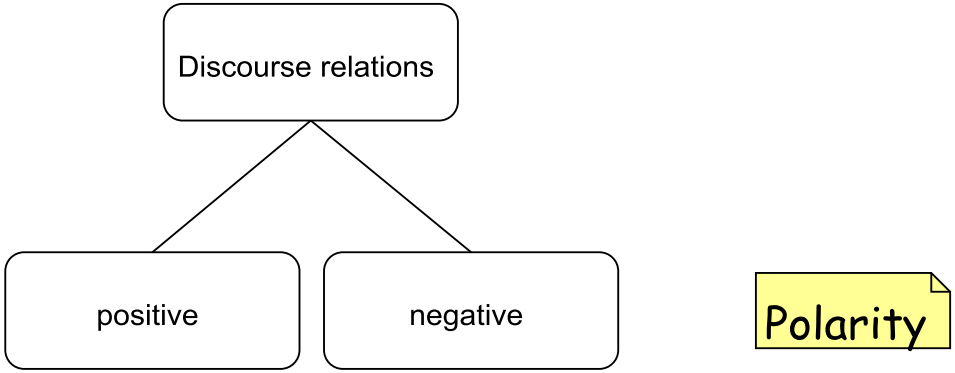


Taxonomy

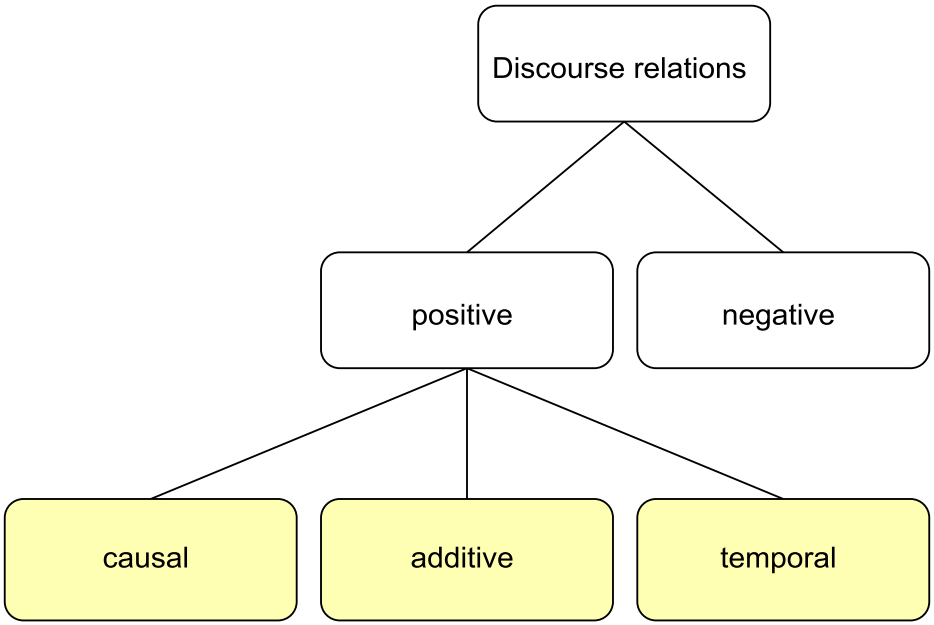
Discourse relations

Universiteit Utrecht

Taxonomy



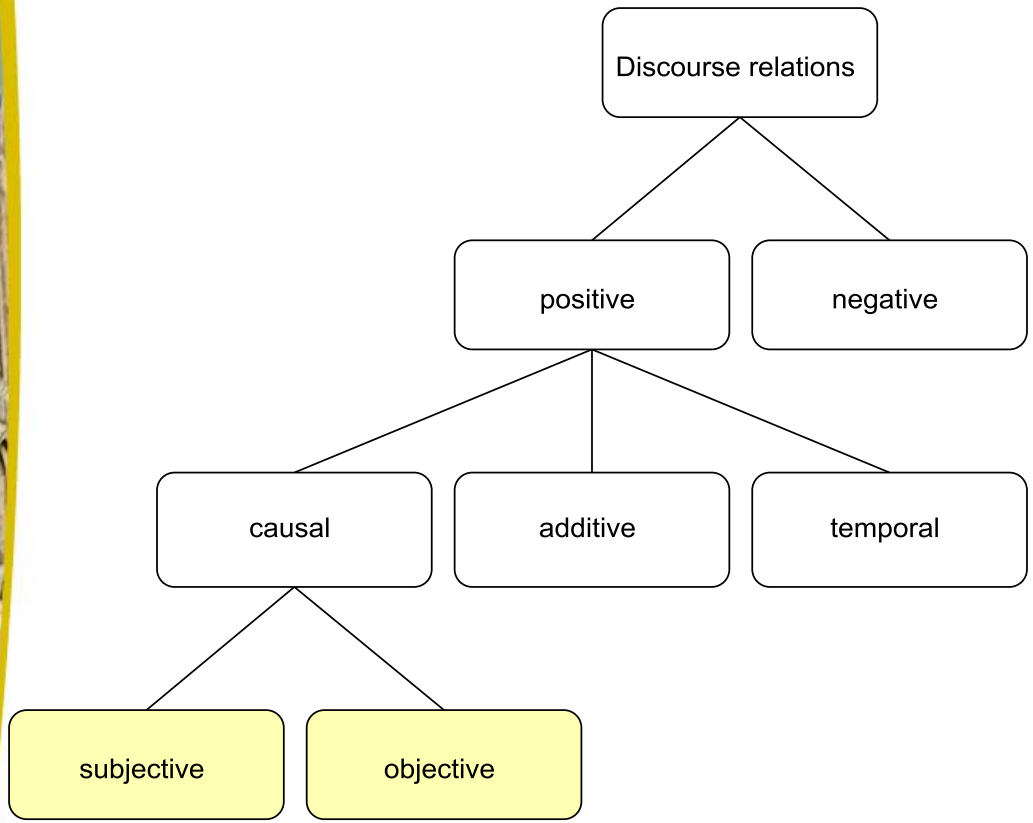
Taxonomy



Polarity

Basic operation

Taxonomy

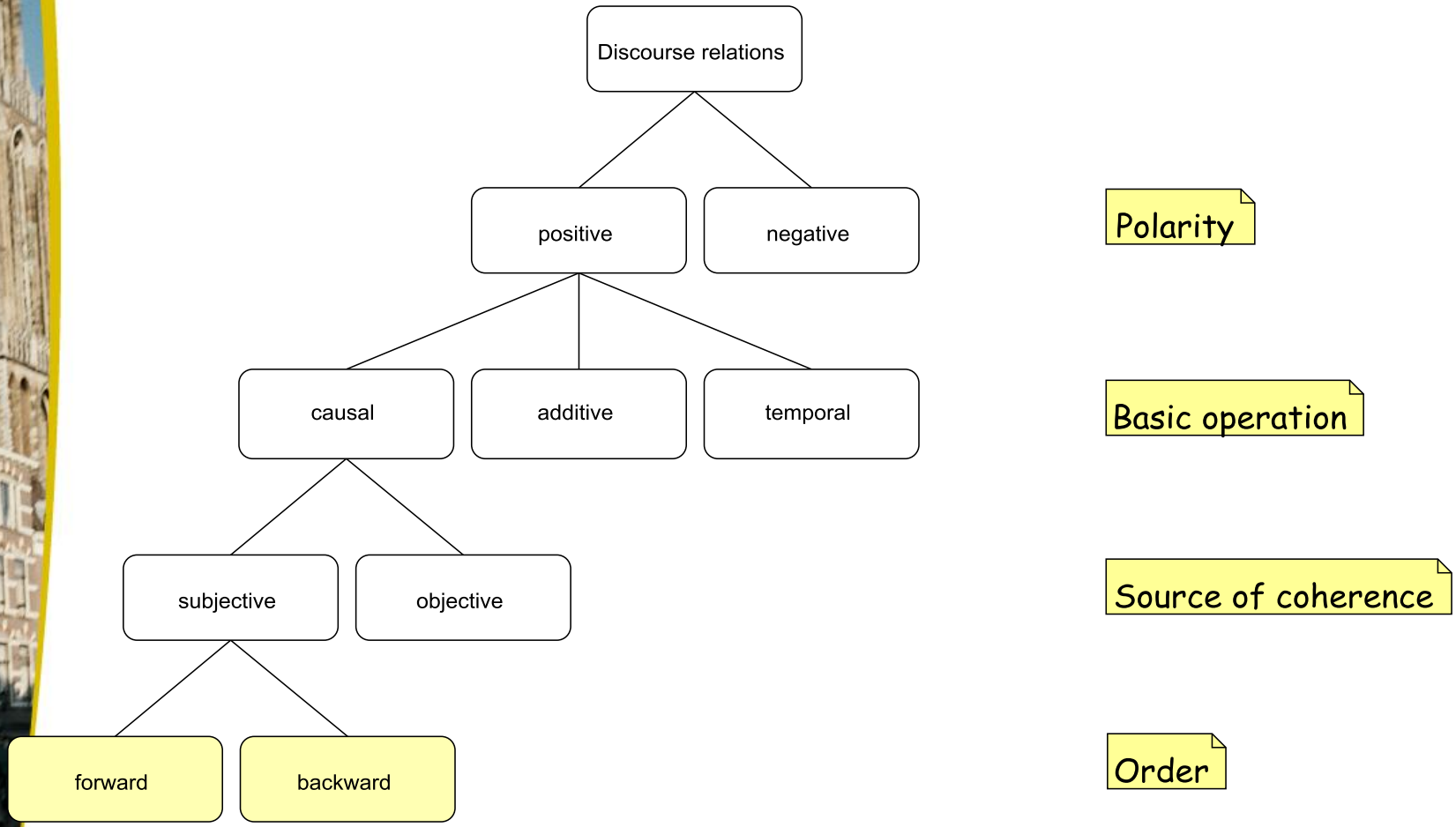


Polarity

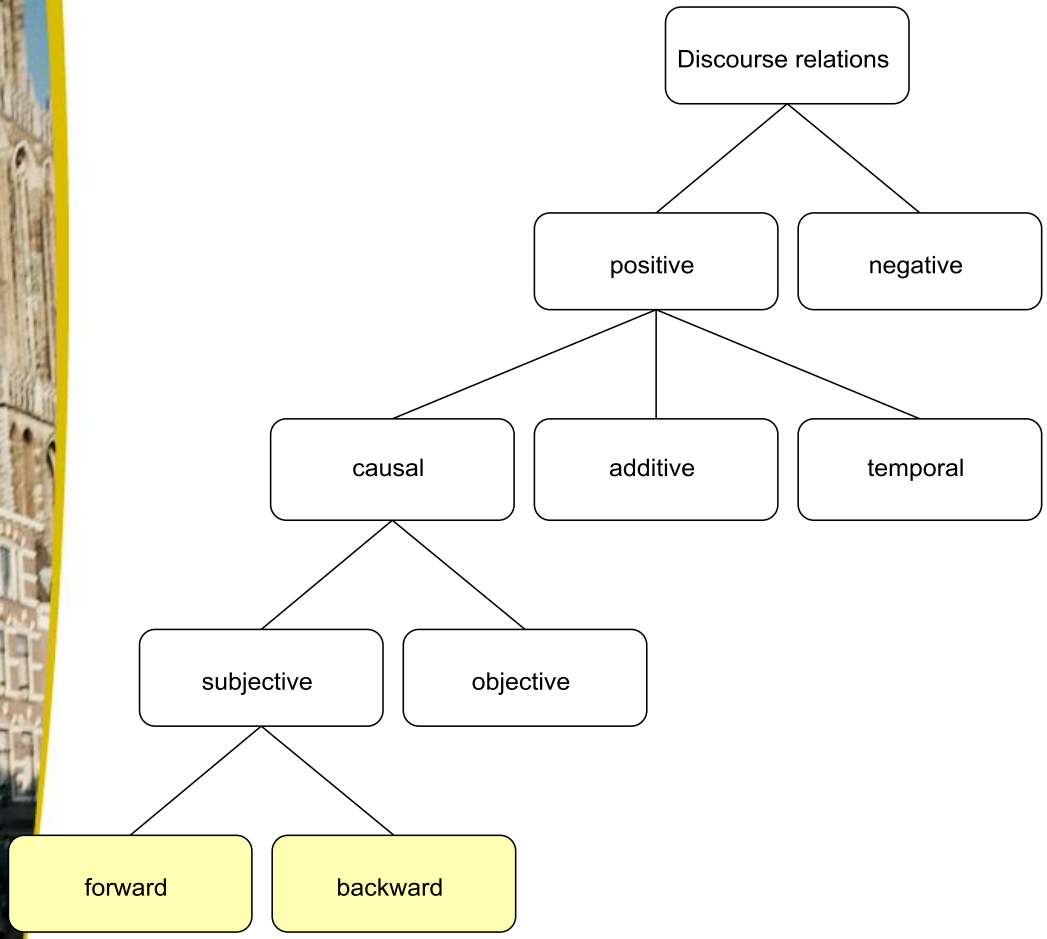
Basic operation

Source of coherence

Taxonomy


Discourse relations
positive
negative
causal
additive
temporal
subjective
objective
forward
backward
Polarity
Basic operation
Source of coherence
Order

Taxonomy



Polarity

Basic operation

Source of coherence

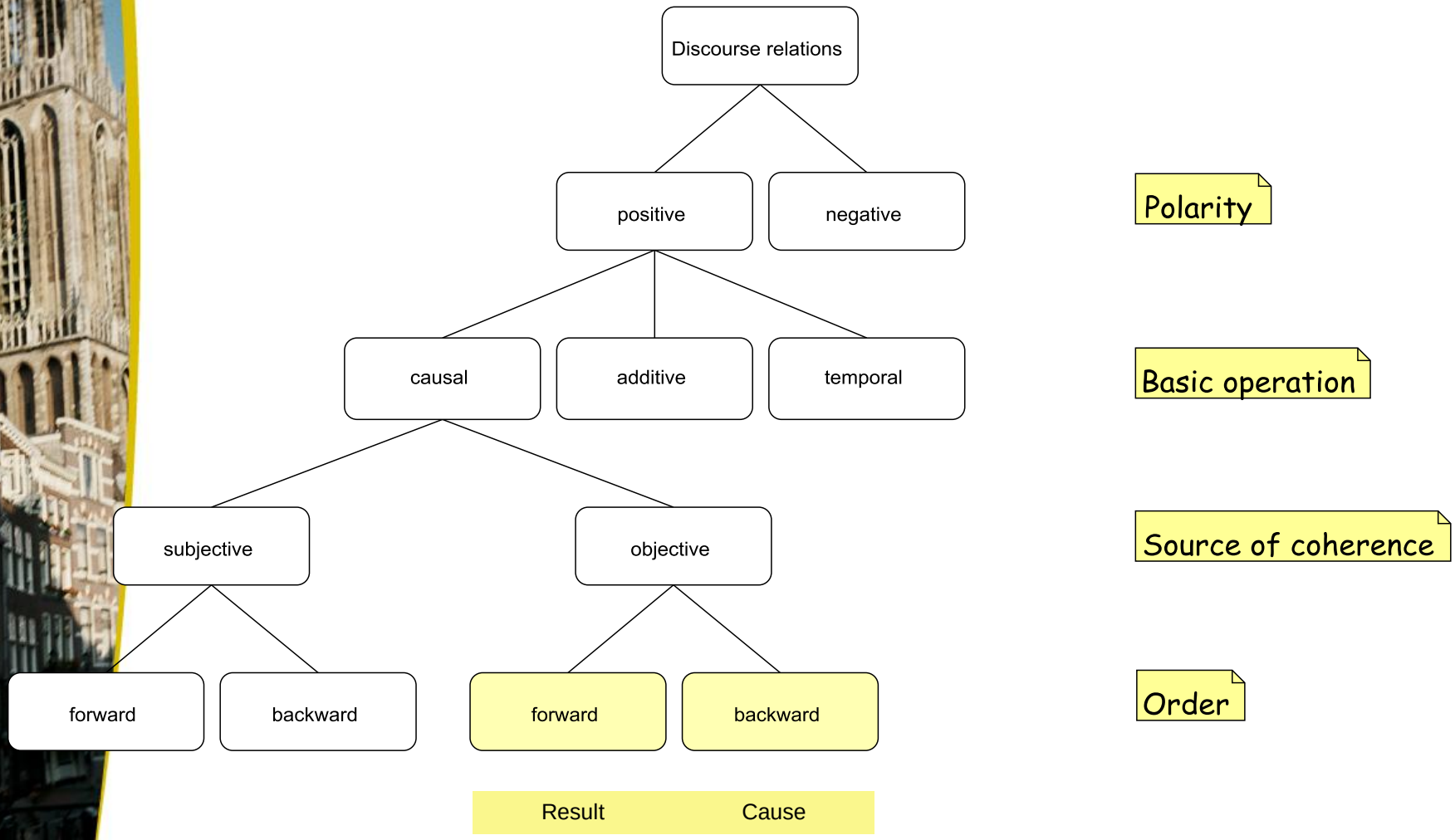
Order

Conclusion Evidence

Justification

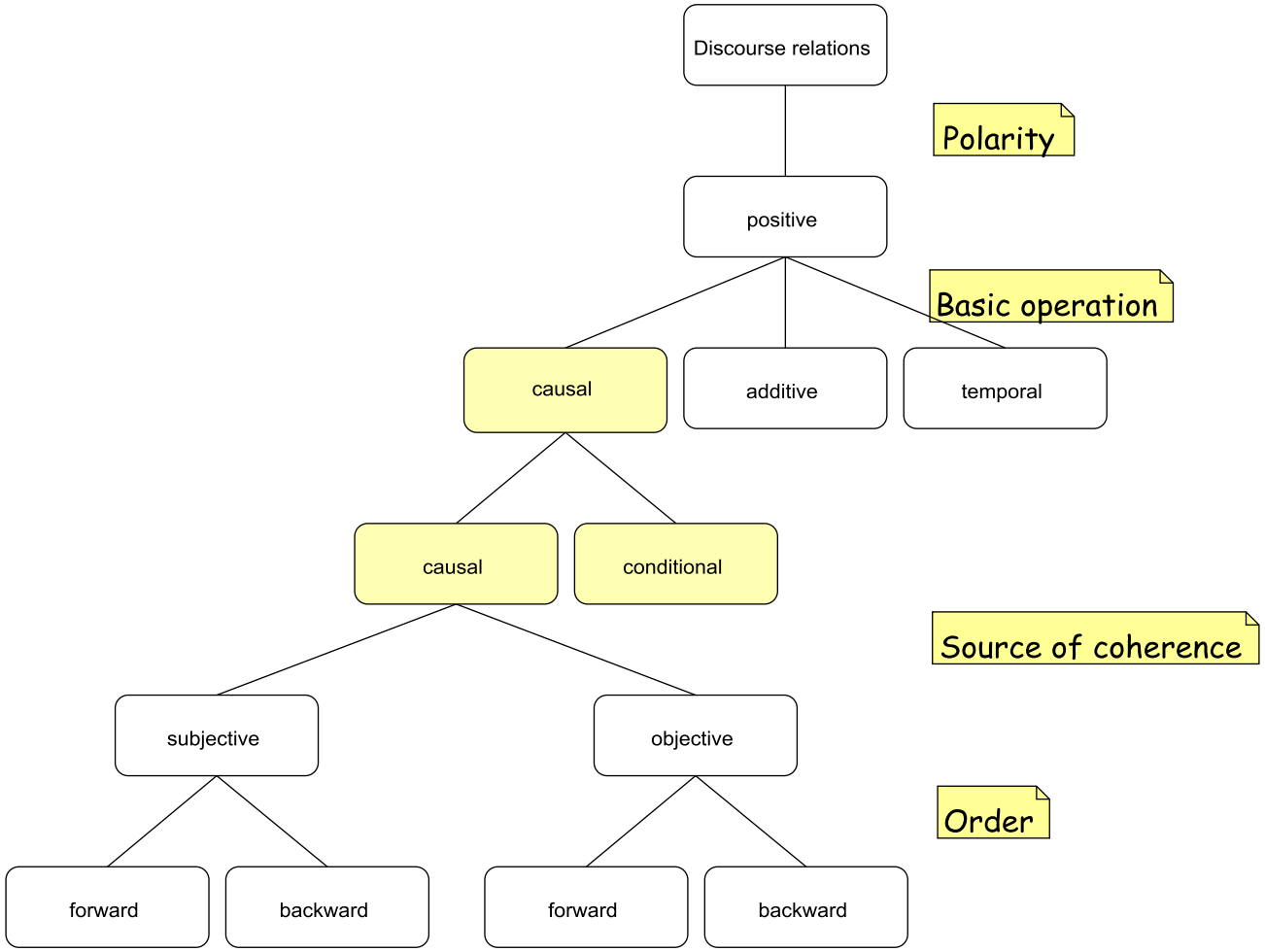
Universiteit Utrecht

Taxonomy

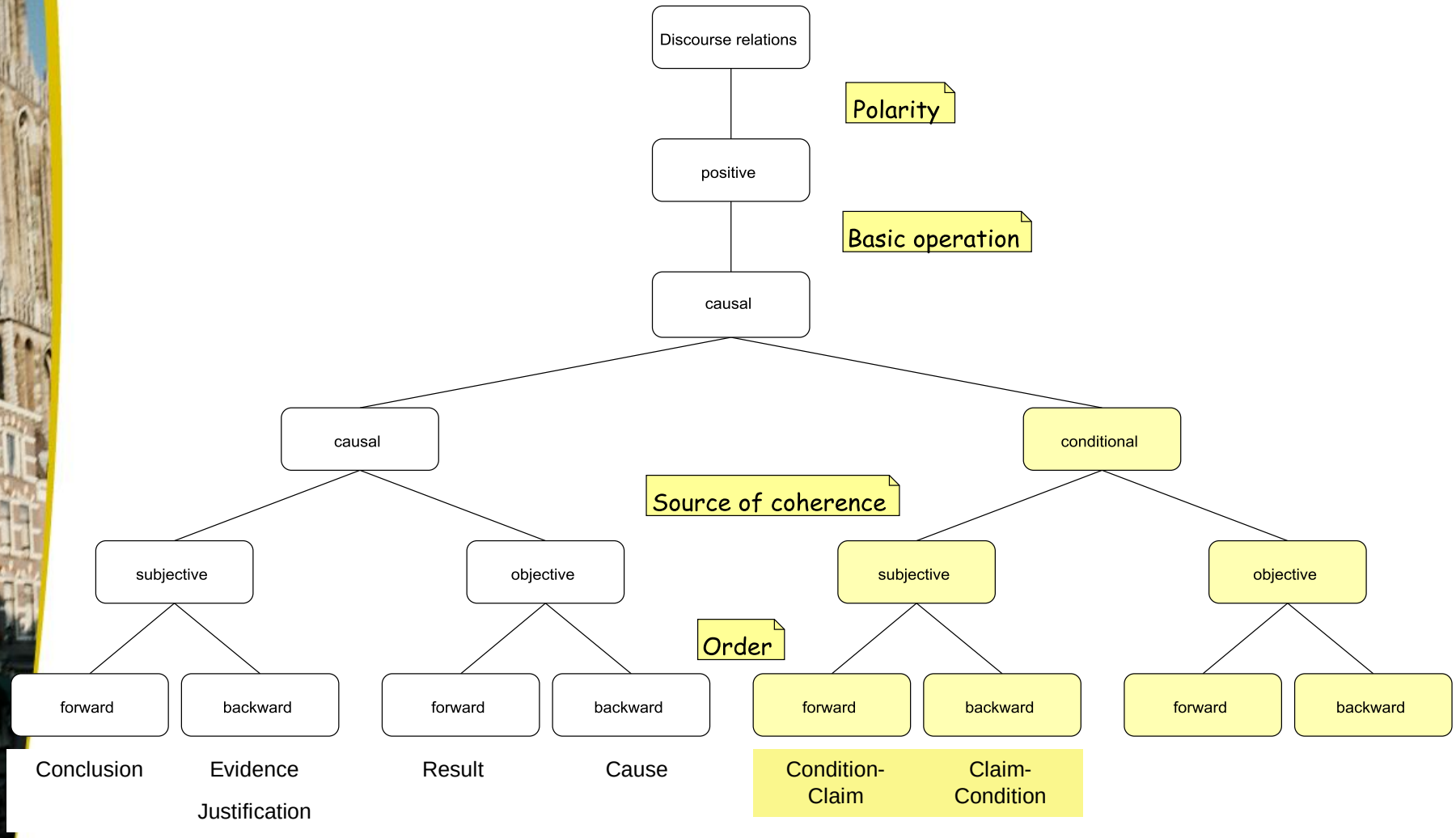


Universiteit Utrecht

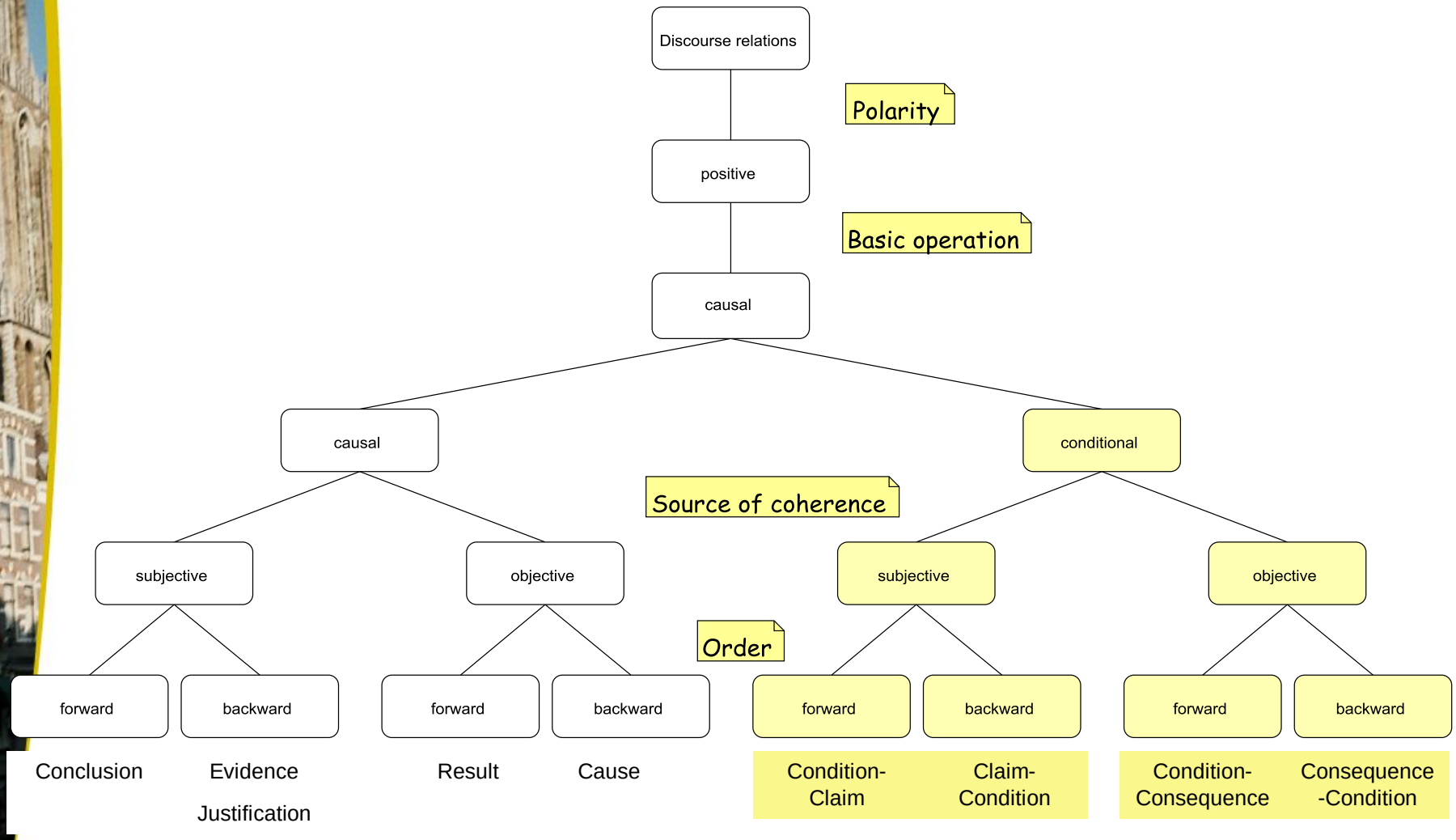
Taxonomy


Discourse relations
Polarity
positive
Basic operation
causal
additive
temporal
causal
conditional
Source of coherence
subjective
objective
Order
forward
backward
forward
backward

Taxonomy

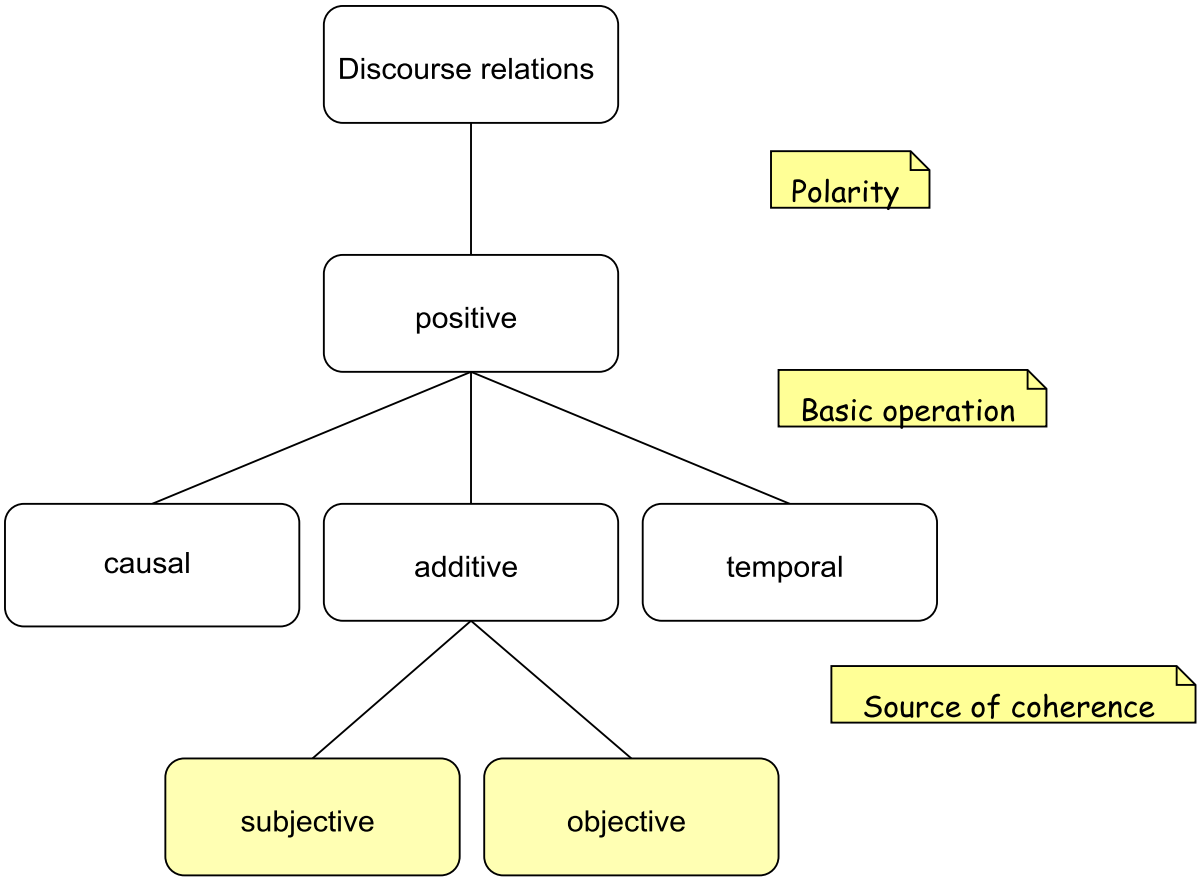

Discourse relations
Polarity
positive
Basic operation
causal
causal
conditional
Source of coherence
subjective
objective
subjective
objective
Order
forward
backward
forward
backward
forward
backward
forward
backward
Conclusion
Evidence
Justification
Result
Cause
Condition-
Claim
Claim-
Condition

Taxonomy

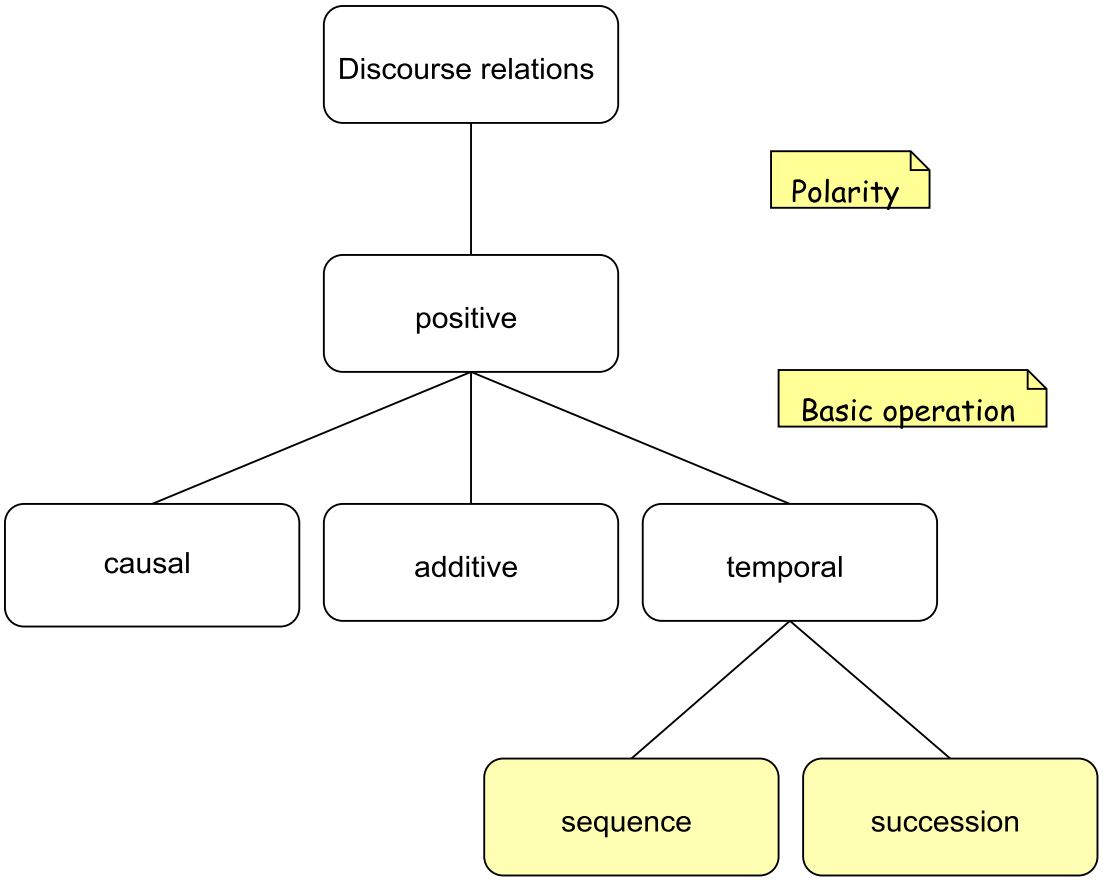


Universiteit Utrecht

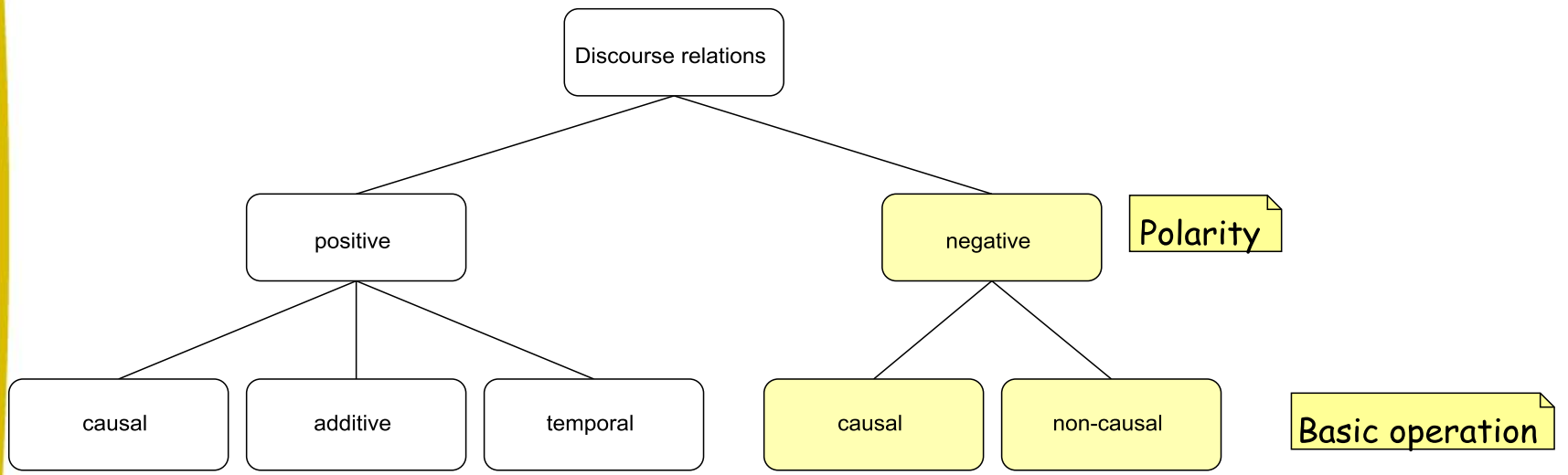
Taxonomy


Discourse relations
Polarity
positive
Basic operation
causal
additive
temporal
Source of coherence
subjective
objective
Order: n.a.
Enumeration
List / Conjunction, Restatement, Alternative, Instantiation

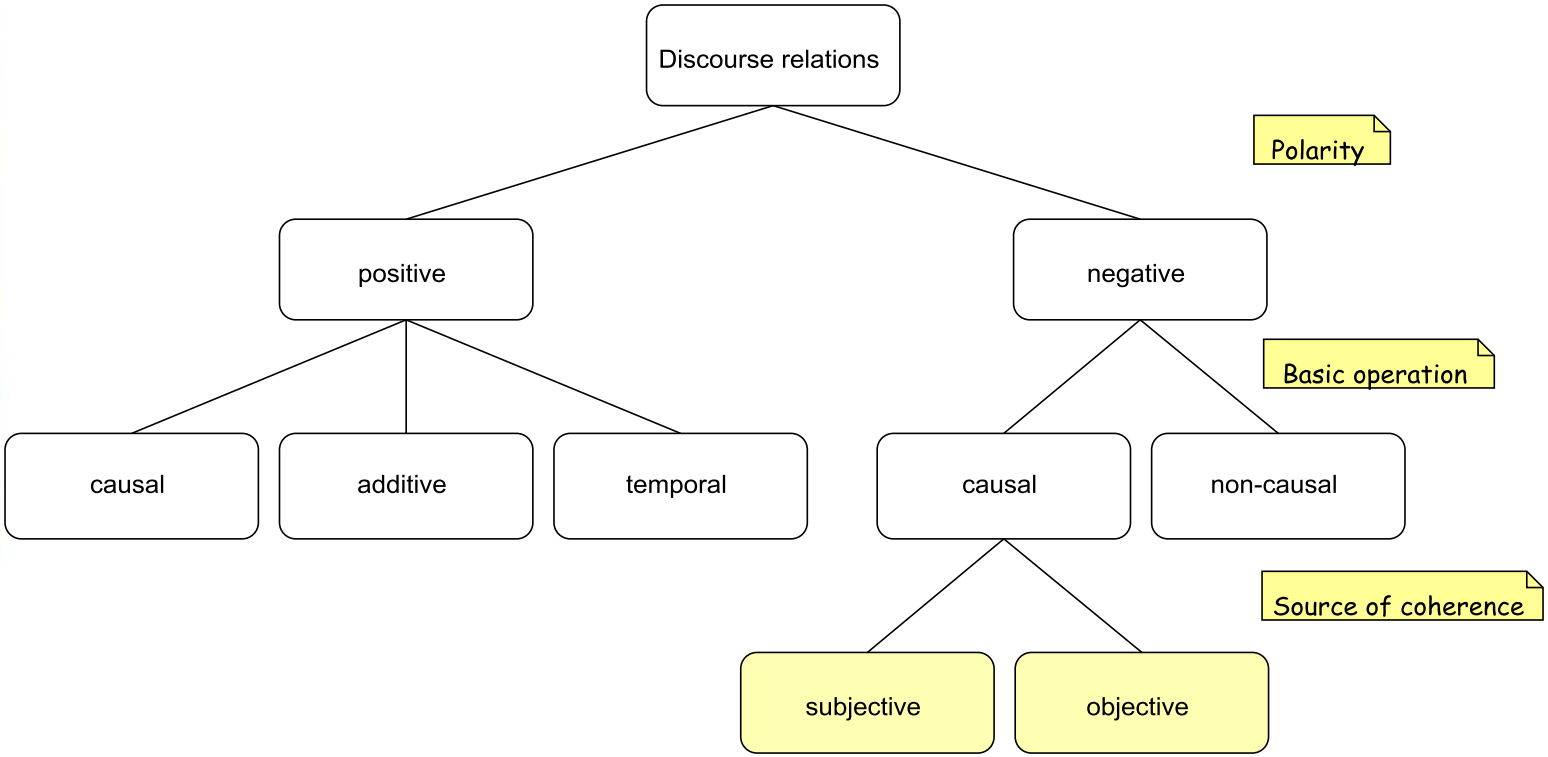
Taxonomy


Discourse relations
Polarity
positive
Basic operation
causal
additive
temporal
sequence
succession

Taxonomy

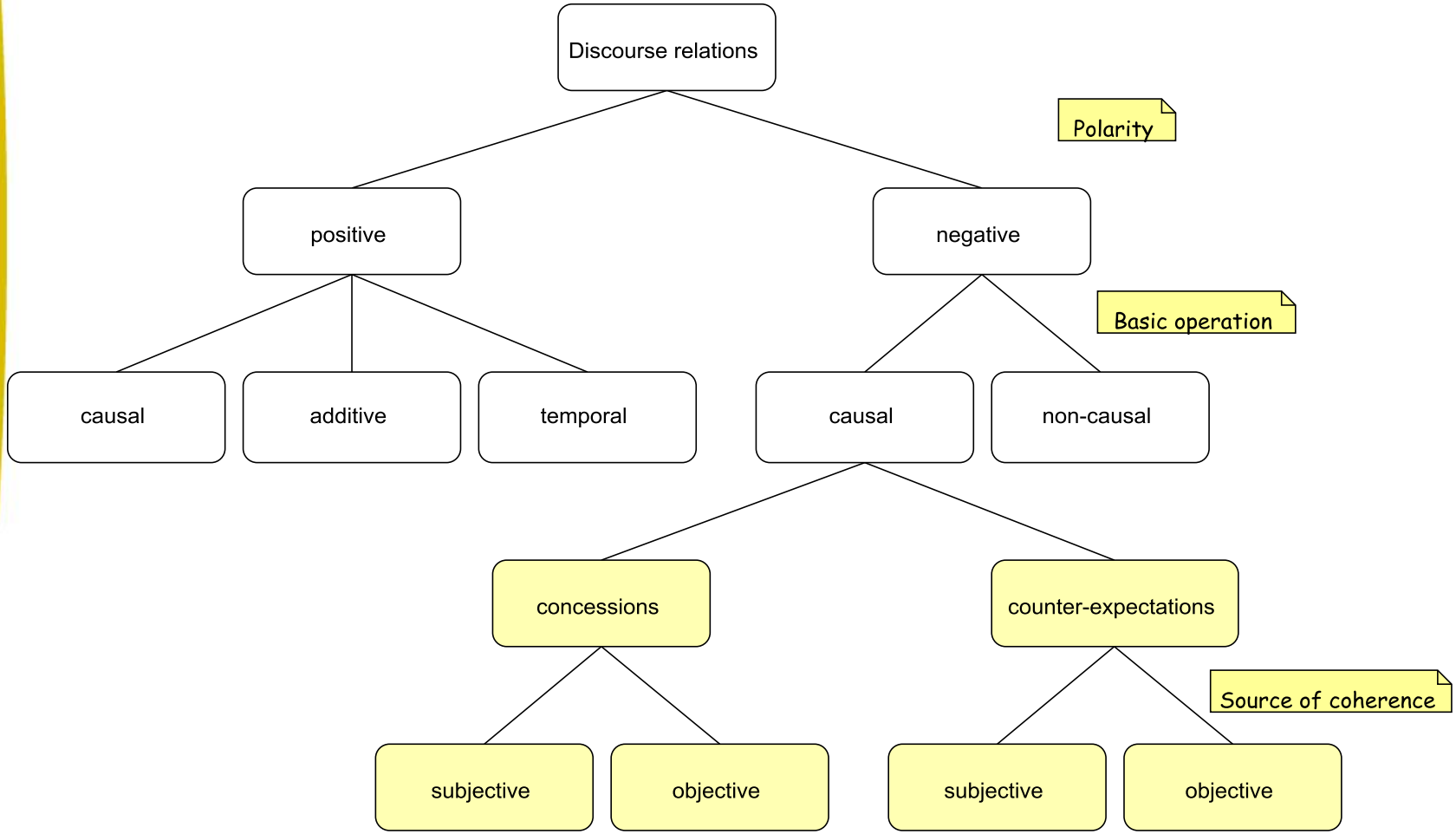

Discourse relations
positive
negative
Polarity
causal
additive
temporal
causal
non-causal
Basic operation

Taxonomy

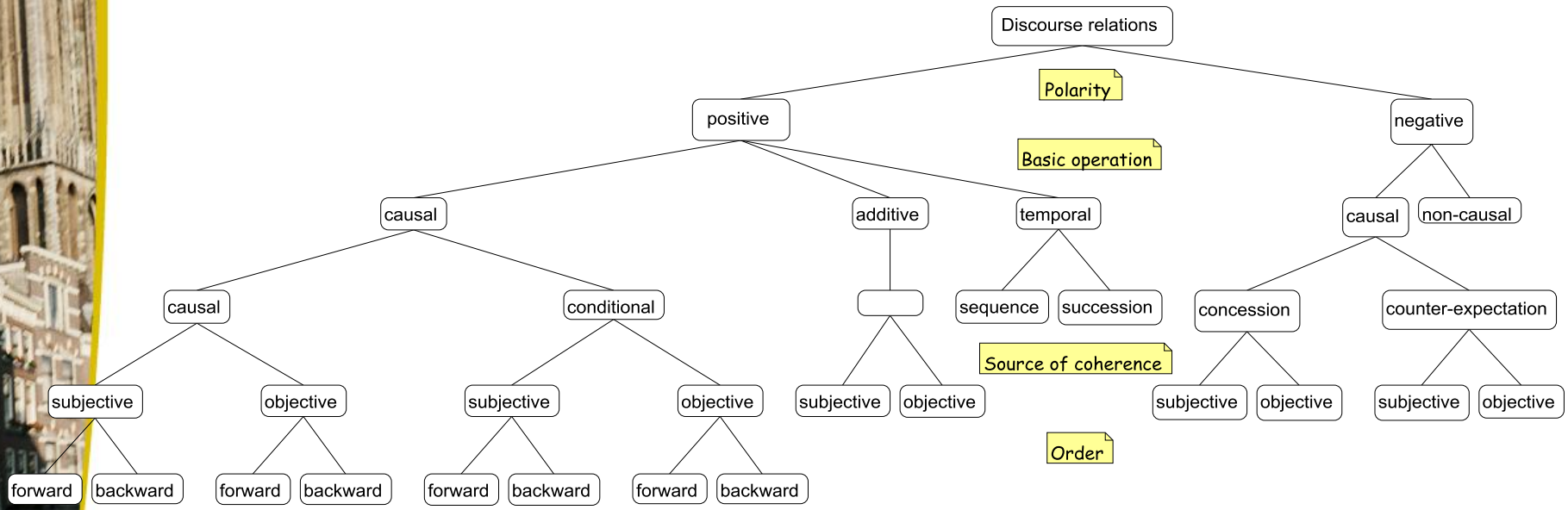


Contrastive Argument-Claim	Contrastive Cause-Consequence
Contrastive Claim-Argument	

Taxonomy


Discourse relations
Polarity
positive
negative
Basic operation
causal
additive
temporal
causal
non-causal
concessions
counter-expectations
Source of coherence
subjective
objective
subjective
objective
Contrastive Argument-Claim
Contrastive Claim-Argument
Contrastive Cause-Consequence

Taxonomy


Discourse relations
Polarity
positive
negative
Basic operation
causal
additive
temporal
causal
non-causal
causal
conditional
sequence
succession
concession
counter-expectation
Source of coherence
subjective
objective
subjective
objective
subjective
objective
subjective
objective
subjective
objective
Order
forward
backward
forward
backward
forward
backward
forward
backward

Why this minimal set of characteristics ?

- Theoretically attractive:
  - Systematic: cross-classification defines categories
  - Empirical evidence from analytical, processing and acquisition studies

- If these basic categories have any cognitive relevance, they should predict learning and interpretation of connectives and relations
- Positive < Negative
- Additive < Causal: English, Dutch
- Objective causal relations processed faster than subjective causal relations (Traxler et al. 1997)

Why a minimal set of characteristics ?

- Descriptive adequacy?
  - Does it work in concrete annotation?
  - Systematic steps might make annotation choices easier
  - Evidence?

New empirical study

- Annotation experiment

- Instead of well-trained experts:
- Are naïve (non-expert, non-trained) annotators capable of annotating coherence relations using the cognitive categories method?

- 20 subjects: 19 advanced BA-students, 1 PhD-student
  - Annotated a corpus of 60 fragments
  - Received a manual and the taxonomy
  - Annotated using an instruction

Material: corpus

- Subjects annotated 60 fragments
- Fragments taken from the DiscAn-corpus, already segmented

- DiscAn-corpus is a relatively causal corpus, with many explicitly marked and positive relations

Material: corpus

- To create a fair distribution, three categories for sentence difficulty were created (easy, medium and difficult)
  - Categories were based on 5 indicators for sentence difficulty:
    1. Sentence length (longer sentences vs. shorter sentences)
    2. Sentence complexity (complex sentences vs. simple sentences)
    3. Marking of the relation (implicit vs. explicit)
    4. Subjectivity (subjective vs. objective)
    5. Negative (negative vs. positive)

Sentence difficulty

Easy sentence:

- John hit his brakes, but his car needed space to stop and that space was not there. The front of the Buick hit the back of the old Chevy in front of him. [John could easily recover from the blow] [because he gripped the steering wheel tightly.] The safety belts did the rest.

Vs.

Difficult sentence

- The day started with a Palestinian attack on a car with colonists. In that attack, two Iraeli's were killed and one was severely injured. Most Palestinian victims were in the area of Jenin. [According to Palestinian sources, Israeli soldiers shot an anti-tank missile at a post of the Palestinian safety services in an area that is completely under the control of the Palestinian authorities. Four cops and a civilian died in that attack.] [The Israeli army says they shot at a group of armed men that were in the area where Palestinian are not allowed to carry guns according to the rules.]

Example of fragment

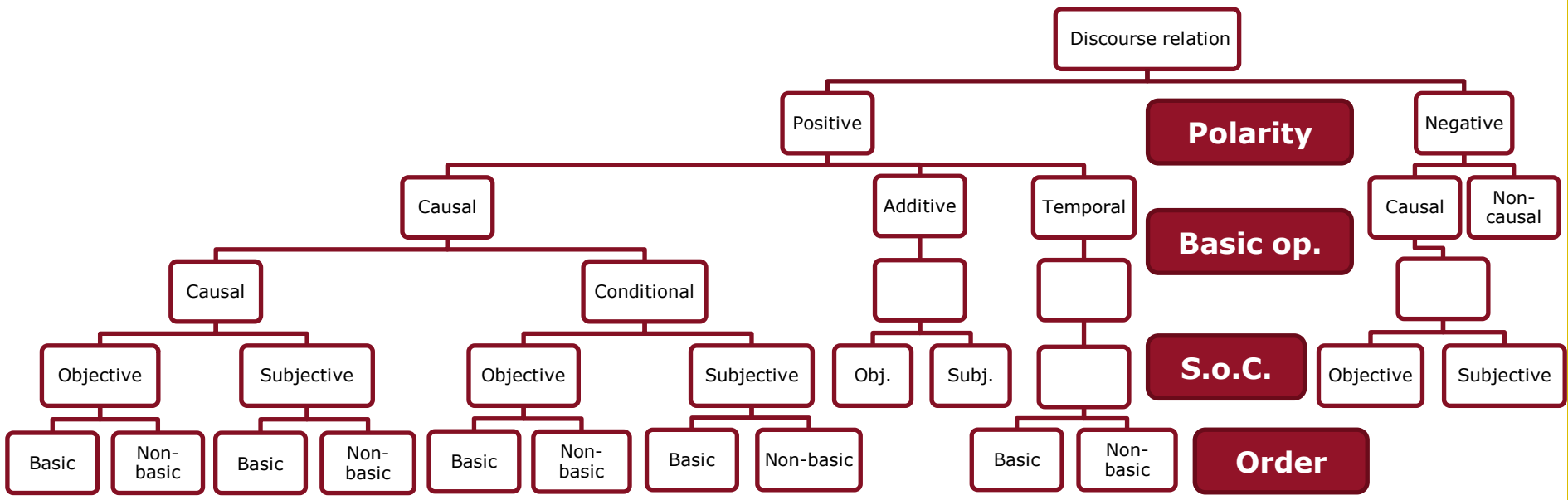
- Fragments were presented in an Excel-file:

Nr	Fragment	Pol	Basic op	S. of coh	Ord
32	"If you want to study well, you have to lock yourself in for three days. Live like a recluse with your Greek books. I'm not going to do that," says Thierry. Instead he will approach this test by calculating. "The translation counts for 50%. [But you know that, based on that, you will never get a five, then you would have to do it perfectly.] [So you look for points in other questions which will earn you a six."]				

Materials: a taxonomy and a manual

Universiteit Utrecht

Materials: a taxonomy and a manual



Material: instructions

Two versions:

- Implicit instruction
  - Relies on annotator's knowledge of the categories
- Explicit instruction
  - Relies on annotator's knowledge of the categories, connective properties, paraphrase tests and substitution tests
- Examples:

Example of implicit instruction

1. Determine the polarity: is the relation positive (‘pos’) or negative (‘neg’)?

2. Determine the basic operation: is the relation causal (‘caus’), additive (‘add’), temporal (‘temp’) or, in the case of negative relations, non-causal (‘non-caus’)? If the relation is causal, is it formulated conditionally (‘caus-cond’)?

(Annotators used the abbreviations while coding)

Example of explicit instruction: substitution test

1. If a relation contains a connective, take this out of the relation. Do take the original connective in consideration while you are interpreting the relation.

2. Can you use *but* to connect the segments?
   - Yes, then the polarity is negative. Fill in 'neg' in Excel for polarity and continue to 1a.
   - No, then the polarity is positive. Fill in 'pos' in Excel for polarity and continue to 2.

(For causal relations *because*; for conditional relations *if*, for additive relations *and*; for temporal relations *when*)

Example of explicit instruction: paraphrase test

2b. Can you paraphrase the relation between S1 and S2 as option A or B below?

A: S1 is the cause, S2 is the consequence

OR

B: S1 is the consequence, S2 is the cause

- Paraphrase A, then the relation has a forward order. Fill in 'for' in Excel for the order. You are done with this relation.
- Paraphrase B, then the relation has a backward order. Fill in 'back' in Excel for order. You are done with this relation.

(*Claim* and *argument* were used for subjective relations)

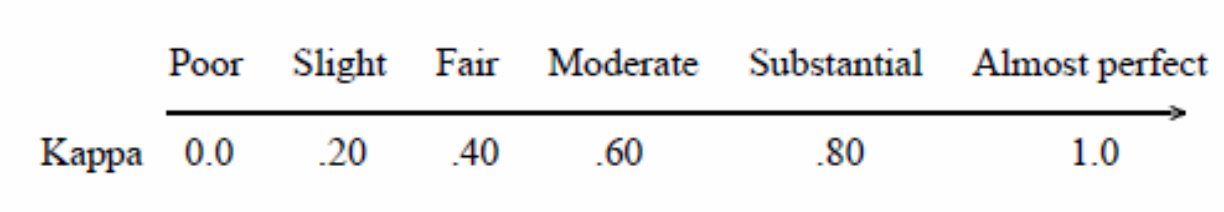
Hypotheses

- Adequate level of agreement with original annotations
- Adequate level of agreement between annotators
- Explicit instruction will result in more agreement than implicit instruction

- Also:
  - Gain insight into effects of sentence difficulty
  - Determine whether some categories lead to more (dis-)agreement than other categories

Processing the data

- Kappa statistics and percentages of agreement
  - Describe kappa scores by using the categories proposed by Landis and Koch (1977)



  - Category of substantial agreement (0,61 < κ < 0.81) allows for tentative conclusions to be drawn
  - Category of almost perfect agreement (κ > 0.81) indicates a reliable method
  - Everything below substantial agreement (κ < 0.61) indicates that the method is not reliable enough

Results

Agreement with original annotations in percentages

Category	Overall (%)	Implicit instruction (%)	Explicit instruction (%)
Polarity	94,7	93,8	95,5
Basic operation	77,1	80,8	73,3
Source of coherence	58,0	56,7	59,3
Order	77,4	79,5	75,3
Average	76.8	77,7	75,9

- Agreement with original annotations for polarity is highest (95%)
- Agreement with original annotations for source of coherence is lowest (58%)
- Implicit condition has more agreement with original annotations on basic operation (80,8%) than explicit condition (73,3%)

Results (cont'd)

Agreement with original annotations: kappa statistics

Categories	**Kappa**	**Level of agreement**
Polarity	0,81	Almost perfect
Basic operation	0,44	Fair
Source of coherence	0,25	Fair
Order	0,66	Substantial

Results (cont'd)

Agreement between annotators in kappa statistics

Category	Overall	Implicit instruction	Explicit instruction	Level of agreement
Polarity	0,68	0,62	0,73	Substantial
Basic operation	0,46	0,5	0,46	Moderate
Source of coherence	0,25	0,28	0,29	Fair
Order	0,61	0,62	0,63	Substantial

- Agreement for polarity is higher in the explicit condition (κ = 0,73) than the implicit condition (κ = 0,62)
- Other differences between the two conditions are negligible

Results (cont'd)

Percentages of agreement with original annotations: sentence difficulty

Category	Easy (%)	Medium (%)	Difficult (%)
Polarity	97,5	95,7	90,8
Basic operation	85,2	78,4	67,5
Source of coherence	57,5	63,2	53,2
Order	76	78,4	77,6

- Agreement for polarity and basic operation decreases when sentence difficulty increases
- Agreement on source of coherence varies with sentence difficulty
- Agreement on order does not show differences

Conclusions

- Hypothesis: there will be an adequate level of agreement with original annotations.
  - → Only for polarity.
  - → Tentative conclusions in the case of the order
  - → Not adequate for basic operation and source of coherence
- Hypothesis: there will be an adequate level of agreement between annotators.
  - → Tentative conclusions in the case of polarity and order
  - → Not adequate for basic operation and source of coherence

Conclusions (cont'd)

- Hypothesis: Explicit instruction will result in more agreement than implicit instruction.
  - → No substantial differences between the two conditions

- Influence of sentence difficulty?
  - → Agreement with original scores for polarity and basic operation decreases when the sentences become more difficult
  - → This effect is not found for source of coherence and order

Discussion

- The results do not show perfect agreement
- Polarity and Order yield considerable amount of agreement
- Source of coherence and Basic Operation are problematic
- We intend to look more closely into least reliable cases

- Instruction: Results are more or less equal for both conditions

- Then again, how high is reliability for competitive proposals?
  - RST: kappa ranging from .6 – 1.0 (Carlson et al. 2002)
  - PDTB: % of agreement ranging from 59.6% – 95.7% (Miltsakaki et al. 2004)
  - Both with expert annotators

Discussion

- The method has not yielded reliable annotations in all respects, with naïve annotators
- Perhaps better results with better trained, expert annotators (who are usually doing this)?

- We have shown how categories can be useful in this context: allows for a systematical, step-wise annotation process
- Which covers the whole set of relations in a theoretically attractive way