Discourse Annotation in TypeCraft

Dorothee Beermann
March, 2013

NTNU – Trondheim
Norwegian University of
Science and Technology

Introduction

Overall Motivation

To create a richly annotated discourse corpus facilitating the analysis of student texts within writing research. The main perspective taken is that of discourse analysis (as opposed to linguistic analysis). A student text is understood as a written discourse, and it is its discourse structure that will be annotated.

Specific Goals

- Introduce a basic terminology for the analysis of discourse structure in form of an discourse annotation schema (DAS)
- Examplify how tags are used by discussing specific texts (Incremental build-up of the DAS)
- Encourage you to provide feedback and extend ou inventory of senses, anchors and chains by annotating more text using the initial DAS

NTNU – Trondheim
Norwegian University of
Science and Technology

8ᵗʰ April technical implementation

What is a discourse relation?

We focus on the informational and semantic discourse relations

we use notions like GOAL, REASON, SUCCESSION etc.

we assume that these relations hold between abstract entities of the type: facts, belief, eventualities, etc. following Asher (1993) who calls these arguments of discource relations *abstract objects.*
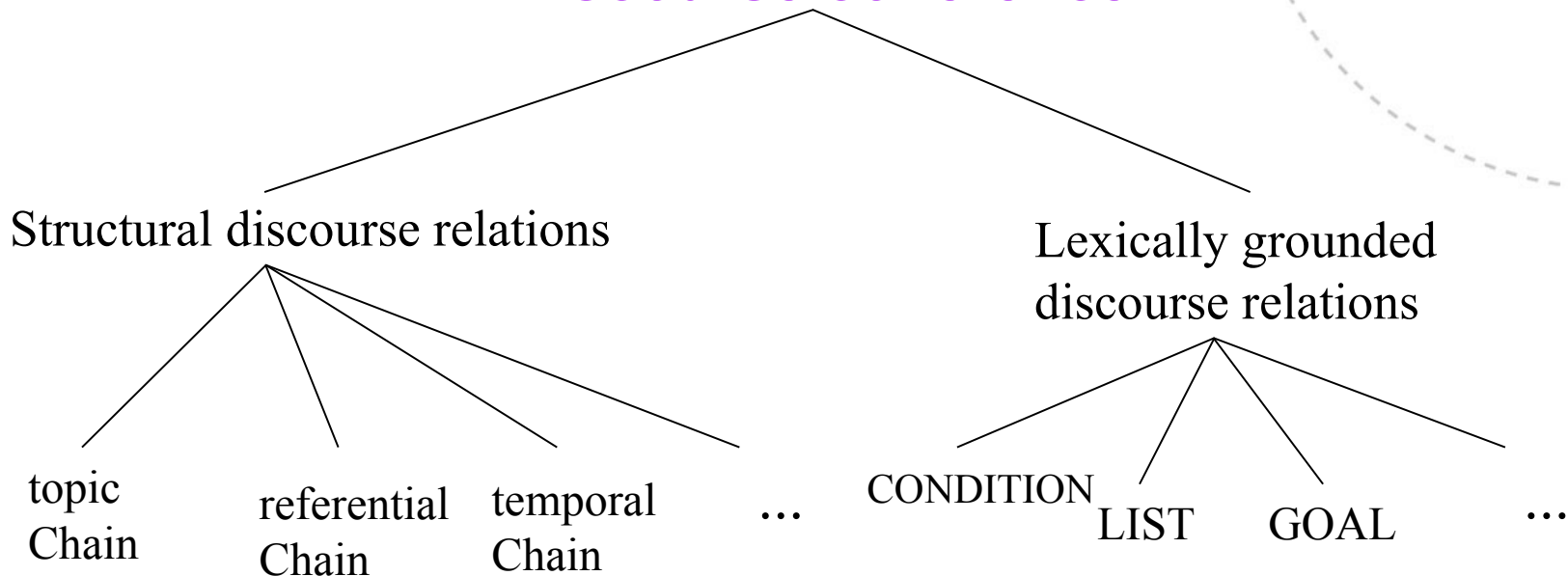
R_{goal} *(motion, activity)*

where *motion is* ARG1 and *activity* ARG2 of the goal relation

NTNU – Trondheim
Norwegian University of
Science and Technology

Types of discourse relations

Discourse coherence

Structural discourse relations

topic Chain   referential Chain   temporal Chain   …

Lexically grounded discourse relations

CONDITION   LIST   GOAL   …

NTNU – Trondheim
Norwegian University of
Science and Technology

How are discourse relations declared?

Structural discourse relations

are 1-to-1 or 1-to-many relations which can be inferred from anaphor-antecedent relations. They can be typed using the grammatical nature of the anaphor-antecedent relation. There are temporal, topic, referential, lexical chains and so forth.

Lexically-grounded discourse relations

- Relations can be grounded in **lexical elements or in punctuation. These are explicit lexically-grounded relations**

- Where lexical elements are absent, relations may be **inferred. These are implicit lexically-grounded relations** (Sense as well as its lexical trigger are implicit).

- **at this point it seems that l-g relations are mostly transitive relations**

NTNU – Trondheim
Norwegian University of
Science and Technology

Where are discourse relations declared?

Textual structural dependencies

- Definitely not unbounded. Domain: paragraph and up. Also here a notion of structure assumed (agains Halliday).

In the literature the assumption seems to be that discourse relations hold primarily between adjacent components, not so sure that that really holds. However that may be, we will assume that

Lexical grounded structural dependencies

- Lexically-triggered discourse relations hold sentence internal and connect adjacent sentences. The triggers are mainly conjunctions and adverbs

NTNU – Trondheim
Norwegian University of
Science and Technology

Tagging triggers using...

Lexical-grounded discourse structure

- Cohesion in Discourse (Halliday & Hasan)
- Discourse Lexicalized TAG (Webber, Joshi, Stone, Knott)

Structural discourse relations

- Discourse GraphBank (Wolf & Gibson)

Lexical-grounded and structural discourse relations require different annotation means.

In addition, I would like to suggest that one annotates *Anchors* and *Situations*

NTNU – Trondheim
Norwegian University of
Science and Technology

Halliday and Hasan (1976)

H&H associate discourse relations with **conjunctive elements**:

- Coordinating and subordinating conjunctions

- Conjunctive adjuncts, including

  - adverbs such as *but*, *so*, *next*, *accordingly*, *actually*, *instead*, etc.

  - prepositional phrases (PPs) such as *as a result*, *in addition*, etc.

  - PPs with *that* or other referential item such as *in addition to that*, *in spite of that*, *in that* case, etc.

Each such element conveys a cohesive relation between

- its matrix sentence and
- a presupposed predication from the surrounding discourse

NTNU – Trondheim
Norwegian University of
Science and Technology

Lexically-grounded discourse structure

Sense	Category	Lexeme	Anchors
CONDITION	adverb	hvis	
GOAL		for å	*indefinite temporal anchor:*
SEQUENCE	conjunction	og	'en dag'
LIST	conjunction	og	*definite temporal anchor:*
PRECEDENCE		etterpå	'de flerste dage etter middag'
INSTANTIATION	expression	for exempel	
SUPPLEMENT	expression	alias	
DISJUNCTION	conjunction	eller	

NTNU – Trondheim
Norwegian University of
Science and Technology

Structural discourse relations

locative/temporal chains

conditional chains

situational chains

referential chains

lexical chains

frame chains

NTNU – Trondheim
Norwegian University of
Science and Technology

Bibliography

- Nicolas Asher (1993). *Reference to Abstract Objects in Discourse*. Kluwer Academic Publishers.

- Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Rashmi Prasad, Aravind Joshi, Bonnie Webber (2005). Attribution and the (non-)alignment of the Syntactic and Discourse Arguments of Connectives. *In Proceedings of the ACL Workshop on Frontiers in Corpus Annotation II: Pie in the Sky*.

- Katherine Forbes-Reilly, Bonnie Webber, Aravind Joshi (2006). Computing Discourse Semantics in D-LTAG. *Journal of Semantics* 23, pp. 55-106.

- Michael Halliday, Ruqaiya Hasan (1976). *Cohesion in English*. London: Longman.

- The PDTB-Group (2006). The Penn Discourse Treebank 1.0. Annotation Manual. *IRCS Technical Report IRCS-0601*. University of Pennsylvania