

Computational Discourse

Section 21.1 - 21.2 in Textbook

hannes@ru.is

T-725-MALV

Discourse

- Language does not normally consist of isolated, unrelated sentences, but instead of **collocated, structured, coherent groups of sentences**
- Such a coherent structured group of sentences is a **discourse**
- Many **different kinds** including monologues (books, news, etc.) and dialogues (including computers as well as humans)

Defining Discourse Analysis

- **Linguistics** tends to focus on isolated sentences and their formal structure, regardless of use
- **Discourse Analysis** looks at what makes discourse **coherent in light of its use**

What is Language Used for?

- Language serves not one single purpose.
Useful to think of two major categories of use:
 - For **transmitting information**
(Transactional / Propositional / Referential / Ideational / Descriptive)
 - For **managing social relationships**
(Interactional / Relational / Emotive / Interpersonal / Social-Expressive)

Transactional Use

- Is the mediation of **information** the most important use?
 - Speaker describes a fact or a point of view
 - Speaker describes an action to be formed
- **Culture is maintained** this way. We build on transmitted knowledge

Interactional Use (1 of 2)

- Traditional linguistics focus less on this use
- **Sociology and sociolinguistics** examine for example how language is used to initiate and manage social relationships (e.g. role of smalltalk)
- **Conversational Analysis** examines in particular how language is used to coordinate turns, define participant roles and in general frame the overall language activity

Interactional Use (2 of 2)

- **Context Analysis** broadens the study even further by looking at the role of nonverbal behavior in conducting interaction
- When we look closer, we see that much of daily language use is interactional

Interactive Use (examples)

A: "Hi!" B: "Hi!"

A: "mm..excuse me..." B: "Yes?"

A: "Damn it's cold today!" B: "Freezing"
(cold weather obvious)

A: "...he's pretty good." B: "Yeah, he's pretty good."

- Not just spoken text, but also in written media like blogs, letters, cards and ads.

Discourse Functions vs. Devices

- When you say that language is being used for a specific purpose, regardless of whether it's transactional or interactional, it is important to distinguish between:
 - The **Discourse Function** being performed (essentially an abstract goal).
 - The **Discourse Device** being used to carry out that function (essentially a concrete action).

Discourse Devices

- Note that a **Discourse Device** can be any **concrete action**, including a particular choice of words or a linguistic structure, and visible behavior like head nods or drawing a picture.

Functions and Devices: Examples

- "Coke is good, but Pepsi is better"
(**Function**: contrast, **devices**: coordinating conjunction, comparative form)
- "Anyways.. it all worked out"
(**Function**: Wrapping up, **device**: cue word)

Discourse Analysis (1 of 2)

- In the field of **linguistics**, a language is described by grammar rules that always have to be true for that language. One counter example would eradicate a rule.
- In the field of **discourse analysis**, a discourse is described by the kinds of linguistic and behavioral devices employed to perform communicative function.

Discourse Analysis (2 of 2)

- The mapping between functions and devices is seldom described by fixed rules, but rather **patterns that seem to emerge** in context
- The dynamic environment has great impact on actual language use
- **Discourse Analysis** therefore provides a kind of **a tool box for analyzing and understanding discourse**, rather than a set of rules such as those describing the world of physics

Coherence

- The elements of a coherent discourse must fit together in a **meaningful** way.
 1. John hid Bill's car keys. He was drunk.
 2. ??John hid Bill's car keys. He likes spinach.
- Hearers strive to find the meaning. This is clearly important. Speakers should help!

Cohesion

- Cohesion is the use of certain **linguistic devices** to link or **tie together** textual units
 - **Lexical cohesion**: Based on relations between words (identical, synonym, hypernym..)
 - **Non-lexical cohesion**: Use of anaphora (referring expressions)

Coherence vs. Cohesion

- **Cohesion:** Refers to the way textual units are linked together (a glue)
- **Coherence:** Refers to the meaning relation between the units

Discourse Structure

- Speakers structure their discourse in some logical fashion and the resulting discourse structure can aid in its processing

Example Discourse

1. A: So you have the engine assembly finished.
2. A: Now attach the rope to the top of the engine.
3. A: By the way, did you buy gasoline today?
4. B: Yes. I got some when I bought the new lawn mower wheel.
5. B: I forgot to take my gas can with me, so I bought a new one.
6. A: Did it cost much?
7. B: No, and I could use another anyway to keep with the tractor.
8. A: OK
9. A: Have you got it attached yet?

Discourse Structure

- The pronoun it in sentence 9 **refers** to an entity that was last mentioned 7 sentences earlier!
- New objects have been discussed in between that could have matched this pronoun.
- Simply taking the closest object that matched the pronoun would have failed to resolve this.
- In fact, **no linear representation of this text structure could provide a good solution.**

Discourse Segmentation

- Even though a good segmentation of the discourse is required, people **do not agree** on the method to use or what kind of structure to build (e.g. linear, hierarchical, etc).
- There is **no exact definition** of a segment, only the feeling that certain sentences seem to fit well together (as shown by experiments with subjects).

Discourse Structure

- Humans easily understand what is going on in the example text.
- Sentences 3-7 are simply a **digression** that doesn't have anything to do with the rope.
- In sentence 8 the producer indicates **the end** of the digression with an **"OK"**.
- Therefore line 9 is interpreted in context of line 2, but not in the context of the part that has concluded.

Discourse Segmentation

- This explanation requires the concept of a **Discourse Segment**
- Roughly, a segment corresponds to a part in the discourse **related by topic**
- Segmentation may rely on things like **lexical cohesion** (e.g. word similarity) and **discourse markers** (cue words) (e.g. "by the way" and "ok")

Coherence Relations

- After we have identified the segments that make up the discourse structure, what can we say about how they are **meaningfully connected** to one another?
- The relationship between segments is referred to as a **coherence relation**
- **Rhetorical Structure Theory** (RST) defines 23 different kinds of relations (e.g. evidence, elaboration, contrast, etc.)

Coherence Assignment

- **Automatic coherence relation assignment** is quite hard to accomplish, but still attempted
- Sometimes called **Discourse Parsing**
- A shallow algorithm may use cue phrases:
 1. Identify the cue phrases in the text
 2. Segment text using the cue phrases
 3. Classify relations based on cue phrases (connectives, e.g. "because")