



Tuebingen Interdisciplinary Corpus of Ambiguity Phenomena

TInCAP - Ein interdisziplinäres Korpus zu Ambiguitätsphänomenen

Wiltrud Wagner, Jutta Hartmann, Lisa Ebert, Gesa Schole, Peter Gietz, Susanne Winkler

Beispielannotationen



RELEVANT PART	zwei Radler
PARAPHRASE 1	zwei Radfahrer
PARAPHRASE 2	zwei Biermischgetränke
COMM. LEVEL	outermost level
DIMENSION	PS ⁺ RS ⁺
TRIGGER	element
RANGE	complex
RELATION	unrelated
PHENOMENON	homonymy

Annotation: Stefanie Kern

"[...] I've lost count of the times I've nearly been killed –"
"Twenty-seven," said Twoflower.
"What?"
"Twenty-seven times," said Twoflower helpfully. "I worked it out. But you never actually have."
"What? Worked it out?" said Rincewind, who was beginning to have the familiar feeling that the conversation had been mugged.
"No, been killed. Doesn't that [...]"
Terry Pratchett: The Light Fantastic, 1986, p. 119

RELEVANT PART	But you never actually have.	But you never actually have.
PARAPHRASE 1	But you never actually have been killed.	But you never actually have been killed.
PARAPHRASE 2	But you never actually have worked it out.	But you never actually have worked it out.
COMM. LEVEL	innermost level	outermost level
DIMENSION	PS ⁻ RS ⁻	PS ⁺ RS ⁰
TRIGGER	complex element	complex element
RANGE	group compound	group compound
RELATION	unrelated	unrelated
PHENOMENON	ellipsis	ellipsis
CONNECTION	change of communicative level	

Annotation: Wiltrud Wagner

Phänomen(bereich)

- Ambiguität, verstanden als Doppel- oder Mehrdeutigkeit, in ihren verschiedensten Formen
- verschiedene Disziplinen: Sprachwissenschaft (Anglistik, Germanistik, Romanistik), Literaturwissenschaft (Anglistik, Germanistik), Latinistik, Medienwissenschaft, Rhetorik, Theologie, Psychologie, Jura
- Beispiele vielfältig hinsichtlich Umfang und Medium

Herausforderung

- interdisziplinäre **Ambiguitätsannotation**
 - profitabel für alle beteiligten Disziplinen
 - Anforderungen der Disziplinen
 - Komplexitätsreduktion
- Terminologie** für disziplinenübergreifende Kommunikation
 - präzise
 - nicht disziplingebunden

Annotationsschema

Example and Metadata

QUOTE: Beispiel (Text, Bild, etc.)

PRIMARY SOURCE/ CITED FROM: Bibliographische Angaben der Quellen

QUOTE TYPE: z.B. spoken, written, pictorial, ...

QUOTE SUBTYPE: z.B. joke, narrative text, drawing, ...

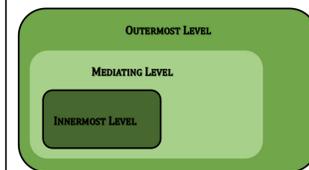
PERIOD OF QUOTE: Ursprungsjahrhundert des Beispiels

LANGUAGE OF QUOTE: Sprache des Beispiels

RELEVANT PART: der ambige Ab- oder Ausschnitt des Quote

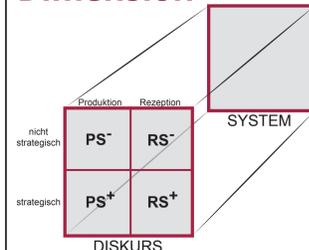
PARAPHRASES: Erläuterung der Lesarten

Communicative Level



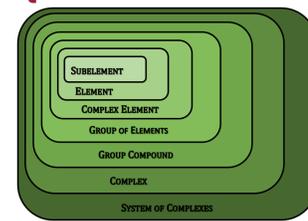
- Angabe der Kommunikationsebene, die für die jeweilige Annotation relevant ist
- Nicht jede Ebene ist bei jedem Beispiel vorhanden

Dimension



- Wird die Ambiguität strategisch produziert?
- Wird die Ambiguität strategisch rezipiert?

Quantitative Annotation



SUBELEMENT: z.B. Phonem, Graphem, Morphem

ELEMENT: z.B. Wort

COMPLEX ELEMENT: z.B. Phrase, Figur

GROUP OF ELEMENTS: z.B. Satz, Figurengruppe

GROUP COMPOUND: z.B. Text-/Diskurs-/Redeabschnitt; Bild

COMPLEX: z.B. Text; Diskurs; Bild + Rezeptionsumstände)

SYSTEM OF COMPLEXES: z.B. thematisch/strukturell/funktionell verbundene Texte/Diskurse/Bilder im Vergleich

- TRIGGER:** Wo wird die Ambiguität ausgelöst?
- RANGE:** Wie weit reicht ihr Effekt/ihre Relevanz?

Qualitative Annotation

TYPE OF PARAPHRASE RELATION

- unrelated:** distinkte, nicht von einander abgeleitete Lesarten (z.B. Homonyme)
- related:** distinkte, von einander abgeleitete Lesarten (z.B. Polyseme)
- open:** multiple, gleichzeitig mögliche Lesartenvarianten (z.B. vage Ausdrücke)

PHENOMENON

Disziplinspezifische Angabe des relevanten Phänomens

Connected Annotations

Verbindung von Annotationen unter Angabe des Grunds

- change of communicative level
- addition
- combined annotations

Technische Umsetzung

Daten

- Datenmodell XML basiert
- Erste Dateneingaben mit XML Editor
- LDAP als Datenbanktechnologie bot sich an, wegen Kompatibilität zu XML (Objekte, Hierarchien), wegen feingranularer Zugriffskontrolle und wegen standardisiertem Netzwerkprotokoll
- Verschiedene Datenarten (Stammdaten, Annotationen, bibliographische Daten, etc.) werden in eigenen Objekten gespeichert

Nachhaltigkeit

- Langfristige Speicherung der Daten durch Abbildung in XML
- soweit möglich TEI-konforme Annotation der Daten
- Erweiterung der TEI-Annotation durch eigenes Schema für die inhaltliche Annotation
- Einbindung von Multimediadateien über Links und Dateipfade
- Import in / Export aus LDAP des gesamten Korpus bzw. von Subkorpora

Benutzerinterface

- Ermöglicht Dateneingabe und Datenretrieval
- Verschiedene Objekte werden zu einem virtuellen Eintrag zusammengefasst mit z.B. beliebig vielen Annotationen
- Jedes einzelne Objekt lässt sich über Zugriffskontrolle schützen
- Beliebig komplexe Abfragen sind grundsätzlich möglich, sodass neue Forschungsfragen beantwortet werden können