Anaphoricity in Connectives: A Case Study on German

Manfred Stede and Yulia Grishina

Applied Computational Linguistics

University of Potsdam / Germany

Overview

- Introduction
- Anaphoric connectives in German
- Case study: *demzufolge*
- Toward disambiguation and resolution
- Outlook

Introduction

- A *connective* signals a coherence relation between two spans of text
  - *Because I'm ill I won't come to the party.*
  - *I'm ill. Thus I won't come to the party.*
- An *event anaphor* picks up an ‚abstract object' antecedent
  - *Sue couldn't come to the party. That disappointed Jim.*

- Some connectives also work as event-anaphors (Webber et al. 2003). In the PDTB corpus, 9% of Arg1's are not adjacent to the connective or to the Arg2
  - *[Tom didn't go to the café.]$_{Arg1}$ It would close soon anyway. [He chose to sit at the beach]$_{Arg2}$ [instead]$_{conn}$.*

Introduction

- Some connectives have an explicitly-anaphoric morpheme
  - *therefore*, *whereby*
  - many more in German!
- Some of these German anaphoric connectives also have additional non-connective readings, where they act as nominal anaphors

- => Overall, a considerable problem of disambiguating and resolving antecedents

Anaphoric Connectives in German

- DiMLex: Machine-readable lexicon of German connectives (Stede 2002)

274 connectives (Scheffler/Stede, LREC 16)

Basic technical approach:
„Theory-neutral" rich source lexicon in XML
mapped via XSLT to specific application resources
  - Language generation in „Polibox" (LISP) (Stede 02)
  - Discourse parsing (Prolog) (Hanneforth et al. 03)
  - Various HTML views for the human user

„Connective"

- Definition (cf. Pasch et al. 2003)
  - closed-class words
  - non-inflectable
  - semantics: two-place relation
  - join two eventualities that could be expressed as full clauses

- Syntactic categories:
  - conjunctions, coordinating and subordinating
  - certain adverbials
  - certain prepositions: *despite, due to, ...*

Anaphoric connectives in DiMLex

- Explicit anaphoric morphemes: 79 connectives (29%)
  - da- (21)   *dadurch*
  - -dessen (17)   *infolgedessen*
  - wo-/wes- (11)   *weswegen*
  - hier- (7)   *hierdurch*
  - -dem (7)   *trotzdem*
  - dem- (6)   *demnach*
  - des- (4)   *deswegen*
  - -dann (3)   *sodann*
  - -dies (2)   *überdies*
  - dessen- (1)   *dessenungeachtet*

Anaphoric connectives in DiMLex

- 40/79 also have non-connective readings as a nominal anaphor
  - second function: relative pronoun, discourse particle, verb particle, …

  - *[Sie schenkte mir ein Buch,]*$_{Arg1}$ *[womit]*$_{conn}$ *[sie mir einen großen Gefallen tat.]*$_{Arg2}$
    ‚She gave me a book, whereby she did me a big favour.'
  - *Sie schenkte mir ein Buch, womit ich nichts anfangen konnte.*
    ‚She gave me a book, with which I could not do anything.'

Case study: *demzufolge*

- Reading 1: nominal anaphor
  - contracted form of „dem zufolge"
  - (i) Introducing a relative clause
    - *Ich las ein Buch, demzufolge die Welt in diesem Jahr untergehen wird.*
      ‚I read a book according to which the world will collapse this year.'
  - (ii) Free adverbial
    - *Ich habe ein interessantes Buch gelesen. Demzufolge wird die Welt in diesem Jahr untergehen.*
      ‚I read an interesting book. According to it the world will collapse this year.'

Case study: *demzufolge*

- Reading 2: connective (Cause-Result)
  - adverbial that can appear in 3 different positions

    - Vorfeld (pre-field)
      [Peter war der beste Torschütze.]$_{Arg1}$
      [Demzufolge]$_{conn}$ *[bekam er den Pokal.]*$_{Arg2}$
    - Mittelfeld (middle-field)
      (...) Er bekam demzufolge den Pokal.
    - Nullstelle (zero position)
      (...) Demzufolge: Er bekam den Pokal.

Case study: *demzufolge*

- Reading 2: connective (Cause-Result)
  - adverbial that can appear in 3 different positions

    - Vorfeld (pre-field)
      [Peter war der beste Torschütze.]$_{Arg1}$
      [Demzufolge]$_{conn}$ *[bekam er den Pokal.]*$_{Arg2}$
    - Mittelfeld (middle-field)
      (…) Er bekam demzufolge den Pokal.
    - Nullstelle (zero position)
      (…) Demzufolge: Er bekam den Pokal.

=> Readings 1 and 2 cannot be easily distinguished with surface-based methods

Corpus study

- 140 instances of *demzufolge* from www.dwds.de
  - zeit50: from print/online editions of weekly newspaper
  - kernel90: from ‚Kernkorpus20', a mixed-genre corpus of 20th-century German
- Window of three sentences; sentence 2 contains *demzufolge*

Corpus study

- For an initial overview, one author annotated kernel90: antecedents and their syntactic types
  - NP: 42 (47%)
    - *demzufolge* as relative pronoun ('according to which'): 33 (37%)
    - *demzufolge* in other function ('therefore'): 9 (10%)
  - VP ('therefore'): 19 (21%)
  - S ('therefore'): 29 (32%)
    - different S-types (see paper)

Corpus study

- For an initial overview, one author annotated kernel90: antecedents and their syntactic types
  - NP: 42 (47%)
    - *demzufolge* as relative pronoun (,according to which'): 33 (37%)
    - *demzufolge* in other function (,therefore'): 9 (10%)
  - VP (,therefore'): 19 (21%)
  - S (,therefore'): 29 (32%)
    - different S-types (see paper)

Balance between antecedent types and between readings (translations)
=> there is no simple majority-based disambiguation

Corpus study: annotator agreement

- IAA for class, connective sense (PDTB taxonomy), argument spans
- One author + two trained annotators
- all 50 instances from zeit50
- sense tags: all from PDTB + non-conn + missing context

Corpus study: annotator agreement

- Three annotators => 150 pairs of annotations
- 103 pairs (69%) completely identical

- Senses:
  - 25 pair disagreements (21 on non-/conn)
  - missing context was used only twice
  - cause-result: 39
  - specialization: 4
- Sense-labeling as 4-way classification task
  Fleiss-kappa for three raters = 0.55

- Arguments:
  - 32 pair disagreements on Arg1 span
  - 18 pair disagreements on Arg2 span

Toward disambiguation and resolution

- For the 40 explicitly-anaphoric connectives, need to
  - disambiguate the reading: non-/conn
  - resolve the arguments or antecedents

- Pilot study: Does POS tagging help?

POS tagging for disambiguation?

- kernel90 data set
- clevertagger (part of ParZu parser, Sennrich et al. 09)
- tagger of MATE tools (Bohnet 10)
- trained on different treebanks with slightly different tagsets
  - PROAV = PROP (pronominal adverb)

POS tagging for disambiguation?

antecedent	PROAV PROP	KOUS	PRELS	other
NP (relpro)	22		6	5
	11	19		3
NP (other)	9			
	8			1
VP	19			
	19			
S	24			
	23			2

Table 1: kernel90 dataset: POS tags assigned to *demzufolge* by the parsers MATE (first row in a cell) and ParZu (second row).

Summary and outlook

- Open question: Do explicitly-anaphoric connectives behave differently from non-explicitly-anaphoric ones?
  - disambiguation non-/conn reading
  - finding arguments/antecedents
  - sense disambiguation
- German: 79 explicitly-anaphoric connectives, 40 of which also have a non-connective reading
- Pilot study on *demzufolge*
  - corpus, agreement study
  - POS tagging helps only to small extent
- Next: Other connectives – check for differences to demzufolge, build classes

thank you!

overflow

Corpus study
with fictitious short examples for illustration

- kernel90: antecedents and their syntactic types
  - NP: 42 (47%)
    - *demzufolge* as relative pronoun ('according to which'): 33 (37%)
      Ich las ein Buch, demzufolge die Welt untergehen wird.
    - *demzufolge* in other function ('therefore'): 9 (10%)
      Es gab viele Hunde und demzufolge viel Gebell.
  - VP ('therefore'): 19 (21%)
    Welche Kinder gesund sind und demzufolge mitfahren dürfen, entscheidet die Lehrerin.
  - S ('therefore'): 29 (32%)
    Fast alle werden mitkommen. Demzufolge brauchen wir zwei Busse.
    - different S-types (see paper)

```xml
<entry id="k173" word="während">
    <syn>
        <cat>subj</cat>
        <sem>
            <coherence_relations>
                <synchronous />
                <contrast />
            </coherence_relations>
        </sem>
    </syn>
    <syn>
        <cat>praep</cat>
        <praep>
            <ante>1</ante> <post>0</post> <circum>0</circum>
            <case>gen</case>
        </praep>
        <sem>
            <coherence_relations>
                <synchronous />
            </coherence_relations>
        </sem>
    </syn>
</entry>
```

Why „specialization"?

- This relation can be compatible with cause-result
- *[Im ARD-Deutschlandtrend liegt Merkel in der Wählergunst deutlich hinter ihren möglichen Herausforderern Steinbrück und Steinmeier.]*Arg1
  *[Bei einer Direktwahl des Regierungschefs würde sie [demzufolge]*conn *im Duell gegen Steinbrück zurzeit mit 37 zu 48 Prozent klar unterliegen.]*Arg2
- 'In the ARD poll, Merkel clearly lags behind her challengers Steinbrück and Steinmeier. In a direct election of the chancellor, she would thus currently lose to Steinbrück with 37 against 48 percent.'