

# Connective-based Local Coherence Analysis

A Lexicon for Recognizing Causal Relationships

Manfred Stede

Applied Computational Linguistics  
Universität Potsdam  
Germany

## Overview

- Text parsing + Local coherence analysis
- Complex connectives in German
- A lexicon of (causal) connectives
- Toward automatic local coherence analysis

## Text parsing: A simple case

Because well-formed XML does not permit raw less-than signs and ampersands, if you use a character reference such as `&#60;`; or the entity reference `&lt;`; to insert the `<` character, the formatter will output `&lt;`; or perhaps `&#60;`;

## Text parsing: A simple case

**Because** well-formed XML does not permit raw less-than signs and ampersands, **if** you use a character reference such as `&#60;`; **or** the entity reference `&lt;`; **to** insert the `<` character, the formatter will output `&lt;`; **or** perhaps `&#60;`;

## Text parsing: A simple case

Because well-formed XML does not permit raw less-than signs and ampersands, if you use a character reference such as &#60; or the entity reference &lt; to insert the < character, the formatter will output &lt; or perhaps &#60;.

Because A, if B or C to D, E or F.

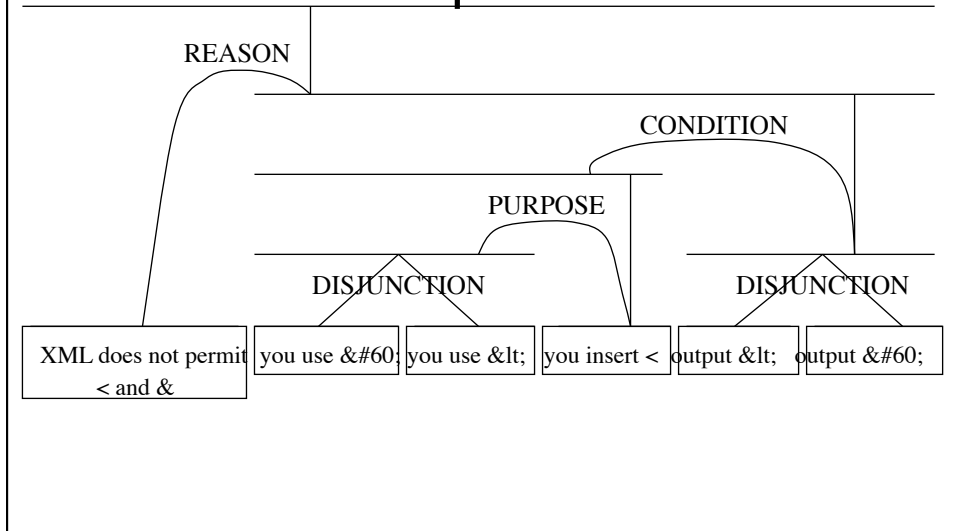
## Text parsing: A simple case

Because well-formed XML does not permit raw less-than signs and ampersands, if you use a character reference such as &#60; or the entity reference &lt; to insert the < character, the formatter will output &lt; or perhaps &#60;.

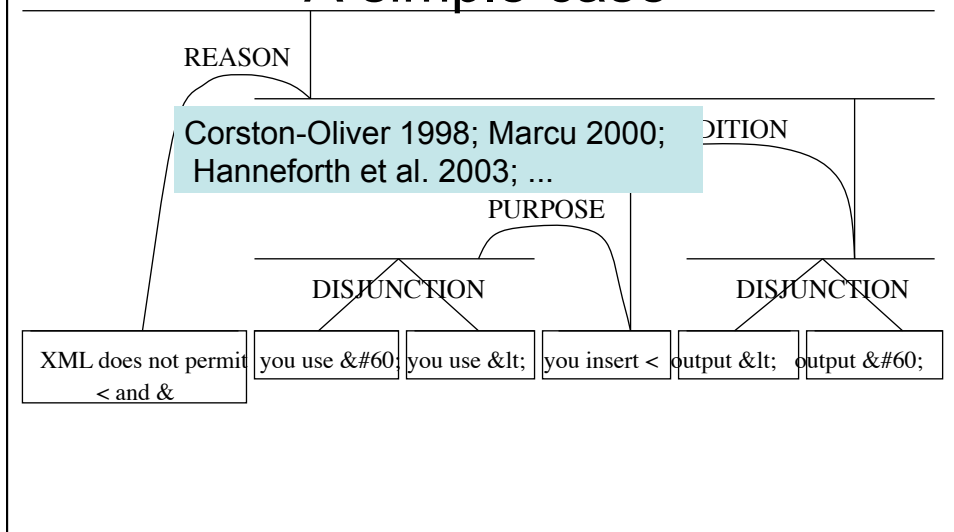
Because A, if B or C to D, E or F.

(Because A, (if (B or C) to D)), (E or F))).

## Text parsing: A simple case



## Text parsing: A simple case



## Connectives: Complications

- Segment embedding
  - We will, because time is short, discuss only one project.
- Long-distance antecedents
  - Meine Firma ist verschuldet. Ich fliege nach Rom, weil ich **deswegen** Urlaub habe.  
*My company is in debt. I'm flying to Rome, because **due to that** I'm on holiday.*
- Within/-out scope of focus particle or negation
  - Ich fahre mit dem Auto, **weil** ich auf Reisen gern rauche.  
*I go by car, **because** I like to smoke on my trips.*
  - Ich fahre **nur** mit dem Auto, **weil** ich auf Reisen gern rauche.  
*I **only** go by car, **because** I like to smoke on my trips.*

## Connectives: Complications

- Ambiguity 1: connective or no connective?
  - Wir sollten zu Abend essen, **denn** es wird spät.  
*We should have dinner, **for** it's getting late.*
  - Was ist **denn** das für ein Abendessen?  
*What sort of dinner is that?*
  - **Da** es spät wird, sollten wir jetzt Abendessen.  
***Since** it's getting late, we should have dinner now.*
  - Ich kenne ein gutes Restaurant. **Da** sollten wir abendessen.  
*I know a good restaurant. **There** we should have dinner.*
- Ambiguity 2: which coherence relation?
  - **If** Tom brings the food, **then** I'll bring the wine.
  - At eight, Tom will bring the food. **Then** I'll bring the wine.

## Connectives: Complications

- Absence: Only 30% of relations signalled by a connective  
(Schauer/Hahn 1998)
- Syntax as replacement: tense, aspect, mood, modality; constructions  
*Had I not forgotten the car keys, we would be in London by now.*
- The slippery slope to assorted open-class lexical items
  - *The car didn't start. **Thus** we took the train.*
  - *Consequently, we took the train.*
  - *As a consequence, we took the train.*
  - *As a dreadful consequence, we took the train.*
  - *The consequence was that we took the train.*
  - *This prompted us to take the train.*
- Multiple connectives
  - *John noticed that she lied, **but then** he married her **anyway**.*

## Connectives: Complications

- Complex connectives
  - ***Either** we take the car, **or** I won't join you on the trip.*
  - *Ich fahre nur mit dem Auto, **weil** ich auf Reisen **nämlich** gern rauche.  
*I only go by car, because I like to smoke on my trips.*  
(note: with *nämlich*, the *nur* is likely to have narrow scope)*
  - *Ich würde ihn nicht heiraten, **wenn** er **auch** reich ist.  
*I wouldn't marry him, **even though** he's rich.**
  - *Er ist **dadurch** zu seinem Geld gekommen, **dass** er täglich die Börsenkurse studiert.  
*He made his money **for that reason that** he monitors the stock exchange on a daily basis.**
  - *Sie hat ihn **deshalb** geheiratet, **weil** er soviel Geld hat.  
*She **therefore**, married him, **because**, he has so much money.**

- Local coherence analysis:
  - For an individual instance of a relation, decide on the ranges of the spans and on the relation conveyed
  - Here:  
consider only relations that are signalled explicitly by a connective

- Local coherence analysis:
  - For an individual instance of a relation, decide on the ranges of the spans and on the relation conveyed
  - Here:  
consider only relations that are signalled explicitly by a connective
- Text parsing:
  - construct a relation-based tree for the complete text
  - for longish text: structural decisions very difficult for human annotators
  - for longish text: cognitively plausible?
  - construct application-dependent structures?

## Overview

- Text parsing + Local coherence analysis
- Complex connectives in German
- A lexicon of (causal) connectives
- Toward automatic local coherence analysis

## Connectives

- 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, ...*
- NO: phrasemes like *for this reason*
- NO: affixes like *-wise / -halber, -wegen*
  - *Moneywise, the new job made no sense for Tom.*
  - *Ihretwegen kann ich nicht zur Party kommen.*  
,Because of her I cannot come to the party.'



## Complex connectives

- 1 word  
*since, although, especially, ...*
- >1 continuous words  
*due to, in particular, for example, ...*
- >1 discontinuous words  
*if .. then, but .. anyway, ...*

## Group W: Components within a clause

- W1: circumpositions
  - *Kannst Du mir **um** unserer Freundschaft **willen** einen Gefallen tun?*  
,Can you do me a favour **for** our friendship's **sake**?‘
  - ***Auf** ihren Rat **hin** habe ich mir einen Mac gekauft.*  
,**On** her advice I bought a Mac.‘

## Group W: Components within a clause

- W2: conjunction + adverbial form a signal non-compositionally
  - *Ich nehme diesen Mac, wenn er auch ziemlich teuer ist.*  
,I take this Mac if it is also quite expensive.'  
,I take this Mac, even though it's quite expensive.'

## Group W: Components within a clause

- W3: {adverbial | conjunction} + adverbial signal a single relation (+ additional facet)
  - *Das Wetter war schlecht, aber ich bin trotzdem in den See gesprungen.*  
,The weather was bad, but I jumped into the lake anyway.'
  - *Ich fliege nach Berlin, weil dort nämlich ein Bruder von mir wohnt.*  
,I'm flying to Berlin because a brother of mine is living there.'
  - *Ich bin dann anschließend noch nach Berlin geflogen.*  
,I then subsequently flew to Berlin.'

## Group W: Components within a clause

- W3: {adverbial | conjunction} + adverbial signal a single relation (+ additional facet)
  - (a) *Meine Firma ist verschuldet.* (b) *Ich fliege nach Rom, weil* (c) *ich nämlich Urlaub habe.*  
 ‚(a) My company is in debt. (b) I'm flying to Rome, (c) because I'm on holiday.‘
  - (a) *Meine Firma ist verschuldet.* (b) *Ich fliege nach Rom,* (c) *weil ich deswegen Urlaub habe.*  
 ‚(a) My company is in debt. (b) I'm flying to Rome, (c) because due to that I'm on holiday.‘

## Group A: Components across clauses

- A1: ‚split‘ connectives
  - *entweder .. oder* (‚either .. or‘)
  - *wenn .. dann* (‚if .. then‘)
  - *einerseits .. andererseits* (‚on the one hand ..‘)
- Complications:
  - One part can be optional, and the other a ‚simple‘ connective in its own right
  - Some of these connectives offer variants:  
*wenn .. dann* = *wenn .. so*
- ¬ .. *sondern* (‚¬ .. but‘) / ¬ .. *noch* (‚¬ .. nor‘)
  - *Er hat den Aufsatz nicht selbst erdacht, sondern geklaut.*  
 ‚He did not invent the essay himself but stole it.‘

## Group A: Components across clauses

- A2: adverbial + minor clause
- *Dadurch, dass mein Auto kaputt ist, kann ich heute nicht zum Treffen kommen.*  
,Due to my car being broken I can't come to the meeting today.'
- *Er ist insofern kein Experte, als er das Fach nie studiert hat.*  
,He is no expert insofar as he never studied the subject.'
- *Sie hatte so lange gewartet, dass ihr die Füße weh taten.*  
,She had waited so long that her feet hurt.'

## Group A: Components across clauses

- A2: adverbial + minor clause
- *Wir haben uns auf einer Party getroffen. Ich habe ihn dadurch kennengelernt.*  
,We ran into each other at a party. I got to know him through that.'
- *Wir haben uns auf einer Party getroffen. Ich habe ihn dadurch kennengelernt, dass wir auf der Tanzfläche übereinander gestolpert sind.*  
,We ran into each other at a party. I got to know him through the fact that we stumbled over each other on the dancefloor.'

## Group A: Components across clauses

- A3: „redundant“ cataphoric adv + conjunction
- *Ich bin deswegen nach Hause gefahren, weil es geregnet hat.*  
,I went home for the reason that it rained‘
- (a) *Es wurde sehr kalt.* (b) *Ich bin deswegen schnell nach Hause gefahren,* (c) *weil ich ja auch keinen Pullover hatte.*  
,(a) It got very cold. (b) I thus drove home quickly, also because I didn't have a sweater.‘

## Overview

- Text parsing + Local coherence analysis
- Complex connectives in German
- A lexicon of (causal) connectives
- Toward automatic local coherence analysis

## DiMLex Discourse Marker Lexicon

- Version 1: 135 German connectives (Stede 03)
- Basic technical approach:  
„Theory-neutral“ rich source lexicon in XML  
mapped via XSLT to specific application resources
  - Language generation in „Polibox“ (LISP) (Stede 02)
  - Rhetorical parsing prototype (Prolog) (Hanneforth et al. 03)
  - Various HTML views for the human user

## DiMLex Discourse Marker Lexicon

- Version 2: currently built in joint project with IDS Mannheim
- Focus on 45 *causal* connectives (in the widest sense) (Klein 1981)
  - Event causation: *The vase broke because it fell to the floor*
  - Explanation: *For many people, German is difficult as a second language because of the case system*
  - Argument: *We should nominate Obama by acclamation, because it will send a strong signal of unity*
- Causal coherence relations used in our system
  - Volitional-Cause
  - Nonvolitional-Cause
  - Purpose
  - Pragmatic-Claim

## DiMLex Discourse Marker Lexicon

- Orthography and syntax
- Disambiguation info:  
non-/connective
- Semantics/Pragmatics
  - Coherence relation(s)
  - Disambiguation info
  - Role linking
  - Idiosyncratic features

## DiMLex Discourse Marker Lexicon

- |   |  |
|---|--|
| • <b>Orthography</b> and syntax           | <i>aufgrund</i>  |
| • Disambiguation info:<br>non-/connective | <otype="simple" ctype="cont" onr="k2-1"><br>aufgrund   |
| • Semantics/Pragmatics                    | <otype="complex" ctype="cont" onr="k2-2"><br>auf Grund |
| – Coherence relation(s)                   |  |
| – Disambiguation info                     | <otype="simple" ctype="cont" onr="k2-3"><br>Aufgrund   |
| – Role linking                            |  |
| – Idiosyncratic features                  | <otype="complex" ctype="cont" onr="k2-4"><br>Auf Grund |

## DiMLex Discourse Marker Lexicon

- |   |  |
|---|--|
| <ul style="list-style-type: none"> <li>• Orthography and <b>syntax</b></li> <li>• Disambiguation info:<br/>non-/connective</li> <li>• Semantics/Pragmatics             <ul style="list-style-type: none"> <li>– Coherence relation(s)</li> <li>– Disambiguation info</li> <li>– Role linking</li> <li>– Idiosyncratic features</li> </ul> </li> </ul> | <p>syntactic category:<br/>coord-conj, subord-conj<br/>preposition<br/>prepadv, pronadv<br/>...</p> <p>order of spans (w/ corpus frequency):<br/>int-ext<br/>ext-int<br/>ext-int-ext</p> <p>average length of spans (from corpus)</p> <p>with focus particle: yes/no</p> <p>can be correlate: yes/no</p> <p>can have correlate: yes/no</p> |
|---|--|

## DiMLex Discourse Marker Lexicon

- |   |   |
|---|---|
| <ul style="list-style-type: none"> <li>• Orthography and <b>syntax</b></li> <li>• Disambiguation info:<br/>non-/connective</li> <li>• Semantics/Pragmatics             <ul style="list-style-type: none"> <li>– Coherence relation(s)</li> <li>– Disambiguation info</li> <li>– Role linking</li> <li>– Idiosyncratic features</li> </ul> </li> </ul> | <p><i>dadurch</i></p> <p>prepadv<br/>vorfeld:1<br/>mittelfeld:1<br/>nacherst:0<br/>nachfeld:1<br/>nullstelle:0<br/>nachnachfeld:0<br/>satzklammer:0</p> <p><i>Dadurch ist es geschehen.<br/>Es ist dadurch geschehen.<br/>Es ist geschehen dadurch.</i></p> |
|---|---|





# DiMLex

## Discourse Marker Lexicon

- Orthography and syntax *daher*
- Disambiguation info: non-/connective
  - Semantics/Pragmatics
    - Coherence relation(s)
    - Disambiguation info
    - Role linking
    - Idiosyncratic features

<pro value="90" ref="k5v2">  
\$. \$\$/PROAV

<pro value="90" ref="k5v1">  
VVFIN \$\$/PROAV

<con value="99" ref="k5v1 k5v2">  
\$\$/PROAV \$, {'dass'}/KOUS

<con value="95" ref="k5v2">  
\$. \$\$/PROAV .\* {'kommen' 'ruehren'} .+ \$,  
{'dass'}/KOUS

<con value="99" ref="k5v1">  
\$\$/PROAV \$.

# DiMLex

## Discourse Marker Lexicon

- Orthography and syntax *schließlich*
- Disambiguation info: non-/connective
  - Semantics/Pragmatics
    - Coherence relation(s)
    - Disambiguation info
    - Role linking
    - Idiosyncratic features

Pragmatic-Claim  
ext=effect; int=cause

- PRO
  - Tense=present & Aktionsart=state - 99
  - Parallel sentence structure - 80
  - Preceding segment: illocution=claim - 90
- CON
  - Verb elided - 99
  - Position=Vorfeld, following PP or NP - 99
  - Collocation „und schließlich" - 80

# DiMLex

## Discourse Marker Lexicon

- Orthography and syntax *schließlich*
- Disambiguation info: non-/connective Temporal-Sequence  
ext=preceding; int=subsequent
- **Semantics/Pragmatics**
  - Coherence relation(s)
  - Disambiguation info
  - Role linking
  - Idiosyncratic features
- PRO
  - Position=in temporal minor clause - 99
  - LinOrder= minor clause, tense=pluperfect - main clause - 80
- CON
  - Mood=subjunctive - 70
  - Tense=present - 70
  - Modality=„subjective“ - 70
  - Position=Mittelfeld - 60

## Complex connectives in DiMLex

- Basic idea:  
One component is *primary*, the other *secondary*
- Primary component is used to „anchor“ the complex connective in the text
- In lexicon entry for primary component, specify a search procedure for the secondary component
- Criteria for primaries
  - the component that is obligatory
  - the component that is structurally more prominent
  - the component that occurs less often in corpora

## Complex connectives in DiMLex

- W1: circumpositions
  - primary: postposition (rare)
  - **um** der Gesundheit **willen**
  - Left:NP & Left:"um"
- W2/W3: conjunction + adverbial
  - primary: conjunction (structural prominence)
  - ..., **weil** ich **nämlich** Urlaub habe
  - Right2EndOfClause:non-anaporic-adv/CAUSE

## Complex connectives in DiMLex

- A1: split connectives
  - primary: obligatory component
  - **Wenn** das Wetter gut ist, **dann** nehmen wir das Boot.  
*If the weather is nice, then we take the boat.*
  - Right:Clause\* & Right:"dann"
- A2: adverbial - *dass/als* clause
  - primary: adverbial (rare)
  - Er ist **insofern** kein Experte, **als** er das Fach nie studiert hat.  
*He is no expert insofar as he never studied the subject.*
  - Right:Clause & Right:"als"

## Complex connectives in DiMLex

- A3: „redundant“ cataphoric adv + conjunction
  - primary: conjunction (obligatory, structural prominence)
  - Ich bin **deswegen** nach Hause gefahren, **weil** es geregnet hat.  
*I went home **for the reason that** it rained.*
  - Left:Clause & Within:adv-anaphoric-CAUSE

## Overview

- Text parsing + Local coherence analysis
- Complex connectives in German
- A lexicon of (causal) connectives
- **Toward automatic local coherence analysis**

## A corpus of (signalled) causes and reasons

- 200 hotel reviews from a website (doyoo.de)
- All potential causal connectives labelled automatically (using DiMLex), then filtered manually
- 1200 instances of causal connectives
- Annotation instructions: how to decide on the spans related by the connective

## A corpus of (signalled) causes and reasons

- *The beach was very nice. Because it was cleaned every morning, there was hardly any garbage. We went there basically every day.*

## A corpus of (signalled) causes and reasons

- *The beach was very nice. **Because** it was cleaned every morning, there was hardly any garbage. We went there basically every day.*

## A corpus of (signalled) causes and reasons

- *The beach was very nice. **Because** [it was cleaned every morning.]<sub>C</sub> [there was hardly any garbage.]<sub>E</sub> We went there basically every day.*

## Complications

- <...> *And also, the staff was quite friendly.*<sub>C</sub>  
**Thus** *[we altogether enjoyed our holidays.]*<sub>E</sub>  
Next time, we'll <...>
- *[The beach was not very pleasant,]*<sub>E</sub> **as** *[it was,]*<sub>Ca</sub> I just have to say this here, *[utterly littered with remains of picnics.]*<sub>Cb</sub>
- *[We reached the hotel late]*<sub>E</sub> **due to** *[the flight's delay]*<sub>C1</sub> and also **because**<sub>Conn2</sub> *[it took so long to find a cab.]*<sub>C2</sub>

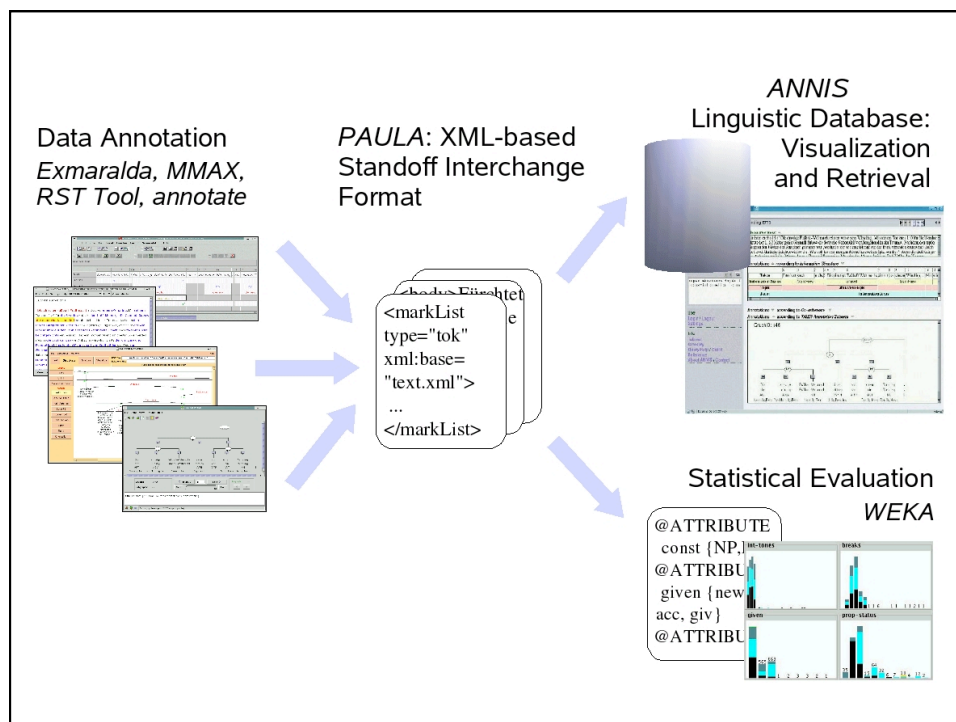
## A corpus of (signalled) causes and reasons

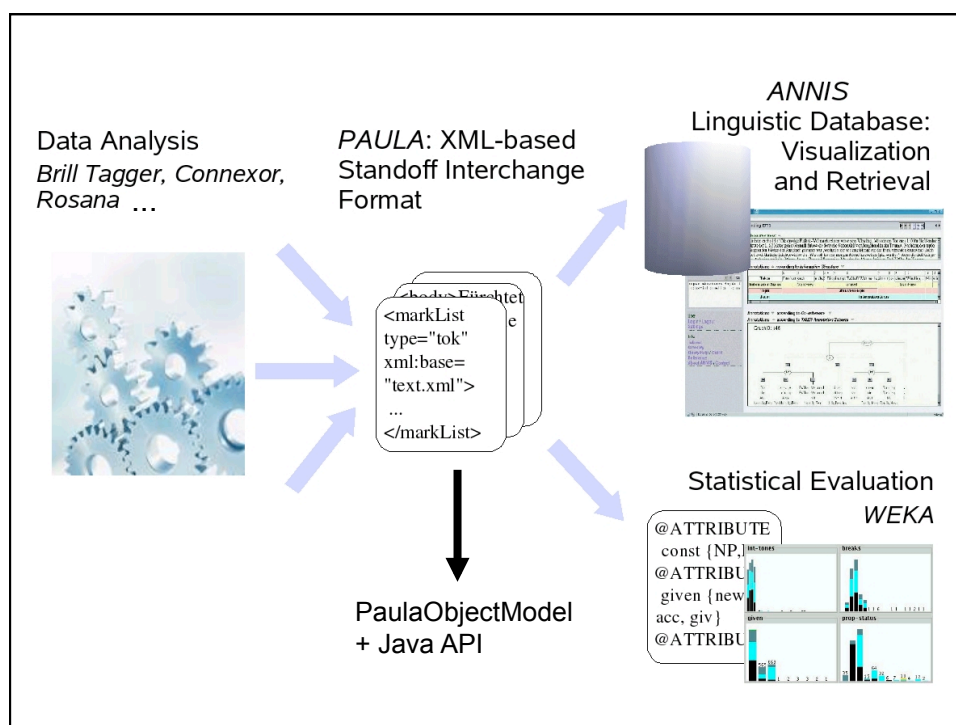
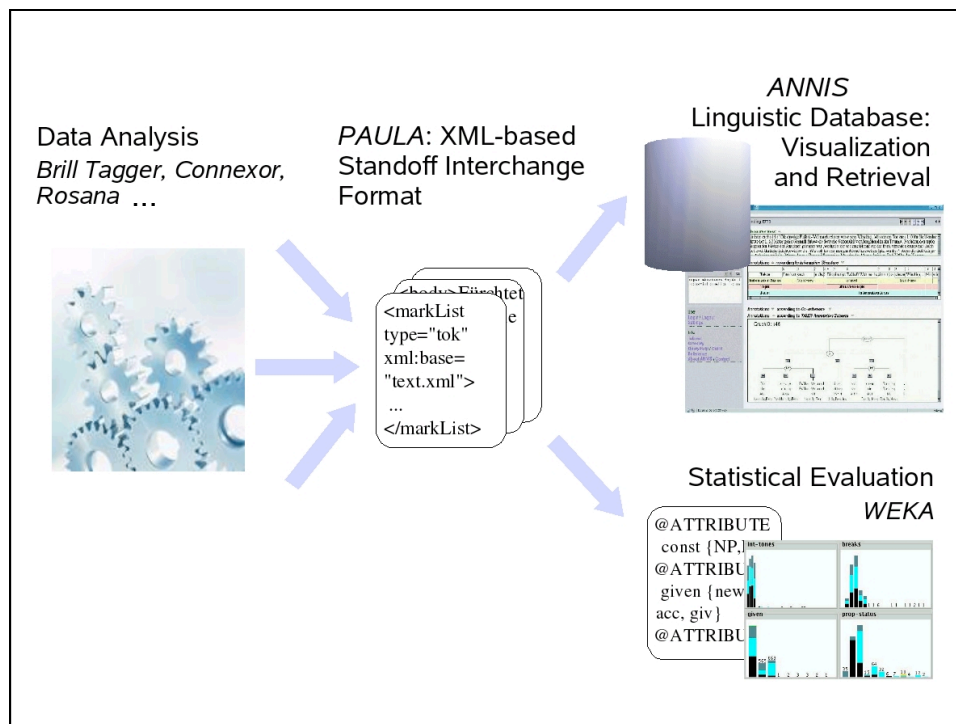
- Besides, annotators decided on the illocutions underlying the spans:
  - **report**: *The beach was not far away*
  - **report-self**: *We went to the beach every day*
  - **ident**: *I felt sorry for the staff*
  - **ident-judge**: *We all were not keen on the pool*
  - **ident-assume**: *The staff probably was bored*
  - **commit**: *I'll come back next year*
  - **direct**: *Avoid the city center around lunchtime*
  - **hypothetic-situation**: **In order to** *be on time, go...*



# Exploiting the corpus

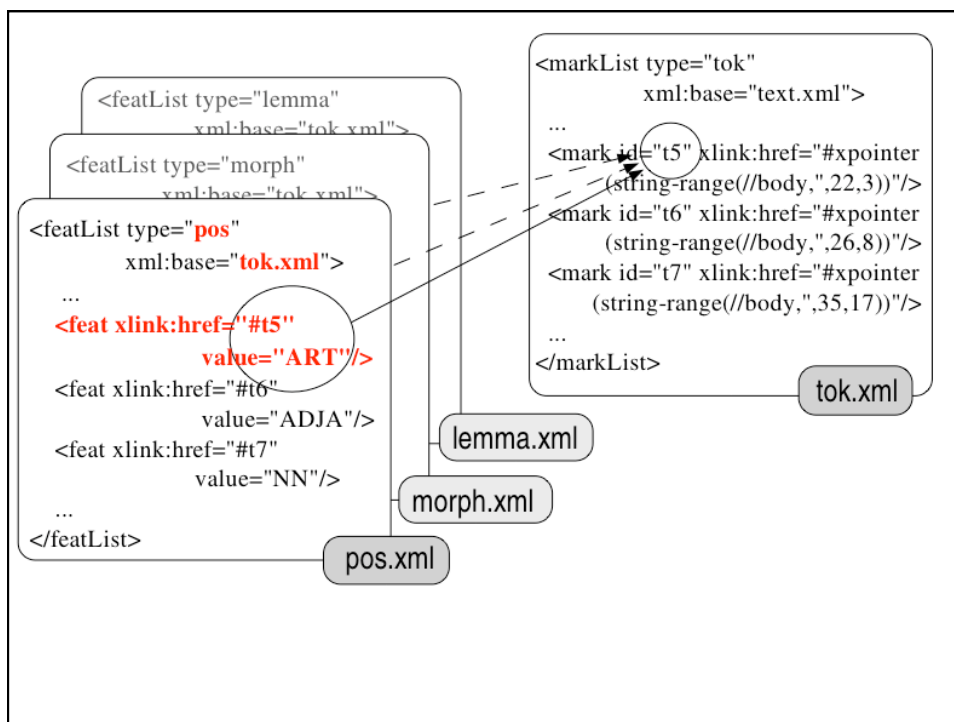
- Frequency counts for DiMLex:
  - order of spans
  - length of external spans
  - etc.
- Overall distribution of causal connectives (genre-specific?)
- Correlations between connective choice and illocutions?
- ...

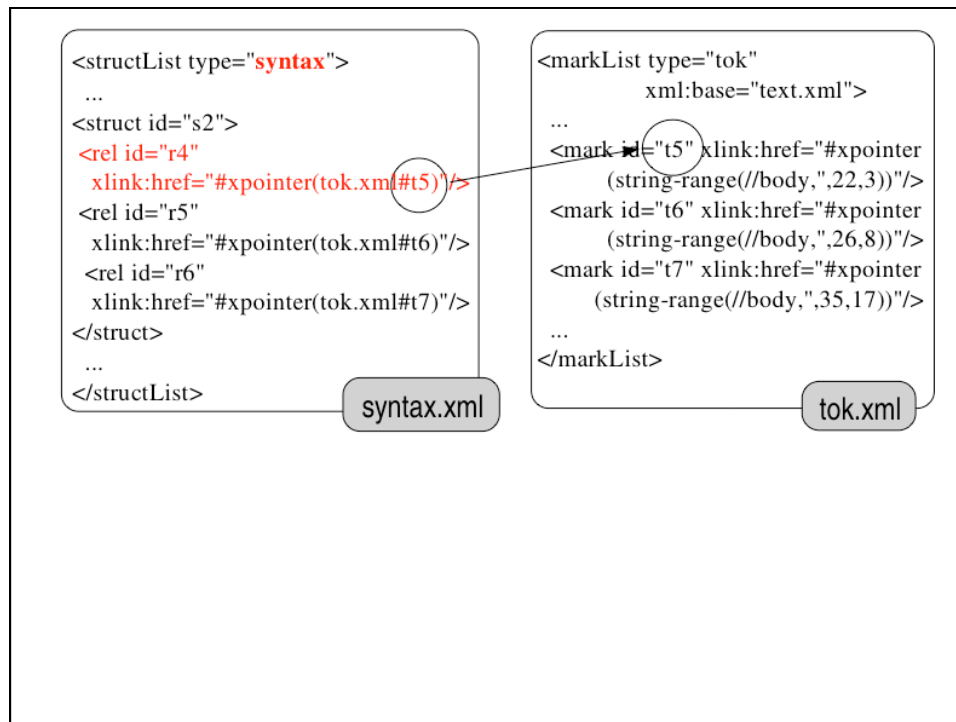




# PAULA

- Potsdamer AUstauschformat für Linguistische Annotation  
*Potsdam exchange format for linguistic annotation*
- Generic standoff XML annotation format
- Compatible with LAF (Ide et al. 07)
- Distribute separate annotation layers to separate files
- Employ a general *merge* script to unify separate layers into a „standard“ inline-XML file
- Various conversion scripts to/from PAULA available on our web site





## The LCA procedure

1. Identify connectives
2. Identify arguments of connectives (aka minimal discourse units)
3. Assign relations to connectives

# The LCA procedure (1)

- Identify connectives
  - Token matching between lexicon and text:  
attach labels to tokens in text
    - a) potential single connective
    - b) potential primary component of complex connectives
  - For (b), run the search procedure for secondary component
    - If found, run disambiguation rules from lexicon entry
    - If positive, remove the „single connective“ labels
  - Run disambiguation rules on single connectives



Suplex

https://intsfb.sfb632.uni-potsdam.de/homes/d1/summar/suplex-demo/view.php?b

Suplex Demo BöllStiftung web.de ZEIK Apple Amazon UniPo CL-Demos

Suplex filmrezension.de - onlin...

### Text

h[1][date] s[1] November 2002

h[2][tagline|title] s[2] Hoffnung in den Zeiten der Einsamkeit

h[3][tagline|title] s[3] Lantana

p[4][describe] s[4] Ein blühender Lantana-Strauch in der Sonne bildet den Anfang des Films . s[5] Orchestriert von dem Zirpen der Zikaden gleitet die Kamera langsam hinab in die scheinbar undurchdringliche Schönheit der Hibisken , um sich nach einer Fahrt durch die Schwärze im Geisterwald der stacheligen und dichten Stiele der Pflanze wiederzufinden . s[6] Eine surrende Fliege auf der Suche nach Verwesung begleitet den langsamen Blick auf eine Frauenleiche . s[7] Nahtlos geht die Szenerie über in einen Liebesakt . s[8] Als der Mann und die Frau später auseinandergehen , erklärt sie ihm , sie habe ihren Ohrring verloren . s[9] Der Zuschauer wird vermuten , dass er Beobachter einer Affäre war , die durch den verhängnisvollen Fund des verräterischen Beweisstückes an das Licht kommen wird . s[10] Fatale Begierde und tödliche Leidenschaft ? s[11] Musterhafte Assoziationen eines Thrillers werden freigesetzt und später jedoch kühl liegen gelassen . s[12] Die ködernde Kriminalhandlung wird sich lediglich beiläufig als roter Faden durch das episodische Psychodrama des australischen Regisseurs Ray Lawrence ziehen .

p[5][describe] s[13] Lantana dreht sich vor allem um vier Ehepaare . s[14] Da ist die Psychologin Dr. Valerie Somers und der Rechtsprofessor John Knox , die den Tod ihrer elfjährigen Tochter nicht verkraften können und sich über ihrer Trauer verloren haben . s[15] Der Polizist Leon Zat und seine Frau Sonja haben sich entfremdet . s[16] Während Sonja versucht , ihrem Mann wieder näher zu kommen , landet Leon im Bett mit Jane . s[17] Die hat ihren Mann Pete verlassen , weil sie herausfand , dass sie ihn nicht mehr liebt . s[18] James Nachbarn , Nik und Paula , bilden das einzige anscheinend glückliche Paar . s[19] Auch wenn Paula sich zuweilen genötigt sieht , mit eindeutigen Drohungen dafür zu sorgen , dass dies so bleibt : " Wenn du unsere Ehe aufs Spiel setzt , dann

### Source

tmp/out/backbone/text.backbor

load

<< | backbone | corefs | debug

### Annotations

☒ Sentences ID

☒ Paragraph/Header ID

☒ 50% sentences

☐ Connectives

☒ Tokens with POS:

☐ DET\_Def

☐ DET\_Indef

☐ N

☐ PREP

☐ PRON\_Dem

☐ PRON\_Pers

☐ PRON\_Poss

☐ PRON\_Rel

☐ Coreference chains:

☐ Lantana

☐ der Suche nach Verwesung

☒ der Mann

☐ der Mann und die Frau

☐ die Frau

☐ ihren Ohrring

☐ einer Affäre

☐ Die ködernde Kriminalhandlung

## Connective Disambiguation

Suplex

https://intsfb.sfb632.uni-potsdam.de/homes/d1/summar/suplex-demo/view.php?bb=tmp/out/bac

T-Online eMail Center LingMail WEB.DE ZEIK Sammlung Ar. r de Ganay DSLR-Forum LispMCL BSCW DissHyperTextLinguistik LEO ACL Anthology Amazon

### Text

Auch wenn die All-Inclusive-Anlage keine Wünsche offen lässt , sollte man die Insel auch erkunden , welche wirklich sehr reizvoll ist . Per gemieteten PKW oder mit geführter Tour . Hierzu will ich aber nicht weiter ausfuern , das ja um die Hotelanlage und nicht Grand Canaria geht . All das hat natürlich auch seinen Preis . Aber keine Sorge , denn mit bezahlten 1023,00 EURO für zwei Wochen finde ich es sehr preiswert . Denn wenn man sich mal ausrechnet , was man sonst alleine für Getränke und Speisen nebenbei ausgegeben hätte , wäre man sicher locker auch auf diesen Betrag gekommen und hätte ihn vermutlich weit überschritten , wenn man sich ein Hotel mit geringerem Grundpreis gesucht hätte .

### Connectives

☐ Number Expressions

☐ << Subjectivity

☒ << Brill: PoS, DC

☐ \$,

☐ \$.

☐ ADJA

☐ ADJD

☐ ADV

☐ APPR

☐ ART

☐ CARD

☒ DCno

☐ KON

☐ KOUS

☐ NE

☐ NN

☐ PDAT

☐ PDS

☐ PIAT

## The LCA procedure (2)

- Identify minimal discourse units  
(for the time being, only where connectives have been found)
  - Based on dependency parsing layer, attribution layer, and connective layer
  - Prepositions/Circumpositions: mark associated NP as internal arg and embedding clause as external arg
  - Subordinating conjunctions: mark embedded clause as internal arg and matrix clause as external arg
  - Coordinating conjunctions: (similar)
  - Adverbials: mark host clause as internal arg; for external arg, prefer span length learned from corpus, but respecting logical document structure (do not cross paragraph breaks unless necessary)

## The LCA procedure (3)

- Identify relations
  - For each relation stored with the connective in DiMLex, run disambiguation rules and accumulate weights
  - assign winning roles to internal and external arguments
- *Marianne entschied sich für das New Beetle Cabrio. Sie wollte schließlich ein richtiges Sommerauto.*
- *Mary chose the New Beetle convertible. Finally/After all, she wanted a real summer car.*
- Pragmatic-Claim: Effect / Cause
- Temporal sequence: Pre / Succ

## Summary

- *Local coherence analysis*: a potentially difficult sub-problem of text parsing
- Classification of complex connectives (for German)
- Representing knowledge of connectives in a computational lexicon that holds both linguistic and corpus-statistical information
- Corpus annotated with causal relations
- (Sketch of) Procedure for automatic local coherence analysis

## Credits

Heike Bieler (text analysis workbench), André Herzog (causality corpus), Kristin Irsig (DiMLex2), Uwe Küssner (connective analysis module), Andreas Peldszus (causality corpus, annotation guidelines)