

# Zur Annotation kausaler Konnektoren in Texten

André Herzog   Florian Hofmann   Andreas Peldszus  
Manfred Stede

Institut für Linguistik, Universität Potsdam

Konvens 2008

# Einstieg: Kohärenzrelationen. . .

- ... beschreiben den sinnvollen inhaltlichen Zusammenhalts der Struktur eines Textes.
- ... bestehen zwischen benachbarten Textsegmenten oder rekursiv zwischen den sich daraus ergebenden größeren Einheiten.
- ... können unterschiedlich "stark" markiert sein.

# Einstieg: Kohärenzrelationen. . .

- ... beschreiben den sinnvollen inhaltlichen Zusammenhalts der Struktur eines Textes.
- ... bestehen zwischen benachbarten Textsegmenten oder rekursiv zwischen den sich daraus ergebenden größeren Einheiten.
- ... können unterschiedlich "stark" markiert sein.

# Einstieg: Kohärenzrelationen. . .

- ... beschreiben den sinnvollen inhaltlichen Zusammenhalts der Struktur eines Textes.
- ... bestehen zwischen benachbarten Textsegmenten oder rekursiv zwischen den sich daraus ergebenden größeren Einheiten.
- ... können unterschiedlich "stark" markiert sein.

# Einstieg: Kohärenzrelationen

temporal, durch Konnektor markiert

Peter räumte den Tisch ab. **Dann** ging er schlafen.

additiv, phrasal markiert

Keiner der Kritiker hat eine konstruktive Lösung des Problems vorgeschlagen. **Hinzu kommt, dass** es vor einem Jahr von den meisten noch nicht einmal als Problem erkannt wurde.

kausal, nicht markiert

Das gestrige Gewitter brachte außergewöhnlich starke Niederschläge. Vielerorts waren die Straßen überschwemmt.

# Einstieg: Kohärenzrelationen

temporal, durch Konnektor markiert

Peter räumte den Tisch ab. **Dann** ging er schlafen.

additiv, phrasal markiert

Keiner der Kritiker hat eine konstruktive Lösung des Problems vorgeschlagen. **Hinzu kommt, dass** es vor einem Jahr von den meisten noch nicht einmal als Problem erkannt wurde.

kausal, nicht markiert

Das gestrige Gewitter brachte außergewöhnlich starke Niederschläge. Vielerorts waren die Straßen überschwemmt.

# Einstieg: Kohärenzrelationen

temporal, durch Konnektor markiert

Peter räumte den Tisch ab. **Dann** ging er schlafen.

additiv, phrasal markiert

Keiner der Kritiker hat eine konstruktive Lösung des Problems vorgeschlagen. **Hinzu kommt, dass** es vor einem Jahr von den meisten noch nicht einmal als Problem erkannt wurde.

kausal, nicht markiert

Das gestrige Gewitter brachte außergewöhnlich starke Niederschläge. Vielerorts waren die Straßen überschwemmt.

# Perspektive

- Automatische Ableitung von Kohärenzrelationen aus beliebigen Texten (Text-Parsing, Textverstehen)
- Ein Anwendungsgebiet spez. für kausale Kohärenz:  
Vom "opinion mining" zu einer Begründungsanalyse  
(nicht nur Meinungen sondern auch Gründe extrahieren)

# Perspektive

- Automatische Ableitung von Kohärenzrelationen aus beliebigen Texten (Text-Parsing, Textverstehen)
- Ein Anwendungsgebiet spez. für kausale Kohärenz:  
Vom "opinion mining" zu einer Begründungsanalyse  
(nicht nur Meinungen sondern auch Gründe extrahieren)

# Beschränkungen

- auf *kausale* Kohärenzrelation
- auf *Konnektor*-markierte Kohärenzrelation
- auf *lokale* Analyse der Kohärenzrelation

# Beschränkungen

- auf *kausale* Kohärenzrelation
- auf *Konnektor*-markierte Kohärenzrelation
- auf *lokale* Analyse der Kohärenzrelation

# Beschränkungen

- auf *kausale* Kohärenzrelation
- auf *Konnektor*-markierte Kohärenzrelation
- auf *lokale* Analyse der Kohärenzrelation

# Übersicht

- 1 Einstieg
- 2 Entwurf von Annotationsrichtlinien
- 3 Experimentelle Evaluation der Richtlinien
- 4 Vorstellung des Korpus
- 5 Ausblick

# Der Gegenstand der Annotation und seine Anforderungen an die Richtlinien

- Instruktionen für

- ① *Segmentierung des Textes in minimale begründungsrelevante Segmente*
- ② *Verknüpfung der Segmente und Konnektoren*
- ③ *Klassifikation der Segmente*

Die Badelandschaft des Hotels ist zufriedenstellend. Nur der Pool wurde schnell dreckig, **weswegen** ein Poolboy angestellt wurde. Wir haben ihn jedoch nie arbeiten sehen.

# Der Gegenstand der Annotation und seine Anforderungen an die Richtlinien

- Instruktionen für
  - 1 *Segmentierung* des Textes in minimale begründungsrelevante Segmente
  - 2 *Verknüpfung* der Segmente und Konnektoren
  - 3 *Klassifikation* der Segmente

Die Badelandschaft des Hotels ist zufriedenstellend. [ Nur der Pool wurde schnell dreckig ] , **weswegen** [ ein Poolboy angestellt wurde ] . Wir haben ihn jedoch nie arbeiten sehen.

# Der Gegenstand der Annotation und seine Anforderungen an die Richtlinien

- Instruktionen für
  - 1 *Segmentierung* des Textes in minimale begründungsrelevante Segmente
  - 2 *Verknüpfung* der Segmente und Konnektoren
  - 3 *Klassifikation* der Segmente

Die Badelandschaft des Hotels ist zufriedenstellend.

[ Nur der Pool wurde schnell dreckig ] ,

weswegen [ ein Poolboy angestellt wurde ] .

Wir haben ihn jedoch nie arbeiten sehen.

# Der Gegenstand der Annotation und seine Anforderungen an die Richtlinien

- Instruktionen für
  - 1 *Segmentierung* des Textes in minimale begründungsrelevante Segmente
  - 2 *Verknüpfung* der Segmente und Konnektoren
  - 3 *Klassifikation* der Segmente

Die Badelandschaft des Hotels ist zufriedenstellend.

[ Nur der Pool wurde schnell dreckig ] ,

weswegen [ ein Poolboy angestellt wurde ]

Wir haben ihn noch nie arbeiten sehen.

illokution: report  
performativ: nein  
meta: nein

# Der Gegenstand der Annotation und seine Anforderungen an die Richtlinien

- Instruktionen für
  - 1 *Segmentierung* des Textes in minimale begründungsrelevante Segmente
  - 2 *Verknüpfung* der Segmente und Konnektoren
  - 3 *Klassifikation* der Segmente

Die Badelandschaft des Hotels ist zu befriedigend.

[ N...ool wurde schnell dreckig ] ,

weswegen [ ein Poolboy angestellt ]

Wir haben ihn noch nie arbeiten sehen.

illokution: report  
performativ: nein  
meta: nein

illokution: report  
performativ: nein  
meta: nein

# Auswahl der Konnektoren

- Basierend auf den Arbeiten der AG Konnektoren vom IDS Mannheim wurden folgende Konnektoren zur Vormarkierung ausgewählt:

also, aufgrund, da, dadurch, daher, damit, darum, dementsprechend, demnach, denn, deshalb, deswegen, doch, durch, halber, nämlich, schließlich, so, somit, so...dass, um, wegen, weil, weshalb

# Vorgriff: Textsorte

- Wunsch: hohe Dichte kausaler Konnektoren, viele Bewertungen (beschreiben, urteilen und begründen)
- Auswahl: 250 Hotelrezensionen vom deutschen Meinungsportal `dooyou.de`
- Nachteil: teilweise mangelhafte Rechtschreibung, Interpunktion und Grammatik
- Chance: Erfahrung sammeln mit dem Normalfall von Benutzer-generierten Texten (Robustheit der Richtlinien)

# Vorgriff: Textsorte

- Wunsch: hohe Dichte kausaler Konnektoren, viele Bewertungen (beschreiben, urteilen und begründen)
- Auswahl: 250 Hotelrezensionen vom deutschen Meinungsportal `dooyou.de`
- Nachteil: teilweise mangelhafte Rechtschreibung, Interpunktion und Grammatik
- Chance: Erfahrung sammeln mit dem Normalfall von Benutzer-generierten Texten (Robustheit der Richtlinien)

# Vorgriff: Textsorte

- Wunsch: hohe Dichte kausaler Konnektoren, viele Bewertungen (beschreiben, urteilen und begründen)
- Auswahl: 250 Hotelrezensionen vom deutschen Meinungsportal `dooyou.de`
- Nachteil: teilweise mangelhafte Rechtschreibung, Interpunktion und Grammatik
- Chance: Erfahrung sammeln mit dem Normalfall von Benutzer-generierten Texten (Robustheit der Richtlinien)

# Vorgriff: Textsorte

- Wunsch: hohe Dichte kausaler Konnektoren, viele Bewertungen (beschreiben, urteilen und begründen)
- Auswahl: 250 Hotelrezensionen vom deutschen Meinungsportal `dooyou.de`
- Nachteil: teilweise mangelhafte Rechtschreibung, Interpunktion und Grammatik
- Chance: Erfahrung sammeln mit dem Normalfall von Benutzer-generierten Texten (Robustheit der Richtlinien)

# Vorgriff: Textsorte

## Beispiel

couchtisch: zu klein und zu kurz, daher unbrauchbar.

## Beispiel

Für alle, welche nicht Sonnenanbeter oder Hitze-Fanatiker sind, sollten doch lieber in der Vor- oder Nachsaison fliegen, da es ansonsten vorkommen kann, das es Temperaturen jenseits der 40° hat , wie es bei uns der Fall war.

# Vorgriff: Textsorte

## Beispiel

couchtisch: zu klein und zu kurz, daher unbrauchbar.

## Beispiel

Für alle, welche nicht Sonnenanbeter oder Hitze-Fanatiker sind, sollten doch lieber in der Vor- oder Nachsaison fliegen, da es ansonsten vorkommen kann, das es Temperaturen jenseits der 40° hat , wie es bei uns der Fall war.

# Herausforderungen an die Richtlinien

## Denkbar einfachster Fall:

[ Der Pool wurde schnell dreckig ] . **Deswegen** [ wurde ein Poolboy angestellt ] .

## Nominalisierungen:

Der Poolboy reagierte nicht auf [ meine Beschwerde ] **wegen** [ des dreckigen Pools ] .

## Koordination:

[ Der Strand war abgelegen ] und **deswegen** [ nicht so überlaufen ] .

## Diskontinuität:

[ Es kann ] **wegen** [ der Einflugschneise ] [ zu Fluglärm kommen ] .

# Herausforderungen an die Richtlinien

## Denkbar einfachster Fall:

[ Der Pool wurde schnell dreckig ] . **Deswegen** [ wurde ein Poolboy angestellt ] .

## Nominalisierungen:

Der Poolboy reagierte nicht auf [ meine Beschwerde ] **wegen** [ des dreckigen Pools ] .

## Koordination:

[ Der Strand war abgelegen ] und **deswegen** [ nicht so überlaufen ] .

## Diskontinuität:

[ Es kann ] **wegen** [ der Einflugschneise ] [ zu Fluglärm kommen ] .

# Herausforderungen an die Richtlinien

## Denkbar einfachster Fall:

[ Der Pool wurde schnell dreckig ] . **Deswegen** [ wurde ein Poolboy angestellt ] .

## Nominalisierungen:

Der Poolboy reagierte nicht auf [ meine Beschwerde ] **wegen** [ des dreckigen Pools ] .

## Koordination:

[ Der Strand war abgelegen ] und **deswegen** [ nicht so überlaufen ] .

## Diskontinuität:

[ Es kann ] **wegen** [ der Einflugschneise ] [ zu Fluglärm kommen ] .

# Herausforderungen an die Richtlinien

## Denkbar einfachster Fall:

[ Der Pool wurde schnell dreckig ] . **Deswegen** [ wurde ein Poolboy angestellt ] .

## Nominalisierungen:

Der Poolboy reagierte nicht auf [ meine Beschwerde ] **wegen** [ des dreckigen Pools ] .

## Koordination:

[ Der Strand war abgelegen ] und **deswegen** [ nicht so überlaufen ] .

## Diskontinuität:

[ Es kann ] **wegen** [ der Einflugschneise ] [ zu Fluglärm kommen ] .

# Herausforderungen an die Richtlinien

## Mehrfachbegründungen:

**Weil** [ uns kalt war ] , [ gingen wir zurück in die geheizte Lobby ] und [ tranken heißen Tee ] .

## Doppelt vorkommende Konnektoren:

[ Wir buchten All Inclusive ] **da** [ wir uns keinen Kopf ums Geld machen wollten ] und **da** [ es nicht viel mehr kostete ] .

## Mehrere unterschiedliche Konnektoren:

**Aufgrund** [ der stinkenden Mülltonnen draußen ] und **weil** [ der Ausblick so grässlich war ] , [ kann ich von diesem Zimmer nur abraten ] .

# Herausforderungen an die Richtlinien

## Mehrfachbegründungen:

**Weil** [ uns kalt war ] , [ gingen wir zurück in die geheizte Lobby ] und [ tranken heißen Tee ] .

## Doppelt vorkommende Konnektoren:

[ Wir buchten All Inclusive ] **da** [ wir uns keinen Kopf ums Geld machen wollten ] und **da** [ es nicht viel mehr kostete ] .

## Mehrere unterschiedliche Konnektoren:

**Aufgrund** [ der stinkenden Mülltonnen draußen ] und **weil** [ der Ausblick so grässlich war ] , [ kann ich von diesem Zimmer nur abraten ] .

# Herausforderungen an die Richtlinien

## Mehrfachbegründungen:

**Weil** [ uns kalt war ] , [ gingen wir zurück in die geheizte Lobby ] und [ tranken heißen Tee ] .

## Doppelt vorkommende Konnektoren:

[ Wir buchten All Inclusive ] **da** [ wir uns keinen Kopf ums Geld machen wollten ] und **da** [ es nicht viel mehr kostete ] .

## Mehrere unterschiedliche Konnektoren:

**Aufgrund** [ der stinkenden Mülltonnen draußen ] und **weil** [ der Ausblick so grässlich war ] , [ kann ich von diesem Zimmer nur abraten ] .

# Herausforderungen an die Richtlinien

## Variable Reihenfolge:

Sowohl **wegen** [ der ruhigen Lage ] [ kann ich das Hotel empfehlen ] ,  
als auch **weil** [ man dort behandelt wird wie ein König ] .

## Korrelate:

[ Ich habe mir ] *nur* **deswegen** [ ein Tattoo machen lassen ] , **weil** [ ich betrunken war ] .

## Konnektoren mit weitem Skopus:

... *viele Gründe vorher*... **Daher** [ gebe ich diesem Hotel 10 von 10 Punkte ] !

# Herausforderungen an die Richtlinien

## Variable Reihenfolge:

Sowohl **wegen** [ der ruhigen Lage ] [ kann ich das Hotel empfehlen ] ,  
als auch **weil** [ man dort behandelt wird wie ein König ] .

## Korrelate:

[ Ich habe mir ] *nur* **deswegen** [ ein Tattoo machen lassen ] , **weil** [ ich betrunken war ] .

## Konnektoren mit weitem Skopus:

... *viele Gründe vorher*... **Daher** [ gebe ich diesem Hotel 10 von 10 Punkte ] !

# Herausforderungen an die Richtlinien

## Variable Reihenfolge:

Sowohl **wegen** [ der ruhigen Lage ] [ kann ich das Hotel empfehlen ] ,  
als auch **weil** [ man dort behandelt wird wie ein König ] .

## Korrelate:

[ Ich habe mir ] *nur* **deswegen** [ ein Tattoo machen lassen ] , **weil** [ ich betrunken war ] .

## Konnektoren mit weitem Skopus:

... *viele Gründe vorher*... **Daher** [ gebe ich diesem Hotel 10 von 10 Punkte ] !

# Unterschiedliche Verwendungsweisen von kausalen Verknüpfungen

*Semantisch:* Beziehung zwischen zwei Ereignissen "in der Welt"

Es hat Frost gegeben, weil die Temperatur über Nacht stark gesunken ist.

*Pragmatisch:* Beziehung zwischen zwei Sprechakten

Es hat Frost gegeben, denn die Dahlien sind ganz schwarz.

- Beschreiben und Begründen unterscheiden zu können, ist nicht zuletzt für eine Argumentationsstruktur notwendig.
- Der Konnektor allein liefert aber oft nicht genug Evidenz.
- Ergo: Illokutionsindikatoren beachten!

# Unterschiedliche Verwendungsweisen von kausalen Verknüpfungen

*Semantisch:* Beziehung zwischen zwei Ereignissen "in der Welt"

Es hat Frost gegeben, weil die Temperatur über Nacht stark gesunken ist.

*Pragmatisch:* Beziehung zwischen zwei Sprechakten

Es hat Frost gegeben, denn die Dahlien sind ganz schwarz.

- Beschreiben und Begründen unterscheiden zu können, ist nicht zuletzt für eine Argumentationsstruktur notwendig.
- Der Konnektor allein liefert aber oft nicht genug Evidenz.
- Ergo: Illokutionsindikatoren beachten!

# Unterschiedliche Verwendungsweisen von kausalen Verknüpfungen

*Semantisch:* Beziehung zwischen zwei Ereignissen "in der Welt"

Es hat Frost gegeben, weil die Temperatur über Nacht stark gesunken ist.

*Pragmatisch:* Beziehung zwischen zwei Sprechakten

Es hat Frost gegeben, denn die Dahlien sind ganz schwarz.

- Beschreiben und Begründen unterscheiden zu können, ist nicht zuletzt für eine Argumentationsstruktur notwendig.
- Der Konnektor allein liefert aber oft nicht genug Evidenz.
- Ergo: Illokutionsindikatoren beachten!

# Unterschiedliche Verwendungsweisen von kausalen Verknüpfungen

*Semantisch:* Beziehung zwischen zwei Ereignissen "in der Welt"

Es hat Frost gegeben, weil die Temperatur über Nacht stark gesunken ist.

*Pragmatisch:* Beziehung zwischen zwei Sprechakten

Es hat Frost gegeben, denn die Dahlien sind ganz schwarz.

- Beschreiben und Begründen unterscheiden zu können, ist nicht zuletzt für eine Argumentationsstruktur notwendig.
- Der Konnektor allein liefert aber oft nicht genug Evidenz.
- Ergo: Illokutionsindikatoren beachten!

# Unterschiedliche Verwendungsweisen von kausalen Verknüpfungen

*Semantisch:* Beziehung zwischen zwei Ereignissen "in der Welt"

Es hat Frost gegeben, weil die Temperatur über Nacht stark gesunken ist.

*Pragmatisch:* Beziehung zwischen zwei Sprechakten

Es hat **wohl** Frost gegeben, denn die Dahlien sind ganz schwarz.

- Beschreiben und Begründen unterscheiden zu können, ist nicht zuletzt für eine Argumentationsstruktur notwendig.
- Der Konnektor allein liefert aber oft nicht genug Evidenz.
- Ergo: Illokutionsindikatoren beachten!

# Illokutionen

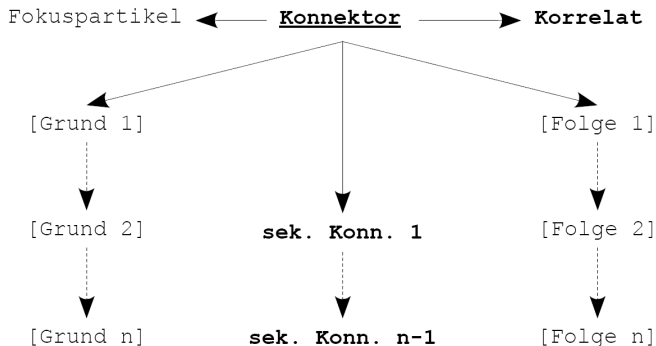
- Inventar von Illokutionen nach (Schmitt, 2000) mit leichten Modifikationen:

# Illokutionen

- Inventar von Illokutionen nach (Schmitt, 2000) mit leichten Modifikationen:

Illokutionstyp	Beschreibung
Reportiva <sub>neut</sub>	Der Autor berichtet von vergangene Ereignissen.
Reportiva <sub>ident</sub>	Der Autor berichtet und stellt sich dabei in den Mittelpunkt.
Identifikativa	Der Autor beschreibt eigene Gefühle und Wünsche.
Evaluativa	Der Autor nimmt eine Wertung (gut/schlecht) vor.
Estimativa	Der Autor stellt eine Vermutung auf.
Kommissiva	Der Autor verpflichtet sich, etwas zu tun.
Direktiva	Der Autor empfiehlt oder fordert zu einer Handlung auf.
Hypothetische Situation	Der Autor nennt einen nicht-realisierten, angestrebten Sachverhalt.

# Beschreibung der Richtlinien

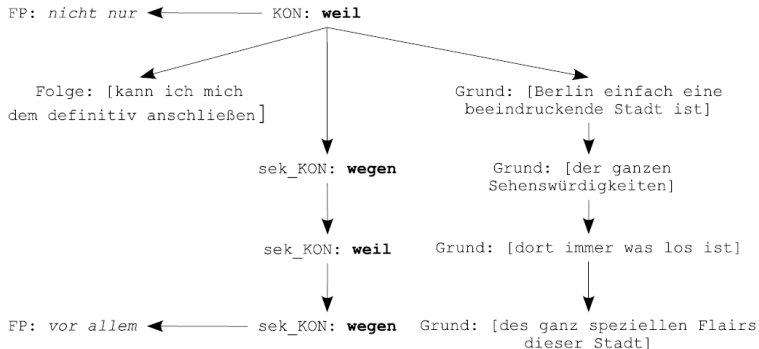


# Beschreibung der Richtlinien: Ein Beispiel

## Ein komplexeres Beispiel

Ein Kurzbesuch in Berlin lohnt sich immer , stand auf meinem handlichen Reiseführer . Und wenn ich so auf meinen Kurz-Trip zurückschaue , [ kann ich mich dem definitiv anschließen ] . *Nicht nur* , **weil** [ Berlin einfach eine beeindruckende Stadt ist ] , wegen [ der ganzen Sehenswürdigkeiten ] oder weil [ dort immer was los ist ] , sondern *vor allem* wegen [ des ganz speziellen Flairs dieser Stadt ] .

# Beschreibung der Richtlinien: Ein Beispiel



# Rekapitulation

- Identifikation von Konnektoren ist gegeben
- Segmentierung der begründungsrelevanten Textpassagen
- Verknüpfung von Konnektoren und Segmenten
- Klassifikation der Segmente nach

# Rekapitulation

- Identifikation von Konnektoren ist gegeben
- Segmentierung der begründungsrelevanten Textpassagen
- Verknüpfung von Konnektoren und Segmenten
- Klassifikation der Segmente nach

# Rekapitulation

- Identifikation von Konnektoren ist gegeben
- Segmentierung der begründungsrelevanten Textpassagen
- Verknüpfung von Konnektoren und Segmenten
- Klassifikation der Segmente nach
  - Illokutionstyp
  - Vorhandensein einer performativen Formel
  - Modalität

# Rekapitulation

- Identifikation von Konnektoren ist gegeben
- Segmentierung der begründungsrelevanten Textpassagen
- Verknüpfung von Konnektoren und Segmenten
- Klassifikation der Segmente nach
  - 1 Illokutionstyp
  - 2 Vorhandensein einer performativen Formel
  - 3 Meta-Bezug auf Texterstellungshandlung oder Medium

# Rekapitulation

- Identifikation von Konnektoren ist gegeben
- Segmentierung der begründungsrelevanten Textpassagen
- Verknüpfung von Konnektoren und Segmenten
- Klassifikation der Segmente nach
  - 1 Illokutionstyp
  - 2 Vorhandensein einer performativen Formel
  - 3 Meta-Bezug auf Texterstellungshandlung oder Medium

# Rekapitulation

- Identifikation von Konnektoren ist gegeben
- Segmentierung der begründungsrelevanten Textpassagen
- Verknüpfung von Konnektoren und Segmenten
- Klassifikation der Segmente nach
  - 1 Illokutionstyp
  - 2 Vorhandensein einer performativen Formel
  - 3 Meta-Bezug auf Texterstellungshandlung oder Medium

# Rekapitulation

- Identifikation von Konnektoren ist gegeben
- Segmentierung der begründungsrelevanten Textpassagen
- Verknüpfung von Konnektoren und Segmenten
- Klassifikation der Segmente nach
  - 1 Illokutionstyp
  - 2 Vorhandensein einer performativen Formel
  - 3 Meta-Bezug auf Texterstellungshandlung oder Medium

# Experimentelle Evaluation der Richtlinien

- Zwei Probanden annotieren jeweils 11 Texten (insg. 78 Konnektoren).
- Annotationswerkzeug: MMAX2 (EML Heidelberg)
- Freie Einarbeitung durch zwei Video-Tutorials und die Annotationsrichtlinien (13 S.): ca. 90 min
- Dauer der eigentlichen Annotation: 3 bis 3,5 h.

# Experimentelle Evaluation der Richtlinien

- Zwei Probanden annotieren jeweils 11 Texten (insg. 78 Konnektoren).
- Annotationswerkzeug: MMAX2 (EML Heidelberg)
- Freie Einarbeitung durch zwei Video-Tutorials und die Annotationsrichtlinien (13 S.): ca. 90 min
- Dauer der eigentlichen Annotation: 3 bis 3,5 h.

# Experimentelle Evaluation der Richtlinien

- Zwei Probanden annotieren jeweils 11 Texten (insg. 78 Konnektoren).
- Annotationswerkzeug: MMAX2 (EML Heidelberg)
- Freie Einarbeitung durch zwei Video-Tutorials und die Annotationsrichtlinien (13 S.): ca. 90 min
- Dauer der eigentlichen Annotation: 3 bis 3,5 h.

# Experimentelle Evaluation der Richtlinien

- Zwei Probanden annotieren jeweils 11 Texten (insg. 78 Konnektoren).
- Annotationswerkzeug: MMAX2 (EML Heidelberg)
- Freie Einarbeitung durch zwei Video-Tutorials und die Annotationsrichtlinien (13 S.): ca. 90 min
- Dauer der eigentlichen Annotation: 3 bis 3,5 h.

# Experiment: Evaluation

hrev\_019 / markable\_9

	man	sollte	wegen	Durchfallgefahr	nicht	das	Leitungswasser	zum	Waschen	benutzen
<b>Annotator1</b>	F1: dir	K1	G1: rep	F1: dir						
<b>Annotator2</b>	F1: dir	K1	G1: rep	F1: dir						

# Experiment: Evaluation

hrev\_019 / markable\_9

man sollte wegen Durchfallgefahr nicht das Leitungswasser zum Waschen benutzen

**Annotator1** F1: dir K1 G1: rep F1: dir

**Annotator2** F1: dir K1 G1: rep F1: dir

hrev\_099 / markable\_4

Das Green House , also das Hauptrestaurant , hatte aufgrund der enormen Deckenhöhe ein wenig den Charme einer Bahnhofshalle

**Annotator1** F1: jud F1: jud K1 G1: rep F1: jud

**Annotator2** F1: jud K1 G1: rep F1: jud

# Experiment: Evaluation der Segmentierung

- Fehlerklassen für die quantitative Evaluation der Segmentierung:
  - 1 Disjunktheit
  - 2 Inklusion/Überlappung
  - 3 Untersegmentierung
  - 4 Tilgung
  - 5 Typverletzung

# Experiment: Evaluation der Segmentierung

- Fehlerklassen für die quantitative Evaluation der Segmentierung:

- 1 Disjunktheit
- 2 Inklusion/Überlappung
- 3 Untersegmentierung
- 4 Tilgung
- 5 Typverletzung



# Experiment: Evaluation der Segmentierung

- Fehlerklassen für die quantitative Evaluation der Segmentierung:

- 1 Disjunktheit
- 2 Inklusion/Überlappung
- 3 Untersegmentierung
- 4 Tilgung
- 5 Typverletzung



# Experiment: Evaluation der Segmentierung

- Fehlerklassen für die quantitative Evaluation der Segmentierung:
  - 1 Disjunktheit
  - 2 Inklusion/Überlappung
  - 3 Untersegmentierung
  - 4 Tilgung
  - 5 Typverletzung



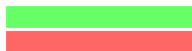
# Experiment: Evaluation der Segmentierung

- Fehlerklassen für die quantitative Evaluation der Segmentierung:
  - 1 Disjunktheit
  - 2 Inklusion/Überlappung
  - 3 Untersegmentierung
  - 4 Tilgung
  - 5 Typverletzung



# Experiment: Evaluation der Segmentierung

- Fehlerklassen für die quantitative Evaluation der Segmentierung:
  - 1 Disjunktheit
  - 2 Inklusion/Überlappung
  - 3 Untersegmentierung
  - 4 Tilgung
  - 5 Typverletzung



# Experiment: Evaluation der Segmentierung

- 34 der 78 Konnektoren wurden übereinstimmend segmentiert (keine Fehlerklasse)
- 30 Konnektoren fielen in eine Fehlerklasse, 11 in zwei, 3 in drei.

# Experiment: Evaluation der Segmentierung

- 34 der 78 Konnektoren wurden übereinstimmend segmentiert (keine Fehlerklasse)
- 30 Konnektoren fielen in eine Fehlerklasse, 11 in zwei, 3 in drei.

# Experiment: Evaluation der Segmentierung

- 34 der 78 Konnektoren wurden übereinstimmend segmentiert (keine Fehlerklasse)
- 30 Konnektoren fielen in eine Fehlerklasse, 11 in zwei, 3 in drei.

Fehlerklasse	Vorkommen
Inklusion/Überlapp.	36
Tilgung	7
Untersegmentierung	6
Disjunktheit	4
Typverletzung	2

# Experiment: Evaluation der Illokutionszuweisung

## Vergleichbarkeits-Kriterium

Zwei Segmente sind bzgl. ihrer Illokutionszuweisung vergleichbar

- (1.) im Fall einer geringfügigen Inklusion/Überlappung oder
- (2.) im Fall einer Untersegmentierung in illokutions-gleiche Teilstimente oder
- (3.) bei identischen Segmenten.

- prozentuale Grad an Übereinstimmung: 0,71
- kappa-Werts (Erwartung 0,43): 0,492
- nach Kombination von *Reportiva* und *Evaluativa* kappa-Wert: 0,575

# Experiment: Evaluation der Illokutionszuweisung

## Vergleichbarkeits-Kriterium

Zwei Segmente sind bzgl. ihrer Illokutionszuweisung vergleichbar

- (1.) im Fall einer geringfügigen Inklusion/Überlappung oder
- (2.) im Fall einer Untersegmentierung in illokutions-gleiche Teilsegmente oder
- (3.) bei identischen Segmenten.

- prozentuale Grad an Übereinstimmung: 0,71
- kappa-Werts (Erwartung 0,43): 0,492
- nach Kombination von *Reportiva* und *Evaluativa* kappa-Wert: 0,575

# Experiment: Evaluation der Illokutionszuweisung

## Vergleichbarkeits-Kriterium

Zwei Segmente sind bzgl. ihrer Illokutionszuweisung vergleichbar

- (1.) im Fall einer geringfügigen Inklusion/Überlappung oder
- (2.) im Fall einer Untersegmentierung in illokutions-gleiche Teilsegmente oder
- (3.) bei identischen Segmenten.

- prozentuale Grad an Übereinstimmung: 0,71
- kappa-Werts (Erwartung 0,43): 0,492
- nach Kombination von *Reportiva* und *Evaluativa* kappa-Wert: 0,575

# Experiment: Evaluation der Illokutionszuweisung

## Vergleichbarkeits-Kriterium

Zwei Segmente sind bzgl. ihrer Illokutionszuweisung vergleichbar

- (1.) im Fall einer geringfügigen Inklusion/Überlappung oder
- (2.) im Fall einer Untersegmentierung in illokutions-gleiche Teilsegmente oder
- (3.) bei identischen Segmenten.

- prozentuale Grad an Übereinstimmung: 0,71
- kappa-Werts (Erwartung 0,43): 0,492
- nach Kombination von *Reportiva* und *Evaluativa* kappa-Wert: 0,575

# Experiment: Evaluation der Illokutionszuweisung

## Vergleichbarkeits-Kriterium

Zwei Segmente sind bzgl. ihrer Illokutionszuweisung vergleichbar

- (1.) im Fall einer geringfügigen Inklusion/Überlappung oder
- (2.) im Fall einer Untersegmentierung in illokutions-gleiche Teilsegmente oder
- (3.) bei identischen Segmenten.

- prozentuale Grad an Übereinstimmung: 0,71
- kappa-Werts (Erwartung 0,43): 0,492
- nach Kombination von *Reportiva* und *Evaluativa* kappa-Wert: 0,575

# Experiment: Evaluation der Illokutionszuweisung

## Vergleichbarkeits-Kriterium

Zwei Segmente sind bzgl. ihrer Illokutionszuweisung vergleichbar

- (1.) im Fall einer geringfügigen Inklusion/Überlappung oder
- (2.) im Fall einer Untersegmentierung in illokutions-gleiche Teilsegmente oder
- (3.) bei identischen Segmenten.

- prozentuale Grad an Übereinstimmung: 0,71
- kappa-Werts (Erwartung 0,43): 0,492
- nach Kombination von *Reportiva* und *Evaluativa* kappa-Wert: 0,575

# Experiment: Evaluation der Illokutionszuweisung

## Vergleichbarkeits-Kriterium

Zwei Segmente sind bzgl. ihrer Illokutionszuweisung vergleichbar

- (1.) im Fall einer geringfügigen Inklusion/Überlappung oder
- (2.) im Fall einer Untersegmentierung in illokutions-gleiche Teilsegmente oder
- (3.) bei identischen Segmenten.

- prozentuale Grad an Übereinstimmung: 0,71
- kappa-Werts (Erwartung 0,43): 0,492
- nach Kombination von *Reportiva* und *Evaluativa* kappa-Wert: 0,575

# Experiment: Konsequenzen

- Überarbeitung der Annotationsrichtlinien
  - Präzisierung des Minimalitätskriteriums für Segmente
  - Ausführlichere Beschreibung der Illokutionstypen

# Experiment: Konsequenzen

- Überarbeitung der Annotationsrichtlinien
  - Präzisierung des Minimalitätskriteriums für Segmente
  - Ausführlichere Beschreibung der Illokutionstypen

# Experiment: Konsequenzen

- Überarbeitung der Annotationsrichtlinien
  - Präzisierung des Minimalitätskriteriums für Segmente
  - Ausführlichere Beschreibung der Illokutionstypen

# Vorstellung des Korpus

- Annotation des gesamten Korpus (knapp 190000 Tokens, ca. 1100 Konnektoren, ca. 2350 Segmente)
- Entwicklung eines praktischen Korpus-Viewer
- Richtlinien, Werkzeuge und Korpus zu finden unter:  
<http://www.ling.uni-potsdam.de/~stede/kausalannot.html>

# Vorstellung des Korpus

- Annotation des gesamten Korpus (knapp 190000 Tokens, ca. 1100 Konnektoren, ca. 2350 Segmente)
- Entwicklung eines praktischen Korpus-Viewer
- Richtlinien, Werkzeuge und Korpus zu finden unter:

<http://www.ling.uni-potsdam.de/~stede/kausalannot.html>

# Vorstellung des Korpus

- Annotation des gesamten Korpus (knapp 190000 Tokens, ca. 1100 Konnektoren, ca. 2350 Segmente)
- Entwicklung eines praktischen Korpus-Viewer
- Richtlinien, Werkzeuge und Korpus zu finden unter:

<http://www.ling.uni-potsdam.de/~stede/kausalannot.html>

# Ausblick

- Integration des Korpus in ANNIS (Demo in der Postersession)
- Automatische Erkennung von Begründungen in MOTS
- Erweiterung des Schemas um detaillierte Konnektor-Klassifikation
- Beschreibung und Annotation von nicht-Konnektor-markierten Kausalzusammenhängen
- Anwendung der bestehenden Annotationsrichtlinien auf englische Texte und Konnektoren

# Ausblick

- Integration des Korpus in ANNIS (Demo in der Postersession)
- Automatische Erkennung von Begründungen in MOTS
- Erweiterung des Schemas um detaillierte Konnektor-Klassifikation
- Beschreibung und Annotation von nicht-Konnektor-markierten Kausalzusammenhängen
- Anwendung der bestehenden Annotationsrichtlinien auf englische Texte und Konnektoren

# Ausblick

- Integration des Korpus in ANNIS (Demo in der Postersession)
- Automatische Erkennung von Begründungen in MOTS
  
- Erweiterung des Schemas um detaillierte Konnektor-Klassifikation
- Beschreibung und Annotation von nicht-Konnektor-markierten Kausalzusammenhängen
- Anwendung der bestehenden Annotationsrichtlinien auf englische Texte und Konnektoren

# Ausblick

- Integration des Korpus in ANNIS (Demo in der Postersession)
- Automatische Erkennung von Begründungen in MOTS
  
- Erweiterung des Schemas um detaillierte Konnektor-Klassifikation
- Beschreibung und Annotation von nicht-Konnektor-markierten Kausalzusammenhängen
- Anwendung der bestehenden Annotationsrichtlinien auf englische Texte und Konnektoren

# Ausblick

- Integration des Korpus in ANNIS (Demo in der Postersession)
- Automatische Erkennung von Begründungen in MOTS
  
- Erweiterung des Schemas um detaillierte Konnektor-Klassifikation
- Beschreibung und Annotation von nicht-Konnektor-markierten Kausalzusammenhängen
- Anwendung der bestehenden Annotationsrichtlinien auf englische Texte und Konnektoren

Vielen Dank für die Aufmerksamkeit!