

# Zur Annotation von kausalen Verknüpfungen in Texten

Andreas Peldszus, André Herzog, Florian Hofmann, Manfred Stede

**Zusammenfassung.** Kausale Verknüpfungen sind in Texten in verschiedenen pragmatischen Verwendungsweisen anzutreffen. Ihre automatische Erkennung stellt eine Teilaufgabe des “Text-Parsing” dar, bei dem für einen Text eine Struktur aus Kohärenzrelationen und Textsegmenten abgeleitet werden soll. Wir stellen eine Korpusstudie vor, bei der anhand ausführlicher Annotationsrichtlinien die Grund–Folge Beziehungen in Produktbesprechungen aus einer Internet-Seite zu erkennen waren, und diskutieren Perspektiven für die automatische Erkennung.

## 1 Einführung und Überblick

Im Laufe der 1990er Jahre entstanden, vor allem motiviert durch die weite Verbreitung der Rhetorical Structure Theory (Mann und Thompson 1988), eine Reihe von Arbeiten zur automatischen Erkennung von Diskursstruktur; stellvertretend seien hier für das Japanische Sumita et al. (1992) und für das Englische Corston-Oliver (1998), Marcu (2000), Hanneforth et al. (2003) genannt. Hier wurde jeweils das Ziel verfolgt, zu einem Text eine Baumstruktur abzuleiten, deren Blattknoten durch die (zunächst zu identifizierenden) “minimalen Diskurseinheiten” und deren interne Knoten durch Kohärenzrelationen gebildet werden, welche minimale Einheiten und dann rekursiv auch größere Textabschnitte miteinander verbinden. Für das Deutsche haben in jüngerer Zeit Lungen et al. (2006) einen entsprechenden Ansatz beschrieben. Diese Aufgabe des automatischen ‘Text-Parsing’ ist aus drei Gründen sehr kompliziert:

1. Auch für menschliche Annotatoren ist die Identifikation einer RST-artigen Struktur schwierig. Zwar berichten Carlson et al. (2003) für die Annotation von Texten des *Wall Street Journal* – eine eher “einfache” Textsorte – recht gute Inter-Annotator-Übereinstimmung, doch dies auf der Grundlage sehr expliziter Annotationsrichtlinien, die ca. 80 Kohärenzrelationen verwenden und für Zweifelsfälle jeweils eine bestimmte *default*-Entscheidung nahelegen.
2. Die Präsenz einer Kohärenzrelation ist im Text oftmals nicht an der Oberfläche erkennbar. Nach Schauer und Hahn (2001) sind nur etwa 1/3 aller Relationen in Texten durch einen Konnektor markiert. Konnektoren wiederum können

sehr spezifisch (z.B. *obwohl*) aber auch sehr vage (z.B. *und*) sein. Das bedeutet, dass die Zuweisung einer Relation in hohem Maße interpretationsabhängig ist und damit nur begrenzt mit oberflächennahen Methoden nachgebildet werden kann.

3. Bereits die Aufgabe der *Segmentierung*, also der Teilung des Texts in die minimalen Einheiten, die der relationalen Analyse zugrunde liegen, ist alles andere als einfach. Von Jasinskaja et al. (2007) wurden Segmentierungsrichtlinien (für menschliche Annotatoren) für das Deutsche vorgeschlagen, die die Bandbreite des Problems verdeutlichen; Längen et al. (2006) beschreiben einen automatischen Ansatz.

Auch hinsichtlich der theoretischen Merkmale der Diskursstruktur besteht zur Zeit wenig Konsens. Während die RST von einer (sehr verarbeitungsfreundlichen) Baumstruktur ohne kreuzende Kanten ausging, argumentieren Wolf und Gibson (2005), dass allgemeinere Graphen zur Repräsentation von Textstruktur erforderlich sind. Einen Mittelweg schlagen Webber et al. (2003) ein, indem sie zwischen ‘strukturellen’ und ‘anaphorischen’ Konnektoren unterscheiden: Die Analyse der ersteren ergebe eine Baumstruktur, die als ‘Gerüst’ der Textrepräsentation betrachtet wird. Die anaphorischen Konnektoren werden demgegenüber auf einer separaten Beschreibungsebene behandelt.

Aus all diesen Gründen wenden wir uns in diesem Papier einem einfacheren Problem zu und konzentrieren uns allein auf die lokale Behandlung bestimmter Kohärenzrelationen. Uns interessiert also nicht die Ableitung einer kompletten Baumstruktur für den Text, sondern nur die Generierung von Konnektor-bezogenen Hypothesen für einzelne Vorkommen von Relationen. Konkret beschäftigen wir uns mit Begründungszusammenhängen, also im weiten Sinne mit der Familie der *kausalen* Kohärenzrelationen. Um eine Arbeitsgrundlage für die Implementierung eines “Begründungserkenners” zu erhalten, haben wir ein Korpus mit begründenden Textpassagen zusammengestellt, Annotationsrichtlinien zur Markierung von Begründungszusammenhängen entworfen und nach einer Pilotstudie überarbeitet, und anschließend die Richtlinien in einem kontrollierten Experiment mit zwei Annotatoren getestet. Das Korpus und die Richtlinien sind über eine Webseite verfügbar.<sup>1</sup> Im Folgenden geben wir zunächst einen Überblick über kausale Kohärenzrelationen (Abschnitt 2) und beschreiben dann unsere Annotationsrichtlinien (Abschnitt 3), wobei einige Komplikationen bei der Behandlung authentischer Texte zur Sprache kommen werden. Abschnitt 4 beschreibt Design, Durchführung und Evaluation des Annotations-experiments, und Abschnitt 5 skizziert Perspektiven für die automatische Erkennung von Begründungen.

---

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

## 2 Kausale Verknüpfungen im Text

Wenn wir ‘Kausalität’ hier als Oberbegriff verwenden, so lassen sich darunter drei Formen der Verwendung kausaler Verknüpfungen in Sprechhandlungen subsumieren: das *Begründen*, das *Erklären*, und das *Argumentieren* (s. z.B. Klein 1987). Für alle lassen sich “prototypische” Verwendungen in entsprechenden Kontexten angeben, doch bei der Durchsicht von Korpusbelegen zeigt sich schnell, dass oft eine eindeutige Zuordnung kaum möglich ist. Etwas einfacher wird das Bild, wenn wir der Literatur zu Kohärenzrelationen folgen und eine Zweiteilung annehmen: *Pragmatische* Relationen bezeichnen danach einen Zusammenhang zwischen zwei Sprechhandlungen des Autors, während *semantische* Relationen einen kausalen Zusammenhang in “der Welt” beschreiben (s. z.B. Sanders et al. 1992). Diese Unterscheidung wird auch in der Konnektoren-Literatur häufig herangezogen; so hat bereits Pasch (1989) darauf hingewiesen, dass mit *denn* eine Behauptung begründet und mit *weil* (jedenfalls in seiner “traditionellen” Verwendung als Subjunktor) stattdessen ein kausaler Zusammenhang lediglich beschrieben wird:

- (1) a. Es hat wohl Frost gegeben, denn die Dahlien sind ganz schwarz.  
 b. ??Es hat wohl Frost gegeben, weil die Dahlien ganz schwarz sind.

Auch dies ist zwar nur eine Tendenzregel, doch zweifellos ist die Unterscheidung auch im Hinblick auf automatische Analyse bedeutsam: Es wäre für bestimmte Textsorten sehr nützlich, wenn Begründung und Beschreibung voneinander unterschieden werden könnten. Da Konnektoren allein aber keine zuverlässige Information liefern, muss man sich darüber hinaus mit *Illokutionsindikatoren* beschäftigen, wenn eine zumindest partielle automatische Analyse gelingen soll. Im obigen Beispiel (1) ist es die Modalpartikel *wohl*, die den ersten Satz als Hypothese ausweist.

Ein Inventar von Illokutionen als Verfeinerung der Sprechakt-Taxonomie von Searle (1976) wurde von Schmitt (2000) vorgeschlagen. Wir wählen es als Ausgangspunkt, um für unsere Zwecke (also mit dem Fokus auf Kausalität) folgende Illokutionstypen zu unterscheiden:

- Reportiva<sub>neur</sub>: Der Autor berichtet von vergangenen Ereignissen.
- Reportiva<sub>self</sub>: Der Autor berichtet über eigene Handlungen.
- Identifikativa: Der Autor beschreibt seine eigenen Gefühle oder Wünsche.
- Evaluativa: Der Autor nimmt eine Wertung vor.
- Estimativa: Der Autor stellt eine Vermutung auf.
- Kommissiva: Der Autor verpflichtet sich selbst, etwas zu tun.

- Direktiva: Der Autor empfiehlt oder fordert zu einer Handlung auf.
- Hypothetische Situation: Der Autor nennt einen nicht-realisierten Sachverhalt, wie im Antezedens eines *wenn-dann* Satzes oder in *um...zu* Konstruktionen.

Hiermit lässt sich eine feinkörnigere Analyse erzielen als mit einer Bezeichnung des Relationstyps (die ja keine spezifische Aussage über den Status der beiden Relata macht). Allerdings muss empirisch überprüft werden (und diesen Schritt hatte Schmitt nicht unternommen), ob Versuchspersonen in der Lage sind, die Unterscheidungen nachzuvollziehen und in Texten hinreichend übereinstimmend zu annotieren.

### 3 Annotationsrichtlinien

Aufgabe der Annotationsrichtlinien war es also, die Annotatoren mit Anweisungen und Beispielen für die Segmentierung des Textes in minimale begründungsbezogene Einheiten und deren Klassifizierung vertraut zu machen. Das hierfür entwickelte Schema sieht als grundlegende Einheiten der Annotation *Konnektoren*, *Segmente* (Gründe und Folgen) und *Fokuspartikeln* vor. Deren Zusammenspiel wird in Abbildung 1 schematisch dargestellt und im Folgenden nach einer kurzen Vorstellung der verschiedenen Einheiten an einem umfangreicheren Beispiel erläutert. Im Interesse der Überschaubarkeit haben wir uns entschlossen, die *Identifikation* von Konnektoren aus dem Experiment heraus zu halten; stattdessen wurden alle kausalen Konnektoren im Annotationsschema bereits vormarkiert. Die Liste von 32 Konnektoren wurde aus den Arbeiten der AG Konnektoren am IDS Mannheim (siehe Pasch et al. 2003)<sup>2</sup> gewonnen. Für dieses Experiment haben wir uns also bewusst auf *Konnektormarkierte* Kausalzusammenhänge beschränkt; solche, die an der Oberfläche gar nicht oder anderweitig signalisiert sind (durch Verben, Substantive, Phraseme) wurden zunächst nicht betrachtet.

#### 3.1 Das Annotationsschema und dessen Einheiten

Der *Konnektor* fungiert als Zentrum des Begründungszusammenhangs. Er verweist auf die ihm zugeordneten Segmente, auf eventuelle Sekundärkonnektoren, vorhandene Korrelate und – falls gegeben – auf eine Fokuspartikel, die sich direkt auf ihn bezieht. Da in dem Experiment die Konnektoren schon vormarkiert waren, bedurfte es in den Richtlinien keiner weiteren Hinweise zur Identifikation der zu annotierenden Konnektoren. Auch eine inhaltliche Klassifikation des Konnektors ist an dieser

---

2. Einen guten Überblick vermittelt auch die Website <http://www.ids-mannheim.de/gra/konnektoren/>.

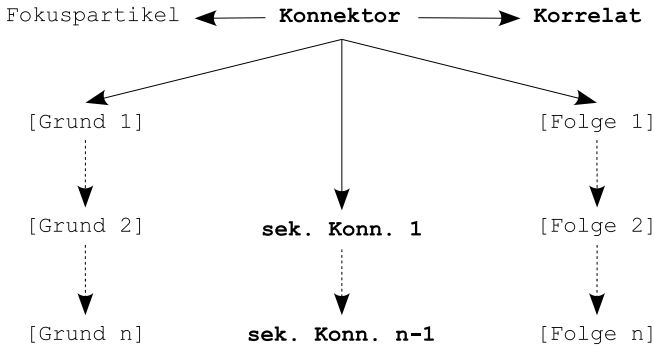


Abbildung 1. Komponenten des Annotationsschemas

Stelle nicht von den Annotatoren gefordert worden. (Einer entsprechenden Erweiterung des Schemas steht aber technisch nichts im Wege.)

Was die Segmente betrifft, verweist der Konnektor selbst nur auf *einen* Grund und *eine* Folge. In einem Begründungszusammenhang werden aber oft mehrere Gründe und Folgen angegeben, und diese sollen einzeln identifiziert werden. Um dem Konnektor mehrere Segmente zuzuordnen, werden diese untereinander verkettet, d.h. jede Einheit zeigt auf ihr Typ-gleiches Nachfolgeelement. Die Annotatoren waren gehalten die Segmente dem Textfluss folgend von links nach rechts zu verketten und den Konnektor auf das erste Glied der Kette zeigen zu lassen.

- (2) **Aufgrund** [ des unerträglichen Gestanks ] , wegen [ der schlechten Lage ] und weil [ ich dort zweimal bestohlen wurde ] , [ werde ich dieses Hotel nicht mehr besuchen ] .

Beispiel (2) verdeutlicht nicht nur mehrfache Gründe, sondern auch das mehrfache Auftreten von Konnektoren innerhalb einer Begründung. Diese eröffnen aber nicht jeder für sich einen separierbaren Begründungszusammenhang, sondern führen nur den zuerst aufgespannten fort. Um die Einheit der Begründung nicht auf mehrere Konnektoren aufzuteilen, die selbst nur einen Teil der Begründung beitragen und nicht als eigenständiger Konnektor verstanden werden könnten, entschieden wir uns für ein Schema, in dem der erste vorkommende Konnektor einer Begründung als der zentrale angesehen wird, während die (eventuell) darauf folgenden zugehörigen Konnektoren als *Sekundärkonnektor* markiert werden. Diese unterscheiden sich bezüglich der Implementierung des Schemas nicht vom zentralen Konnektor. Jedoch sollten die Annotatoren ihnen keine Segmente zuordnen, damit die Verkettung von Gründen und Folgen ununterbrochen bleibt (siehe Abbildung 1).

Auch die Annotation von Korrelaten ist im Schema vorgesehen. Der Konnektor zeigt dabei auf den redundanten Teil. Vorgreifend ist aber zu sagen, dass in den Experimenten die Annotation von Korrelaten noch nicht gefordert war.

Die Annotatoren hatten zudem die Aufgabe, im Text *Fokuspartikeln* zu markieren, insofern diese sich direkt auf den Konnektor beziehen. Von einem theoretisch naiven Begriff von Fokuspartikel ausgehend, wurden den Annotatoren als Beispiele hierfür *nur, sogar, auch, nicht, vor allem* etc. gegeben. Dieses Merkmal mag auf den ersten Blick nicht zentral für unsere Fragestellung sein, es scheint uns aber aus mehreren Gründen sinnvoll, diese Information in der Annotation nicht unberücksichtigt zu lassen. Denn zum einen sind damit bestimmte Fälle des Anführens eines Nicht-Grundes leichter identifizierbar (Beispiel: *Nicht wegen des Frostes sind die Dahlien ganz schwarz.*). Außerdem wird allgemein hervorgehoben, wann Autoren semantisch gesehen mit Gründen als Alternativen<sup>3</sup> umgehen (Beispiel: *Auch deswegen rate ich von billigem Kunstdünger ab.*). Beides könnte für eine automatische Erkennung hilfreich sein.

Ein erheblicher Teil der Richtlinien beschäftigt sich mit der Segmentierung des Textes in begründungsrelevante minimale Einheiten. Den Annotatoren wurden dazu zwei grundlegende Regeln genannt: 1. Minimiere die Außengrenzen des Segments, und 2. Vermeide diskontinuierliche Segmente. Beide Regeln wurden an vielen Beispielen für verschiedene Situationen erläutert. Als Normalfall werden Segmente beschrieben, die entweder vollständige Haupt- und Nebensätze (z.B. *weil [wir schon müde waren]*), Nominalisierungen (z.B. *wegen [des schlechten Wetters]*) oder bei Zwecksetzungen Infinitive (z.B. *um [uns etwas abzulenken]*) sind. Zudem wurden die Annotatoren mit Beispielen auf den Umgang mit sloganhaften Verkürzungen und kompakten bzw. elliptischen Konstruktionen vorbereitet.

Die zweite größere Aufgabe des Experiments war die Klassifikation der Segmente. Die Richtlinien beschreiben und exemplifizieren die oben aufgeführten Illokutionstypen. Sollte ein Annotator zwischen zwei Illokutionstypen schwanken, war er angehalten sich an den oberflächlichen Merkmalen und vor allem am Verb zu orientieren. Daneben sollen die Annotatoren für jedes Segment festhalten, ob es ein Performativum enthält, welches als Indiz für die Illokution gilt, und schliesslich, ob der Autor sich auf die Texterstellungshandlung selber (wie in *Dazu kann ich leider nichts sagen.*) oder auf das Medium bezieht (so z.B. in *Da es nun ja endlich eine Extrarubrik für dieses Hotel gibt ...*).

---

3. Wir orientieren uns dabei an einem Begriff von Alternativen im Sinne von Rooth (1992).

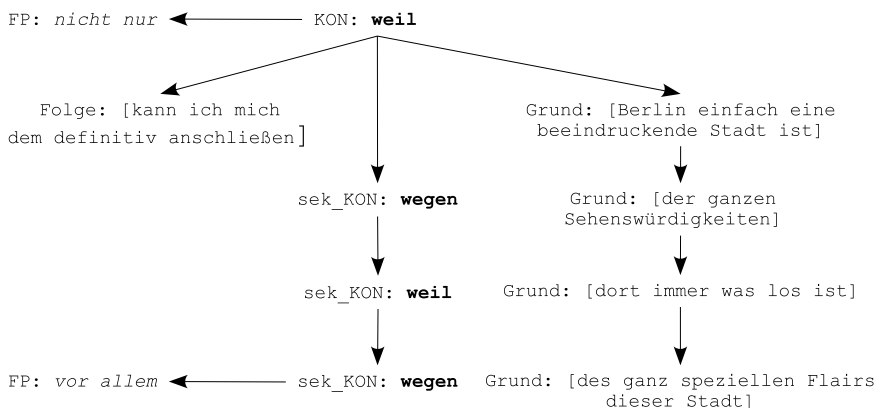


Abbildung 2. Schematische Darstellung der Analyse von Beispiel (3)

### 3.2 Ein komplexeres Beispiel

Ein komplexeres (selbst konstruiertes) Beispiel soll das angestrebte, also zu den Richtlinien konforme Resultat der Annotation veranschaulichen:

- (3) Ein Kurzbesuch in Berlin lohnt sich immer , stand auf meinem handlichen Reiseführer . Und wenn ich so auf meinen Kurz-Trip zurückschauen , [ 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 ] .

Die Segmentierung und die Identifikation des zentralen Konnektors und der Fokuspartikel ist schon durch Klammerung, Unterstreichung und Kursivdruck angezeigt. Die Verknüpfung und Kettenbildung der Elemente wird in Abbildung 2 dargestellt.

## 4 Experiment und Evaluation

Die Güte der Richtlinien und die Komplexität der Annotationsaufgabe haben wir durch ein kontrolliertes Experiment untersucht. Im Folgenden charakterisieren wir kurz die behandelte Textsorte, dann stellen wir das Design des Experiments vor und erläutern das Evaluationsergebnis.

#### 4.1 Textsorte

Maßgeblich für die Wahl der Textsorte war zum Einen, dass der Texttyp ‘argumentativ’ sein sollte, um kausale Konnektoren in möglichst verschiedenen Verwendungen anzutreffen. Zum anderen sollten die Texte im Hinblick auf eine automatische Erkennung ‘authentisch’ und auch anwendungsbezogen von Interesse sein. Die Wahl fiel auf Produktbesprechungen der Website [dooyoo.de](http://dooyoo.de), insbesondere Besprechungen von Hotels, da diese sehr häufig sowohl beschreibende als auch bewertende Anteile aufweisen, weshalb wir das gewünschte Spektrum der Verwendungen von Kausal-konnektoren hier erwarten konnten.

Insgesamt haben wir 250 solcher Besprechungstexte zufällig ausgewählt. Naturgemäß gilt für diese Textsorte, dass die deutsche Grammatik nicht durchgehend beherrscht wird, und dass stilistisch mit einer gewissen Variationsbreite zu rechnen ist. Dies macht eine perspektivische automatische Verarbeitung nicht einfacher, stellt jedoch schlicht den “Normalfall” für Blog-/Webforum-artige Texte dar.

#### 4.2 Design des Annotationsexperiments

Als Vorstufe zur Annotation des gesamten Korpus führten wir ein Annotationsexperiment mit zwei Annotatoren durch, die mit der Materie nicht vertraut waren. Unser Ziel war es zu überprüfen, inwieweit unsere Richtlinien ausreichen und wie (nicht-) übereinstimmend die beiden Annotatoren diese auf die Annotation der Texte anwenden. Für die Annotation selbst benutzten wir das Werkzeug MMAX2<sup>4</sup>, welches uns erlaubt, ein Annotationsschema mit geringem Aufwand zu erstellen und schon vorhandene Annotationen zu importieren. Dies war wichtig, weil wir vorab die kausalen Konnektoren halbautomatisch markiert hatten. Dann erstellten wir ein den Richtlinien entsprechendes Annotationsschema für MMAX2 und importierten das Korpus in das Werkzeug.

Zum Testen des MMAX2-Schemas, der markierten Konnektoren und des ersten Entwurfs der Richtlinien führten wir mit drei Versuchspersonen vor dem eigentlichen Experiment eine Pilotstudie durch. Den Annotatoren wurde die Funktionsweise von MMAX2 erklärt, dann haben sie die Richtlinien gelesen und es konnten Fragen dazu gestellt werden. Für die Annotation der zufällig ausgewählten Texte hatten die Versuchspersonen 90 min. Zeit. Ihre Aufgabe war es, zu jedem vormarkierten Konnektor Grund (oder Gründe) und Folge/n zu markieren und mit dem Konnektor in Relation zu setzen. Zudem sollten sie die markierten Konnektoren überprüfen und gegebenenfalls die falsch markierten löschen. Optional hatten sie die Möglichkeit, neue kausale

---

4. <http://mmax2.sourceforge.net>



Konnektoren zu identifizieren. Die Illokutionszuweisung für jedes Segment wurde in dieser Pilotstudie vorerst ausgelassen, um die Segmentierung kontrolliert testen zu können.

Mit den Erkenntnissen bzw. Erfahrungen, die wir mit dieser informellen Annotation gewonnen hatten, überarbeiteten wir die Richtlinien sowie das MMAX2-Annotationsschema<sup>5</sup> und überprüften noch einmal die markierten Konnektoren im Korpus. In direkter Vorbereitung des Annotationsexperiments wurden 11 Texte mit insgesamt 78 Konnektoren ausgewählt, die möglichst viele unterschiedliche Konnektoren enthielten. Um möglichst gleiche Bedingungen für die Annotatoren zu schaffen wurden zwei Trainingsvideos für die Annotation mit MMAX erstellt, die einerseits die Funktionsweise von MMAX erläutern, andererseits in das Annotationsschema einführen.

Zu Beginn des eigentlichen Experiments, in dem nun auch die Illokutionszuweisung vorgenommen werden sollte, bekamen die Probanden die vollständigen Annotationsrichtlinien und die Trainingsvideos ausgehändigt. Die Einarbeitung dauerte bei beiden Annotatoren rund 90 Minuten. Während des Annotationsvorgangs konnten die Annotatoren dann keine Rückfragen mehr stellen, Videos und Richtlinien waren die einzigen Informationsquellen. Für die Annotation selbst wurden drei bis dreieinhalb Stunden benötigt.

#### 4.3 Evaluation

Zur Erleichterung der qualitativen Durchsicht der Ergebnisse der unterschiedlichen Annotatoren wurde ein Skript entwickelt, welches die Daten zusammenfasst und von allen Konnektoren eine übersichtliche Gegenüberstellung im HTML-Format generiert. Abbildung 3 zeigt exemplarisch die Darstellung einer übereinstimmenden und einer nicht übereinstimmenden Annotation.

Diese erste qualitative Durchsicht der Ergebnisse läßt auf zwei Schwachstellen in den Annotationsrichtlinien schließen. Erstens besteht zwischen den Annotatoren Uneinigkeit darüber, welches Segment in einer Zwecksetzung mit *um...zu* der Grund und welches die Folge sein soll. Die gewünschte Zuordnung wurde in den Richtlinien nicht explizit gemacht. Zweitens ist von einem Annotator das Minimalitätskriterium bei der Segmentierung so ernst genommen worden, dass er bei jeder in einem Grund vorkommenden Aufzählung von Nominalen jedes einzelne Element als Grund markierte. In den Richtlinien wurde zwar beschrieben, welches Textmaterial ein Segment nicht enthalten sollte, um noch als minimal zu gelten, nicht aber,

---

5. Das Schema ist ebenfalls erhältlich auf <http://www.ling.uni-potsdam.de/~stede/kausallannot.html>

hrev_245 / markable_8				
	Schon allein	durch	seine weitläufige	Bauweise verdient es einen fetten Pluspunkt
<b>Anno.1</b>	FP	K1	G1: report	F1: eval
<b>Anno.2</b>	FP	K1	G1: report	F1: eval
hrev_085 / markable_4				
	Normal sollte einen das nicht ärgern	Weil	wenn man das nicht will	macht man das nicht
<b>Anno.1</b>		K1	G1: report	F1: report
<b>Anno.2</b>	F1: direct	K1	G1: report	

Abbildung 3. Ausschnitt der HTML-Ausgabe des Aufbereitungs-Skripts: übereinstimmende Annotation

welches es mindestens enthalten müsste, um noch als eigenständiges Segment zu gelten.

Im zweiten Schritt sollte die von den Annotatoren durchgeführte Segmentierung des Textes verglichen und quantitativ die “Uneinigkeit” zwischen den Annotatoren ausgewertet werden. Dazu legten wir folgende fünf Fehlerklassen fest. Für jeden Konnektor wurde gezählt, wie oft welche Fehlerklasse bei seinen Segmenten auftrat.

- Fehlerklasse (0) - Disjunktheit: Die jeweils markierten Segmente teilen kein Textmaterial.
- Fehlerklasse (i) - Inklusion/Überlappung: Die Segmente enthalten im Wesentlichen dasselbe Textmaterial, die Grenzen variieren um wenige Worte.
- Fehlerklasse (ii) - Untersegmentierung: Das Segment eines Annotators wurde vom anderen als Kette mehrerer Segmente annotiert, d.h. in mehrere Teilsegmente unterteilt.
- Fehlerklasse (iii) - Tilgung: Das Segment eines Annotators hat keine Entsprechung beim anderen.
- Fehlerklasse (iv) - Typverletzung: Die Grund-Folge-Zuweisung ist vertauscht.

Im Experiment annotierten die Versuchspersonen 34 der 78 Konnektoren übereinstimmend. 30 Konnektoren fielen in eine der o.g. Fehlerklassen, 11 Konnektoren in jeweils zwei. Erwartungsgemäß trat Fehlerklasse (i) (Inklusion/Überlappung) mit insgesamt 36 Vorkommen am häufigsten auf. Die Fehlerklassen (ii) (Untersegmentierung) mit 6 Vorkommen und (iv) (Typverletzung) mit 2 Vorkommen sind auf die schon in der qualitativen Durchsicht identifizierten aber behebbaren Schwachstellen der Annotationsrichtlinien zurückzuführen. Die Fehlerklassen (0) (Disjunktheit) mit 4 Vorkommen und (iii) (Tilgung) mit sogar 7 Vorkommen erfordern noch eine genauere Untersuchung der potenziellen Ursachen, da hier eine erhebliche Uneinigkeit vorliegt.

Schließlich wurde für die Evaluation der Illokutionszuweisungen für jedes sich inhaltlich entsprechende Segmentpaar die Zuweisung verglichen. Zwei Segmente galten dabei als sich inhaltlich entsprechend, wenn sie im Wesentlichen dasselbe Textmaterial enthalten und die Grenzen nur um wenige Wörter variieren. Der resultierende Grad übereinstimmender Zuweisungen beträgt ca. 0,71. Da die prozentuale Übereinstimmung nur bedingt aussagekräftig ist, entschieden wir uns für eine Berechnung des kappa-Wertes. Die zu seiner Berechnung notwendige statistisch erwartete Übereinstimmung beträgt 0,43, dies liefert einen kappa-Wert von 0,492. Auffällig ist ein hoher Verwechslungsgrad zwischen Reportiva<sub>neut</sub> und Evaluativa. Da sich diese Illokutionstypen semantisch nahe stehen, kombinierten wir für eine zweite Auswertung beide Kategorien, was den kappa-Wert auf (immer noch nicht befriedigende) 0,575 erhöht.

Im Anschluss an die Evaluation haben wir die Richtlinien abermals überarbeitet, um die gefundenen Schwachstellen zu beheben, einen Schwerpunkt bildete dabei die ausführlichere Dokumentation der Illokutionszuweisung, die für die Annotatoren im Experiment problematisch war. Anschließend wurde das Korpus von Annotatoren, die mit den Richtlinien sehr gut vertraut waren, vollständig annotiert.

## 5 Fazit und Perspektiven

Unsere hier vorgestellte Arbeit umfasst die Erstellung eines Korpus' von argumentativen Texten, die Desambiguierung und Markierung der darin vorkommenden kausalen Konnektoren, die Entwicklung und Überarbeitung von Annotationrichtlinien für die Segmentierung und Klassifikation der Relata anhand der Ergebnisse von zwei Annotationsexperimenten und mündet in der Bereitstellung des vollständig annotierten Korpus.

Der gesamte Prozess der Erstellung und mehrfachen Überarbeitung der Richtlinien anhand authentischer Texte hat gezeigt, dass die auf den ersten Blick relativ "einfach" anmutende Aufgabe der Identifikation kausaler Verknüpfungen mit verschiedenen Komplikationen behaftet ist. Dies ist nur zum Teil der nicht ganz problemlosen Textsorte geschuldet; auch in "traditioneller" Schriftsprache dürften die meisten Probleme auftreten. Unsere Richtlinien reflektieren diese Phänomene, die ähnlich auch für andere Typen von Kohärenzrelationen gelten dürften.

Für das Korpus besteht ein nächster Schritt darin, auch anderweitig markierte sowie nicht-markierte Kausalverknüpfungen zu annotieren, was freilich ein erheblich aufwändigeres Prozedere bedeutet, da die Annotatoren die Texte dann komplett durcharbeiten müssen, anstatt sich auf vormarkierte Konnektoren zu konzentrieren. Dies stellt dann eine "Vorstufe" zur kompletten rhetorischen Analyse dar.

Eine automatische Erkennung von Kausalverknüpfungen wird aber einstweilen nur anhand der Konnektoren möglich sein, daher betrachten wir das entstandene Kor-

pus als wichtige Ressource für entsprechende Entwicklungen. Die Arbeitsschritte bestehen in der Erkennung und Desambiguierung potenzieller Konnektoren (z.B. *da* / *daher*: Konnektor oder Ortsadverbial?), der Erkennung zusammengehöriger Konnektoren<sup>6</sup> und der Identifikation der Grund-/Folge-Segmente. Gerade der letztgenannte Schritt wird kaum “perfekt” lösbar sein (ein *darum* kann sich problemlos auf eine lange Satzfolge beziehen), sondern nur mit heuristischen/statistischen Verfahren, wofür sorgfältig annotierte Korpora eine unabdingbare Voraussetzung darstellen.

## Literaturverzeichnis

- Carlson, Lynn, Daniel Marcu, und Mary Ellen Okurowski (2003). Building a discourse-tagged corpus in the framework of rhetorical structure theory. In: Jan van Kuppevelt und Ronnie Smith (Hrsg.), *Current Directions in Discourse and Dialogue*, Dordrecht: Kluwer.
- Corston-Oliver, Simon (1998). *Computing of Representations of the Structure of Written Discourse*. Doktorarbeit, University of California at Santa Barbara.
- Hanneforth, Thomas, Silvan Heintze, und Manfred Stede (2003). Rhetorical parsing with underspecification and forests. In: *Proc. of HLT-NAACL*, 31–33, Edmonton/Canada.
- Jasinskaja, Katja, Jörg Mayer, Jutta Boethke, Annika Neumann, Andreas Peldszus, und Kepa Joseba Rodríguez (2007). Discourse tagging guidelines for german radio news and newspaper commentaries, ms., Universität Potsdam.
- Klein, Josef (1987). *Die konklusiven Sprechhandlungen*. Tübingen: Niemeyer.
- Lüngen, Harald, Henning Lobin, Maja Bärenfänger, Mirco Hilbert, und Csilla Puskas (2006). Text parsing of a complex genre. In: Bob Martens und Milena Dobрева (Hrsg.), *Proc. of the Conference on Electronic Publishing (ELPUB 2006)*, Bansko, Bulgaria.
- Lüngen, Harald, Csilla Puskas, Maja Bärenfänger, Mirco Hilbert, und Henning Lobin (2006). Discourse segmentation of german written text. In: Tapio Salakoski, Filip Ginter, Sampo Pyysalo, und Tapio Phikkala (Hrsg.), *Proceedings of the 5th International Conference on Natural Language Processing (FinTAL 2006)*, Berlin/Heidelberg/New York: Springer.
- Mann, William und Sandra Thompson (1988). Rhetorical structure theory: Towards a functional theory of text organization. *TEXT* 8:243–281.
- Marcu, Daniel (2000). *The theory and practice of discourse parsing and summarization*. Cambridge/MA: MIT Press.
- Pasch, Renate (1989). Adverbialsätze – kommentarsätze – adjungierte sätze. eine hypothese zu den typen der bedeutungen von ‘weil’, ‘da’ und ‘denn’. In: Wolfgang Motsch (Hrsg.), *Wortstruktur und Satzstruktur*, Linguistische Studien des ZISW: Reihe A – Arbeitsberichte 194, 141–158, Berlin: Akademie der Wissenschaften der DDR.
- Pasch, Renate, Ursula Brauße, Eva Breindl, und Ulrich Herrmann Waßner (2003). *Handbuch der deutschen Konnektoren*. Berlin/New York: Walter de Gruyter.
- Rooth, Mats (1992). A theory of focus interpretation. *Natural Language Semantics* 1(1):75–116.
- Sanders, Ted, Wilbert Spooren, und Leo Noordman (1992). Toward a taxonomy of coherence relations. *Discourse Processes* 15:1–35.

---

6. Siehe dazu (Stede und Irsig 2008) (u.a. Korrelate: *Wir blieben deshalb am Strand, weil das Wetter immer besser wurde*)

- Schauer, Holger und Udo Hahn (2001). Anaphoric cues for coherence relations. In: Galia Angelova (Hrsg.), *Proc. of the Conference on Recent Advances in Natural Language Processing (RANLP)*, 228–234.
- Schmitt, Holger (2000). *Zur Illokutionsanalyse monologischer Texte*. Frankfurt: Peter Lang.
- Searle, John R. (1976). A classification of illocutionary acts. *Language in Society* 5:1–23.
- Stede, Manfred und Kristin Irsig (2008). Complex connectives: Complications for local coherence analysis. In: *Proceedings of the Workshop "Constraints in Discourse"*, Potsdam.
- Sumita, K., K. Ono, T. Chino, T. Ukita, und S. Amano (1992). A discourse structure analyzer for Japanese text. In: *Proceedings of the International Conference on Fifth Generation Computer Systems*, 1133–1140.
- Webber, Bonnie, Matthew Stone, Aravind Joshi, und Alistair Knott (2003). Anaphora and discourse structure. *Computational Linguistics* 29(4):545–587.
- Wolf, Florian und Edward Gibson (2005). Representing discourse coherence: a corpus-based study. *Computational Linguistics* 31(2):249–287.

