

# Tagging kausaler Relationen

Yannick Versley

Diplomarbeit für das Fach Informatik

Betreut von

Prof. Christopher Habel,  
Arbeitsbereich Wissens- und Sprachverarbeitung

Prof. Wolfgang Menzel,  
Arbeitsbereich Natürlichsprachliche Systeme

---

**Zusammenfassung** In dieser Diplomarbeit geht es um kausale Beziehungen zwischen Ereignissen und Erklärungsbeziehungen zwischen Ereignissen, bei denen kausale Relationen eine wichtige Rolle spielen. Nachdem zeitliche Relationen einerseits ihrer einfacheren Formalisierbarkeit und andererseits ihrer gut sichtbaren Rolle in der Grammatik (Tempus und Aspekt, zeitliche Konjunktionen) wegen in jüngerer Zeit stärker im Mittelpunkt des Interesses standen, soll hier argumentiert werden, dass kausale Beziehungen und die Erklärungen, die sie ermöglichen, eine wichtigere Rolle im Kohärenzgefüge des Textes spielen. Im Gegensatz zu “tiefen” Verfahren, die auf einer detaillierten semantischen Repräsentation des Textes aufsetzen und infolgedessen für unrestringierten Text m. E. nicht geeignet sind, wird hier untersucht, wie man dieses Ziel erreichen kann, ohne sich auf eine aufwändig konstruierte Wissensbasis verlassen zu müssen.

**abstract** Causal relations between events and explanatory relations among these events, where the causal relations play an important role, are the main topic of the present diploma thesis. After temporal relations between events have been more in the focus of interest recently because of both being easier to formalize and playing a visible role in grammar (notably the effects of time and aspect, as well as temporal conjunctions), I will argue that causal relations and the explanations they provide play the greater role in the coherence of a text. In contrast to “deep” approaches that rely on a fine-grained semantic representation of the text and by consequence can be unsuitable for unrestricted text, I will investigate how to reach this goal without requiring an expensive hand-coded knowledge base.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>5</b>
<b>2</b>	<b>Kausalität: Kognitionswissenschaftliche Grundlagen</b>	<b>11</b>
2.1	Ereignisse . . . . .	12
2.2	Theorien zur Kausalität . . . . .	14
2.3	Kausale und andere Erklärungen . . . . .	20
2.3.1	Kausalbeziehungen in Conceptual Dependency . . . . .	22
2.3.2	Inuis Planbox-Modell . . . . .	23
2.3.3	Vergleich . . . . .	24
<b>3</b>	<b>Kausalität im Text</b>	<b>25</b>
3.1	Kausalität und Temporalität . . . . .	27
3.2	Diskursstruktur . . . . .	30
3.2.1	Lokale Diskursstruktur . . . . .	35
<b>4</b>	<b>ein Tagger für kausale Relationen</b>	<b>39</b>
4.1	Textmuster . . . . .	40
4.1.1	Schwach überwachte Lernverfahren . . . . .	46
4.2	Unüberwachte Klassifikation von Diskursrelationen . . . . .	47
4.3	Untersuchung zu Marcu und Echihabi . . . . .	49
4.3.1	Systemaufbau . . . . .	51
4.3.2	Analyse . . . . .	57
<b>5</b>	<b>Zusammenfassung und Ausblick</b>	<b>67</b>

---

# Kapitel 1

## Einleitung

Können Computer Texte verstehen? Auf diese Frage hört man heute, genauso wie vor 10 oder 20 Jahren, ein optimistisches Ja, von anderen ein “Noch nicht, aber bald” und von wieder anderen ein “tun sie nicht, werden sie nicht und sollen sie auch nicht.”

Sicherlich hängt die Antwort auf obengenannte Frage davon ab, welche Texte man meint, und was man unter Verstehen genau versteht. Vollständiges Verstehen eines Textes im Sinne der Leistung, die ein der betreffenden Domäne kundiger Muttersprachler, der im Vollbesitz seiner geistigen Kräfte ist, im wachen Zustand vollbringen kann, ist noch nicht ansatzweise zu erwarten. Vielleicht aber reicht auch ein partielles Verständnis des Textes. Eine pragmatische Sichtweise hierzu ist, sich eine Aufgabe herauszusuchen, bei der man, wenn ein Mensch sie löst, annimmt, dass er dabei den Inhalt von Texten (zumindest ansatzweise) versteht.

In der Sprachsignalverarbeitung (d.h. dem Bereich, der sich damit beschäftigt, aus einem digitalisierten Sprachsignal die Folge an Worten herauszufinden, die der Sprecher geäußert haben wollen könnte) sowie im Bereich des Information Retrieval (wo es gilt, zu einer Suchanfrage aus einer großen Dokumentensammlung passende Dokumente oder Teile von Dokumenten zu finden), ist die Evaluation bereits seit längerem ein fester Bestandteil und hat zum Teil die Wahrnehmung der vorherrschenden Probleme wesentlich beeinflusst, insbesondere, da dadurch verschiedene und verschiedenartige Systeme vergleichbar werden; im Bereich der Sprachsignalverarbeitung wäre da die etwa Einsicht, dass stochastische Verfahren zunächst bessere Ergebnisse liefern als ein aus rein linguistischen Gesichtspunkten aufgebauter Erkennen.

Auf der anderen Seite steht allerdings die Einsicht, dass das Einbeziehen bestimmter linguistischer Information zusammen mit stochastischer Modellierung mit weniger Daten bessere Ergebnisse erzielen kann als ein “uninformiertes” stochastisches Modell. Schaut man aber genau hin, so stellt man fest, dass beim Tying von Teilmodellen<sup>1</sup> phonetische Information verwendet wird, um das Fehlen von

---

<sup>1</sup>dem Zusammenfassen von (Di-,Tri-)Phonmodellen oder Teilen davon, um für die Parameterschätzung ausreichend viele Beispiele zur Verfügung zu haben

Daten auszugleichen.

Im Bereich des Sprachverstehens scheint man zunächst weit entfernt zu sein von einer solchen Evaluation — schließlich ist die Frage, wie die Bedeutung des Texts, so wie sie ein kompetenter Leser erfassen würde, zu repräsentieren sei, bisher nur ansatzweise beantwortet und immer noch ein interessanter Forschungsgegenstand.

Bisherige Evaluationen konzentrierten sich deshalb entweder auf Teilaspekte wie die Desambiguierung von Wörtern in einem Text (Senseval) oder die Auflösung von Anaphern (Pronomen und definitiver Beschreibungen), oder formulieren ein Problem, dessen Lösung entweder textuell ist oder eine einfache formale oder semi-formale Beschreibung. Als Vertreter der ersten Gattung sei das Zusammenfassen von Texten genannt, oder die Beantwortung von Fragen entweder zu einem Text oder einer größeren Textsammlung. Zur zweiten Gattung würde etwa das Einordnen von Text in vordefinierte Kategorien gehören, oder das Füllen von Slots, wie sie aus einem Datenbankschema stammen könnten (Relationsextraktion).

Die dritte Message Understanding Conference (MUC-3) war die erste größer angelegte Evaluation, die sich um das Problem der Relationsextraktion drehte (Lehnert und Sundheim, 1991). Aufgabe war es, aus Zeitungsartikeln zu terroristischen Anschlägen Slots wie “Perpetrator” und “Physical Target” zu füllen (was zugegebenermaßen keine besonders ausgefeilte Darstellung ist, aber in Hinsicht auf die gewählte Domäne ausreicht). Mit wechselnden Domänen (und wachsender Komplexität der Ergebnisbeschreibung) wurde dieser Ansatz unter anderem in weiteren MUC-Konferenzen verfolgt. Die noch Ende der 80er Jahre populären Ansätze zum Textverstehen, mit kognitiv adäquaten Hypothesen und expliziter Repräsentation des nötigen Wissens und der gemachten Inferenzen den Verstehensprozess nachzubilden (vgl. (Charniak und McDermott, 1985)) wurden dabei von Systemen verdrängt, die auf die entsprechende Domäne zugeschnitten waren und mit einem Minimum an Repräsentation auskamen (sogenannte “flache” Systeme).

Mittlerweile gibt es keine Evaluationen zur Relationsextraktion mehr, die gewonnenen Erkenntnisse und Techniken bleiben für neuere Aufgaben und Methoden nützlich bzw. sind teils notwendige Voraussetzung, wobei in heutigen Systemen diese sogenannten “flachen” Verfahren durch komplexere Module ergänzt werden.

Bereits die erste MUC-Konferenz machte klar, dass es beim Aufbau solcher größerer Systeme auch um einen sinnvollen Einsatz der verfügbaren Arbeitskraft gehen muss. So schreibt etwa Lehnert (1994) über das System der Universität Massachussetts, das bei der MUC-3 großen Erfolg hatte:

The graduate students who implemented the UMass/MUC-3 system had no desire to ever build anything like it again. Their labor was time consuming and tedious. They established the viability of the UMass approach relative to other approaches, but with a human labor factor that threw into question the practicality of the technology.

Bei solchen größeren Systemen ergibt sich also neben der Evaluierbarkeit die Wichtigkeit von Techniken wie Modularisierung und Wiederverwendung von Kom-

ponenten, die es auch außerhalb der Sprachverarbeitung erlauben, große Systeme mit vertretbarem Aufwand zu erstellen.

Eine Besonderheit des Language Engineering (so wird dieser Teil der Computerlinguistik genannt, der sich eher mit Entwurf, Architektur und Entwicklung befasst) im Vergleich zum normalen Software Engineering ist, dass andere wiederverwendbare Daten, die nicht Bestandteil von Programmen sind, wie Lexika, Corpora und Ontologien, unter dem Sammelbegriff *Language Resources* zusammengefasst, eine wichtige Rolle in der Systemkonstruktion spielen.

Corpora etwa, also Texte, die mit zusätzlicher Information annotiert sein können, können zur Validierung/Evaluation von Systemen benutzt werden, genauso wie zur (Weiter-)Entwicklung von Systemen oder zum Trainieren statistischer Komponenten.

Im Bereich des Sprachverstehens im weiteren Sinne gibt es neben Corpora/Annotationsschemata zur Wortdisambiguierung (SemCor, Senseval), den Daten der Evaluationsaufgaben zur Relationsextraktion (mehrere Message Understanding Conferences und einige unter dem Titel Automatic Content Extraction) sowie zur Textzusammenfassung (Document Understanding Conference, TIPSTER) und zur Beantwortung von Fragen (das Remedia- und das CBC4Kids-Corpus für Einzeldokumente, die Corpora der TREC- und CLEF-Evaluationen für Dokumentensammlungen) auch Corpora, in denen abstrakte Information zur Bedeutung (d.h. solche, die keine unmittelbare Bedeutung für eine im mindesten realitätsnahe Aufgabe) haben, namentlich TimeML (und das TimeBank-Corpus) zur Annotation zeitlicher Relationen zwischen den in einem Text erwähnten Ereignissen und (unter anderem) die Rhetorical Structure Theory (und den Corpus von (Carlson et al., 2001)), die mit den bedeutungstragenden Textspannen eine baumförmigen Struktur aufbauen.

Beiden gemeinsam ist, dass die vorgeschlagene Annotation eine Zwischenstufe darstellt, die (so die Befürworter) nützlich für eine semantische/pragmatische Verarbeitung des Textes ist. Vielleicht gibt es noch mehr Ähnlichkeiten, zumal (Mani und Pustejovsky, 2004) ein Modell vorschlagen, das die baumförmige Strukturierung eines Diskursmodells mit zeitlichen Relationen verbindet.

Gibt es ein semantisches/pragmatisches Gegenstück zur Grammatikalität? Schauen wir, ob das, was man gemeinhin unter der Kohärenz eines Textes versteht, ähnliches leistet wie die Grammatik auf syntaktischer Ebene.

- Ähnlich wie es grammatische und ungrammatische Äusserungen gibt, gibt es Äusserungen, die kohärent oder inkohärent wirken:
  - (1.1) a. Die Katze sitzt auf dem Bett.  
b. Auf Bett Katze dem sitzt.
  - (1.2) a. Jan hat lange Haare. Er spielt in einer Rockband.  
b. Jan fliegt von Paris nach Moskau. Er mag Spinat.
- Die Phrasenstrukturbäume der Syntax haben, so wird allgemein angenommen, ihre Entsprechung in den Diskursbäumen, dazu mehr in Kapitel 3.2.

Die Möglichkeit der Strukturzuweisung ist hierbei gleichermaßen das Hauptkriterium für die Akzeptabilität.

- Durch die zugewiesene Struktur bekommt der Leser zusätzliche Information, die über jene, die in den Einzelstücken enthaltene hinausgeht.

(1.3) Mann beißt Hund / Hund beißt Mann

(1.4) Robert wurde tot aufgefunden. Der Mörder wurde nicht erkannt / Der Idiot hat den Finger in die Steckdose gesteckt.

Während bei der Syntax durch Wortstellung und Funktionsmorpheme nur wenige Mehrdeutigkeiten übrigbleiben, muss bei der Rekonstruktion der Diskursstruktur auf semantische Information und allgemeines Weltwissen zurückgegriffen werden. Darüberhinaus argumentiert Givón (1995), dass die Strukturierung des Textes nur ein Nebenprodukt von Kohärenz ist, die selbst eher eine Eigenschaft des Produktions- oder Verstehensprozesses ist und mehr mit der mentalen Struktur des Inhalts zu tun hat. Für diese Sicht sprechen auch Ergebnisse von (Rosé et al., 1995), die unter bestimmten Bedingungen ein Abweichen von baumförmiger Strukturierung feststellten, oder das in (Walker, 2000) vertretene Modell, das ohne Diskursstruktur als konstruktiven Bestandteil auskommt.

Für eine stärkere Berücksichtigung der mentalen Struktur spricht sich auch Peter Clark (Clark et al., 2001; Clark, 2001) aus, der postuliert, dass beim Verstehen von Text nicht einzelne Fakten herangezogen werden, sondern ganze Aspekte des Wissens über etwas, sogenannte Sichten (Views), die stark an die von Schank (1982) postulierten “Memory Organization Packets” erinnern.

Eine detaillierte und in Bezug auf die Domäne vollständige Repräsentation des zum Verstehen benötigten Weltwissens, wie sie bei klassischen “tiefen” Systemen und zum Teil auch bei SDRT (Lascarides und Asher, 1993; Asher et al., 1995; Asher und Lascarides, 1996) vorausgesetzt wird, ist für unrestringierten Text mit den gegenwärtig zur Verfügung stehenden Methoden nur unvollständig zu erreichen. Koreferenzketten etwa von Personen bieten sich als Alternative an, reichen zur Etablierung der Kohärenz jedoch nicht aus (vgl. Beispiel 1.2). Vielversprechender wäre es, anstatt konkreter Objekte und Personen (von Lyons (1977) 1st-order entities genannt) auch Ereignisse (2nd-order entities) oder Fakten (3rd-order entities) mit einzubeziehen. Da Ereignisse wenig und Fakten noch weniger untersucht sind, bietet es sich an, sich zunächst auf Ereignisse zu konzentrieren und die Art, wie diese Ereignisse im Diskurs aufgegriffen werden zu untersuchen und eventuell in einem Annotationsschema zu formalisieren.

Wie erwähnt, gibt es zur Annotation zeitlicher Relationen zwischen den Ereignissen im Text bereits ein Annotationsschema namens TimeML, das sich als Standard zur Annotation zeitlicher Information im Text herausgebildet hat.

Während frühere Annotationsschemata wie das zur MUC-6 verwendete sich darauf beschränkten, die Textspannen von Zeitausdrücken zu markieren, wurde dieses im TIDES-Schema (Ferro et al., 2000) um eine einheitliche Darstellung für Zeitausdrücke ergänzt. Das Schema von TimeML (Pustejovsky et al., 2002)



erfasst schließlich auch Ereignisse, die im Text erwähnt werden und ihre zeitlichen Beziehungen zueinander und zu den Zeitausdrücken.

Das Finden von Zeitausdrücken und das Konvertieren in eine einheitliche Darstellung (wobei dann bei “am Freitag” und ähnlichen Ausdrücken aus dem Kontext geschlossen werden muss, um welchen Freitag es sich handelt) sowie das Verknüpfen von Zeitausdrücken und Ereignissen wird von “state-of-the-art”-Systemen (Filatova und Hovy, 2001; Schilder und Habel, 2001) gut beherrscht, während das in-Beziehung-Setzen von Ereignissen untereinander – in den meisten Texten kommen nur wenige Zeitausdrücke vor, so daß diese nicht zur Anordnung verwendet werden können – ein Problem darstellt, an dem derzeit noch geforscht wird; (Mani et al., 2003) erreichen mit ihrem Tagger, der Ereignisse in Nachrichtentexten partiell anordnet, einen F-Wert<sup>2</sup> von 0,75.

Der Zusammenhang der Ereignisse in einem Text ist jedoch nicht rein zeitlicher Natur; vielmehr, zur Etablierung von Kohärenz sind zeitliche Bezüge weniger geeignet als solche, die (in einem weiteren Sinne) kausal sind. Solche kausalen Bezüge lassen dann möglicherweise auch zuverlässigere Schlüsse auf zeitliche und räumliche Anordnung der Ereignisse zu als man ohne jede Repräsentation von Kausalität erwarten dürfte.

Kausale Bezüge spielen ausser für die zeitlichen Anordnung auch bei einer wichtigen Klasse von Erklärungen eine Rolle; Erklärungsbeziehungen sind wichtige Bausteine des Diskurses, für die bisher keine überzeugende Formalisierung gefunden wurde. Die Sicht, Erklärungen seien Antworten auf Warum-Fragen greift zu kurz, da Warum-Fragen und ihre Antworten sich einer Formalisierung bisher zumindest teilweise widersetzen.

Heutige Systeme für (allgemeines) Question Answering können Fakten-Fragen (welches ist der längste Fluss der Erde, wann wurde Margret Thatcher geboren etc.) recht gut beantworten, haben bei Warum-Fragen und bei Wie-Fragen jedoch Schwierigkeiten.

(Anand et al., 2000) bemerken hierzu, dass solche Fragen sich von Fakten-Fragen unter anderem dadurch unterscheiden, dass sie durch ganze Verbphrasen oder (Teil-)Sätze beantwortet werden müssen und beschreiben einen “Clausal Tagger”, der für Erklärungen, Arten, etwas zu tun und Meinungen relevante Textstellen (also solche, die als Antworten auf Wie- oder Warum-Fragen in Frage kommen) anhand von lexikalischen, syntaktischen oder diskursbasierten Wortmustern markiert.

Leider sind Wortmuster, die eindeutig auf Kausalität bzw. kausale Erklärungen hinweisen, in Texten verhältnismäßig selten. Girju (2002) untersuchte hierzu, wie kausative Verben in Texten als Antworten für Ursachenfragen gefunden werden können.

---

<sup>2</sup>Evaluiert man einen Tagger, so benutzt man üblicherweise 3 Statistiken: *Precision*, den Anteil der richtig getaggtten Annotationen an allen vom Tagger getaggtten Annotationen, *Recall*, den Anteil der richtig getaggtten Annotationen an den zu taggenden Annotationen und den *F-Wert*, den man als  $\frac{2PR}{P+R}$  errechnet. Sind Precision und Recall gleich, so hat der F-Wert ebenfalls diesen Wert, ist entweder Precision oder Recall gleich 0, so ist der F-Wert ebenfalls 0.

In dieser Arbeit soll zunächst in Kapitel 2 untersucht werden, in welche Beziehung zusammengehörige Ereignisse  $A$  und  $B$  im Text stehen, wenn  $B$  die Antwort auf “Warum  $A$ ” (oder “Wie  $A$ ”) darstellt, sowie der Zusammenhang zwischen der Relation der Ereignisse und der sie beschreibenden Diskurssegmente betrachtet werden. Knott (1994) teilt die Relationen zwischen Diskurssegmenten (Diskursrelationen) ein in *semantische* Diskursrelationen, also Relationen zwischen den Ereignissen und *pragmatische* Diskursrelationen, die beim Planen des Diskursaufbaus eine Rolle spielen.

In Kapitel 3 soll erörtert werden, wie Kausalität und kausale Erklärungen im Text markiert sind und wie sie zur Kohärenz des Textes (im Sinne einer Diskursstruktur) beiträgt. Zusätzlich wird überlegt, wie zeitliche Markierung oder Indizien für die Diskursstruktur gegebenenfalls als Informationsquellen für kausale Zusammenhänge genutzt werden können.

Schließlich soll ein Kapitel 4 überlegt werden, aus welchen vorhandenen, anzupassenden oder neu zu entwickelnden Komponenten ein System bestehen könnte, das kausale Relationen in Texten findet und annotiert.

In Kapitel 5 soll anhand der gefundenen Ergebnisse zusammenfassend überlegt werden, welche der betrachteten Ansätze geeignet sein könnten, beziehungsweise wie sie anzupassen wären, um einen solchen Kausal-Tagger zu realisieren, und welche grundsätzlichen Probleme zu erwarten wären.

## Kapitel 2

# Kausalität: Kognitionswissenschaftliche Grundlagen

Die Anbindung neuer Äußerungen an den Diskurs wird im Allgemeinen als in-Beziehung-Setzen der neuen Information mit Information aus dem Diskurskontext beschrieben, die entsprechenden Beziehungen als Diskurs- oder Kohärenzrelationen (Hobbs, 1979) bezeichnet. Kausale Erklärungen, bei (Mann und Thompson, 1987) wiederzufinden als Diskursrelationen *cause*, *volitional cause*, *result* und *volitional result*, gehören zweifellos zu den prototypischen Vertretern der Menge der Diskursrelationen, wobei in den dazugehörigen Beschreibungen Kausalität (und Finalität) als Primitive gebraucht werden. Andererseits ist anzumerken, dass Kausalität – anders als Zeit, die als linearer Ablauf konsensträchtigt und modelltheoretisch “sauber” zu formalisieren ist – in einer auf klassischer Prädikatenlogik basierenden Theorie nicht in einer einfachen und einleuchtenden Weise unterzubringen ist und in logischen Diskurstheorien wie SDRT (Lascarides und Asher, 1993; Asher et al., 1995; Asher und Lascarides, 1996) eher stiefmütterlich behandelt wird.

In diesem Kapitel soll es zunächst darum gehen, wie Kausalität im Rahmen einer logischen Theorie modelliert werden kann. Hierzu ist es hilfreich, sich die Unterschiede zwischen der Art, wie eine Wissenschaft wie die Physik (in der Kausalität nur eine untergeordnete Rolle spielt) mit etwas umgeht und der Art, wie Menschen ohne den Einsatz von Hilfsmitteln es üblicherweise tun, vergegenwärtigt. Während mathematische Modelle in der Physik im Allgemeinen darauf basieren, möglichst genaue Information über das gesamte Geschehen zu modellieren, sind die von Menschen benutzten kognitiven Modelle (denen wir unsere logischen Modelle nachempfinden wollen) meist aufgabenbezogen und müssen mit der Tatsache zurechtkommen, dass die Information über den betrachteten Weltausschnitt stets lückenhaft und gleichzeitig eher qualitativer als quantitativer Natur ist. Es geht also darum, mit der verfügbaren Information in Bezug auf die Aufgabe möglichst effektiv umzugehen.

Der Preis dieser Informations-Ökonomie ist – neben einer geringeren Genauigkeit der Voraussagen, die für übliche Probleme nicht schädlich ist – die Möglichkeit des Irrtums. Nichtmonotone Logiken (d.h. solche, die Schlüsse auch im Vorhandensein von Ausnahmen zulassen, die eine Revision der Hypothese notwendig machen könnten) verhelfen uns durch dieses aussen-vor-lassen der Ausnahmen in den meisten Fällen zu mehr Wissen, das (normalerweise) die Kosten der gemachten Irrtümer überwiegt.

Shoham (1990) schreibt dazu:

It is not cost effective to reason explicitly, even about all those factors which we *do* understand. Batteries go dead very infrequently, and so it is a waste of time to examine regularly (whether mentally or physically) the condition of the battery before we predict that the car will start.

And dead batteries are not an extreme example: How about a banana stuck in the tail pipe (an example from McCarthy, 1980) or a nuclear explosion in the vicinity, both of which will prevent (!)<sup>1</sup> the motor from starting. Surely we do not want to check *these* conditions each time we start the car.

## 2.1 Ereignisse

Zunächst soll hier kurz dargelegt werden, was Ereignisse und Zustände (zusammen als Eventualitäten bezeichnet) sind und wie sie in TimeML (Pustejovsky et al., 2002; Saurí et al., 2004) verwendet werden, da die kausalen Beziehungen, auf die weiter unten näher eingegangen werden soll, Relationen zwischen Eventualitäten sind, und um später den Nutzen einer kausalen Erweiterung von TimeML erörtern zu können.

Der Annotationsleitfaden von TimeML (Saurí et al., 2004) beschreibt Eventualitäten wie folgt:

We consider “events” a cover term for situations that *happen* or *occur*. Events can be punctual (1-2) or last for a period of time (3-4). We also consider as events those predicates describing *states* or *circumstances* in which something obtains or holds true (5). Not all stative predicates will be marked up, however. See subsection 2.1.3, on events belonging to the class ‘STATE’.

Hierbei sei angemerkt, dass das, was bei TimeML als *event* bezeichnet wird, hier Eventualität genannt wird, während der Term Ereignis für Eventualitäten gebraucht wird, die keine Zustände sind.

---

<sup>1</sup>das Ausrufezeichen hier stammt von Shoham. Tatsächlich bekommen wir bei dieser Theorie der Kausalität eine Theorie der Verhinderung, die für sich genommen recht kompliziert wäre, gratis dazu. Siehe (Crouch et al., 2001) zu einer (wesentlich komplizierteren) Lösung, Verhinderungs-Aussagen in einer Montague-artigen Semantik unterzubringen.

## 2.1 Ereignisse

---

Ereignisse sind klar begrenzt, während Zustände als solches nicht individuiert sind: nimmt man einen (zeitlichen) Teil eines Zustandes, so ist es immer noch der gleiche Zustand, während ein Teil eines Ereignis im Allgemeinen ein verschiedenes Ereignis ist. Dies ist parallel zu der Unterscheidung zwischen Objekten (wie einem Ring) und Material (Wasser, das Gold, aus dem der Ring gemacht ist).

Zustände können durch Angaben zur Quantität (er hat eine halbe Stunde lang Rasen gemäht, er ist 2 Kilometer gelaufen) oder eines Zielpunkts (er hat den Rasen vor dem Haus gemäht, er ist zur Tankstelle gelaufen) als Ereignisse verwendet werden, Ereignisse durch Iteration (Er hat stundenlang Nägel in die Wand geschlagen, er ist jeden Morgen zur Tankstelle gelaufen) als Zustände.

In TimeML werden nur zeitlich begrenzte (und damit individuierte) Zustände annotiert, während Ereignisse, die durch Iteration zu einem unbegrenzten Zustand werden, nicht annotiert werden. Folgendes Beispiel aus den Annotationsrichtlinien illustriert dies:

(2.1) Businesses are emerging on the Internet so quickly that no one, including government regulators, can keep track of them.

Das Auftauchen (*emerging*) ist als Iteration einer nicht eingegrenzten Vielzahl solcher Ereignisse zu verstehen und hat deshalb auch keine klare zeitliche Begrenzung. Deshalb wird *emerging* genauso wie *keep* nicht als Event annotiert.

Bei der Negation eines Ereignisses (er hat keinen Nagel in die Wand geschlagen) ergibt sich im Allgemeinen ein Zustand, der durch das nicht-Auftreten dieses Ereignisses charakterisiert ist (Herweg, 1991).

TimeML behandelt negierte Ereignisse – zumindest soweit aus (Saurí et al., 2004) ersichtlich – so, dass ein instantiiertes Ereignis annotiert wird, dieses aber durch Verknüpfung mit dem als “signal” annotierten *not* als negiert markiert. Im Fall von

(2.2) John did *not* leave on Monday but on Tuesday

wird das Ereignis zu *leave* zweimal instantiiert, einmal zeitlich mit dem Zeitausdruck für *Monday* verknüpft und als negiert markiert und einmal zeitlich mit dem Zeitausdruck für *Tuesday* verknüpft.<sup>2</sup>

Diese Lösung birgt Probleme mit dem Quantorenskopos in der semantischen Interpretation (Pustejovsky et al., 2003). Der Autor dieser Diplomarbeit hält – zumindest in Kontexten, die für eine Individuierung des negierten Ereignisses sprechen – eine Behandlung von negierten Ereignissen als Individuen für durchaus sinnvoll in dem Sinne, dass dieses Individuum einer Entität in einer möglichen, aber nicht tatsächlichen Welt im mentalen Modell des Sprechers oder des Hörers entspricht.

Wenn das Auftreten eines Ereignisses in einem Kontext (normativ) gefordert ist, sorgt das Auftreten des Kontextes für eine Individuierung des nicht-Auftretens des Ereignisses, genauso wie ein expliziter Bezug zu einem Ereignis dies ermöglicht. So ist etwa folgendes möglich:

---

<sup>2</sup>vgl. (Saurí et al., 2004) S. 39,46,63

- (2.3) Johann hat dreimal die Tür nicht zugemacht und damit ein Durcheinander auf meinem Schreibtisch verursacht.

Das Offenlassen der Tür scheint hier eine Handlung, oder zumindest ein Ereignis zu sein. Warum?

Johann hat die Tür aufgemacht, ist hindurchgegangen und niemand hat sie seitdem wieder zugemacht. Um die Tür zu passieren, ist es notwendig, die Tür aufzumachen, so dass man das Aufmachen der Tür nicht als Grund anführen kann. Die normative Erwartung, dass die Tür möglichst immer geschlossen sein soll, führt nun zu einem neuen Handlungstyp, dem Offenlassen der Tür, der darin besteht, dass jemand, der die Tür aufmacht und sie passiert, sich davon entfernt, ohne sie vorher wieder zu schliessen. Dadurch wird es möglich, einem anormalen Zustand eine Handlung zuzuweisen, die ihn regelmässig bewirkt.

Dieser Bezug zur Negation ist insofern von Bedeutung, als bei der (kausalen) Beziehung zwischen Ursache und Wirkung die Negation von Ereignissen eine Rolle spielt, etwa in der Form, dass, ohne die Ursache auch die Wirkung nicht eingetreten wäre. Dieses Nicht-Eintreten (bei  $A$  etwa  $\bar{A}$ ) sollte hierbei im Sinne einer Kontrastmenge<sup>3</sup> verstanden werden, wie folgendes Beispiel aus (Schaffer, 2003) zeigen soll:

- (2.4) Sokrates starb, weil er am Abend den Schierlingsbecher zu sich nahm.

- (2.5) Sokrates starb, weil er *am Abend* den Schierlingsbecher zu sich nahm

Eine Theorie, in der  $A$  und  $\bar{A}$  muss den beiden Schierlingsbecher-Aufnahme-Ereignissen verschiedene Typen zuweisen, während eine quaternäre Theorie aufgrund der unterschiedlichen Betonung unterschiedliche Kontrastmengen annehmen kann.

## 2.2 Theorien zur Kausalität

Lewis (1986) zitiert Hume mit einer doppelten Definition für Kausalität: “We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or, in other words, where, if the first object had not been, the second never had existed” (Hervorhebung Lewis).

Der ersten Definition entspricht die Regularitätsanalyse, die Kausalität mit Bezug auf Gesetzmässigkeiten definiert; die zweite Definition, die zunächst ohne Bezug auf Gesetzmässigkeiten auskommt, ist die, auf der Lewis seine Erklärung von Kausalität durch Counterfactuals basiert.

Zunächst zur Regularitätsanalyse. Lewis formuliert dies auf folgende Weise:

Sei  $C$  die logische Aussage, dass das Ereignis  $c$  existiert/eintritt und  $E$  jene, dass das Ereignis  $e$  existiert/eintritt. Dann ist nach der vorgestellten Analyse  $c$  eine Ursache von  $e$ , gdw.

---

<sup>3</sup>deren Extension nicht unbedingt das logische Komplement der Extension des Negierten sein muss

- $C$  und  $E$  beide wahr sind
- es eine nicht-leere Aussagenmenge  $\mathcal{L}$  von Gesetzen und
- eine Menge  $\mathcal{F}$  wahrer Einzelaussagen gibt, so dass
- $\mathcal{L}$  und  $\mathcal{F}$  zusammen  $C \rightarrow E$  implizieren
- $\mathcal{L}$  und  $\mathcal{F}$  zusammen  $E$  nicht implizieren
- $\mathcal{F}$  alleine nicht  $C \rightarrow E$  impliziert.

Laut Lewis besitzt dieser Ansatz Probleme, wenn es gilt, zwischen kausalen und anderen Beziehungen zwischen Ereignissen zu unterscheiden, insbesondere:

**Umkehr** Wenn Ereignisse des Typs von  $e$  regelmäßig mit Ereignissen des Typs von  $c$  auftreten, könnte ebenfalls  $e$  die Ursache und  $c$  die Wirkung sein<sup>4</sup>.

**Epiphänomene** Wenn  $c$  nicht nur  $e$ , sondern auch  $e'$  verursacht, könnte beispielsweise  $e$  als Ursache von  $e'$  angesehen werden.

**Präemption** Es gibt Fälle, in denen ein Ereignis  $e_1$  von einem Ereignis  $e_2$  folgt, und zwischen den beiden eine kausale Verbindung besteht (seien sie durch mehrere Ursache-Wirkung-Beziehungen verbunden oder  $e_1, e_2$  sind von den Typen  $E_1, E_2$  und  $E_1$ -Ereignisse verursachen  $E_2$ -Ereignisse), und in denen  $e_1$  dennoch nicht die Ursache von  $e_2$  ist.

Nehmen wir an, Jan macht einen Spaziergang im Gebirge. Ein Felsen rollt heran und macht dabei ein Rumpelgeräusch, das Jan veranlasst, dem Felsen auszuweichen.

Ist nun der herabrollende Felsen der Grund dafür, dass Jan nicht vom Felsen erschlagen wurde? Dies würde man verneinen, denn ohne den Felsen wäre die Gefahr, vom Felsen überrollt zu werden, nicht gegeben.

Diese Probleme nimmt Lewis zum Anlass, die zweite Definition aufzugreifen und Kausalität durch Counterfactuals (also Aussagen der Form “wäre  $A$  der Fall, so wäre  $B$  der Fall”, kurz  $A \Box \rightarrow B$ ) zu definieren.

Nehmen wir also  $O(e)$  als die Aussage, dass  $e$  eintritt, würden wir die Aussage “ $c$  hat  $e$  verursacht” als

$$(O(c) \Box \rightarrow O(e)) \wedge (\neg O(c) \Box \rightarrow \neg O(e))$$

ausdrücken.

Was bedeutet nun aber dieses  $A \Box \rightarrow B$  aus modelltheoretischer Sicht?

Nehmen wir eine (diskrete<sup>5</sup>) Menge möglicher Welten an, von denen wir uns in einer bestimmten, nennen wir sie einmal  $w$  befinden, und eine Art Metrik zwischen

---

<sup>4</sup>Dies passiert nicht, wenn man eine zeitliche Gerichtetheit von Ursache und Wirkung annimmt. Lewis tut dies nicht.

<sup>5</sup>Es funktioniert auch, wenn sich die Welten häufen dürfen, ist aber dann etwas komplizierter

den Welten, so dass wir Aussagen der Form  $w_1 \sqsubset_w w_2$  (für “ $w_1$  liegt näher an  $w$  als  $w_2$ ”) machen können.

Nehmen wir vereinfachend an, dass zu einer Menge  $W' \subseteq W$  von Welten immer eine Menge  $\min_{\sqsubset_w} W'$  von nächsten Welten existiert dann gilt  $A \square \rightarrow B$  genau dann, wenn

$$\min_{\sqsubset_w} \{w \in W | w \models A\} \subseteq \{w \in W | w \models B\}$$

also in allen nächsten  $A$ -Welten  $B$  gilt. Analog kann man  $A \diamond \rightarrow B$  dafür schreiben, dass in mindestens einer nächsten  $A$ -Welt  $B$  gilt.

Haben wir damit unser Problem, oder zumindest die Teilprobleme mit der Vertauschung von Ursache und Wirkung und der Epiphänomene, gelöst? Immerhin könnte man doch sagen, dass die nächste Welt, die wie Wirkung nicht enthält, die Ursache ebenfalls nicht enthält, und ähnliches bei Epiphänomenen postulieren. Noch schlimmer, wie in (Menzies, 2001) erwähnt, könnte eine naive Interpretation der Counterfactuals seltsame Aussagen hervorbringen.

Nehmen wir etwa folgende Aussage: “Wenn Nixon auf den roten Knopf gedrückt hätte, hätte es einen Atomkrieg gegeben.” Eine Welt, in der der rote Knopf einen Wackelkontakt hat und der Erde ein Atomkrieg erspart wurde, wäre der unsrigen Welt sicherlich viel ähnlicher, als eine, in der der tatsächlich ein Atomkrieg stattfindet.

Lewis’ Position hierzu ist, daß die Abweichungen bestimmte Gesetzmässigkeiten (Naturgesetze) beachten müssen. Mit seinen Worten:

To get rid of an actual event  $e$  with the least over-all departure from actuality, it will normally be best not to diverge at all from the actual course of events until just before the time of  $e$ . The longer we wait, the more we prolong the spatiotemporal region of perfect match between our actual world and the selected alternative. Why diverge sooner rather than later? Not to avoid violations of laws of nature. Under determinism *any* divergence, soon or late, requires some violation of the actual laws.

Die Interpretation von Counterfactuals ist also nicht trivial, sondern muss vor dem Hintergrund einer Theorie geschehen, die für Vorhersagen sorgt.

Nun zu Shoham (1990). Shoham greift den weiter oben verworfenen Regularitätsansatz in leicht veränderter Form wieder auf, entwickelt die Idee der zugrundeliegenden Gesetzmässigkeiten aber weiter.

Nehmen wir folgendes Beispiel: Dreht man den Zündschlüssel im Schloss, so springt (normalerweise) der Motor an. Er springt natürlich nicht an, wenn die Batterie alle ist, kein Benzin mehr im Tank, oder eine Banane im Auspuff steckt. Diese Dinge prüft man aber im Allgemeinen niemals nach und verschwendet (normalerweise) auch keinen Gedanken daran. Mit einer epistemischen Logik (dazu gleich mehr) könnte man das wie folgt ausdrücken:

$$\square \text{Zündschlüssel-drehen} \wedge \diamond \text{Situation-normal-17} \rightarrow \square \text{Motor-springt-an}$$



Mit anderen Worten, wenn wir glauben, dass jemand den Zündschlüssel gedreht hat und für möglich halten, dass keine Kobolde am Werk waren etc., dann müssen wir ebenfalls glauben, dass der Motor anspringt.

Wollen wir das Ganze etwas formaler definieren, so besteht ein Modell für diese Logik aus einer Menge von Welten, in denen jeweils atomare Aussagen wahr sind oder nicht, aus einer Zugänglichkeitsrelation, die festlegt, welche Welten von einer bestimmten Welt aus “sichtbar” sind, und einer “aktuellen” Welt  $w$ .  $\Box\varphi$  heisst dann, dass in allen “sichtbaren” Welten  $\varphi$  gilt,  $\Diamond\varphi$  (als Abkürzung für  $\neg\Box\neg\varphi$ ) heisst, dass in einer sichtbaren Welt  $\varphi$  gilt.

Normalerweise muss man sich bei einer solchen Logik Gedanken über die Axiomatisierung der Zugänglichkeitsrelation machen (etwa ob sie reflexiv sein müssen soll). Da wir nur logische Kombinationen aus Formeln der Form  $\Box\varphi$  bzw.  $\Diamond\varphi$  (wobei  $\varphi$  keine Modaloperatoren enthält), können wir uns dies zunächst sparen.

Wir stellen ein Ereignis/einen Zustand hier als Aussage der Form  $H(t_1, t_2, p)$  dar,  $p$  ist ein Prädikat, das das Ereignis oder den Zustand identifiziert. Bei einem Ereignis heisst dies, dass es an  $t_1$  angefangen und an  $t_2$  geendet hat, bei einem Zustand würde es heissen, dass der Zustand mindestens über das Intervall von  $t_1$  bis  $t_2$  gedauert hat. Der letzte Punkt sei der *späteste Zeitpunkt* einer solchen Aussage, dies wird auf naheliegende Weise auf Formeln erweitert.

Eine *kausale Aussage* sei nun ein Satz der Form

$$\Phi \wedge \Theta \rightarrow \Box\varphi$$

wobei

- $\varphi$  ein basaler Satz (d.h. ohne Modaloperatoren und ohne Zeitvariablen) ist, dessen spätestester Zeitpunkt  $t_1$  sei.
- $\Phi$  eine Konjunktion von Sätzen der Form  $\Box\varphi_i$  ist, wobei  $\varphi_i$  ein atomarer basaler Satz mit einem spätesten Zeitpunkt  $t_i$  ist, so dass  $t_i < t_1$  gilt.
- $\Theta$  ist eine Konjunktion von Sätzen der Form  $\Diamond\varphi_j$  ist, wobei  $\varphi_j$  ein atomarer basaler Satz und mit einem spätesten Zeitpunkt  $t_j$  ist, so dass  $t_j < t_1$  gilt.

Hierbei dürfen  $\Phi$  und/oder  $\Theta$  leer (d.h. immer wahr) sein. Ist  $\Phi$  leer, so heisst die Aussage *Grenzbedingung*, ansonsten *kausale Regel*.

Wir möchten nun Aussagen aus dem naheliegendsten (d.h. bevorzugten) Modell (bzw. Modellen) ableiten können. Hierfür definieren wir uns eine Relation der *zeitlichen Ahnungslosigkeit* (chronological ignorance)  $\Box_{ci}$ ; im Gegensatz zu vorher gibt es nur eine davon, und sie ist über *Modelle* und nicht über *Welten* definiert. Ein Modell  $M_1$  wird dann gegenüber einem anderen ( $M_2$ ) bevorzugt, wenn sie sich bis zu einem Zeitpunkt  $t_0$  nicht unterscheiden und wenn man an  $t_0$  in  $M_2$  mehr Fakten “weiss” (d.h. es gibt ein  $\varphi(t_0)$  mit  $M_1 \not\models \Box\varphi(t_0)$  und  $M_2 \models \Box\varphi(t_0)$ ).<sup>6</sup>

---

<sup>6</sup>In Shohams Modell wird angenommen, dass Zeitpunkte, oder zumindest interessante Zeitpunkte sich nicht häufen.

Natürlich wäre es noch schöner, wenn wir ein Verfahren hätten, das uns möglichst schnell ein (modulo Äquivalenz) eindeutiges Modell liefert. Hierfür definiert Shoham zunächst weitere Anforderungen an die zugrundeliegende Theorie:

Eine *kausale Theorie*  $\Psi$  ist eine Menge kausaler Aussagen, so daß

- es einen Zeitpunkt  $t_0$  gibt, so dass für alle Grenzbedingungen  $\Theta \rightarrow \Box\varphi$  in  $\Psi$  gilt, dass der späteste Zeitpunkt  $t_1$  von  $\varphi$  vor  $t_0$  liegen muss.
- es keine zwei Aussagen in  $\Psi$  gibt, so dass die eine  $\Diamond H(t_1, t_2, p)$  auf der linken Seite enthält und die andere auf ihrer linken Seite  $\Diamond \neg H(t_1, t_2, p)$  enthält
- Wenn  $\Phi_1 \wedge \Theta_1 \rightarrow \Box H(t_1, t_2, p)$  und  $\Phi_2 \wedge \Theta_2 \rightarrow \Box \neg H(t_1, t_2, p)$  beide in  $\Psi$  vorkommen, so ist  $\Phi_1 \wedge \Theta_1 \wedge \Phi_2 \wedge \Theta_2$  unerfüllbar.

Dann gilt (laut Shoham, dort Theorem 1) folgendes:

Ist  $\Psi$  eine kausale Theorie, dann

- hat  $\psi$  ein zeitlich ahnungslosestes Modell und
- für je zwei zeitlich ahnungsloseste Modelle  $M_1$  und  $M_2$  und jede basale Aussage  $\varphi$  gilt  $M_1 \models \Box\varphi$  gdw  $M_2 \models \Box\varphi$ .

Wir können ein Modell konstruieren, indem wir mit der frühesten Grenzbedingung anfangen und das Modell (bzw. die  $\Box\varphi$ -Aussagen) der Zeitachse entlang weiterstricken.

An dieser Stelle sei angemerkt, dass sowohl die von Lewis (informell) vorgeschlagene Methode zur Evaluation von Counterfactuals als auch Shohams Verfahren zur Modellkonstruktion in zeitlicher Richtung vorgehen. Dies mag daran liegen, dass wir uns – mit den Worten von Hawking gesprochen – an die Zukunft nicht erinnern können.

Nun können wir mit einer kausalen Theorie Voraussagen machen (wobei “Rückwärts”-Kausalität hier per Konstruktionsprinzip ausgeschlossen ist und das Problem der Epiphänomene auf die kausale Theorie abgewälzt wird). Im Gegensatz zu Lewis, der allgemeine Aussagen der Form “ $X$  verursacht  $Y$ ” als unsinnig verwirft, sind bei Shoham Aussagen, die dieser Form ähnlich sind, die Grundbausteine.

Shoham definiert anhand seiner Theorie zu Vorhersagen nun Prädikate zu kausalen Aussagen.

DCAUSES ( $H(t_1, t_2, p), H(t_3, t_4, q), \Phi$ ) gilt, wenn es einen Satz  $H(t_5, t_6, r)$  gibt, so dass

- der Satz

$$\Box H(t_1, t_2, p) \wedge \Diamond H(t_5, t_6, r) \rightarrow \Box(t_3, t_4, q)$$

in  $\Phi$  vorkommt und

- es nicht der Fall ist, dass

$$\Phi, \Box H(t_1, t_2, p) \models_{\Box ci} \Box \neg H(t_5, t_6, r)$$

letzteres verhindert selbst-präemptive Ursachen (etwa eine Zündung, die die Batterie kurzschliesst und damit ein Starten des Motors wiederum verhindert).

Nun zu tatsächlichen Ursachen:  $DCAUSED(H(t_1, t_2, p), H(t_3, t_4, q), \Phi)$  gilt, wenn es einen Satz  $H(t_5, t_6, r)$  gibt, so dass

- beide Bedingungen oben ( $DCAUSED$ ) erfüllt sind
- und  $\Phi \models_{\square_{ci}} \square H(t_1, t_2, p) \wedge \diamond H(t_5, t_6, r)$  gilt

Wie versprochen bekommen wir eine Theorie des Verhinderns gratis dazu:

$DPREVENTS(H(t_5, t_6, r), H(t_3, t_4, q), \Phi)$  gilt, wenn es einen Satz  $H(t_1, t_2, p)$  gibt, so dass  $\square H(t_1, t_2, p) \wedge \diamond \neg H(t_5, t_6, r) \rightarrow \square H(t_3, t_4, q)$  ein Satz aus  $\Phi$  ist.

$DPREVENTED(H(t_5, t_6, r), H(t_3, t_4, q), \Phi)$  gilt, wenn es einen Satz  $H(t_1, t_2, p)$  gibt, der obige Bedingung erfüllt und  $\Phi \models_{\square_{ci}} \square H(t_1, t_2, p) \wedge \neg \diamond H(t_3, t_4, q)$  gilt.

Wir haben nun Kausalität als Artefakt einer qualitativen Theorie dargestellt, die (nicht-monotone) Voraussagen ermöglicht. Folgende Fragen sind jedoch noch offen:

- Woher soll diese qualitative Theorie kommen?
- Was ist mit Aussagen wie “Rauchen verursacht Krebs”? Es ist sicherlich nicht (sachlich) richtig, anzunehmen, dass alle oder die überwiegende Mehrzahl der Raucher später Krebs bekommt.

Zunächst zur Frage, woher wir unsere Vorhersage-Theorie nehmen.

Aus pragmatischer Sicht ist eine Vorhersage-Theorie notwendig, um die Folgen von Handlungen einzuschätzen bzw. Möglichkeiten der Beeinflussung aufzuzeigen. So zitiert ([Woodward, 2001](#)) von Wright:

... to think of a relation between events as causal is to think of it under the aspect of (possible) action. It is therefore true, but at the same time a little misleading to say that if  $p$  is the cause of  $q$ , then if I could produce  $p$  I could bring about  $q$ . For that  $p$  is the cause of  $q$ , I have endeavored to say here, *means* that I could bring about  $q$ , if I could do (so that)  $p$ .

Diese Sichtweise liefert uns sicherlich ein Kriterium, um Epiphänomene auszuschließen. Stellen wir uns vor, dass ein Tiefdruckgebiet das Fallen des Barometers verursacht, und kurz darauf ein Sturm heraufzieht. Es gibt einen Zusammenhang zwischen dem Barometer und dem Sturm, aber die naheliegendsten Methoden, das Fallen des Barometers zu verhindern (etwa, den Zeiger zu fixieren oder das Barometer in eine Kompressionskammer zu legen) würden nichts am Sturm ändern.

In diesem Sinne ist die Aussage, Rauchen verursache Krebs derart sinnvoll, dass man, wenn man keinen Krebs bekommen will, dies zu erreichen versuchen kann, indem man nicht raucht.

Andererseits, und hier liegt die Grenze der Erklärungsfähigkeit des Manipulierbarkeits-Ansatzes, gibt es keine (naheliegenden) Methoden, etwas an dem heranahendem Tiefdruckgebiet zu ändern.

Änderungen an Tiefdruckgebieten müssen in Gedankenexperimente verlegt werden, die auf vergangene Erfahrungen zurückgehen. Mit den Häufigkeiten aus vergangenen Beobachtungen können wir versuchen, Wahrscheinlichkeiten oder zumindest Häufigkeiten für Ereignisse zu ermitteln, wie man es in der Medizin und Soziologie tut, wenn man eine kausale Verbindung glaubhaft machen möchte.

Naheliegender ist etwa

$$P(B|A) > P(B)$$

bzw.  $P(B \wedge A) > P(B \wedge \neg A)$ , um Problemen mit bedingten Wahrscheinlichkeiten aus dem Weg zu gehen. Wir müssen zusätzlich eine Gerichtetheit in der Zeit voraussetzen, weil Umkehrungen sonst stets möglich sind.

Die Probleme, die man dann mit Epiphänomenen bekommt und die Lösungen, die dafür in Frage kommen, sind in (Hitchcock, 2002) beschrieben, worauf hier aus Platzgründen verwiesen sei.

### 2.3 Kausale und andere Erklärungen

In vielen Kontexten wird aus den kausalen Faktoren, die die Entstehung eines Ereignisses beeinflusst haben, einer als “die” Ursache bezeichnet, während die anderen als Nebenbedingungen gelten.

Im Falle einer Hungersnot etwa könnte als “die” Ursache entweder die vorherrschende Trockenheit gelten oder die Unfähigkeit der örtlichen Regierung, auf die Trockenheit zu reagieren. Die Auswahl der Erklärung beeinflusst die evaluative/normative Sicht auf das Ereignis (Verantwortung).

Neben der eigentlichen Kausalität (also den oben besprochenen Beziehungen zwischen Ereignissen) existieren weitere Relationen, die zur Erklärung von Ereignissen bzw. mit Ereignissen geeignet sind.

Bei der eigentlichen Kausalität wird eine Eventualität mit einer anderen verknüpft:

(2.6) Johann fiel von der Brücke, weil er geschubst wurde.

(2.7) Johann fiel von der Brücke, weil er betrunken war.

(2.8) Weil er die Schnapsflasche geleert hatte, war Johann betrunken.

Eine kausale Verknüpfung, bei der die Ursache ein Ereignis ist, bleibt bestehen, wenn man die Sätze nicht mit *weil*, sondern mit *nachdem* verknüpft, dessen Funktion primär zeitlich ist.

(2.9) Nachdem Johann geschubst wurde, fiel er von der Brücke.

(2.10) Nachdem er die Schnapsflasche geleert hatte, war Johann betrunken.

### 2.3 Kausale und andere Erklärungen

---

Handlungen können durch Ziele, Geisteszustände, vorhergegangene Ereignisse, aber auch zukünftige Ereignisse erklärt werden:

- (2.11) Johann sprang von der Brücke, um sich abzukühlen.
- (2.12) Johann sprang von der Brücke, weil er eine Abkühlung brauchte.
- (2.13) Johann sprang von der Brücke, weil ihn seine Freundin verlassen hatte.
- (2.14) Johann kauft sich einen neuen Anzug, weil sein Bruder morgen heiratet.

Erklärungen müssen sich nicht auf Ereignisse beziehen, sondern können zwischen beliebigen Aussagen stehen:

- (2.15) Der Fisch ist gefährlich, weil er keine natürlichen Feinde hat.

Auf eine Handlung sind oft mehrere verschiedene Sichten möglich, etwa folgende Beschreibungen ein und desselben Ereignis (nach einem Beispiel aus (Allen, 1984)):

- Paul hat auf dem Klavier in rhythmischer Abfolge schwarze und weisse Tasten gedrückt.
- Paul hat Golliwogg's Cakewalk einigermaßen fehlerfrei gespielt.
- Paul hat Marie aufgeweckt.

oder auch:

- Er hat den Wäschetrockner benutzt.
- Er hat die Wäsche getrocknet.

Zwischen diesen Beschreibungen besteht oft so etwas wie eine hierarchische Beziehung:

- (2.16) Paul hat Marie aufgeweckt, indem er Golliwogg's Cakewalk gespielt hat.
- (2.17) # Paul hat Golliwogg's Cakewalk gespielt, indem er Marie aufgeweckt hat.

Allen nennt diese Beziehung nach Goldman *generates*. Inui (2004) bezeichnet diese Beziehung als *means*.

### 2.3.1 Kausalbeziehungen in Conceptual Dependency

In der Conceptual Dependency-Theorie (Schank, 1972; Schank und Abelson, 1977) wird zwischen einer Anzahl verschiedener Kausalbeziehungen unterschieden:

**r** Eine Handlung ändert einen Zustand

**E** ein Zustand ermöglicht eine Handlung/ein Ereignis

**dE** ein Zustand verhindert eine Handlung/ein Ereignis

**I** ein Zustand/eine Handlung erzeugt einen Geisteszustand

**R** ein Geisteszustand ist Grund für eine Handlung

dazu werden folgende Abkürzungen eingeführt:

**rE** Eine Handlung ergibt einen Zustand, der eine Handlung erlaubt

**IR** eine Handlung oder ein Zustand erzeugt einen Geisteszustand, der Grund für eine Handlung ist.

Nehmen wir einige Beispiele und schauen, ob wir sie mit obigen Relationen analysieren können:

(2.18) Johann fiel von der Brücke, weil er betrunken war.

(2.19) Johann sprang von der Brücke, weil seine Frau ihn verlassen hatte.

(2.20) Johann räumt seinen Schreibtisch auf, weil er morgen in den Urlaub fährt.

Beispiel 2.18 ist trivial, falls wir die Unterscheidung zwischen Ereignissen und Handlungen einmal ausser Acht lassen:

Johann betrunken  
↓ **E**  
Johann fällt von der Brücke

Beispiel 2.19 können wir mit der Abkürzung **IR** beschreiben:

Johanns Frau verlässt ihn  
↓ **IR**  
Johann springt von der Brücke

Das letzte Beispiel (2.20) ist etwas schwieriger, weil das in-den-Urlaub fahren hypothetisch ist. Versuchen wir es so:

Johann plant, in den Urlaub zu fahren  
↓ **I**  
Johann möchte, dass sein Arbeitsplatz aufgeräumt ist  
↓ **R**  
Johann räumt seinen Schreibtisch auf

dry-laundry-in-the-sun(\$actor,\$laundry)

<i>precondition:</i>	weather(sunny)
<i>effect:</i>	get-dry(\$laundry)
<i>decomposition:</i>	hang(\$actor,\$laundry)

Abbildung 2.1: Beispiel eines Plan-Operators aus (Inui, 2004)

Eine ähnliche Konstruktion eines Geisteszustandes ist auch im folgenden Beispiel aus (Lascarides und Asher, 1993) möglich:

(2.21) Max öffnete die Tür. Es war stockdunkel im Zimmer.

Lascarides und Asher analysieren dies als *Background*, also daß der zweite Satz Umstände aufführt, unter denen das Ereignis aus dem ersten Satz passiert – die genaue Beziehung der Ereignisse wird klar, wenn man versucht, dies anhand der Conceptual Dependency-Relationen zu analysieren.

Max öffnet die Tür [zum Zimmer]	im Zimmer ist es stockdunkel
↓ <b>E</b>	↓ <b>I</b>
Max sieht, dass es im Zimmer stockdunkel ist	

Diese Analyse erlaubt Schluss auf Erzählperspektive; Wäre Max blind, so würde der Text inkohärent.

### 2.3.2 Inuis Planbox-Modell

Inui (2004) geht von einer Struktur aus, wie sie für das Planen von Handlungen postuliert wird:

Die Relation zwischen der Vorbedingung und der Gesamthandlung nennt Inui *precond*, die zwischen Gesamthandlung und Nachbedingung *effect*, die zwischen Gesamthandlung und Dekomposition *means*. Kausale Relationen zwischen zwei Ereignissen, die nicht im Kontext einer Handlung stehen, nennt er *cause*.

Inui weist besonders darauf hin, dass solche kausalen Axiome, nicht nur in Kausalbeziehungen eine Rolle spielen, sondern auch bei anderen Konstruktionen, wie in folgenden Beispielen:

- (2.22)
- a. The laundry dried well today *because* it was sunny.
  - b. The laundry dried well, *though* it was not sunny.
  - c. *If* it was sunny, the laundry could dry well.
  - d. The laundry dried well *because of* the sunny weather.

Inui (2004) stellt auch ein Verfahren vor, um die japanische Konjunktion *tame* in Bezug auf diese Einteilung zu disambiguieren. Hierzu sei auf Kapitel 4 verwiesen.

### 2.3.3 Vergleich

diese Überschrift ist doof

intuitiv entspricht Schanks **E**-Relation Inuis *precond*-Beziehung beziehungsweise der *enables*-Beziehung von Shoham, **r** Inuis *result* bzw. Shohams *dcauses*.

Die **I** und **R**-Relationen von Schank finden sich bei Inui und Shoham nicht wieder, da diese Intentionalität nicht modellieren.

Fragt man nach dem Grund für eine Handlung, ist meist eine Verbindung vom Typ **IR** gefordert, beim Grund für einen Zustand oder ein Ereignis eher **r** bzw. **E** oder **rE**. Sind zwei Handlungen desselben Agenten mit **E** verbunden, so kann die zweite Handlung als Grund für die erste gedeutet werden.

- (2.23) a. Sie stieg ins Auto und fuhr davon  
b. Das Ventil öffnete sich und der Kesseldruck sank.
- (2.24) a. \* Sie stieg ins Auto. Dadurch fuhr sie davon.  
b. Das Ventil öffnete sich. Dadurch sank der Kesseldruck.
- (2.25) a. Sie stieg ins Auto, um davonzufahren.  
b. \* Das Ventil öffnete sich, um ...

Nicht immer jedoch sind **E**-Relationen auch als kausale Erklärungen zu verstehen. Schauen wir uns einmal folgenden Ausschnitt aus einem Beispiel von (Schank und Abelson, 1977) an:

John hält eine offene Bierdose	der Stuhl kippt um
↓ <b>E</b>	↓ <b>E</b> ↓ <b>E</b>
Bier spritzt auf den Stuhl	John fällt auf den Boden

Fragt man, warum Bier auf den Stuhl gespritzt ist, so sind beide Fakten (das Halten einer offenen Bierdose und das Umkippen des Stuhls) für sich alleine genommen nicht ausreichend, als Erklärung würden entweder beide Ereignisse oder (als “der” Grund) das Umkippen des Stuhls genommen; während bei Handlungen **E**-Relationen in der Regel als Hintergrundbedingungen zu verstehen sind, stellen sie bei anderen Ereignissen Ursachen dar. Bei der Auswahl “des” Grundes scheint wiederum eine graduelle Version der Unterscheidung zwischen Ursachen und Hintergrundbedingungen eine Rolle zu spielen – es wäre zu untersuchen, welche Gesetzmässigkeiten diesem graduellen Unterschied zugrundeliegen.



## Kapitel 3

# Kausalität im Text

Ein System, das kausale Verbindungen zwischen den im Text evozierten Ereignissen finden soll, das aber im Unterschied zu einem Menschen nicht über allzu detailliertes Wissen über diese Ereignisse verfügt, ist auf Vermutungen angewiesen, die mit Indizien wie dem Vorhandensein (oder auch nicht) und der Art einer Markierung im Text zumindest teilweise verifiziert oder falsifiziert werden können/müssen.

Folgendes wäre denkbar:

- die Relationen sind explizit durch Konnektoren oder andere Marker markiert. Diese Marker können entweder eindeutig kausal sein oder mehrdeutig sein.
- die Relation ergibt sich durch die Stellung im Text, möglicherweise zusammen mit Wissen, das hierüber Plausibilitätsurteile erlaubt.
- die Relation besteht zwischen im Text erwähnten Ereignissen, deren Beziehung auch durch die Stellung im Text nicht erschlossen werden kann.

Schauen wir uns angestrebte Anwendungsdomänen an, so wird klar, dass die Anforderungen in Hinsicht auf die Ausbeute eines Taggers für kausale Relationen (oder den Teil des Systems, der diese Aufgabe miterfüllt) recht unterschiedlich sind und damit auch der Grad, zu dem man sich auf die (möglichst eindeutige) Markiertheit der kausalen Relationen verlassen darf.

Einerseits könnte ein solches System dazu benutzt werden, Weltwissen über eine Domäne zu sammeln, wie es Inui (2004) oder Torisawa (2003) tun. Da ohnehin davon auszugehen ist, dass die meisten Fakten nicht in der Domäne auftauchen, andererseits aber auf diese Weise mehr Fakten gefunden werden können als ein menschlicher Experte sich auf Anhieb ausdenken kann, ist hier ein Ansatz, der sich wie der von Inui hauptsächlich auf Diskursmarker stützt und damit eine gute Präzision erreicht, gut geeignet.

Anders wäre es bei einem System, das Diskursrelationen erkennen soll. Marcu und Echihabi (2002) erwähnen, dass in einem mit Diskursrelationen annotierten

Korpus (Carlson et al., 2001) nur 61 von 238 (Contrast) bzw. 79 von 307 (Explanation/Evidence) markiert sind. Ein System, das sich hier allein auf Diskursmarker verlässt wäre also nicht erfolgreich.

Bei einem System, das für Question-Answering benutzt werden soll, sollte man ausnutzen, dass bei den Fragen zum Text evidente Zusammenhänge (die im Text nicht markiert sind und demnach schwieriger zu finden) eine weniger große Rolle spielen als solche, die nicht-evident sind und eine Markierung durch Diskursmarker bzw. kausative Verben fordern. Allerdings sind nicht alle Antworten auf Kausalrelationen derartig markiert, sondern zum Teil auch benachbart, aber nicht markiert oder an nicht benachbarten Stellen im Text (wie hängen die dann in der Diskursstruktur zusammen?). Für Question-Answering auf grossen Dokumentensammlungen kann man hoffen, dass aufgrund von Redundanz zumindest eine Antwort gefunden werden kann, die eindeutig markiert ist, in speziellen Domänen sowie für Reading Comprehension kann man nicht davon ausgehen und sollte daher auch die schwierigeren Fälle mit einbeziehen.

Hier ein Beispiel<sup>1</sup>, das diese Unterscheidung illustrieren soll:

- (3.1) WERNER NIEFER, Chef von Mercedes-Benz, soll nach einem von ihm verschuldeten Verkehrsunfall im Mai 1990 in Rom nun eine Strafe von 80 000 Mark zahlen.
- (3.2) Die Stuttgarter Staatsanwaltschaft erklärte am Mittwoch, sie habe einen entsprechenden Strafbefehl wegen fahrlässiger Körperverletzung beantragt.
- (3.3) Niefer hatte im Mai 1990 in Rom einen Reisebus gefahren und den Wagen nach Darstellung der Staatsanwaltschaft nicht mehr rechtzeitig gestoppt, als ihm Fußgängerinnen entgegenkamen.
- (3.4) Deshalb sei eine deutsche Touristin von dem Bus erfaßt und am Bein erheblich verletzt worden.
- (3.5) Die Staatsanwaltschaft teilte weiter mit, die umfangreichen Ermittlungen hätten einen hinreichenden Tatverdacht wegen weiterer Delikte wie alkoholbedingte Straßenverkehrsgefährdung, unerlaubtes Entfernen vom Unfallort und Fahren ohne Fahrerlaubnis nicht ergeben.

Wir finden den expliziten Marker *deshalb* in Satz 3.4, auch den zeitlichen Marker *nach*, der im Kontext von 3.1 einen kausalen Zusammenhang markiert.

Die Beziehung zwischen Ermittlungen und Tatverdacht (der hier negiert ist) und durch das Kausalverb *ergeben* vermittelt wird, ist ebenfalls kausaler Natur.

Die Beziehung zwischen dem Verkehrsunfall (3.1) und dem Strafbefehl (3.2) ist nicht markiert. Hier scheint entscheidend zu sein, dass die Beziehung zwischen Strafe und Strafbefehl dem Leser klar ist.

<sup>1</sup>Negra-Corpus, Sätze 17327ff

### 3.1 Kausalität und Temporalität

---

Wie früher erwähnt, müssen *weil* und *deshalb* nicht immer (im engeren Sinne) kausal, sondern kann allgemein Erklärungen markieren<sup>2</sup>:

- (3.6) a. Die Sozialdemokratie gehört jedoch deshalb unwiederbringlich der Vergangenheit an, *weil* sie noch immer meint, die Krise des Wohlfahrtsstaates einzig und allein auf dem Feld der Ökonomie bewältigen zu können.
- b. "Man produziert Unsinn, *weil* die Leute Unsinn wollen - und die Leute wollen Unsinn, weil sie nichts anderes mehr kennen."
- c. Die Straßenbahnen der Linie 19 rauschen bald schneller durch Bockenheim, *weil* die Ampelschaltungen besser auf ihre Fahrt abgestimmt werden.
- (3.7) a. Das Gesicht stelle beim Säugling die größte Fläche unbedeckter Haut dar und sei deshalb für die Wärmeregulierung entscheidend.
- b. Die Polizei war losgelöst von der allgemeinen Verwaltung, und die Bewohner fühlten sich als "Mußpreußen".  
Der Haß gegen das Preußentum richtete sich *deshalb* zwangsläufig gegen das Polizeipräsidium und deren führende Persönlichkeiten.

Während Beispiel 3.6c etwas mit Kausalität zu tun hat und Beispiele 3.6b und 3.7b sich auf Gründe beziehen, sind 3.6a und 3.7a Begründungen und nicht Verknüpfungen zwischen Ereignissen.

### 3.1 Kausalität und Temporalität

Zeitliche Marker werden oft benutzt, um kausale Beziehungen auszudrücken:

- (3.8) Wir haben in Holland Urlaub gemacht. *Nun* sind wir pleite.
- (3.9) *Nach* unserem Urlaub waren wir pleite.
- (3.10) "Wie vor den Kopf geschlagen" fühlten sich die Eltern, *als* der Landrat beschloss, die asbesthaltigen Baustoffe, die inzwischen auch in den Trennwänden festgestellt wurden, schon in den Sommerferien beseitigen zu lassen.
- (3.11) Der Schuhmacher Julius Lieske (20) wurde *nach* mühsamen Ermittlungen der Tat überführt und drei Monate später in Limburg durch das Beil hingerichtet.
- (3.12) *Nachdem* in den Niederlanden, Großbritannien und Neuseeland die Öffentlichkeit über die Risiken der Bauchlage aufgeklärt wurde, ging der Plötzliche Säuglingstod um ein bis zwei Drittel zurück.<sup>3</sup>

---

<sup>2</sup>Negra-Corpus, Sätze 1372,1955,3264;1323,174/175

<sup>3</sup>Negra-Corpus, Sätze 16471,179,1310

Dies lässt sich einerseits so erklären, daß zeitliche Marker die zeitliche Relation zwischen den erwähnten Eventualitäten einschränken und die geschlossene zeitliche Relation Einfluß darauf hat, ob der Leser eine kausale Verbindung annimmt oder nicht.

Ist die zweite Eventualität ein Zustand, so wird angenommen, dass dieser Zustand mit der ersten Eventualität zeitlich überlappt (*temporal overlap*). Um obiges Beispiel wieder aufzugreifen, würde

(3.13) Wir haben in Holland Urlaub gemacht. Wir waren pleite.

so verstanden werden, dass das Pleitesein während des Urlaubs gegolten hat.

Amsili und Rossari (1998) untersuchten, wie sich verschiedene Kombinationen von Anordnung und Tempus (Passé simple vs. Imparfait) sowie des Konnektors “donc” (deshalb) auf Akzeptabilität einerseits und den Ausdruck von Kausalität andererseits auswirken.

Ein Ergebnis ist, dass bei der Kombination von Passé simple (PS) und Imparfait (IMP) Kausalität (hier: Explanation als Diskursrelation) geschlossen wird, falls das beschriebene Ereignis nicht vollendet sein muss, um die andere Eventualität zu bewirken:

- (3.14) a. Jean attrapa une contravention. Il roulait trop vite.  
*Jan hat sich einen Strafzettel eingefangen. Er fuhr zu schnell*
- b. #Jean tomba. La branche cassait.  
*Jan ist gefallen. Der Ast brach ab.*

Amsili und Rossari führen diesen Unterschied darauf zurück, dass im zweiten Fall die zweite Eventualität abgeschlossen sein muss, um als Ursache für die erste in Frage zu kommen (*accomplished causality*), während sie es im ersten Fall nicht sein muss (*non-accomplished causality*, Lewis (1986) spricht von “piecemeal causation”).

Diesen Unterschied kann man im Deutschen mit den zeitlichen Markern *dabei* und *kurz darauf* beobachten. Bei der *non-accomplished causality* wird die kausale Lesart mit *kurz darauf* nicht angenommen.

- (3.15) a. Max reparierte sein Auto. Dabei machte er sich dreckig.  
b. (?) Der Zweig brach. Dabei fiel Jan.

- (3.16) a. # Max reparierte sein Auto. Kurz darauf machte er sich schmutzig.  
b. Der Zweig brach. Kurz darauf fiel Jan.

Eine andere Sicht auf die kausale Wirkung zeitlicher Marker wäre die, dass zeitliche Marker die zwei Eventualitäten in eine Beziehung setzen, die nicht allein in der zeitlichen Beziehung der Eventualitäten liegt. So lässt sich *als*, das Herweg (1991) als “vage Simultaneität” (d.h. als Überlappung) analysiert, auch einsetzen, um Eventualitäten zu verknüpfen, die zwar aufeinander bezogen sind, sich aber (zumindest nach üblicher Analyse) nicht zeitlich überlappen. Während

### 3.1 Kausalität und Temporalität

---

(3.17) \*Als Peter das Zimmer betrat, verließ er es wieder.

nicht geht, sind

(3.18) a. Als Maria Peter erblickte, erschrak sie.

b. Als Peter sein Examen mit einer Fete feierte, lud er auch Maria ein

möglich, obwohl sich die Ereignisse nur mit etwas “pragmatischer Toleranz” (Herweg) als überlappend ansehen lassen<sup>4</sup>.

Bras et al. untersuchten den Einfluss der Konnektive “puis” (dann) und “un peu plus tard” (kurz darauf) auf die Diskursstruktur haben.

Wenn eine verzögerte Ursache ausgeschlossen ist, kann *un peu plus tard* Result ausschließen, in ähnlicher Weise wie von Amsili und Rossari beobachtet.

(3.19) a. Max trébucha. Il tomba et se cassa la jambe.

*Max stolperte. Er fiel und brach sich das Bein.*

b. Max trébucha. **Puis** il tomba et se cassa la jambe.

*Max stolperte. Dann fiel er und brach sich das Bein.*

c. Max trébucha. **Un peu plus tard**, il tomba et se cassa la jambe.

*Max stolperte. Kurz darauf fiel er und brach sich das Bein.*

Interessant ist, dass sich trotz der gleichen zeitlichen Beziehung ein unterschiedliches Verhalten bezüglich der Zulässigkeit von Result als Diskursrelation. *puis* drückt das direkte Aufeinanderfolgen von Ereignissen aus und erlaubt Narration als einzige Diskursrelation, wogegen *un peu plus tard* Result im Allgemeinen nicht ausschließt:

(3.20) a. L'acide tomba dans le mélange. Une explosion se produisit.

*Die Säure fiel in das Gemisch. Es gab eine Explosion*

b. L'acide tomba dans le mélange. **Puis** une explosion se produisit.

*Die Säure fiel in das Gemisch. Dann gab es eine Explosion.*

c. L'acide tomba dans le mélange. **Un peu plus tard** une explosion se produisit.

*Die Säure fiel in das Gemisch. Kurz darauf gab es eine Explosion.*

Ähnlich verhält sich *dann* im Deutschen:

(3.21) Wir sind nach Holland gefahren. Dann waren wir pleite.

Die Verbindung mit *dann* legt nicht nahe, dass das Pleitesein etwas mit dem Nach-Holland-Fahren zu tun hat.

*Nachdem*, das laut Herweg (1991) ein Ereignis im Nachzustand eines anderen lokalisiert, kann ebenfalls verwendet werden, um einen rein pragmatischen (nicht-kausalen) Erklärungszusammenhang darzustellen<sup>5</sup>:

<sup>4</sup>Beispiele 76a,76b,77a aus (Herweg, 1991)

<sup>5</sup>Negra-Corpus, Satz 2755

- (3.22) Nachdem die Stadtbevölkerung in den letzten Jahren immer wieder erfreut feststellen durfte, dass der als so trübe verrufene Main beispielsweise für eine ganze Reihe gefiederter Freunde ein durchaus attraktives Revier ist, könnten auch die Yachtlenker das ihrige dazu tun, um solche (wiederentstehenden) Paradiese nicht zu gefährden.

Erklärt wird die Forderung, die Yachtlenker mögen das ihrige dazu tun, die Erklärungsbeziehung ist aber eher pragmatisch als kausal; das Erklärende ist nicht das Perzeptions-Ereignis der Stadtbevölkerung, sondern der Zustand des Mains.

*Nachdem* muss nicht unbedingt eine Erklärungsbeziehung markieren, es ist auch eine rein zeitliche Lesart möglich, die eine andere Kohärenzrelation nahelegt:<sup>6</sup>

- (3.23) Nachdem das Anwesen 1865 in den Besitz der Stadt Bensheim übergegangen war, stellte diese die großen Räume zunächst den evangelischen Bürgern als Betsaal zur Verfügung.
- (3.24) Nachdem die Rückenlage traditionell als die natürliche Schlafposition des Säuglings gegolten hatte, wurde ab den frühen siebziger Jahren in der medizinischen Fachliteratur die Bauchlage propagiert.

Das erste Beispiel (3.23) enthält zwar eine *E*-Relation, das *zunächst* schließt aber eine finale Lesart aus. Im darauffolgenden Text geht es um die weitere Nutzung des Anwesens, durch den *Nachdem*-Nebensatz wird der Nachzustand des ingressiven Besitzübergangs-Ereignisses thematisiert und der Hauptsatz ist eine Elaboration dieses Zustands.

Im zweiten Beispiel dient das *nachdem* vor allem dazu, einen Kontrast (Rückenlage-Bauchlage) zu markieren.

## 3.2 Diskursstruktur

Die Sätze des Textes und die darin beschriebenen Ereignisse/Zustände/Fakten stehen nicht für sich allein, sondern sind in einer bestimmten Weise strukturiert. Allgemein wird angenommen, dass diese Struktur baumförmig ist und die Integration der Sätze des Textes in eine solche Struktur einen wesentlichen Bestandteil dessen ausmacht, was als Kohärenz des Textes bezeichnet wird.

Die Art, wie diese Struktur (auf Seite des Sprechers) erzeugt wird, ist vom Zusammenspiel von Intentionen auf der einen Seite (Grosz und Sidner, 1986) und Konventionen der inhaltlichen Strukturierung auf der anderen Seite vorgegeben, die partielle Struktur, die der Hörer aufbaut<sup>7</sup> stellt einen wichtigen Teil des Kontexts dar, in dem weitere Sätze des Diskurses interpretiert und eingeordnet werden.

---

<sup>6</sup>Negra-Corpus, Sätze 2083, 1327

<sup>7</sup>Es gibt Modelle wie das in (Walker, 1998), die gängige Diskursphänomene erklären, ohne so etwas wie eine Diskursstruktur vorauszusetzen. Dies ist aber eher die Ausnahme als die Regel.

So ist die Zugänglichkeit der Referenten von Pronominalanaphern (Er, Sie) und bestimmten Formulierungen, die Situationen aufgreifen (“auf diesem Wege”, “des-halb”) sehr stark durch diesen Kontext geprägt, die von (anaphorischen) definiten Beschreibungen wie “der Präsident” oder Referenzen auf Situationen/Fakten, wie etwa “als sie sich küssten” nicht so stark.

Die Zugänglichkeit von Referenten hängt mit der Anbindung des Satzes an die bereits vorhandene partielle Diskursstruktur zusammen, die, so nehmen etwa (Asher und Lascarides, 1996) an, weitere Inferenzen erlaubt. Es wird angenommen, dass für diese Anbindung die Knoten am “rechten Rand” des Diskursbaumes zugänglich sind (Webber, 1991); bei der Auflösung definitiver Beschreibungen passiert es möglicherweise, dass der Knoten, über den die Anbindung geschieht, in den rechten Rand aufgenommen oder auf andere Weise zugänglich gemacht wird. Webber et al. (2001) vertreten die Ansicht, dass Diskursmarker wie *trotzdem* oder *dann* als Anaphern funktionieren und damit (auf ähnliche Weise wie bei den Nominalanaphern vermutet) eine Diskursstruktur hervorbringen, die bei einem nicht-anaphorischen Modell (in dem *trotzdem* etc. ähnlich wie konjunktionale Diskursmarker behandelt werden) zu nicht-projektiven Diskursstrukturen führen würde.

In Ergänzung der Beschränkung auf den rechten Rand zeigt Walker (1996), dass Verdrängungseffekte ältere Knoten im Diskurs unzugänglich machen können, so dass beim “Aufstieg” zu diesen Knoten zusätzliche Äusserungen gemacht werden müssen, die aus Sicht eines reinen “Stapel”-Modells informationell überflüssig sind.<sup>8</sup>

Die Relationen zwischen den Knoten des Diskursgraphen (Diskursrelationen) werden in der Literatur sowohl als rein intentional (Grosz und Sidner, 1986) als auch allein mit Bezug auf die beschriebenen Entitäten (Asher und Lascarides, 1996) beschrieben.

Wilensky (1994) argumentiert, dass die Diskursrelationen im wesentlichen semantische Relationen zwischen den Objekten des Diskurses (die dann nicht nur Eventualitäten, sondern auch Propositionen umfassen). Dies stützt er damit, dass Argumente, Thesen, Kontraste etc. auch ausserhalb eines (textuellen) Diskurses nützlich und bedeutsam sind.

Während sich Mann und Thompson (1987) und (Hovy, 1990) für eine offene Menge von Diskursrelationen aussprechen, die insbesondere durch die Konventionen der Domäne bestimmt sind, vertreten (Knott, 1994) und (Asher und Lascarides, 1996) eine abgeschlossene (und relativ kleine) Menge von Diskursrelationen.

Die Bedeutung des Diskurses enthält mehr Information als die einzelnen Sätze für sich genommen hätten, wie das folgende (mittlerweile klassische) Beispiel zeigt:

- (3.25) a. Max fell. John pushed him.  
b. Max fell. John helped him up.

---

<sup>8</sup>in einer dynamischen Sichtweise entspricht ein Baum, an dessen rechten Rand sich weitere Knoten anlagern können, einem Stapel, von dem Knoten heruntergenommen und dann neue hinzugefügt werden.

- $\pi_0$  ADC Names Stephen Mitchell Vice President of National Accounts
- $\pi_1$  MINNEAPOLIS, Jan 30, 2003 (BUSINESS WIRE) — ADC (Nasdaq: ADCT; www.adct.com) today announced the appointment of Stephen Mitchell as Vice President of National Accounts.
- $\pi_2$  In this role, Mitchell has overall responsibility for driving ADC's channel sales strategies, including channel development in North America
- $\pi_3$  He will also have responsibility for several service provider accounts including Sprint and Alltel as well as ADC's government, original equipment manufacturer (OEM) and broadcast markets.
- $\pi_4$  Based in the company's Richardson, Texas, office, Mitchell reports to Jay Hilbert, ADC's senior vice president of Global Sales and Marketing.

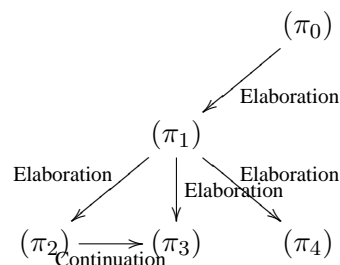


Abbildung 3.1: Diskursgraph aus (Asher et al., 2004)



## 3.2 Diskursstruktur

---

Im ersten Beispiel wird der zweite Satz als kausale Erklärung und das Ereignis als zeitlich davor liegend eingeordnet, im zweiten Beispiel wird das Ereignis des zweiten Satzes als zeitlich danach liegend interpretiert.

Schaut man sich eine Definition aus (Mann und Thompson, 1987) an, so wird deutlich, dass diese zu einem Henne-Ei-Problem führen können, wenn man gleichzeitig die Diskursrelation und die Folgerungen daraus raten will:

tut man das wirklich?

relation name:	VOLITIONAL CAUSE
constraints on N	presents a volitional action or else a situation that could have arisen from a volitional action
constraints on S:	none
constraints on the N+S combination:	S represents a situation that could have caused the agent of the volitional action in N to perform that action; without the presentation of S, R might not regard the action as motivated or know the particular motivation; N is more central to W's purposes in putting forth the N-S combination than S is.
the effect:	R recognizes the situation presented in S as a cause for the volitional action presented in N
locus of the effect:	N and S

Dieses Problem haben Lascarides und Asher (1993) versucht, mit einer stärkeren Formalisierung der Beziehungen zwischen den Bestandteilen der Diskursstruktur zu lösen. Sie unterscheiden zwischen folgenden Diskursrelationen:

- Explanation( $\alpha, \beta$ ): das Ereignis aus  $\beta$  erklärt, warum das Ereignis aus  $\alpha$  passiert ist (möglicherweise indem es es bewirkt hat)
- Elaboration( $\alpha, \beta$ ): das in  $\beta$  erwähnte Ereignis ist Teil des in  $\alpha$  beschriebenen (bzw. dessen Vorbereitungsphase)
- Narration( $\alpha, \beta$ ): das in  $\beta$  beschriebene Ereignis ist eine Konsequenz aus (aber nicht im engeren Sinne bewirkt von) den in  $\alpha$  beschriebenen Ereignis.
- Background( $\alpha, \beta$ ): Der Zustand aus  $\beta$  beschreibt die Umstände, unter denen das Ereignis in  $\alpha$  stattfindet.
- Result( $\alpha, \beta$ ): Das durch  $\alpha$  beschriebene Ereignis verursachte das Ereignis oder den Zustand, der in  $\beta$  beschrieben ist.

Diese Relationen werden mit Hilfe einer quantorenfreien Defaultlogik axiomatisiert.

$\langle \tau, \alpha, \beta \rangle$  bedeutet, dass im Diskursgraphen  $\tau$  die neue DRS  $\beta$  an die schon vorhandene DRS  $\alpha$  angebunden wird.

Wenn an  $\beta$  einen Zustand beschreibt, so wird angenommen, dass das Ereignis aus  $\alpha$  diesen überlappt.

$$\langle \tau, \alpha, \beta \rangle \wedge \text{state}(\text{me}(\beta)) > \text{overlap}(\text{me}(\alpha), \text{me}(\beta))$$

Narration ist die normalerweise angenommene Beziehung:

$$\langle \tau, \alpha, \beta \rangle > \text{Narration}(\alpha, \beta)$$

Die Relation legt eine zeitliche Reihenfolge fest:

$$\Box(\text{Narration}(\alpha, \beta) \rightarrow \text{me}(\alpha) \prec \text{me}(\beta))$$

Die beiden DRSen haben ein gemeinsames Thema<sup>9</sup>

$$\Box(\text{Narration}(\alpha, \beta) \rightarrow \exists \gamma. (\text{topic}(\alpha) = \gamma \wedge \text{topic}(\beta) = \gamma \wedge \alpha \neq \gamma \wedge \beta \neq \gamma))$$

Asher et al (1995) erweitern dies mit der Forderung, dass der Nachzustand der in  $\alpha$  beschriebenen Eventualität  $e_\alpha$  und der Vorzustand der in  $\beta$  beschriebenen Eventualität  $e_\beta$  (zeitlich) überlappen müssen:

$$\text{Narration}(\alpha, \beta) \rightarrow O(\text{post}(e_\alpha), \text{pre}(e_\beta))$$

Explanation wird angenommen, wenn eine kausale Beziehung zwischen den Ereignissen besteht:

$$\langle \tau, \alpha, \beta \rangle \wedge \text{cause}(\text{me}(\beta), \text{me}(\alpha)) > \text{Explanation}(\alpha, \beta)$$

Explanation schliesst Narration aus:

$$\Box(\text{Explanation}(\alpha, \beta) \rightarrow \neg \text{Narration}(\alpha, \beta))$$

Zusätzlich soll das Thema des erklärenden Satzes das Erklärte sein:

$$\Box(\text{Explanation}(\alpha, \beta) \rightarrow \text{topic}(\beta) = \alpha)$$

Elaboration wird angenommen, wenn  $\beta$  eine vorbereitende Handlung von  $\alpha$  beschreibt, verhält sich ansonsten ähnlich wie Explanation.

$$\langle \tau, \alpha, \beta \rangle \wedge \text{prep}(\text{me}(\beta), \text{me}(\alpha)) > \text{Elaboration}(\alpha, \beta)$$

$$\Box(\text{Elaboration}(\alpha, \beta) \rightarrow \neg \text{Narration}(\alpha, \beta))$$

$$\Box(\text{Elaboration}(\alpha, \beta) \rightarrow \text{topic}(\beta) = \alpha)$$

Wird  $\beta$  an  $\alpha$  angebunden, und die beiden beschriebenen Eventualitäten überlappen sich, wird als Relation Background angenommen:

$$\langle \tau, \alpha, \beta \rangle \wedge \text{overlap}(\text{me}(\alpha), \text{me}(\beta)) > \text{Background}(\alpha, \beta)$$

Result wird angenommen, wenn eine kausale Verbindung der zwei Eventualitäten plausibel ist, etwa bei Temperaturänderungen und dem Verbiegen von Bi-metallstreifen. Bras et al (2001) nennen dies *permissible cause*:

$$(\langle \tau, \alpha, \beta \rangle \wedge \text{Permissible} - \text{Cause}(\alpha, \beta)) > \text{Result}(\alpha, \beta)$$

<sup>9</sup>Was das Thema(topic) einer Aussage ist, scheint nicht ausreichend formalisiert. Dies ist umso bemerkenswerter, da im weiter unten vorgestellten Ansatz von Polanyi et al ebenfalls kommentarlos eine Oberthema-UntertHEMA-Relation verwendet wird.

### 3.2.1 Lokale Diskursstruktur

Wir haben bisher als “lokale” Sicht auf den Aufbau des Diskurses die kennengelernt, dass der zu integrierende Satz an einem Knoten am rechten Rand des bisher aufgebauten Teilgraphen angebunden wird, ohne uns dazu Gedanken zu machen, wie dies aus einer rein lokalen Perspektive aussieht. Hier sollen nun zwei Ansätze vorgestellt werden, die *nur* von einer lokalen Perspektive ausgehen und ein Ansatz, der zeigt, wie diese Ansätze mit einem “globalen” Diskursmodell zusammenhängen könnten.

Die Theorie des Centering (Grosz et al., 1995) geht von der Verteilung von Pronomen aus und postuliert, dass jeder Satz durch den Kontext — hier im Wesentlichen der Satz vorher — ein *backward looking center* bekommt, das als einzige Entität als Pronomen realisiert werden kann. Die Entitäten der aktuellen Äusserung bilden dann die Menge der *forward looking center*, aus denen das nächste *backward looking center* nach bestimmten Kriterien — darunter Realisierung als Pronomen sowie die syntaktische Rolle im Satz — ausgewählt wird.

Folgendes Beispiel aus (Grosz et al., 1995) sei hier wiedergegeben:

- (3.26) a. John went to his favorite music store to buy a piano.  
b. He had frequented the store for many years.  
c. He was excited that he could finally buy a piano.  
d. He arrived just as the store was closing.
- (3.27) a. John went to his favorite music store to buy a piano.  
b. It was a store John had frequented for many years.  
c. He was excited that he could finally buy a piano.  
d. It was closing just as John arrived.

Weil Beispiel 3.26 das Center stets beibehält, während es in 3.27 hin- und her-springt, wird 3.26 als intuitiv kohärenter empfunden.

(Analysiert man das Ganze aus globaler Sicht, so ist in 3.27 Äusserung (b) unter (a) eingeordnet, während (c) an (a) eingeordnet wird und (b) verdrängt, so dass der Klavierladen in (d) nicht als Pronomen realisiert werden kann.)

Kameyama et al.(1993) wenden die Centering-Idee auf das Problem der Referenzzeit an. In einem Diskurs wie dem folgenden, von Kameyama et al. aus (Webber, 1988) zitiert

- (3.28) a. John went over to Mary’s house.  
b. On the way, he had stopped by the flower shop for some roses.  
c. Unfortunately, they failed to cheer her up.  
c’. He picked out 5 red ones, 3 white ones and 1 pale pink.

sieht man, dass nach (b) zwei Referenzzeitpunkte verfügbar sind: einerseits der nach dem Aussuchen, auf den sich (c') bezieht und andererseits der alte Referenzzeitpunkt aus (a), auf den sich (c) bezieht. Die Entscheidung, welcher Zeitpunkt für den nächsten Satz als *temporal backward center* fungiert, wird dabei durch ein *Temporal Centering Principle* und ein *Temporal Property Principle* bestimmt:

Temporal Centering Principle (TCP)

When  $\varphi_i$  evokes a set of unresolved discourse reference intervals  $\tau_1, \dots, \tau_n$ , each  $\tau_k$  is resolved against the current attentional state as follows

- *if* causal axioms generate a hypothesis, take it,
- *else if* the TPP generates a hypothesis, take it,
- *else*  $\tau_1, \dots, \tau_n$  are new to the discourse.

After resolution,  $\tau_1, \dots, \tau_n$  are in the  $TCf_i$  in the output context of  $f_i$ .

Temporal Property-sharing Principle (TPP)

Resolve the current discourse interval  $\tau_i$  against a discourse reference interval  $\tau_j$  in the attentional state. By default,  $\tau_i$  and  $\tau_j$  should share the same *property* reflecting the temporal position in the logical form.

mit "share the same *property*" ist gemeint, dass Past-Referenzen ähnlich wie vorhergegangene Past-Referenzen aufgelöst werden.

Centering/Temporal Centering bieten eine gute Erklärung, wie pronominale Referenzen und Tempus-Referenzen lokal aufgelöst werden, der Zusammenhang zu einem globalen Diskursmodell ist jedoch zunächst nicht ersichtlich. Neben dem Ansatz von (Walker, 1998), der Centering auf naheliegender Weise mit einem Cache-Modell des Diskurses verbindet und hier aus Platzgründen nicht näher vorgestellt werden soll, sei hier der Ansatz von Polanyi et al. erwähnt, die von einem hierarchischen Diskursmodell ausgehen und zeigen, wie die Relationen von Theme und Rheme aufeinanderfolgender Sätze dazu benutzt werden können, den jeweils nächsten Satz ins Diskursmodell einzuordnen.

(Polanyi et al., 2003) stellen einen Ansatz vor, wie aus den Beziehungen von Theme und Rheme der rechten unteren Kante des Diskursgraphen und des anzubindenden Knotens die Art der Anbindung inferiert werden kann.

Entscheidend für diesen Ansatz ist, daß zwischen den Objekten des Diskurses so etwas wie Ähnlichkeit ( $\approx$ ) oder Subordination ( $\downarrow$ ) gefunden werden kann.

Theme-Theme chaining

- $\theta_u \approx \theta_d, \rho_u \downarrow \rho_d$ : Subordination zu  $d$   
(John) $_{\theta}$  is a nice looking guy. (He) $_{\theta}$  has blond hair.
- $\theta_u \approx \theta_d, \rho_u \not\downarrow \rho_d$ : Subordination zu anderem Knoten, Koordination mit  $d$   
(John) $_{\theta}$  is a nice looking guy. (He) $_{\theta}$  works for a bank.

### 3.2 Diskursstruktur

---

- $\theta_u \downarrow \theta_d$ : Subordination

*(John and Bill)<sub>θ</sub> are nice looking guys.*

*(John)<sub>θ</sub> works for a bank.*

- $\theta_u \not\approx \theta_d, \theta_d \downarrow \theta', \theta_u \downarrow \theta'$ : *u* und *d* von  $\theta'$  subordiniert.

*(John and Bill)<sub>θ</sub> are nice looking guys.*

*(John)<sub>θ</sub> works for a bank. (Bill)<sub>θ</sub> plays tennis*

#### Theme-Rheme chaining

- $\theta_u \downarrow \rho_d$ : Subordination

*(John and Bill)<sub>θ</sub> work for a bank in Palo Alto.*

*(Palo Alto)<sub>θ</sub> is a town in Silicon Valley*



## Kapitel 4

# ein Tagger für kausale Relationen

Im vorherigen Kapitel wurde betrachtet, wie und unter welchen Umständen kausale Relationen in einem Text vorkommen können und welche Relationen für Erklärungen (d.h. Antworten auf Wie- oder Warum-Fragen) benutzt werden können.

Nehmen wir als erste Annäherung an einen funktionierenden Tagger einmal den Tagger, der im QA-System für Reading Comprehension von Anand et al (2000) Erklärungen (Why), Handlungen (How) und Meinungen (Opinion) findet.

Kandidaten für Erklärungen/Begründungen werden mit folgenden Mustern gefunden, wobei auf das Finden solcher Muster das (automatische) Finden von Textstellen zu solchen Mustern weiter unten in diesem Kapitel näher erklärt werden soll. Der Teil, der als Erklärung benutzt werden soll, ist hier jeweils kursiv markiert.

- *because* \*
- *because of (NP)*
- *saying (NP)* \*
- *so* \*
- *(NP)* meant, mit NP ≠ “it” oder “that”
- means that \*
- means *the* \*
- *(S)* This/That could/should/would mean ...
- *for (GERUND)*
- *(S)* That is why ...
- *(S)* So there ...
- *(Infinitiv-VP)* \*

Hierbei bedeutet (*S*) den vorhergegangenen Satz und \* bedeutet den Rest des Satzes (bis zum nächsten Satztrennzeichen).

Anand et al. teilen diese Muster in folgende Kategorien ein:

- Lexical rules: Konjunktionen, Modalverben (means)
- Syntactic rules: um-zu (im Englischen: nicht subkategorisiertes to)
- Discourse rules: nächster Satz greift vorhergehenden diskurs-anaphorisch (That would mean, That is why) auf

Riloff und Thelen (2000) benutzen einen ähnlichen Ansatz, mit *want* und *so/because* als Indizien für Erklärungen.

Während wir vorher über Erklärungsrelationen gesprochen haben, ist hier nur noch von dem Erklärenden (als dem einen Ende der Relationen) gesprochen. Wo ist das andere Ende hin?

Im System von Anand et al. wird eine Erklärung als Antwort auf eine Why-Frage durch (Wortstamm-)Überlappung des gesamten Satzes mit der Frage ermittelt; Riloff und Thelen berücksichtigen ebenfalls den vorhergehenden und darauffolgenden Satz.

## 4.1 Textmuster

Wie findet man automatisch Muster wie die obigen in Texten? Und wie kommt man auf solche Muster? Im folgenden soll veranschaulicht werden, welche Komponenten ein Tagging-System unterhalb der Ebene der Textmuster braucht, welche naheliegenden Alternativen/Erweiterungen zum Ansatz der Textmuster existieren und wie man solche Muster in Texten findet.

Als Beispiel für ein solches Tagging-System soll hier PIE (Prolog Information Extractor)<sup>1</sup> dienen, ein Rahmenwerk für Informationsextraktion, das als Komponente für einen Temporalen Tagger (Schilder et al., 2003) entstand. Das wesentlich bekanntere, in Java geschriebene Rahmenwerk GATE<sup>2</sup> (Cunningham, 2000), das als Vorbild für PIE diente, ist in Java geschrieben und hat einen wesentlich größeren Ressourcenbedarf. Ein wesentlicher Grund, den Kern der Funktionalität des frei verfügbaren GATE-Rahmenwerks zu duplizieren war, dass Prolog sich besser zur deklarativen Formulierung von Problemlösungen eignet als Java (und die Tatsache, dass die vorherige Version des Temporal-Taggers als Prolog-Programm implementiert war).

Der Eingabetext wird mit Hilfe regulärer Ausdrücke in Tokens aufgeteilt (zum Problem der Tokenisierung bzw. des Erkennens von Satzgrenzen sei auf die Übersicht in (Manning und Schütze, 1999) verwiesen). Die gefundenen Tokens werden dann als Fakten des Prädikats `token_arc` gespeichert. Informationen, die sich auf eine

---

<sup>1</sup>Nicht gemeint ist das System der Universität Manitoba bei der MUC 6, das ebenfalls PIE (Principar-based Information Extraction) genannt wurde

<sup>2</sup><http://gate.ac.uk>



## 4.1 Textmuster

---

Textspanne beziehen, werden als Annotationen im Prädikat `arc` gespeichert, bestehend aus einem Prolog-Term sowie dem Anfangspunkt und dem Endpunkt im Text.

Wenn wir etwa die Tokens *Ein*, *einfaches*, *Beispiel* haben und *Ein einfaches* als `b(x)` und *Beispiel* als `c(y)` annotiert ist, dann würde dies etwa so aussehen:

Token	Anfang	Ende	
Ein	1	2	<code>token_arc('Ein', 1, 2).</code>
einfaches	2	3	<code>token_arc('einfaches', 2, 3).</code>
Beispiel	3	4	<code>token_arc('Beispiel', 3, 4).</code>
Annotation	Anfang	Ende	
<code>b(1)</code>	1	3	<code>arc(b(x), 1, 3).</code>
<code>c(2)</code>	3	4	<code>arc(c(y), 3, 4).</code>

Im Vergleich zu einer einfacheren SGML-basierten Annotation hat dies den Vorteil dass überlappende Annotationen möglich sind (im obigen Beispiel etwa zusätzlich eine Annotation für *einfaches Beispiel*).

Weitere Annotationen können durch Regeln hinzukommen, durch Seiteneffekte von Regeln sowie Disambiguierung können Annotationen gelöscht oder verändert werden.

Regeln haben folgende Syntax:

$$\begin{aligned}
 \langle \text{Regel} \rangle &::= \langle \text{Annot} \rangle \text{---} \rightarrow \langle \text{Expr} \rangle \\
 \langle \text{Annot} \rangle &::= \langle \text{Term} \rangle \\
 \langle \text{Expr} \rangle &::= \langle \text{Annot} \rangle | @ \langle \text{Annot} \rangle | [ \langle \text{Token} \rangle , \dots ] \\
 &\quad | \{ \langle \text{Goal} \rangle \} \\
 &\quad | \langle \text{Expr} \rangle , \langle \text{Expr} \rangle | \langle \text{Expr} \rangle ; \langle \text{Expr} \rangle | ? \langle \text{Expr} \rangle \\
 &\quad | \langle \text{Expr} \rangle == \langle \text{Expr} \rangle | < < ( \langle \text{Expr} \rangle ) | > > ( \langle \text{Expr} \rangle ) \\
 &\quad | + \langle \text{Annot} \rangle | - \langle \text{Annot} \rangle | \langle \text{Annot} \rangle / \langle \text{Annot} \rangle
 \end{aligned}$$

Das “;” steht hier (wie in einer DCG-Regel in Prolog) für Sequenz, das “{;” für Alternativen.

Eine Regel `a(B,C)--->b(B),c(C)` würde im Beispiel oben eine Annotation `a(x,y)` mit Anfang 1 und Ende 4 einfügen.

Regeln können auf drei verschiedene Weisen angewandt werden:

**first\_match** bei mehreren überlappenden Vorkommen wird das zuerst gefundene benutzt (default)

**all\_matches** es werden alle gefundenen Vorkommen als neue Annotationen eingetragen

**longest\_match** von mehreren überlappenden Vorkommen wird nur das längste behalten.

Mit *@annot* wird nicht nach einer vorhandenen Annotation *annot* gesucht, sondern die Regel zu *annot* aufgerufen. Auf diese Weise sind etwa rekursive Regeln möglich.

Schliesst man einen Prolog-Aufruf  $\{aufruf\}$  in die rechte Seite der Regel ein, so wird das entsprechende Teilziel beim Aufruf der Regel an der entsprechenden Stelle aufgerufen, ähnlich wie es bei dem entsprechenden DCG-Konstrukt der Fall wäre.

In PIE können Regeln —ähnlich wie bei GATE— Seiteneffekte haben, also auch Annotationen löschen. Ein *+annot* fügt eine Annotation (zusätzlich zu der im Regelkopf) hinzu. Beispielsweise würde eine Regel

$$rule1 \rightarrow a(A) == +part1(A), b, a(A) == +part2(A).$$

dort, wo eine Sequenz  $a(1), b, a(1)$  gefunden wird, neue Annotationen  $part1(1)$  und  $part2(1)$  hinzufügen. Das  $==$  steht hierfür für die Gleichheit der Textspannen, in diesem Fall ist die Textspanne, die als Ergebnis mit dem dem Term  $part1(A)$  annotiert wird, die gleiche wie die (vorher wie nachher) mit dem Term  $a(A)$  annotierte.

Mit einer Regel wie

$$rule2 \rightarrow a(A)/a(-), b(A)/b(-)$$

würden im Falle eines Matches alle anderen  $a(\dots)$  und  $b(\dots)$ -Annotationen gelöscht.

Zusätzlich zu den Tokens werden Part-of-Speech-Tags (kurz POS-Tags) und Lemmata benutzt, wie sie der Stuttgarter TreeTagger (Schmid, 1994) ermittelt und als Terme der Form  $lemma(Lemma, POS-Tag)$  annotiert.

Mit ähnlichen Verfahren wie den für das POS-Tagging kann man einfache Nominal- und Verbalphrasen finden (Chunking) sowie Namen von Personen, Orten und Organisationen (Named Entity Recognition oder kurz NER).

Statt eines POS-Tags besteht die Ausgabe des Taggers für das Finden von Personen- Firmen- und Ortsnamen dann aus Elementen aus einer Menge wie  $O$ (outside),  $B-LOC$ (begin location),  $I-LOC$ (inside location) etc.

Peter	B-PERSON
Lustig	I-PERSON
war	O
gestern	O
in	O
Buenos	B-LOC
Aires	I-LOC
.	O

Ab Mitte/Ende der 90er Jahre entstanden Parser, deren Genauigkeit die Verwendung von Parsebäumen anstatt der flachen Strukturen, die ein Chunker liefert, für bestimmte Zwecke sinnvoll erscheinen lässt.

Khoo et al.(2000) etwa beschreiben ein System, das mit hangeschriebenen Regeln auf einem Dependenzbaum, den der Conexor FDG-Parser liefert, kausale Zusammenhänge finden soll.

Art und Güte der Ergebnisse

Ein Parser, der eine Dependenzstruktur erzeugt, ist hierbei einem Phrasenstrukturbaum vorzuziehen, da die Muster dann einfacher ausfallen.

Neben Dependenzparsern wie FDG (Tapanainen und Järvinen, 1997) und MiniPar (Lin, 1998) für das Englische und CDG (Foth et al., 2004) für das Deutsche wären dies auch TAG-Parser wie XTAG (XTAG Research Group, 2001) oder CG-basierte wie der statistische CCG-Parser (Hockenmaier, 2003) (beide für das Englische). Aus Phrasenstrukturbäumen, wie sie von stochastischen CFG-Parsern wie dem von Collins (1999) (für das Englische) oder dem von Klein und Manning (2001) (für das Englische und das Deutsche) erzeugt werden, kann ebenfalls einigermaßen zuverlässig ein Dependenzbaum erzeugt werden (Daum et al., 2004)<sup>3</sup>, basierend auf Kopfprojektion, wie auch in (Buchholz et al., 1999)<sup>4</sup>, und (Jijkoun und de Rijke, 2004)<sup>5</sup> verwandt.

Um Textmuster für eine Relation zu finden, schlägt Hearst (1992) in seinem Paper zur Erkennung von Hypernymen (Oberbegriffen) folgendes Verfahren vor:

1. Man legt sich auf eine interessierende Relation fest, etwa *Gruppe/Mitglied*
2. Man stellt eine Liste von Termpaaren zusammen, zwischen denen diese Relation besteht, etwa *Dänemark/Land*. Mit dem beschriebenen Verfahren findet man dann iterativ neue. Im günstigsten Fall hat man ein Lexikon oder eine Datenbank, die man als Ausgangspunkt verwenden kann.
3. Man sucht Stellen im Korpus, wo diese Terme syntaktisch benachbart sind und hält die Vorkommen und deren Umgebung im Text fest
4. Man findet die Gemeinsamkeiten dieser Umgebungen und nimmt an, dass häufig vorkommende Muster Indizien für die interessierende Relation sind.
5. Wenn man ein solches Muster gefunden hat (und nachgeprüft, dass es nicht zuviele Terme findet, die nicht in der gewünschten Relation stehen), kann man es benutzen, um mehr Elemente der Relation zu finden und bei 2. weiterzumachen.

Die wenigsten rein textuellen Muster sind eindeutig (d.h. finden nur die beabsichtigten Terme). Es ist daher in den meisten Fällen eine Disambiguierung notwendig, um einerseits von der größeren Abdeckung der nichteindeutigen Muster zu profitieren, andererseits aber die Einbuße bei der Genauigkeit zu verkleinern.

Girju (2003) betrachtete kausative Verben (cause, lead to, stimulate). Aus einem Trainingskorpus von 6523 Relationen, die einem Textmuster NP<sub>1</sub> verb NP<sub>2</sub> entsprachen und von denen 2101 kausal waren, wurden durch ein Lernverfahren

---

<sup>3</sup><http://nats-www.informatik.uni-hamburg.de/view/Papa/PapaDownloads>

<sup>4</sup><http://ilk.kub.nl/~sabine/homepage/software.html>

<sup>5</sup><http://staff.science.uva.nl/~jijkoun/software.html>

(Entscheidungsbäume) Constraints abgeleitet, die angeben, wann ein solches Muster Kausalität vermittelt.

Als Merkmale für den Klassifikator benutzte Girju dabei Kategorien aus WordNet, einer lexikalisierten Ontologie; bei NP<sub>1</sub> und NP<sub>2</sub> wurde jeweils angegeben, ob die entsprechende WordNet-Klasse eine Unterklasse von *entity*, *psychological-feature*, *abstraction*, *state*, *event*, *act*, *group*, *possession* oder *phenomenon* ist. Neben dem Verb *to cause*, das stets eine kausale Verbindung anzeigt, wurde als weitere gute Regel (98% Genauigkeit) gefunden, dass ein kausatives Verb mit einem Objekt, das von *phenomenon* subsumiert wird, eine kausale Relation anzeigt. Insgesamt hatte das System eine Precision von 74% und einen Recall (bezogen auf die betrachteten, durch kausative Verben indizierten kausalen Relationen) von 89%.

Zum Finden der (ambigen) Kausalverben griff Girju hier auf die WordNet-Relation *cause to* zurück (die zu selten ist, um sie direkt zu benutzen), und extrahierte mit diesen Relationen Wortmuster, die in Glossen (Erklärungen zu WordNet-Konzepten) eine kausale Verbindung anzeigen — so besteht zwischen dem Synset *kill/1* und dem von *die/1* eine *cause to*-Relation und die Beschreibung von *kill* beinhaltet *cause to die; put to death, usually intentionally or knowingly*, was *cause* als Kausalverb entlarvt.

Diese Vorgehensweise und die Tatsache, dass zwar bei den in Beziehung gesetzten Nominalphrasen, nicht aber beim Verb, die (transitive) Unterklassenbeziehung zu bestimmten Konzepten als Merkmal für den Entscheidungsbaum benutzt wurde ist ein deutlicher Hinweis darauf, dass WordNet als semantisches Lexikon für Nomen relativ gut zu gebrauchen ist (siehe (Kaplan und Schubert, 2001) zu wesentlichen Kritikpunkten), bei Verben allerdings Schwierigkeiten bereitet.

Bei EuroWordNet (Vossen et al., 1998), einem Projekt, das als Ziel hat, für verschiedene Sprachen (inklusive des Deutschen) eine WordNet-ähnliche Ontologie zu erschaffen und zumindest einen Teil (der etwa dem Umfang von WordNet 1.7 entspricht) sprachübergreifend durch einen “Interlingual Index” zu verbinden, wurde dies besser gelöst, indem eine Hierarchie verwendet wurde, die auch nicht lexikalisierte Konzepte (wie “vertikale Bewegung”) enthält.

Für das Englische gibt es die von Bonnie Dorr (1997) erstellte “LCS database”<sup>6</sup>, die eine Klassifikation der Verben aus WordNet 1.6 in einer Einteilung, die auf Levin (1993) zurückgeht sowie Subkategorisierungsrahmen und Selektionsbeschränkungen (die in WordNet 2.0 ansatzweise ebenfalls enthalten sind) enthält.

Inui (Inui, 2004) entwickelt ein Verfahren zur automatischen Klassifikation eines Paares aus Hauptsatz und Nebensatz, die durch die Konjunktion *tame* verbunden sind. Hierzu bedient er sich einer Unterscheidung zwischen Zuständen (*state of affairs*) und Handlungen (*act*)<sup>7</sup>, wobei eine bestimmte Kombination hiervon ein starkes Indiz für die jeweilige Relation ist.

In der folgenden Tabelle ((Inui, 2004), S. 56) bezeichnet HS den Hauptsatz, NS

<sup>6</sup><http://clipdemos.umiacs.umd.edu/engleslex/download.html>

<sup>7</sup>alles, was keine Handlung ist, fällt bei Inui in die zweite Kategorie, beispielsweise auch Überschwemmungen

## 4.1 Textmuster

den Nebensatz, und  $Act_m$  bzw.  $SOA_m$  bezeichnet die Eventualität des Hauptsatzes, analog  $Act_m$  bzw.  $SOA_m$  die Eventualität des Nebensatzes.

HS	NS	häufigste Relation	
SOA	SOA	<i>cause</i> ( $SOA_s$ , $SOA_m$ )	96%
Act	SOA	<i>effect</i> ( $Act_s$ , $SOA_m$ )	93%
SOA	Act	<i>precond</i> ( $SOA_s$ , $Act_m$ )	90%
Act	Act	<i>means</i> ( $Act_m$ , $Act_s$ )	85%

Da im Japanischen (wie im Deutschen oder Englischen bei Nominalisierungen) das Subjekt oft weggelassen wird, hat Inui zur Einschätzung der Volitionalität einer Eventualität einen Klassifikator erstellt, der diese anhand einer Kombination von Merkmalen ermittelt. Im Vergleich zu der Methode, bei Verben, die in den Trainingsdaten (etwa 1000 Haupt- und Nebensätze mit *tame*) vorkommen, die häufigste Kategorie und sonst Act (als insgesamt häufigste Kategorie) vorherzusagen, bringt der Klassifikator mit einer Kombination aus dem Verb selbst, den Merkmalen aus drei semantischen Lexika für das japanischen, Kasusmarkern, Tempus, Aspekt und Modalität sowie dem Vorhandensein eines Menschen oder einer Organisation als Agent, leichte Verbesserungen – von 0.864 für den Hauptsatz auf 0.888, für den Nebensatz von 0.803 auf 0.885.

Eine Kombination obiger Merkmale sowie einer regelbasierten Methode, zu bestimmen, ob bei Hauptsatz und Nebensatz der Agent übereinstimmt ergibt einen Klassifikator, der für *cause*, *precond* und *means* 80% Recall und 95% Precision erreicht, für die *effect*-Relation bei 30% Recall einen Wert von 90% Precision.

Semantische Lexika wie das von Girju verwendete WordNet oder das EDR-Konzeptlexikon, das Inui (neben zwei anderen semantischen Lexika) benutzt hat sind sehr aufwändig zu konstruieren und erreichen auch dann keine 100%ige Abdeckung. Alternativ zur manuellen Konstruktion solcher Lexika wird ebenfalls über andere Möglichkeiten nachgedacht, von den Worten des Textes zu einer Darstellung zu gelangen, die semantische Ähnlichkeiten besser berücksichtigt. Als Beispiel sei hier ein Verfahren zum statistischen Clustering erwähnt.

Lin und Pantel (2001) stellen ein Verfahren vor, bei dem Ähnlichkeiten nicht zwischen einzelnen Wörtern (hierzu siehe etwa (Lee, 1999)), sondern für Pfade über den Kanten des syntaktischen Dependenzbaumes definiert ist.

Zum Beispiel würde der Ausdruck

$$[N:subj:V] \leftarrow \text{buy} \rightarrow [V:from:N]$$

in einem Satz wie  $X$  buys something from  $Y$  die Relation zwischen  $X$  und  $Y$  ausdrücken.

Um die Ähnlichkeit zwischen zwei solchen Pfaden zu ermitteln, definieren Lin und Pantel zunächst einen Wert für ein Tripel  $(\pi, Slot, w)$ , der größer ist, wenn  $w$  besonders oft an der durch  $\pi$  und  $Slot$  vorgegebenen Stelle vorkommt (wobei  $\pi$  der Pfad ist, der uns interessiert und  $Slot$  entweder  $X$  oder  $Y$  ist, je nachdem ob das

Wort  $w$  an der einen oder anderen Position im Pfad auftaucht:

$$\text{mi}(\pi, s, w) = \log \frac{P(\pi, s, w)}{P(s)P(\pi|s)P(w|s)}$$

Für die Ähnlichkeit zweier Pfade  $\pi_1, \pi_2$  wird zunächst eine Ähnlichkeit zweier Pfade für einen bestimmten Slot definiert, und zwar als mi-Gewicht jener Worte, die mit beiden Pfaden vorkommen geteilt durch das mi-Gewicht aller Worte für  $(\pi_1, s), (\pi_2, s)$ :

$$\text{sim}\left((\pi_1, s), (\pi_2, s)\right) = \frac{\sum_{w \in T(\pi_1, s) \cap T(\pi_2, s)} \text{mi}(\pi_1, s, w) + \text{mi}(\pi_2, s, w)}{\sum_{w \in T(\pi_1, s)} \text{mi}(\pi_1, s, w) + \sum_{w \in T(\pi_2, s)} \text{mi}(\pi_2, s, w)}$$

Die Ähnlichkeit zweier Pfade wird dann als geometrisches Mittel aus dem Ähnlichkeitswert für  $s = \text{SlotX}$  und dem für  $s = \text{SlotY}$  errechnet.

Hier ein Beispiel aus (Lin und Pantel, 2001): gesucht wurden zu einem Pfad aus der Frage ( $X$  manufactures  $Y$ ;  $X$  is author of  $Y$ ) möglichst ähnliche Pfade.

$X$ manufactures $Y$	$X$ is author of $Y$
$X$ makes $Y$	$X$ co-authors $Y$
$X$ produce $Y$	$X$ is co-author of $Y$
$X$ is in $Y$ business	$X$ writes $Y$

Lin und Pantel berichten, dass bei einem Test mit ausgewählten Fragen aus der Fragensammlung der TREC-9 Question Answering-Evaluation bei den ähnlichsten 40 Pfaden zwischen 0% ( $X$  leave  $Y$ ) und 92% ( $X$  manufactures  $Y$ ) eine plausible Paraphrasierung darstellen.

#### 4.1.1 Schwach überwachte Lernverfahren

Maschinelle Lernverfahren sind zur Integration verschiedener Informationsquellen gut einsetzbar. Nicht unterschätzen sollte man allerdings den Aufwand, der für das Annotieren der entsprechenden Datenmenge benötigt wird; ein annotierter Corpus ist zwar im Allgemeinen wiederverwendbar, muss aber eine bestimmte Größe erreichen, wenn er als Trainingscorpus für ein maschinelles Lernverfahren benutzt werden soll.

Bayerl (2004) gibt in ihrem Artikel zur Disambiguierung des Diskursmarkers *wenn* eine Einschätzung von Ng wieder, dass zur Annotation eines für die Disambiguierung von Diskursmarkern genügend großen Korpus ein Aufwand von etwa 16 Mannjahren notwendig sei. Der Konnektor *wenn* scheint hier ein Extremfall zu sein, zumal (Soricut und Marcu, 2003) für die (allgemeine) Aufgabe, aus einen Parsebaum einen entsprechenden Baum aus Diskurseinheiten zu erstellen, mit dem RST-Diskurskorpus (Carlson et al., 2001) bessere Ergebnisse erzielen. Sehr wahrscheinlich ist syntaktische Information jedoch weniger hilfreich, wenn man auch

## 4.2 Unüberwachte Klassifikation von Diskursrelationen

---

an satzübergreifenden Relationen interessiert ist (ebenso im Falle von stark mehrdeutigen Markern wie *wenn*).

Einen Ausweg aus dem Problem der knappen Trainingsdaten sollen Verfahren bieten, die neben annotierten Daten auch nicht annotierte Daten verwenden (schwach überwachte Lernverfahren) und Lernverfahren, die bei der Auswahl zu annotierender Beispiele behilflich sind (aktive Lernverfahren).

Bootstrapping-Verfahren gehören zur Klasse der schwach überwachten Lernverfahren und basieren darauf, dass mit den anfänglich annotierten Daten zwei unterschiedliche Klassifikatoren trainiert werden — besonders gut funktioniert es hier, wenn die verwendeten Merkmale tatsächlich disjunkt sind, im Beispiel von Named Entity Recognition (Collins und Singer, 1999) etwa einerseits der Name selbst und andererseits die Umgebung. Unannotierte Daten werden mit beiden Klassifikatoren getaggt und dann dem jeweils anderen Klassifikator als zusätzliches Trainingsmaterial gegeben, also die Ergebnisse des ersten Klassifikators für den zweiten und umgekehrt.

Im Vergleich zu einem komplett überwachten Training oder dem Erweitern des Trainingsmaterials durch die Ausgaben desselben Taggers ergibt sich hier bei gleicher Datenmenge ein Gewinn an Genauigkeit, laut (Zhou und Goldman, 2000) funktioniert dies sogar, wenn man statt unabhängiger Merkmale zwei hinreichend verschiedene Trainingsverfahren benutzt.

Jones et al. (2003) kombinieren Bootstrapping und aktives Lernen für die Konstruktion eines NER-Klassifikators. Ausgehend von einer Initialisierung mit möglichst häufigen und eindeutigen Worten (*seed words*) werden aus der Menge der Beispiele, für die die zwei Klassifikatoren eine unterschiedliche Klasse vorhersagen, zur manuellen Korrektur ausgewählt, wobei es möglich ist, den Kontext (“operations in . . .”) oder die betreffende Nominalphrase (“Italy”) getrennt voneinander als Beispiele auszuwählen, so dass mit einer Entscheidung des menschlichen Annotators sofort mehrere Beispiele annotiert werden. Jones et al. berichten über signifikante Verbesserung im Vergleich zu alleiniger Verwendung des Bootstrapping-Verfahrens (co-EM).

## 4.2 Unüberwachte Klassifikation von Diskursrelationen

Aus dem Ansatz heraus, arbeitsintensive Annotationsarbeiten zu vermeiden, erscheinen auch Verfahren attraktiv, die ganz ohne oder mit automatisch klassifizierten Beispielen auskommen. Während die Güte eines auf solche Art gewonnenen Klassifikators im Allgemeinen nicht an die eines Klassifikators heranreicht, der mit annotierten Daten trainiert wurde, lassen sich hiermit wesentlich größere Mengen an Trainingsdaten benutzen, bei denen es ansonsten illusorisch wäre, sie auf herkömmliche Art zu annotieren.

Marcu und Echihabi (2002) entwickelten ein Verfahren, um Diskursrelationen zwischen zwei Sätzen zu finden. Für jede der Diskursrelationen (Contrast, Cause/Explanation/Evidence, Condition, Elaboration) wählten sie durch Textmuster

positive Beispiele aus, in denen die Relation durch Diskursmarker (but, because, if/then, for example) explizit markiert sind, dazu kamen noch zwei Klassen No-Relation-Same-Text (zwei Sätze, die im Text mehr als 3 Sätze voneinander entfernt sind) und No-Relation-Different-Texts (zwei Sätze aus verschiedenen Texten).

Dadurch, dass die Diskursmarker, auf Grund deren die Beispiele ausgewählt wurden, aus den Trainingsbeispielen entfernt wurden, sollte der auf diese Weise erhaltene Klassifikator genügend semantische Information mit einfangen, um auch unmarkierte Satzpaare einordnen zu können.

In einem Beispiel wie

- (4.1) a. Such standards would preclude arms sales to states like Libya, which is also currently subject to a U.N. embargo.  
 b. *But* states like Rwanda before its present crisis would still be able to legally buy arms.

sollte etwa das Wortpaar *embargo-legally* einen Hinweis auf die gegensätzliche Polarität der Sätze geben.

Aus einem Corpus mit 41 Millionen Sätzen<sup>8</sup>, den sie den “Raw corpus” nannten, erhielten sie so für jede Klasse zwischen 0,8 Millionen (Cause/Explanation/Evidence) und 3,8 Millionen (Contrast) Beispiele aus. Für ein Satzpaar  $W_1, W_2$  wurden jeweils die Wortpaare  $(w_1, w_2) \in W_1 \times W_2$  (ohne die zur Auswahl benutzten Diskursmarker) benutzt. Der damit trainierte Naive-Bayes-Klassifikator erreichte beim Testen mit zweimal 5000 Beispielen einer der 6 Klassen, die auf die gleiche Weise ausgewählt wurden, zwischen 64% (Contrast vs. No-Rel-same-Text) und 93% (Cause/Explanation/Evidence vs. Elaboration) Genauigkeit.

Der Naive-Bayes-Klassifikator basiert auf der Annahme, daß die Wahrscheinlichkeit eines Wortpaares  $(w_1, w_2)$  nur von der betrachteten Relation abhängt (und nicht auch von anderen Wortpaaren, der Satzlänge oder anderen Dingen).<sup>9</sup>

$$P(r \wedge W_1, W_2) = P(r) \cdot \prod_{w_1 \in W_1, w_2 \in W_2} P(w_1, w_2 | r)$$

Auf einem weiteren Corpus, dem “Bllip corpus”, der aus Artikeln des Wall Street Journal besteht, die von Charniak mit einem statistischen Parser geparkt sind, wurden für den Klassifikator nur Nomen, Verben und Adverbien benutzt<sup>10</sup>. Sie berichten, dass aufgrund der Größe die Ergebnisse für den Raw-Corpus insgesamt besser sind, jedoch die Ergebnisse für den Bllip-Corpus, nämlich eine paarweise Genauigkeit zwischen 58% und 82%, wesentlich besser sind als für eine gleich große Teilmenge des Raw Corpus.

<sup>8</sup>zum Vergleich: der Negra-Corpus enthält ca. 20.000 Sätze

<sup>9</sup>Diese vereinfachende Annahme gilt nur eingeschränkt, weil die Wahrscheinlichkeit für ein Paar (a,b) nicht unabhängig von der für ein Paar (a,c) ist.

<sup>10</sup>aus (Marcu und Echihiabi, 2002) wird nur ungefähr deutlich, welche Wortarten dabei benutzt wurden und welche nicht



### 4.3 Untersuchung zu Marcu und Echiabi

---

Der Ansatz des “Marker-Ratens” scheint durchaus attraktiv zu sein: Lapata und Lascarides (2004) verwendeten einen ähnlichen Ansatz, um (auf dem Bllip-Corpus) den (entfernten) zeitlichen Marker (hierbei wurden *after, as, before, once, since, when, while* sowie *until* verwendet) bzw die ursprüngliche Anordnung zu raten. Bei ersterem erreichten sie mit einer Kombination von Merkmalen 70,7% Genauigkeit für eine einzelne Entscheidung für oder gegen einen Marker (gegenüber einer Baseline von 42,6%).

Anders als Marcu und Echiabi bei ihrem Vorhaben benutzten Lapata und Lascarides neben Wörtern zusätzliche semantische (oder zumindest semantikträchtige) Information. Neben einer Klassifikation der Verben, einerseits die 25 (thematisch aufgedgliederten) *lexicographer files* von WordNet, andererseits in den Kategorien von Levin (1993) wurden Tempus, Modalität und Indizien für Aspektklassen als temporale Signatur und die Argumentenstruktur (also die Anzahl der untergeordneten Nominal-, Verbal- und Präpositionalphrasen) als syntaktische Signatur, dazu noch vorkommende Nomen sowie deren Klassifikation in *lexicographer files* und die Position des betreffenden Nebensatzes (topikalisiert oder nicht) berücksichtigt.

Hierbei wurden für die Klassenbildung bei den Verben eine Wortsinndisambiguierung wie in (Lapata und Brew, 1999) beschrieben vorgenommen.

Torisawa (2003) benutzt ebenfalls einen Ansatz mit einem unüberwachten Lernverfahren. Ausgehend von der Idee, dass Konjunktionen von Verbphrasen wie

(4.2) He drank the liquor and was intoxicated by it.

bei denen die zwei Sätze zwei gemeinsame Rollen-Füller aufweisen, ein gutes Indiz für eine kausale Beziehung sind, also dass man, sieht man regelmässig Sätze wie 4.2, davon ausgehen kann, dass

(4.3) When someone drinks liquor, he will often/normally be intoxicated

eine zulässige Generalisierung ist. Mit einer statistischen Abschätzung, um die Datenknappheit auszugleichen, erreichte das Verfahren auf ungesehenem Text eine Genauigkeit von etwa 50%.

### 4.3 Untersuchung zu Marcu und Echiabi

Der Ansatz von Marcu und Echiabi scheint vielversprechend, da man mit einem solchen Klassifikator möglicherweise genug Information über die lokale Diskursstruktur bekommt, um kausale Relationen auch ohne Diskursmarker zu finden. Da kausale Relationen entweder nicht oder nicht eindeutig (*nachdem* wird ebenfalls benutzt, um eine rein zeitliche Einordnung zu markieren, *weil* ist in Fällen allgemeiner – nicht-kausaler – Erklärungen zu finden), wäre zu hoffen, dass solche Mehrdeutigkeiten mit einem Klassifikator wie dem von Marcu und Echiabi vorgestellten einfacher aufzulösen sind.

Betrachtet man die Diskursrelationen auf einer feineren Ebene, zum Beispiel die kausale und nicht-kausale Verwendung von *weil* so kann es sein, dass bestimmte der feineren Relationen (sei es prinzipiell oder als Präferenz) nicht mit eindeutigen Diskursmarkern markiert werden, so erhält man eine systematische (und wahrscheinlich signifikante) Abweichung zwischen den Trainingsdaten und dem, was in Texten tatsächlich vorkommt. Auf der gröberen Ebene erwähnen Marcu und Echi-habi, dass Elaboration als häufigste Relation sehr selten explizit markiert ist, was beim Testen mit dem Diskurskorpus von Carlson et al zu schlechten Ergebnissen führt.

Die genannten Unterschiede auf der gröberen Ebene liessen sich möglicherweise ausgleichen, während auf anderen Unterschieden beruhende Problem nicht zu beheben wären.

Mit einem Corpus, der mit kausaler Information annotiert ist, liesse sich sicherlich ein besser geeigneter Klassifikator erstellen; die Erstellung eines Corpus mit ausreichendem Umfang wäre allerdings im Rahmen dieser Diplomarbeit zu aufwändig.

Angesichts der riesigen verwendeten Datenmenge scheint es illusorisch, durch die Hinzunahme von noch mehr Daten die Ergebnisse verbessern zu wollen, da bei Datenmengen oberhalb einiger Gigabyte zumindest Beschaffungsschwierigkeiten zu erwarten sind.

Eine Verbesserung wäre demnach nur möglich, wenn man ausser den Wörtern (bzw. Wortpaaren) weitere Merkmale mit einbezieht, oder indem man ein anderes Lernverfahren oder statistisches Modell benutzt.

Die Unabhängigkeitsannahme, von deren Erfüllung die Genauigkeit des Klassifikators abhängt, scheint kaum gegeben zu sein. Andererseits ist die Wahl des Lernverfahrens durch die Datenmenge und die Anzahl der möglichen Parameter stark eingeschränkt.

Der Gebrauch von Lernverfahren, die besser mit spärlichen Daten zurechtkommen, etwa *k*-nächste-Nachbarn-Klassifikatoren (wie sie Daelemans et al. (1996) fürs POS-Tagging benutzen) scheint bei dieser Datenmenge aufgrund der längeren Rechenzeit nicht sinnvoll.

Um so eher scheint es sinnvoll, wie in dem Fall, dass nur Nomen und Verben berücksichtigt werden, durch Reduzierung der Zahl der freien Parameter den Suchraum so zu reduzieren, dass für eine verlässliche Abschätzung genügend Daten vorhanden sind.

Insbesondere scheint es vielversprechend, nicht Worte, sondern semantische Klassen zu benutzen. Durch die geringere Parameterzahl würde einerseits die benötigte Datenmenge kleiner, andererseits würde es möglich, Lernverfahren, die sich im Hinblick auf stochastische Unabhängigkeit günstiger verhalten zu benutzen.

Ein auf dem weiter oben beschriebenen PIE basiertes System zur Extraktion der semantischen Klassen wurde für das Training nicht benutzt, weil es mit 20

Zeilen/sec. <sup>11</sup> zu langsam ist.

Um grössere Datenmengen taggen zu können, wurde ein C-Programm geschrieben, das in WordNet die Synset-Id eines einzelnen Wortes nachschlägt und anhand des WordNet-SUMO-Mappings das entsprechende Konzept der “Standard Upper Ontology” (Niles und Pease, 2001). Da SUMO nur wenige Konzepte (etwa 1100) enthält, bleibt die Anzahl der Paare überschaubar (potentiell etwa 1.200.000 gegenüber 900.000.000 möglichen Wortpaaren bei einem auf 30.000 Wörter beschränkten Wortschatz).

Das C-Programm betrachtet einzelne Wörter sowie benachparte Wortpaare (etwa “come up”, “middle school” oder “New York”) und wählt jeweils den häufigsten Wortsinn aus, wobei Worte, die zu stark mehrdeutig sind (mehr als 20 Einträge, etwa come, play) sowie Worte aus einer Stopwort-Liste ausgenommen sind.

Anhand  
Semcor/Senseval-Testdaten  
schauen, wie oft man damit  
richtig liegt

#### 4.3.1 Systemaufbau

Das Skript `filter` dient dazu, aus dem Textcorpus Beispiele für die betrachteten Relationen herauszufiltern. Von den 4 betrachteten Klassen (cause, elaboration, contrast, non-rel-same-doc) werden die Beispiele für die ersten drei Klassen erzeugt, indem in einem Satz jeweils der Teil links und rechts des gesuchten Wortes genommen wird. Aus

Camacho also testified that the television tabloid, “Hard Copy,”  
had offered to pay him for the story, *but* it was only “some peanuts.

würden dann folgende Sequenzen von Tokens werden:

key="but"			
part1			
	<i>Camacho</i>	<i>also</i>	<i>testified</i>
that	the	<i>television</i>	<i>tabloid</i>
,	"	<i>Hard</i>	<i>Copy</i>
,	had	<i>offered</i>	to
pay	him	for	his
<i>story</i>	,		
part2			
it	was	<i>only</i>	"
some	<i>peanuts</i>		

Verben, Adverben, Nomen und Eigennamen, sind kursiv hervorgehoben

Als Corpora wurden Agenturmeldungen der New York Times ( $\approx 3$  Mio Satzpaare, entsprechend dem NYT-Teil des *English Gigaword*-Corpus des Linguistic Data Consortium) sowie der auch von Marcu und Echiabi verwendete Bllip-Corpus ( $\approx 100\,000$  Satzpaare) benutzt.

---

<sup>11</sup>auf einem iBook G3 mit 600MHz, 384MB Ram

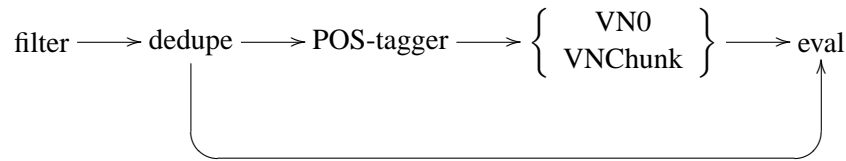


Abbildung 4.1: Schematischer Systemaufbau

Aus praktischen Gründen wurden als Merkmale nur die 30 000 häufigsten Wörter ausgewählt, was etwa 3% der Worte im Text unberücksichtigt lässt. Auch mit dieser Beschränkung würde eine einfache Kookurrenzmatrix etwa 1,8 Gigabyte groß werden. Im Vergleich zu Bigram-Sprachmodellen ist diese Kookurrenzmatrix aber dichter besetzt, so dass der (Speicher-)Overhead die Benutzung von Sprachen wie Perl oder Python hier unpraktisch macht. Für den Paar-Klassifikator wurde deshalb eine in C++ geschriebene Implementation eines Doppel-Hashes (Sedgewick, 1992) benutzt, bei dem durch Benutzung zweier unabhängiger Hashfunktionen ein gutartiges Verhalten beim Einfügen selbst in eine stark gefüllte Tabelle (und ein nicht sehr gutartiges Verhalten beim – hier nicht gebrauchten – Löschen von Einträgen) erreicht wird.

Neben dem von Marcu und Echiabi benutzten Laplace-Glättungsverfahren wurde Good-Turing-Glättung (Manning und Schütze, 1999) sowie ein absolutes Glättungsverfahren verwendet.

Die Grundidee eines absoluten Glättungsverfahrens ist, den von 0 abweichenden Häufigkeiten etwas (nennen wir es  $q$ ) abzuziehen, um diese Wahrscheinlichkeitsmasse auf die nicht gesehenen Ereignisse (hier: Wortpaare) umzuverteilen. In diesem speziellen Fall sollen Wortpaare, die noch mit keiner der betrachteten Relationen vorgekommen sind, überhaupt nicht berücksichtigt werden; die umzuverteilende Wahrscheinlichkeitsmasse soll also der Wahrscheinlichkeit entsprechen, ein Paar zu finden, das noch nicht mit der betrachteten Relation, aber mit irgendeiner Relation vorgekommen ist.

Bezeichne  $N$  die Anzahl unterschiedlicher Paare (Typ) und  $T$  die Gesamtanzahl gesehener Paare (Token) sowie  $N_r$  und  $T_r$  die entsprechenden Zahlen für die betrachtete Relation  $r$ . Bei einem Wortpaar für die Relation  $r$  ist die Wahrscheinlichkeit für ein in  $r$  ungesehenes Paar gleich  $\frac{N_r}{T_r}$  (in  $T_r$  Ereignissen ist  $N_r$  mal etwas Neues aufgetaucht). Die Wahrscheinlichkeit, dass ein Wortpaar für  $r$  noch für keine Relation gesehen wurde, ist

$$\frac{N}{T_r} \cdot \frac{T_r}{T} = \frac{N}{T}$$

In  $T_r$  Ereignissen ist  $N$  mal etwas Neues aufgetaucht, davon zählen wir einen Anteil von  $\frac{T_r}{T}$  als zu  $r$  gehörig.

Die Wahrscheinlichkeit für ein insgesamt gesehenes Paar, das für die Relation  $r$  nicht gesehen wurde, wäre dann, da alle für  $r$  gesehenen Paare auch in irgendeiner

### 4.3 Untersuchung zu Marcu und Echiabi

---

Relation gesehen wurden:

$$1 - \frac{1 - \frac{N_r}{T_r}}{1 - \frac{N}{T}} = \frac{\frac{N_r}{T_r} - \frac{N}{T}}{1 - \frac{N}{T}}$$

Bei Schiefverteilung sind Good-Turing und absolute Glättung besser geeignet als Laplace-Glättung, bei etwa gleich häufigen Relationen (wie cause und elaboration) macht die Auswahl des Glättungsverfahrens keinen Unterschied.

Zunächst wurde geschaut, ob Paare der SUMO-Konzepte, die den Wörtern des Satzes entsprechen, als Merkmale für den Klassifikator besser geeignet sind als die Wortpaare selbst. Durch die grobe Auflösung der SUMO-Klassen scheint allerdings zuviel Information verloren zu gehen.

quantifizieren

Als weitere Alternative wurde geschaut, welche Auswirkung es hat, wenn man Gruppen von Wörtern, die man für die Wortsinndisambiguierung als eine Einheit betrachtet, zusammenfasst, namentlich Nominalgruppen (wie middle school, business executive) sowie Verben und ihre Partikel (come up, play down). Zusätzlich wurden Gruppen von Adverbien zusammengefasst und Indikatoren für den Aspekt eines Verbs (Progressiv, Perfektiv, Passiv) als getrennte Merkmale verwandt. Personennamen (also Gruppen von PN-Tokens, die einen Vornamen oder "Mr./Mrs." enthalten) wurden auf ein einheitliches Merkmal ( $\langle \text{Person} \rangle$ ) abgebildet, das Gleiche gilt für Geldbeträge ( $1000 \$ \Rightarrow \langle \text{money amount} \rangle$ ).

Adjektive wurden entweder als Teil einer Nominalgruppe betrachtet (federal government), oder als eigenständiges Merkmal, wenn sie nach einer Form von *to be* stehen. Bei Verben wird ein eventuell vorhandenes *not* mit zum Verb gruppiert.

Nomen, die ohne Artikel nach einem Verb (take place) oder einer Präposition (behind bars) standen, wurden mit dieser zusammen gruppiert.

Muster:

- NP+ (of NP+)\*
- (this|last|next|previous) (year|quarter|month|week)
- JJ? NN+
- RB+
- [is] JJ
- not? VV RP?

Alternativ wurde eine Variante betrachtet, bei der Adverbien, Verben und Nomen als Einzelwörter verwendet wurden (VNO). Beim Bllip-Corpus wurden zunächst Duplikate der Satzpaare entfernt. Dies war nötig, weil Duplikate einen großen Anteil (30%?) der Sätze ausmachten.

Die Ergebnisse auf dem Bllip-Corpus lassen (ähnlich wie die von Marcu und Echiabi) darauf hoffen, dass sich das Ergebnis mit mehr Daten verbessern lässt, oder mit anderen Worten: das Modell leidet auch bei der verwendeten Datenmenge

an Datenknappheit. Für den Klassifikator bietet es sich an, durch Zusammenfassen ähnlicher Ereignisse oder Verfahren, die ähnliches bewerkstelligen die Lernbarkeit positiv zu beeinflussen. Im Bereich des Information Retrieval (wo es um das Suchen von Dokumenten zu einer Anfrage gibt und die interessierende Relation eine zwischen Dokumenten und Termen ist) gibt es ein Verfahren namens *Latent Semantic Analysis*, das darauf basiert, die Kookurrenzmatrix durch eine mit niedrigerem Rang zu ersetzen, wobei das Ergebnis möglichst wenig verzerrt werden soll.

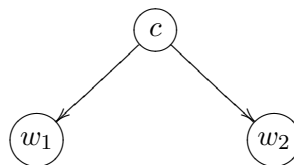
Man schreibt die Kookurrenzmatrix  $M$  als

$$M = U^T S V$$

wobei  $S$  nur in der Diagonale besetzt ist. Setzt man die Werte ins  $S$ , angefangen mit dem betragsmäßig kleinsten, auf 0, so hat die gesamte Matrix einen niedrigeren Rang (und man kann die entsprechenden Spalten von  $U$  bzw.  $V$  weglassen). Man könnte nun für jede Relation die Kookurrenzmatrizen für die Paare durch erwähnte Dimensionalitätsreduktion glätten, das Ergebnis ist aber für den angestrebten Zweck ungeeignet, denn minimiert wird die lineare Abweichung, so dass etwa negative Zahlen herauskommen können.

Hofmann (1999) schlägt für solche Probleme ein statistisches Clusteringverfahren vor, bei dem eine Mixtur von Wortverteilungen modelliert wird, deren Parameter mittels einer Variante des *Expectation Maximization*-Algorithmus ((Dempster et al., 1977), siehe auch (Manning und Schütze, 1999)) iterativ geschätzt werden.

Im entsprechenden statistische Modell (zu weiteren Alternativen für statistische Modelle, um paarweise auftretende Ereignisse zu beschreiben, siehe (Hofmann und Puzicha, 1998)) wird zuerst aus einer Verteilung  $P(c)$  ein Cluster herausgesucht, um dann die Worte des ersten Teils mit der abhängigen Verteilung  $P(w_1|c)$  bzw. die des zweiten Teils mit  $P(w_2|c)$  einzuschätzen.

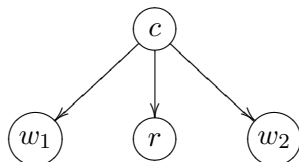


$c$  ist dabei eine “versteckte” Variable, die (insbesondere im Training) nicht beobachtet werden kann, weswegen der EM(Expectation Maximization)-Algorithmus gebraucht wird. Zunächst initialisiert man die Verteilungen  $P(c)$ ,  $P(w_1|c)$ ,  $P(w_2|c)$  mit zufälligen Werten. Mit diesen Verteilungen bekommt man für jedes Datum aus den Trainingsdaten eine Verteilung für  $c$ , die man dann benutzt, um die Verteilungen  $P(c)$ ,  $P(w_1|c)$ ,  $P(w_2|c)$  anhand der Trainingsdaten und der bisherigen Schätzung für das  $c$  jedes Trainingsdatums neu zu schätzen.

### 4.3 Untersuchung zu Marcu und Echiabi

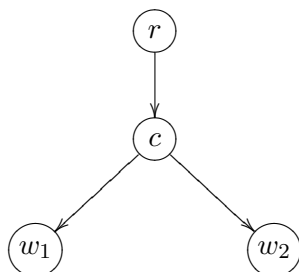
---

In einem ersten Ansatz, dieses Modell zur Klassifikation einzusetzen, wurde die Relation als weitere Zufallsvariable von  $c$  abhängig gemacht:



Eine Implementierung dieses Modells und ein Test mit dem NYT-Corpus ergab, dass die Cluster von den häufigsten Relationen (contrast, samedoc) dominiert wurden, was zu schlechten Ergebnissen führte.

In einem zweiten Modell wurden die Cluster jeweils einer Relation zugeordnet, was in etwa folgendes Modell ergibt:



Auf diese Art können tatsächlich bessere Ergebnisse als mit Paaren erreicht werden (für alle 4 Relationen auf dem NYT-Corpus 50% mit dem Clustering-Modell gegenüber 42% für Paare).

Eine Variation der Anzahl der verwendeten Cluster brachte das (eher unerwartete) Ergebnis, dass die Wahl der Clusteranzahl das Ergebnis nur sehr wenig beeinflusst (siehe Abbildung 4.2). Im Extremfall ( $n = 2$ ) degeneriert das Modell zu einem einfachen Naive-Bayes-Modell mit einer Verteilung für die Worte jeden Teils — Paare spielen hier überhaupt keine Rolle mehr.

Will man eine Verteilung für Paare in das Modell integrieren, so steht dem zunächst wieder die hierfür nicht gültige Unabhängigkeitsannahme im Weg, denn für ein Wort  $a$  aus dem ersten Teil und zwei Wörter  $b_1, b_2$  aus dem zweiten Teil hängen  $P(a, b_1)$  und  $P(a, b_2)$  beispielsweise voneinander ab.

Diese müsste uns bei einem Maximum-Entropy-Modell (im Folgenden MaxEnt-Modell) nicht stören, da die Unabhängigkeitsannahme hierfür nicht gebraucht wird. MaxEnt-Modelle gehören zur Klasse der Loglinearen Modelle, d.h. für ein Ereignis  $x$  wird die Wahrscheinlichkeit als

$$P(x) = \frac{e^{\sum_i f_i(x)\lambda_i}}{Z}$$

berechnet<sup>12</sup>, wobei  $f_i(x)$  bei Vorhandensein eines Merkmals (in unserem Fall könnte das ein Wort oder ein Wortpaar sein) 1 ist und sonst 0;  $Z$  ist eine Normierungskonstante, die so gewählt ist, daß  $\sum_x P(x) = 1$  bleibt.  $\lambda_i$  sind die Parameter des

<sup>12</sup>Aus Gründen der Einfachheit wird hier der nicht der Fall einer bedingten Verteilung, sondern der einfachere Fall einer einzigen Verteilung betrachtet.

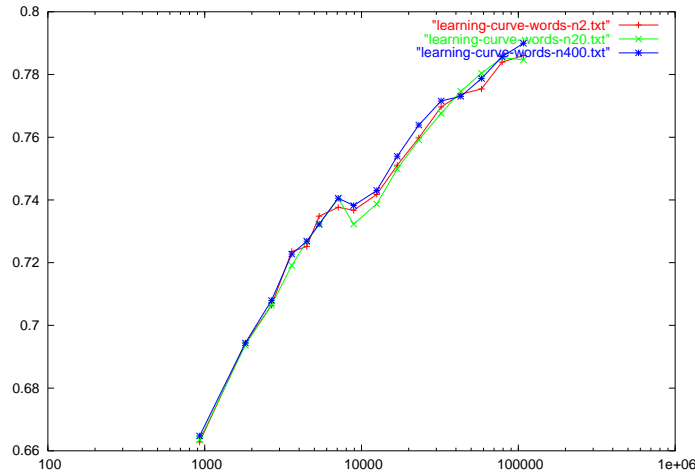


Abbildung 4.2: cause vs. elaboration: Einfluss unterschiedlicher Clusteranzahlen, waagerechte Achse entspricht Menge der Trainingsdaten

Modells und müssen im Training bestimmt werden, in einem Naive-Bayes-Modell würde hier einfach  $\lambda_i := \log P(f_i = 1)$  gesetzt. Bei einem MaxEnt-Modell wird aus der Klasse der Loglinearen Modelle dasjenige ausgewählt, das einerseits die Constraints

$$P(f_i = 1) = \tilde{P}(f_i = 1) \quad \text{für } i = 1, \dots$$

erfüllt ( $\tilde{P}(f_i = 1)$  ist hierbei die relative Häufigkeit von  $f_i$  in den Trainingsdaten) und andererseits die größte Entropie besitzt (was damit äquivalent ist, dass  $P$  den Trainingsdaten die maximale Wahrscheinlichkeit zuweist).

Die Berechnung der  $\lambda_i$  wird üblicherweise mit einem numerischen Approximationsverfahren durchgeführt, das iterativ das (eindeutige) globale Maximum findet, dabei aber relativ langsam konvergiert. (Charniak, 2000) beschäftigt sich mit diesem Problem im Kontext der Parameterschätzung eines generativen Parsing-Modells und argumentiert, dass in vielen Fällen eine einfache erste Annäherung einen großen Teil des Nutzens bringt, den ein MaxEnt-Modell brächte.

In unserem Fall würden wir, hätten wir ein Loglineares Modell, für jedes Wort aus dem ersten Teil, für jedes Wort aus dem zweiten Teil sowie für die Paare einen Parameter haben, wir nennen sie für unser Beispiel einmal  $\lambda_{w_1}$ ,  $\lambda_{w_2}$ ,  $\lambda_{(w_1, w_2)}$ .

Gewünscht wäre, dass im Fall eines aus  $w_1$  bestehenden ersten Teils und eines aus  $w_2$  bestehenden zweiten Teils folgendes gilt:

$$\lambda_{w_1} + \lambda_{w_2} + \lambda_{(w_1, w_2)} = \log P(w_1, w_2)$$



### 4.3 Untersuchung zu Marcu und Echiabi

---

Wenn wir  $\lambda_{w_1} := \log P(w_1)$  und  $\lambda_{w_2} := \log P(w_2)$  setzen, so erhalten wir schließlich

$$\lambda_{(w_1, w_2)} = \log P(w_1, w_2) - \log P(w_1) - \log P(w_2)$$

was im obigen Fall von  $a$  im linken Teil und  $b_1, b_2$  im rechten Teil zu folgendem (gültigen) Ergebnis führen würde:

$$\begin{aligned} \log P(a, b_1, b_2) &= \lambda_a + \lambda_{b_1} + \lambda_{b_2} + \lambda_{a, b_1} + \lambda_{a, b_2} \\ &= \log P(a, b_1) + \log P(a, b_2) - \log P(a) \\ P(a, b_1, b_2) &= P(a) \cdot P(b_1|a) \cdot P(b_2|a) \end{aligned}$$

Mit dieser Methode kann man ein gemeinsames Modell für Einzelworte und Paare (seien es alle Paare oder nur die  $n$  häufigsten) mit oder ohne Clustering erstellen. Auf dem Bllip-Corpus bringt die Hinzunahme der  $n$  häufigsten Paare (für  $n = 30\,000$ ,  $n = 200\,000$ ) als Merkmale keine Verbesserung (siehe Abbildung 4.3). Nimmt man statt der Wortpaare Kategorienpaare (also Paare von SUMO-Konzepten) hinzu, so ist die Verschlechterung weniger ausgeprägt. Integriert man beides in einem Modell, wird das Ergebnis jedoch nicht besser (Abb. 4.4).

#### 4.3.2 Analyse

Was für Information ist es, die der Klassifikator tatsächlich benutzt? Diese Frage ist entscheidend dafür, wie gut es funktionieren würde, ein einmal trainiertes Modell für andere Textsorten zu verwenden, oder (was hier noch wichtiger ist) ob das Modell auch bei unmarkierten Sätzen gut funktionieren würde.

Soria und Ferrari(1998) untersuchten das Problem der Markierung von Diskursrelationen mit zwei Experimenten: im ersten Experiment wurde den Versuchspersonen eine Bilder Geschichte gezeigt, die sie gesprochen oder geschrieben nach erzählen sollten, im zweiten Experiment wurden den Versuchspersonen Sätze präsentiert, bei denen sie die entsprechende (Diskurs-)Relation raten sollten. Hierfür wurde eine Einteilung in additiv (entspricht in etwa Marcus *elaboration*), konsequentuell (entspricht in etwa *cause/explanation/evidence*) und kontrastiv (entspricht *concession*) vorgenommen.

Die kontrastive Relation war im ersten Experiment fast immer (in 90% der Fälle) markiert. Dementsprechend schlechter ist die Erfolgsquote im zweiten Experiment, wenn der Konnektor fehlt.

	additiv	konsequentuell	kontrastiv
mit Konnektor	73%	89%	83%
ohne Konnektor	64%	60%	43%

Die folgenden Tabellen zeigen die Merkmale mit der größten Präzision für den Bllip-Corpus. Hierzu ist anzumerken, dass für den weiter oben betrachteten Klassifikator die Merkmale im Wesentlichen nach Häufigkeit ausgewählt wurden<sup>13</sup>

---

<sup>13</sup>genauer gesagt, dem Produkt der Kullback-Leibler-Divergenz zur Gesamtverteilung und der

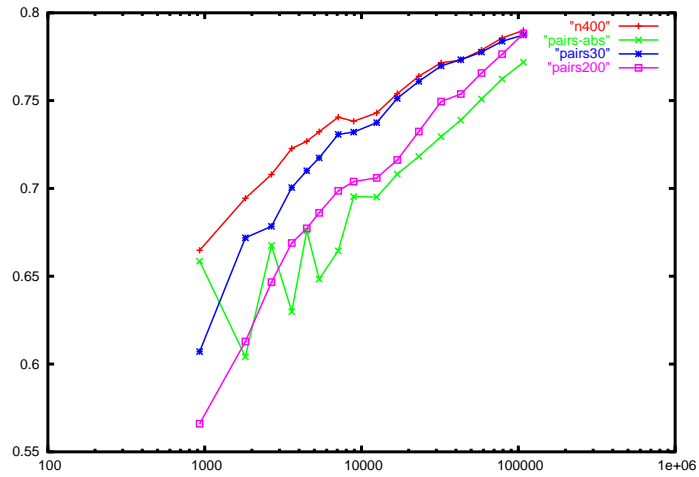


Abbildung 4.3: cause vs. elaboration: Hinzunahme von Paaren

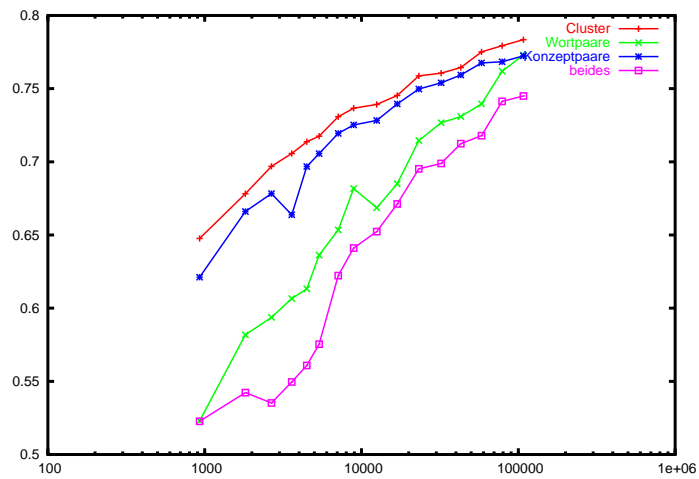


Abbildung 4.4: cause vs. elaboration: Hinzunahme von Paaren und Paaren von Kategorien

### 4.3 Untersuchung zu Marcu und Echiabi

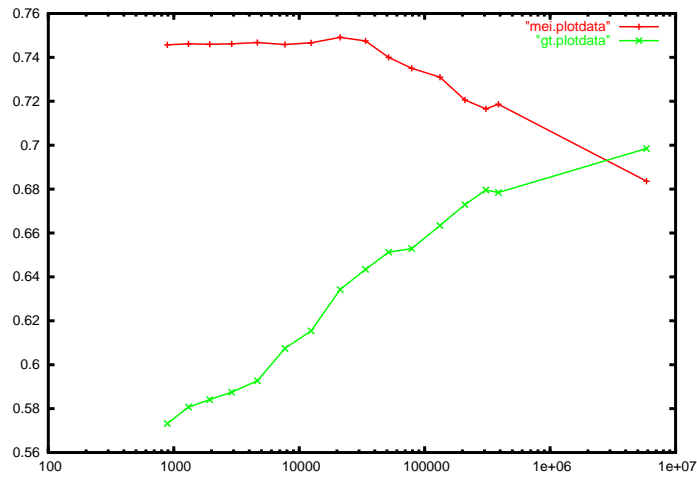


Abbildung 4.5: Naive-Bayes und MaxEnt-inspiriert; waagerechte Achse entspricht Anzahl der berücksichtigten Paare

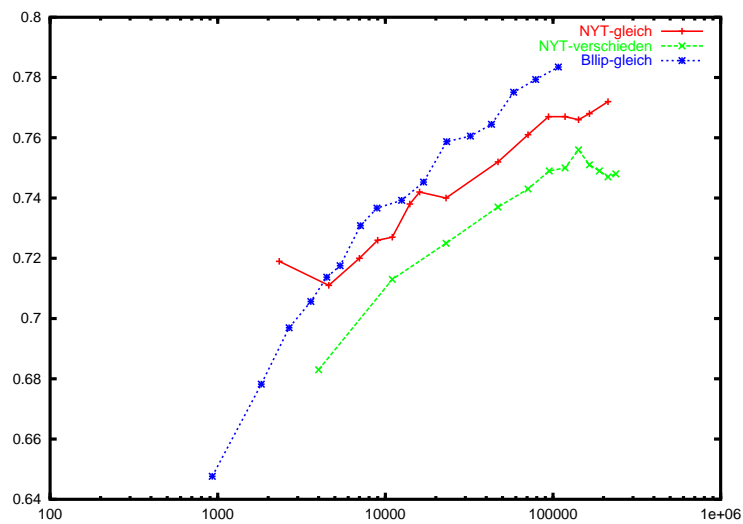


Abbildung 4.6: Genauigkeit des Verfahrens: verschiedene Corpora, Testset aus gleicher Region oder vom Ende

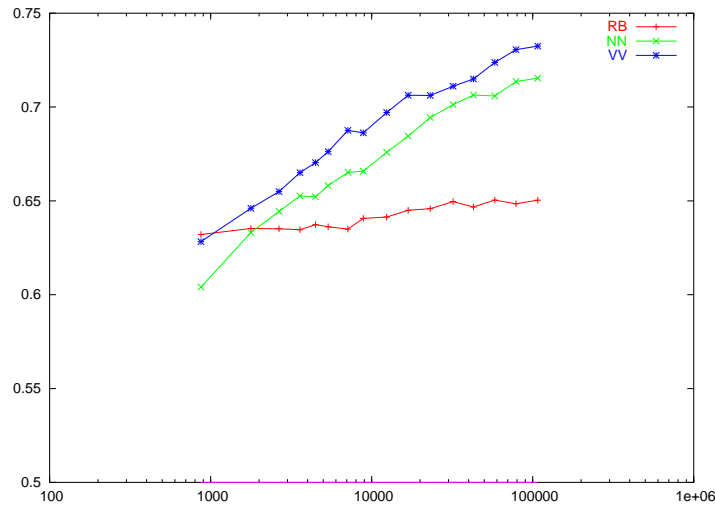


Abbildung 4.7: Genauigkeit allein anhand von Verben (VV), Nomen (NN) und Adverbien (RB)

und nicht (wie es bei Regelinduktion möglicherweise der Fall gewesen wäre) die Präzision.

Bei der *cause*-Relation fällt der Einfluss von Ausdrücken wie “partly/largely/primarily because” auf, der von dieser Fragestellung her uninteressant ist<sup>14</sup>, aber auch, dass Dinge wie *strong yen*, *bad weather* oder *strong demand* typische kausale Erklärungen sind (die andererseits das, was damit erklärt werden nicht immer oder unbedingt nach sich ziehen). Bei den entsprechenden Daten für den NYT-Corpus kommen zusätzlich noch Urteils-Adjektive wie *meaningless*, *noteworthy*, *misleading unnecessary* dazu, die relativ oft (zwischen 74 und 77% der Fälle) vor einem entsprechenden Marker stehen.

Häufigkeit; informationstheoretisch gesprochen entspricht dies der Anzahl der Bits, die bei einer Entropiekodierung der Klassen eingespart worden wären, wenn man das entsprechende Merkmal als Hinweis auf die Klasse nimmt.

<sup>14</sup>möglicherweise ist *partly* ein gutes Hilfsmittel, um weitere kausale Marker zu finden

#### 4.3 Untersuchung zu Marcu und Echiabi

---

concession		
sound/(person)	100.00%	91
say/not say	99.40%	166
point/point	100.00%	81
earlier/down	100.00%	75
say/not elaborate	99.32%	147
say/economist	100.00%	68
adjust/say	100.00%	65
not elaborate/say	100.00%	64
comment/spokesman	100.00%	50
acknowledge/say	98.35%	243
decline/note	100.00%	48
concede/say	98.41%	189
(person)/administration official	100.00%	47
company/not elaborate	100.00%	47
iran-contra/(person)	100.00%	46
perf/not say	100.00%	46
say/further comment	100.00%	46
up/down	98.52%	135
u.s./so far	100.00%	45
year/unchanged	100.00%	44
not disclose/%	100.00%	43
(money amount)/warn	100.00%	43
rise/down	100.00%	43
(person)/privately	100.00%	43

Tabelle 4.1: Paare mit der größten Präzision (Blip ohne same-doc): concession

	cause	
just/not mean	100.00%	59
partly/partly	98.00%	50
edition/error	100.00%	24
partly/market	96.30%	27
partly/use	96.15%	26
partly/cost	95.45%	22
partly/concern	95.45%	22
partly/country	95.24%	21
largely/cost	95.24%	21
largely/loss	95.00%	20
sell/not fit	92.00%	25
partly/company	87.01%	77
partly/sell	90.48%	21
company/uncertainty	87.88%	33
partly/last year	87.50%	24
partly/(city)	83.33%	72
mainly/increase	86.36%	22
primarily/sale	86.36%	22
partly/*prog	82.28%	79
largely/fall	85.71%	21
significant/(person)	84.62%	26
partly/increase	83.87%	31
rating/company	85.00%	20
partly/year	82.22%	45
say/high cost	85.00%	20

Tabelle 4.2: Paare mit der größten Präzision (Bllip ohne same-doc): cause

#### 4.3 Untersuchung zu Marcu und Echiabi

---

elaboration		
current account/trade	100.00%	53
current account/good	100.00%	48
divide/dated	100.00%	43
evenly/dated	100.00%	42
current account/unilateral transfer	100.00%	41
divide/mature	100.00%	41
offering/dated	100.00%	40
evenly/mature	100.00%	40
offering/mature	100.00%	38
tender offer/expire	100.00%	37
fiscal year/oct.	100.00%	34
current account/measure	100.00%	26
nikkei/tuesday	100.00%	26
current account/service	97.96%	49
durable good/year	100.00%	24
filing/% stake	100.00%	23
co./publish	100.00%	22
securities/% stake	100.00%	22
exchange commission/% stake	100.00%	22
federal home loan bank board/regulate	100.00%	21
current account/comprise	100.00%	20
offer/expire	94.90%	157
select/point	95.92%	49
strong yen/make	96.88%	32
accord/track	96.88%	32

Tabelle 4.3: Paare mit der größten Präzision (Bllip ohne same-doc): elaboration

concession		
2/not rule out	100.00%	78
2/further comment	100.00%	73
2/sceptic	100.00%	51
2/premature	100.00%	43
2/not elaborate	98.10%	210
2/meanwhile	100.00%	39
2/unissued	100.00%	34
2/thrill	100.00%	33
2/avail	100.00%	30
2/not as much	100.00%	29
2/unclear	97.35%	113
2/such effort	100.00%	28
2/current plan	100.00%	27
2/Bureau of Statistics	100.00%	27
2/Statistics Sweden	100.00%	27
2/overrule	100.00%	26
1/admittedly	100.00%	25
2/cautiously	100.00%	25
2/caution	96.28%	349
2/wonder	96.38%	221
2/even then	100.00%	24
2/scoff	100.00%	24
1/maybe so	100.00%	24
2/nonetheless	96.88%	96

Tabelle 4.4: Einzelwörter mit der größten Präzision (Blip ohne same-doc): concession



### 4.3 Untersuchung zu Marcu und Echiabi

---

	cause	
1/misstate	91.67%	36
1/partly	83.88%	1452
1/negative implication	86.36%	22
2/slow sale	85.19%	27
1/largely	77.73%	1136
1/significant	74.42%	215
2/strong yen	74.07%	81
1/possible downgrade	76.00%	25
2/death threat	76.19%	21
1/qualified opinion	76.19%	21
1/CreditWatch	73.68%	38
1/precisely	70.23%	131
2/bad weather	73.08%	26
2/strong demand	70.59%	68
1/jeopardy	70.00%	20

Tabelle 4.5: Einzelwörter mit der größten Präzision (Bllip ohne same-doc): cause

elaboration		
2/make computer	100.00%	24
1/26-week bill	95.74%	47
1/Proceed	95.00%	40
2/-owned	91.71%	217
2/unilateral transfer	93.33%	45
1/first section index	95.45%	22
2/class-action status	95.45%	22
2/dated	90.91%	55
1/first section	88.89%	27
2/overnight loan	88.00%	25
1/evenly	83.33%	54
2/comprise	80.90%	89
2/Bankruptcy Code	78.89%	90
2/rename	78.57%	56
1/Atlantis	79.41%	34
2/from creditors	77.63%	76
2/between banks	78.26%	23
2/a.m.	75.51%	49
2/spin off	73.24%	71
2/expire	71.97%	578
2/bankruptcy-law protection	72.88%	59
2/found	71.56%	109
2/not relate	71.43%	84
2/own	70.07%	2185

Tabelle 4.6: Einzelwörter mit der größten Präzision (Bllip ohne same-doc): elaboration

## Kapitel 5

# Zusammenfassung und Ausblick

Anhand bestehender philosophischer und kognitionswissenschaftlicher Theorien wurde gezeigt, wie der Begriff der Kausalität mit naiven/qualitativen Theorien zur Vorhersage von Ereignissen aus einem teilweise bekannten Kontext zusammenhängt. Dieser Ansatz ist insofern reduktionistisch, als er der Kausalität die Existenzberechtigung als Naturphänomen abspricht und ihr “nur” den Status eines wichtigen Artefakt menschlichen Denkens gibt.

Vor dem Hintergrund diesen Ansatzes wurde diskutiert, in welcher Hinsicht ein Annotationsschema für kausale Relationen in Texten anders aufgebaut sein muss als ein Schema für zeitliche Relationen wie es etwa TimeML ist, und der Zusammenhang zwischen zeitlichen und kausalen Relationen zwischen Ereignissen betrachtet.

Als Vorarbeit für einen Tagger, der solche kausalen Relationen automatisch annotiert, wurde betrachtet, ob und wie Kausalität explizit im Text markiert wird. Da explizite Markierung kausaler Relationen zu selten ist, als daß man sich allein darauf verlassen könnte, wurde die Verwendbarkeit von Markern wie danach und nachdem, die primär eine zeitliche Funktion haben, betrachtet sowie untersucht, inwiefern kausale Verbindungen, die im Text nicht explizit gemacht werden, entdeckt werden können.

Es wurde untersucht, ob der Ansatz von (Marcu und Echihabi, 2002) der Paare von semantisch verwandten Worten als Indizien für eine bestimmte Diskursrelation benutzt, durch Hinzunahme weiterer Merkmale verbessert werden kann und zum anderen, ob das vorgestellte Verfahren, sei es insgesamt oder als Hilfe zum Finden solcher Paare semantisch verwandter Worte, für einen Tagger nutzbar gemacht werden könnte. Die hier gemachten Experimente weisen darauf hin, dass bei Corpora in der Größenordnung des Bllip-Corpus Einzelworte als Indizien für Diskursrelationen besser sind als Wortpaare, und dass zumindest die Indizien, die mit hoher Sicherheit für die Anwesenheit eines kausalen Markers (*because, thus, ...*) sprechen, in unmarkierten Sätzen nicht aussagekräftig sind.

Für einen funktionierenden Tagger scheint es am vielversprechendsten, ähnlich wie in (Mani et al., 2003) verschiedene lokale Indizien auszuwerten, um ein je-

weiliges Referenz-Ereignis und die Beziehung zu diesem Ereignis zu ermitteln. Dies könnte man durch eine (aus Qualitätsgründen) von Menschen konstruierte, möglichst einfach gehaltene Ontologie gängiger Ereignistypen komplementieren, deren lexikalische Basis durch Clusteringtechniken erweitert werden kann, um so durch Wissen über Ereignisarten und deren mögliche kausale Beziehungen eine größere Verlässlichkeit auch in Abwesenheit expliziter Markierung zu erreichen.

# Literaturverzeichnis

- J. Allen (1984). Towards a General Theory of Action and Time. *Artificial Intelligence*, 23.
- P. Amsili und C. Rossari (1998). Tense and Connective Constraints on the Expression of Causality. In *Coling'1998*.
- P. Anand, E. Breck, B. Brown, M. Light, G. Mann, E. Riloff, M. Rooth, und M. Thelen (2000). Fun with Reading Comprehension. Technical report, MITRE.
- N. Asher, M. Aurnague, M. Bras, P. Sablayrolles, und L. Vieu (1995). De l'espace-temps dans l'analyse du discours. *Sémiotiques*, 9:11–62.
- N. Asher, P. Denis, J. Kuhn, E. Larson, E. McCready, A. Palmer, B. Reese, und L. Wang (2004). Extracting and Using Discourse Structure to Resolve Anaphoric Dependencies: Combining Logico-Semantic and Statistical Approaches. In *Workshop SDRT, Traitement Automatique des Langues Naturelles 2004*.
- N. Asher und A. Lascarides (1996). Bridging. In R. B. R. van der Sandt und M. Bierwisch, Hrsg., *From Underspecification to Interpretation*, Working Papers of the Institute for Logic and Linguistics.
- P. S. Bayerl (2004). Disambiguierung deutschsprachiger Diskursmarker: Eine Pilot-Studie. *Linguistik Online*, 18(1).
- M. Bras, A. Le Draoulec, und L. Vieu (2001). Temporal Information and Discourse Relations in Narratives: the role of French connectives *puis* and *un peu plus tard*. In *ACL'2001 Workshop on Temporal and Spatial Information Processing*, S. 49–56.
- S. Buchholz, J. Veenstra, und W. Daelemans (1999). Cascaded Grammatical Relation Assignment. In *EMNLP/VLC-99*.
- L. Carlson, D. Marcu, und M. E. Okurowski (2001). Building a discourse-tagged Corpus in the Framework of Rhetorical Structure Theory. In *2nd SIGDIAL Workshop on Discourse and Dialogue, Eurospeech 2001*.
- E. Charniak (2000). A Maximum-Entropy-Inspired Parser. In *HLT/NAACL 2000*.

- E. Charniak und D. McDermott (1985). *Introduction to Artificial Intelligence*, Kapitel 10. Language Comprehension. Addison Wesley.
- P. Clark (2001). Reference Resolution and Views. Working Note 25.
- P. Clark, J. Thompson, K. Barker, J. Fan, B. Porter, D. Tecuci, und P. Yeh (2001). More Thoughts on Views. Working Note 22.
- M. Collins (1999). *Head-Driven Statistical Models for Natural Language Parsing*. PhD thesis, University of Pennsylvania.
- M. Collins und Y. Singer (1999). Unsupervised Models for Named Entity Classification. In *EMNLP/VLC-99*.
- D. Crouch, D. Bobrow, C. Condoravdi, J. Everett, V. Paiva, R. Stolle, und M. van den Berg (2001). Preventing Existence. In *ACM Conference on Formal Ontology in Information Systems*.
- H. Cunningham (2000). *Software Architecture for Language Engineering*. PhD thesis, University of Sheffield. <http://gate.ac.uk/sale/thesis>.
- W. Daelemans, J. Zavrel, P. Berck, und S. Gillis (1996). MBT: A memory-based part of speech tagger generator. In *EMNLP/VLC-96*.
- M. Daum, K. Foth, und W. Menzel (2004). Automatic transformation of phrase treebanks to dependency trees. In *LREC 2004*.
- A. Dempster, N. Laird, und D. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Statistical Society Series B*, 39:1–38.
- B. J. Dorr (1997). Large-Scale Dictionary Construction for Foreign Language Tutoring and Interlingual Machine Translation. *Machine Translation*, 12(4):271–322.
- L. Ferro, I. Mani, B. Sundheim, und G. Wilson (2000). TIDES temporal annotation guidelines. Technical report, The MITRE Corporation.
- E. Filatova und E. H. Hovy (2001). Assigning Time-Stamps to Event-Clauses. In *Workshop on Temporal and Spatial Information Processing at ACL-2001*, Toulouse.
- K. Foth, W. Menzel, und I. Schröder (2004). Robust Parsing with Weighted Constraints. *Natural Language Engineering*.
- R. Girju (2003). Automatic Detection of Causal Relations for Question Answering. In *HLT-NAACL 2003 Workshop on Multilingual Summarization and Question Answering*.
- R. Girju und D. Moldovan (2002). Text Mining for Causal Relations. In *proceedings of the AAAI 2002 Spring Symposium*.

- T. Givón (1995). Coherence in text vs. coherence in mind. In M. A. Gernsbacher und T. Givón, Hrsg., *Coherence in spontaneous text*. John Benjamins.
- B. Grosz, A. Joshi, und S. Weinstein (1995). Centering: A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics*, 2(21):203–225.
- B. Grosz und C. Sidner (1986). Attention, Intentions and the Structure of Discourse. *Computational Linguistics*, 12(3):175–204.
- M. Hearst (1992). Automatic Acquisition of Hyponyms from Large Text Corpora. In *14th International Conference on Computational Linguistics*.
- M. Herweg (1991). Temporale Konjunktionen und Aspekt: der sprachliche Ausdruck von Zeitrelationen zwischen Situationen. *Kognitionswissenschaft*, 2(2):51–90.
- C. Hitchcock (2002). Probabilistic Causation. In E. N. Zalta, Hrsg., *The Stanford Encyclopedia of Philosophy*.
- J. Hobbs (1979). Coherence and Coreference. *Cognitive Science*, 3:67–90.
- J. Hockenmaier (2003). *Data and Models for Statistical Parsing with Combinatorial Categorical Grammar*. PhD thesis, University of Edinburgh.
- T. Hofmann (1999). Probabilistic Latent Semantic Analysis. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI 99)*.
- T. Hofmann und J. Puzicha (1998). Unsupervised Learning from Dyadic Data. Technical Report ICSI TR-98-042, International Computer Science Institute (ICSI), Berkeley.
- E. H. Hovy (1990). Parsimonious and profligate approaches to the question of discourse relations. In *5th International Workshop on Natural Language Generation*.
- T. Inui (2004). *Acquiring Causal Knowledge from Text Using Connective Markers*. PhD thesis, Nara Institute of Science and Technology.
- V. Jijkoun und M. de Rijke (2004). Enriching the Output of a Parser Using Memory-Based Learning. In *ACL 2004*.
- R. Jones, R. Ghani, T. Mitchell, und E. Riloff (2003). Active Learning for Information Extraction with Multiple View Feature Sets. In *ECML 2003 Workshop on Adaptive Text Extraction and Mining*.
- M. Kameyama, R. Passoneau, und M. Poesio (1993). Temporal Centering. In *ACL 1993*.

- A. Kaplan und L. Schubert (2001). Measuring and Improving the Quality of World Knowledge Extracted from WordNet. Technical Report 751, University of Rochester, Computer Science Department.
- C. Khoo, S. Chan, und Y. Niu (2000). Extracting Causal Knowledge from a Medical Database Using Graphical Patterns. In *ACL 2000*.
- D. Klein und C. D. Manning (2001). A Generative Constituent-Context Model for Improved Grammar Induction. In *Proceedings of the 40th Annual Meeting of the ACL*.
- A. Knott (1994). Using Linguistic Phenomena to Motivate a Set of Rhetorical Relations. *Discourse Processes*, 18(1):35–62.
- M. Lapata und C. Brew (1999). Using subcategorization to resolve verb class ambiguity. In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*.
- M. Lapata und A. Lascarides (2004). Inferring Sentence-internal Temporal Relations. In *HLT-NAACL 2004*.
- A. Lascarides und N. Asher (1993). Temporal Interpretation, Discourse Relations and Commonsense Entailment. *Linguistics and Philosophy*, 16(5):437–493.
- L. Lee (1999). Measures of Distributional Similarity. In *37th Annual Meeting of the Association for Computational Linguistics (ACL-99)*, S. 25–32.
- W. Lehnert (1994). Cognition, Computers, and Car Bombs: How Yale Prepared Me for the 1990s. In R. C. Schank und E. Langer, Hrsg., *Beliefs, Reasoning, and Decision Making: Psycho-Logic in Honor of Bob Abelson*, S. 143–173. Lawrence Erlbaum Associates.
- W. Lehnert und B. Sundheim (1991). A performance evaluation of text analysis technologies. *AI Magazine*, Fall:81–94.
- B. Levin (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press.
- D. Lewis (1986). *Philosophical Papers II*, Kapitel Causation. Oxford University Press.
- D. Lin (1998). Dependency-based Evaluation of MINIPAR. In *Workshop on the Evaluation of Parsing Systems*.
- D. Lin und P. Pantel (2001). Discovery of Inference Rules for Question Answering. *Natural Language Engineering*, 4(7):343–360.
- J. Lyons (1977). *Semantics*. Cambridge University Press.



- I. Mani und J. Pustejovsky (2004). Temporal Discourse Models for Narrative Structure. In *ACL Workshop on Discourse Annotation*.
- I. Mani, B. Schiffman, und J. Zhang (2003). Inferring Temporal Ordering in News. In *HLT 2003*.
- W. C. Mann und S. A. Thompson (1987). Rhetorical Structure Theory: A Theory of Text Organization. Technical report, University of Southern California, Information Sciences Institute (ISI).
- C. D. Manning und H. Schütze (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.
- D. Marcu und A. Echihabi (2002). An Unsupervised Approach to Recognizing Discourse Relations. In *ACL 2002*.
- P. Menzies (2001). Counterfactual Theories of Causation. In E. N. Zalta, Hrsg., *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/spr2001/entries/causation-counterfactual/>.
- I. Niles und A. Pease (2001). Toward a Standard Upper Ontology. In C. Welty und B. Smith, Hrsg., *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)*.
- L. Polanyi, M. van den Berg, und D. Ahn (2003). Discourse Structure and Sentential Information Structure. An Initial Proposal. *Journal of Logic, Language and Information*, 12(3). auch: Report FXPAL-PR-03-187.
- J. Pustejovsky, G. Katz, und R. Gaizauskas (2003). Practical Applications of Temporal and Event Reasoning. In *ESSLLI 2003*. Tutorial Slides.
- J. Pustejovsky, R. Saurí, A. Setzer, und B. Ingria (2002). *TimeML Annotation Guidelines*.
- E. Riloff und M. Thelen (2000). A Rule-based Question Answering System for Reading Comprehension Tests. In *ANLP/NAACL-2000 Workshop on Reading Comprehension Tests as Evaluation for Computer-Based Language Understanding Systems*.
- C. Rosé, B. di Eugenio, L. Levin, und C. Van Ess-Dykema (1995). Discourse processing of dialogue with multiple threads. In *ACL-1995*.
- R. Saurí, J. Littman, B. Knippen, R. Gaizauskas, A. Setzer, und J. Pustejovsky (2004). *TimeML Annotation Guidelines*.
- J. Schaffer (2003). The Metaphysics of Causation. In E. N. Zalta, Hrsg., *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/spr2003/entries/causation-metaphysics/>.

- R. Schank (1982). Reminding and Memory Organization: An Introduction to MOPs. In W. Lehnert und M. Ringle, Hrsg., *The State-of-the-Art in Natural-Language Understanding*, S. 455–494. Lawrence Erlbaum Associates.
- R. Schank und R. Abelson (1977). *Scripts, Plans, Goals and Understanding*. Lawrence Erlbaum Associates.
- R. C. Schank (1972). Conceptual Dependency: A Theory of Natural Language Understanding. *Cognitive Psychology*, 3(4):pages 532–631.
- F. Schilder und C. Habel (2001). From temporal expressions to temporal information: Semantic tagging of news messages. In *Workshop on Temporal and Spatial Information Processing at ACL-2001*.
- F. Schilder, C. Habel, und Y. Versley (2003). Temporal Information Extraction and Question Answering: Deriving Answers for When-Questions. In R. Bernardi und M. Moortgat, Hrsg., *Questions and Answers: Theoretical and Applied Perspectives (2nd CologNet-ElsNet Symposium)*. Utrecht Institute of Linguistics OTS.
- H. Schmid (1994). Probabilistic Part-of-Speech Tagging using Decision Trees. In *Proceedings of the International Conference on New Methods in Language Processing*.
- R. Sedgewick (1992). *Algorithmen in C++*. Addison-Wesley.
- Y. Shoham (1990). Nonmonotonic Temporal Reasoning and Causation. *Cognitive Science*, 14(2):213–252.
- C. Soria und G. Ferrari (1998). Lexical marking of discourse relations - some experimental findings. In *COLING 98?*
- R. Soricut und D. Marcu (2003). Sentence Level Discourse Parsing using Syntactic and Lexical Information. In *HLT-NAACL 2003*.
- P. Tapanainen und T. Järvinen (1997). A non-projective dependency parser. In *5th Conference on Applied Natural Language Processing (ANLP'97)*.
- K. Torisawa (2003). An Unsupervised Learning Method for Commonsensical Inference Rules on Events. In R. Bernardi und M. Moortgat, Hrsg., *Questions and Answers: Theoretical and Applied Perspectives (2nd CologNet-ElsNet Symposium)*. Utrecht Institute of Linguistics OTS.
- P. Vossen, L. Bloksma, H. Rodriguez, S. Climent, N. Calzolari, A. Roventini, F. Bertagna, A. Alonge, und W. Peters (1998). The EuroWordNet Base Concepts and Top Ontology. Technical report, EuroWordNet.
- M. Walker (1996). Limited Attention and Discourse Structure. *Computational Linguistics*, 22(2):255–264.

- M. Walker (1998). *Centering Theory in Discourse*, Kapitel Centering, Anaphora Resolution and Discourse Structure. Oxford University Press.
- M. Walker (2000). Toward a Model of the Interaction of Centering with Global Discourse Structure. *Verbum*.
- B. Webber (1988). Tense as Discourse Anaphor. *Computational Linguistics*.
- B. Webber (1991). Structure and Ostension in the Interpretation of Discourse Deixis. *Language and Cognitive Processes*, 6(2):107–135.
- B. Webber, A. Joshi, M. Stone, and A. Knott (2001). Anaphora and Discourse Structure. *Computational Linguistics*.
- R. Wilensky (1994). Discourse, Probability, and Inference. In R. C. Schank und E. Langer, Hrsg., *Beliefs, Reasoning, and Decision Making: Psycho-Logic in Honor of Bob Abelson*, S. 363–387. Lawrence Erlbaum Associates.
- J. Woodward (2001). Causation and Manipulability. In E. N. Zalta, Hrsg., *The Stanford Encyclopedia of Philosophy*.
- XTAG Research Group (2001). A Lexicalized Tree Adjoining Grammar for English. Technical Report IRCS-01-03, IRCS, University of Pennsylvania.
- Y. Zhou und S. Goldman (2000). Enhancing Supervised Learning with Unlabeled Data. In *17th International Conference on Machine Learning*.



# Erklärung

Ich versichere, dass die vorstehende Arbeit selbstständig und ohne fremde Hilfe angefertigt und mich anderer als der im beigefügten Verzeichnis angegebenen Hilfsmittel nicht bedient habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht.

Ich bin mit einer Einstellung in den Bestand der Bibliothek des Fachbereiches einverstanden.

Hamburg, 17. Dezember 2004