

INTERNE REPORTS DER DFG-FORSCHERGRUPPE 437  
„TEXTTECHNOLOGISCHE INFORMATIONSMODELLIERUNG“

---

Projekt A2 „Sekimo“

# Koreferenz, Kospezifikation und Bridging: Annotationschema

Daniela Goecke  
Maik Stührenberg  
Fakultät für Linguistik  
und Literaturwissenschaft  
Universität Bielefeld

[daniela.goecke@uni-bielefeld.de](mailto:daniela.goecke@uni-bielefeld.de)  
[maik.stuehrenberg@uni-bielefeld.de](mailto:maik.stuehrenberg@uni-bielefeld.de)

Anke Holler  
Seminar für Deutsche Philologie  
Georg-August-Universität Göttingen  
[anke.holler@phil.uni-goettingen.de](mailto:anke.holler@phil.uni-goettingen.de)

9. Oktober 2007

## Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>3</b>
<b>2</b>	<b>Vorannahmen zum Annotationsschema</b>	<b>3</b>
2.1	Koreferenz und Kospezifikation . . . . .	3
2.2	Vergleich zu existierenden Annotationsschemata . . . . .	6
<b>3</b>	<b>Generelles Vorgehen</b>	<b>6</b>
3.1	Markierung der relevanten Diskursentitäten . . . . .	7
3.2	Typen von Diskursentitäten . . . . .	9
3.3	Annotation unter Serengeti . . . . .	10
<b>4</b>	<b>Annotation der (diskurs-)semantischen Beziehungen</b>	<b>11</b>
4.1	Annotation von Koreferenz . . . . .	11
4.2	Annotation von Kospezifikation . . . . .	12
4.3	Annotation von indirekter Anaphorik . . . . .	14
<b>5</b>	<b>Übersicht über semantische Relationen</b>	<b>15</b>
5.1	Relationen der Koreferenzbeziehung . . . . .	15
5.2	Relationen der Kospezifikation und indirekter Anaphorik . .	16
5.2.1	Direkte Anaphorik (Kospezifikation) . . . . .	17
5.2.2	Kospezifikation mit Propositionen oder Ereignissen als Bezugsgröße . . . . .	23
5.2.3	Indirekte Anaphorik . . . . .	25
<b>6</b>	<b>Annotationsprozess</b>	<b>29</b>
	<b>Literatur</b>	<b>34</b>

## 1 Einführung

Das Annotationsschema beschreibt die Annotation anaphorischer Relationen im Projekt A2 „Sekimo“ der Forschergruppe 437 „Texttechnologische Informationsmodellierung“. Das Schema basiert auf einem Vorschlag zur Annotation von Koreferenzphänomenen, der für die Annotation von Koreferenzstrukturen im Projekt B1 „HyTex“ entwickelt wurde (vgl. Holler, 2003, Holler u. a., 2004, Holler-Feldhaus, 2004).<sup>1</sup>

Für das Projekt A2 wurde dieses Schema erweitert, um die Annotation indirekter Anaphorik (Bridging-Relationen, vgl. Clark, 1977) zu erlauben. Im Folgenden wird der Schwerpunkt auf die Annotation im Projekt A2 gelegt. Das Dokument gliedert sich wie folgt: Zunächst werden grundlegende Überlegungen zum Annotationsschema dargestellt. Insbesondere wird hier die Unterscheidung der Relationen *Koreferenz*, *Kospezifikation* und *Bridging* erläutert. Die Annotation der semantischen Relationen wird in Abschnitt 4 beschrieben, eine Übersicht der semantischen Relationen ist in Abschnitt 5 dargestellt.

## 2 Vorannahmen zum Annotationsschema

### 2.1 Koreferenz und Kospezifikation

Die Grundidee des Annotationsschemas besteht darin, die Unterscheidung zwischen Kospezifikation (Sidner, 1979) und Koreferenz in der Annotation abzubilden. Während zwei Ausdrücke nur dann koreferieren, wenn sie auf dieselbe Entität in der Welt verweisen, genügt für Kospezifikation, dass ein Ausdruck einen vorangegangenen Ausdruck sprachlich wieder aufgreift. Diese Unterscheidung wird bereits in der im Projekt B1 entstandenen ersten Version des Schemas vorgenommen. Die Erweiterung des Schemas für das Projekt A2 berücksichtigt zusätzlich indirekte Anaphorik<sup>2</sup>; hier ist das

---

<sup>1</sup> Grundlage für den Entwurf für das Projekt B1 war das von Poesio (1998) im Rahmen von MATE vorgeschlagene Annotationsschema, das für die Aufgaben im Projekt B1 erweitert und präzisiert wurde.

<sup>2</sup> In der ersten Version des Schemas des Projektes B1 war die Möglichkeit zur Annotation indirekter Anaphorik durch die Einführung eines Relationstyps Bridging bereits angelegt. Dieser Relationstyp wurde jedoch nicht weiter elaboriert, da Bridging-Relationen im Projekt B1 nicht annotiert werden sollten.

Antezedens einer Anapher nicht explizit realisiert, sondern muss aus dem Kontext erschlossen werden.

In der Annotation wird diese Unterscheidung dadurch umgesetzt, dass Kospezifikation und indirekte Anaphorik jeweils als Beziehung zwischen Diskursentitäten markiert werden (Projekte A2 und B1), während Koreferenz als Beziehung zwischen einer Diskursentität und einem Topic in einer TopicMap ausgezeichnet wird (Projekt B1). Eine TopicMap ist ihrerseits eine Repräsentation der Welt, denn durch die TopicMap wird die Welt bzw. ein Ausschnitt der Welt modelliert. Wie in (Holler-Feldhaus, 2004) erörtert, erlaubt es diese zusätzliche Ebene, den Bezug zu Entitäten in der Welt de facto auch zu annotieren.

Unter dem soeben skizzierten Ansatz lässt sich Kospezifikation als eine horizontale Annotationsbeziehung verstehen, denn sie wird auf ein und derselben Ebene (der textrepräsentationellen Ebene) ausgezeichnet; Koreferenz hingegen stellt eine vertikale Beziehung, also eine Beziehung zwischen verschiedenen Ebenen dar, hier zwischen der textrepräsentationellen Ebene und der weltrepräsentationellen Ebene (der TopicMap). Abbildung 1 veranschaulicht das zugrunde gelegte Text-Welt-Modell.

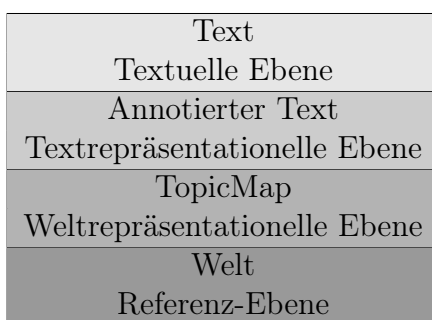


Abbildung 1: Text-Welt-Modell

Es sei noch einmal ausdrücklich darauf hingewiesen, dass Koreferenz formal-semantisch eine Sprache-Welt-Beziehung ist und als Referenzidentität in der Welt interpretiert wird. Koreferenz liegt dann vor, wenn zwei oder mehrere Ausdrücke von den Kommunikationsteilnehmern dazu benutzt werden, um auf *denselben Referenten in der realen Welt* Bezug zu nehmen. Dass hier Koreferenz als Beziehung zu einem Modell der Welt, dem als TopicMap realisierten Wissensnetz, ausgedrückt wird, hat rein praktische Gründe. So

wird Koreferenz annotierbar und die Unterscheidung von Kospezifikation und Koreferenz schlägt sich unmittelbar in der Annotation nieder.

Für die Annotation von Kospezifikation sowie indirekter Anaphorik wird folgende Interpretation zugrunde gelegt: Kospezifikation und indirekte Anaphorik werden innerhalb der textrepräsentationellen Ebene annotiert: ein Ausdruck greift einen vorangegangenen Ausdruck sprachlich wieder auf. Sowohl bei Kospezifikation als auch – noch deutlicher – bei indirekter Anaphorik besteht neben der textuellen Ebene die semantische Interpretation: sprachliche Ausdrücke führen Diskursreferenten (in unserem Fall *Diskursentitäten*) ein oder verweisen auf bereits eingeführte Diskursreferenten (vgl. Webber, 1988); zwischen Diskursreferenten können semantische Relationen bestehen. Abbildung 2 verdeutlicht diesen Zusammenhang.

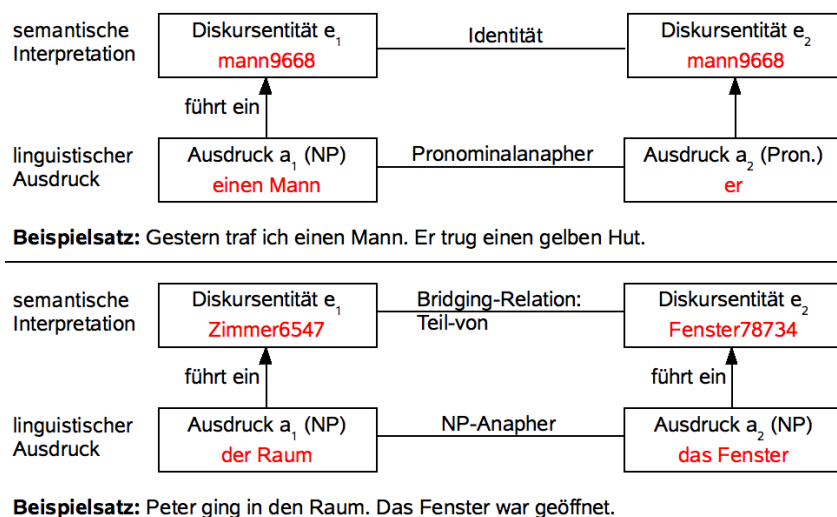


Abbildung 2: Unterscheidung zwischen der Textebene und der Ebene der semantischen Interpretation

Der anaphorische Bezug zwischen sprachlichen Ausdrücken wird in einer XML-Repräsentation als Elemente `cospecLink` oder `bridgingLink` annotiert (s. Abschnitte 4.2 auf Seite 12 und 4.3 auf Seite 14), die (diskurs-)semantische Interpretation findet sich als Relationstyp wieder. Die Art der

anaphorischen Relation (Pronominalanapher vs. NP-Anapher) wird nicht explizit annotiert, sondern ist implizit in den morphologisch-syntaktischen Informationen der linguistischen Ausdrücke kodiert, die Diskursentitäten einführen.

## 2.2 Vergleich zu existierenden Annotationsschemata

In den vergangenen Jahren wurden verschiedene Annotationsschemata für anaphorische Relationen entwickelt, u.a. das UCREL-Schema (Fligelstone, 1992, Garside u. a., 1997), das SGML-basierte MUC-Annotationsschema (Hirschmann, 1997), sowie das XML-basierte MATE/GNOME-Schema (Poesio, 1998, 2004).

Wie im MATE/GNOME-Schema ist das hier vorgestellte Annotationschema XML-basiert und verwendet Standoff-Annotationen (Thompson u. McKelvie, 1997). Eine explizite Unterscheidung von Koreferenz und Kospezifikation ist in Poesio (1998) nicht vorhanden und es wird unter MATE letztlich nur Kospezifikation annotiert.<sup>3</sup> In Poesio (2004) wird auf den Unterschied zwischen Koreferenz und Anaphorik hingewiesen, die Unterscheidung wird im Annotationsschema jedoch nicht durch unterschiedliche Elemente realisiert (vgl. Poesio, 2004, Abschnitt 2.1). Diese Aspekte werden ausführlich in (Holler-Feldhaus, 2004) diskutiert.

Ein weiteres Schema stellen Krasavina u. Chiarcos (2007) vor. Das Schema fokussiert die Annotation mehrsprachiger Korpora (*cross-linguistic anaphoric annotation*, vgl. Krasavina u. Chiarcos, 2007, S. 156); eine explizite Unterscheidung von Koreferenz und Kospezifikation wird nicht angenommen. In diesem Sinne stellt das hier vorgestellte Annotationsschema eine Weiterentwicklung dar.

## 3 Generelles Vorgehen

Datengrundlage für die Korpuserstellung sind die deutschen Korpusdaten des Projektes C1 (47 linguistische Fachtexte) sowie 21 deutsche Zeitungsartikel,

---

<sup>3</sup> Dadurch, dass die Tags unter MATE `coref:x` heißen, wird dieser Punkt verdeckt. Solange nur endophorisch auf Textebene annotiert wird (*horizontal* in Bezug auf Abbildung 1 auf Seite 4), kann Koreferenz nicht adäquat markiert werden, denn Koreferenz ist eine Text-Welt-Beziehung (*vertikal* in Bezug auf Abb.1). Tags zur Markierung einer vertikalen Beziehung sind bei MATE nicht vorgesehen.

die im Projekt A2 zusammengestellt wurden. Das Korpus wird mittels des Versionierungssystems SUBVERSION (SVN) verwaltet. Soweit nicht anders angegeben, beziehen sich die folgenden Beschreibungen auf das Projekt A2; zur Vorbereitung des Korpus für das Projekt B1 siehe Holler (2003).

Für die Annotationen im Projekt A2 wird eine Vorverarbeitung mit dem funktional-abhängigen Parser-Tagger MACHINESE SYNTAX der Firma Connexor Oy durchgeführt. Die morphologisch-syntaktischen Informationen des Machineese Syntax werden auf Wortebene annotiert, XML als Ausgabeformat ist möglich.

Die Annotation wird mehrstufig durchgeführt. In einem ersten Schritt werden alle vorkommenden relevanten Diskursentitäten detektiert und mit einer Diskurs-ID versehen (siehe Abschnitt 3.1). In einem zweiten Schritt werden die relevanten semantischen Beziehungen annotiert, die entweder zwischen Diskursentitäten (Kospezifikation/indirekte Anaphorik) oder zwischen Diskursentitäten und Topics (Koreferenz) etabliert sein können (siehe Abschnitt 4 auf Seite 11).

### 3.1 Markierung der relevanten Diskursentitäten

Das Annotationsschema ist XML-basiert, d. h. die Annotation kann unter Verwendung beliebiger XML-Editoren (z. B. Oxygen<sup>4</sup>) durchgeführt werden. Das Annotationsschema ist in Form einer XML-DTD formalisiert (vgl. Anhang 6 auf Seite 31). Es wird ein Element `de` definiert, um relevante Diskursentitäten zu markieren. Das Element `de` trägt das obligatorische Attribut `deID`, dessen Wert ein *qualified name* (*QName*) sein muss, und ein ebenfalls obligatorisches Attribut `deType`, das den Typ der jeweiligen Diskursentität angibt. Mögliche Werte für `deType` sind in Listing 1 aufgeführt. Das obligatorische Attribut `headRef` hat als Wert die Referenz auf das Kopfnomen der zugrundeliegenden Nominalphrase. Das Attribut `deAuto` wurde eingeführt, um manuell korrigierte Diskursentitäten zu markieren; der Defaultwert `1` wird für automatisch detektierte Diskursentitäten vergeben. Das optionale Attribut `deAnaphoric` gibt für diejenigen Diskursentitäten, die durch das Pronomen *es* eingeführt werden, an, ob die Verwendung anaphorisch ist.

---

<sup>4</sup> <http://www.oxygenxml.com>

### Listing 1: Ausschnitt aus der DTD: Diskursentitäten

```
<!ELEMENT de (cnx-pi_token | de | cnx-pi_sentence)*>

<!ATTLIST de
  deID ID #REQUIRED
  headRef IDREF #REQUIRED
  deType (nom | namedEntity | prop | projProp | evType | ev | state |
  cluster | adv) "nom"
  deAuto (1|0) "1"
  deAnaphoric CDATA #IMPLIED>
```

Für die ID der Diskursentität wird als Format `de_X_Y` vorgeschlagen, wobei  $X$  eine Variable ist, die in Abhängigkeit von der Sorte der jeweiligen Diskursentität als  $n$ ,  $p$ ,  $pp$ ,  $c$ ,  $s$ ,  $e$  oder  $et$  spezifiziert sein kann und  $Y$  eine natürliche Zahl. Man erhält dann z. B. als mögliche IDs `de_n_1` oder `de_e_1`. Im Projekt A2 wird  $n$  als Defaultwert für das Attribut `deType` angenommen, da nur nominale Diskursentitäten annotiert werden, die Infix-Variable  $X$  fällt daher zusammen mit den sie umgebenden Unterstrichen weg.

Als relevante Diskursentitäten gelten alle sprachlichen Ausdrücke, die einen Diskursreferenten im Sinne von Kamp u. Reyle (1993) in die Diskurs- bzw. Textrepräsentation einführen. Es wird empfohlen, in einem ersten Schritt die nominalen (auch *named entity*) Diskursentitäten zu markieren, weil diese im Text leicht zu identifizieren sind und danach die übrigen Entitäten.

Im Projekt A2 werden *nominale* Diskursentitäten auf der Basis der von Machine Syntax annotierten Wortformen automatisch detektiert. Hierbei werden zunächst einfache Diskursentitäten markiert, d. h. Diskursentitäten, die durch eine einfache NP realisiert sind. Aufbauend auf den einfachen Diskursentitäten können auch komplexe Diskursentitäten annotiert werden. Als komplexe NP markiert werden NPs mit Präpositionalphrase (z. B. *der Mann auf dem Bahnsteig*) oder NPs mit NP als Prämodifizierer (z. B. *der dem Bahnsteig stehende Mann*). NPs mit Relativsatz werden *nicht* als komplexe Diskursentität markiert. Restriktive und nicht-restriktive Relativsätze werden gleichermaßen behandelt, die Kospezifikation wird über einen `cospecLink` (vgl. Abschnitt 4.2 auf Seite 12) zwischen den Diskursentitäten von NP und Relativpronomen markiert.

Im folgenden Abschnitt werden die einzelnen Typen von Diskursentitäten dargestellt.

## 3.2 Typen von Diskursentitäten

Im Projekt A2 werden nur Diskursentitäten vom Typ *nom* oder *namedEntity* annotiert.

### deType=„nom“

Diskursentitäten vom Typ *nom* werden durch Nominalphrasen in den Diskurs eingeführt. Nominale Diskursentitäten können automatisch auf der Basis einer syntaktischen Analyse identifiziert werden, im Projekt A2 wird dazu die Parser-Tagger-Ausgabe von MACHINESE SYNTAX genutzt.

Listing 2: Annotationsbeispiel: deType=„nom“

```
<de deID="de202" deType="nom">sie</de>
<de deID="de207" deType="nom">ihre zwei Brüder</de>
<de deID="de324" deType="nom">die Kindertagesstätte</de>
```

### deType=„namedEntity“

Der Wert *namedEntity* beschreibt eine Teilmenge der nominalen Diskursentitäten, nämlich diejenigen nominalen Diskursentitäten, die durch eine *Named Entity Recognition* automatisch detektiert werden können, also solche, die beispielsweise den Klassen *Person*, *Organisation*, *Ortsbeschreibungen* (Länder, Regionen, Städte, Flüsse) zugeordnet werden können.<sup>5</sup> Im Projekt A2 werden zunächst nur Diskursentitäten der Klasse *Person* als Element *de* mit dem *deType* *namedEntity* markiert.

Listing 3: Annotationsbeispiel: deType=„namedEntity“

```
<de deID="de201" deType="namedEntity">Marie Rolfs</de>
<de deID="de256" deType="namedEntity">Herr Peter Müller</de>
```

**N. B.:** Literaturangaben werden als komplexe Diskursentitäten vom Typ *nom* annotiert (vgl. folgendes Listing).

Listing 4: Annotationsbeispiel: Literaturangaben (1)

```
<de deID="de260" deType="nom">
  <de deID="de261" deType="namedEntity">Müller</de> 1997
</de>
```

<sup>5</sup> vgl. [http://www.itl.nist.gov/iaui/894.02/related\\_projects/muc/](http://www.itl.nist.gov/iaui/894.02/related_projects/muc/)

### Listing 5: Annotationsbeispiel: Literaturangaben (2)

```
<de deID="de793" deType="nom">  
  <de deID="de794" deType="namedEntity">Schmidt et al.</de> 1975:29ff  
</de>
```

#### **deType=„prop“**

Der Wert *prop* markiert Propositionen. Alle Deklarativsätze werden so markiert, ebenso alle eingebetteten *dass*-Sätze, die nicht-restriktiven Relativsätze und die Adverbialsätze. Interrogativsätze (auch wenn sie eingebettet sind) dürfen *nicht* als *prop* ausgezeichnet werden.

#### **deType=„projProp“**

Der Wert *projProp* markiert die sog. projektiven Propositionen. Dazu zählen vor allem die Interrogativsätze und die Komplemente von Modaloperatoren.

#### **deType=„evType“**

Der Wert *evType* wird zur Markierung von Ereignistypen (*event-types*) verwendet. Annotiert werden müssen das Vollverb und die zum jeweiligen Ereignistyp dazugehörigen Argumente.

#### **deType=„ev“**

Der Wert *ev* steht für Ereignisse (*events*) und wird für Ereignisverben verwendet. Annotiert wird das entsprechende Ereignisverb.

#### **deType=„state“**

Der Wert *state* soll benutzt werden, um statische Verben auszuzeichnen. Wenn man aus praktischen Erwägungen die Verben nicht so feinkörnig unterscheiden will, kann man diesen Wert auch unterdrücken und alle Verben einheitlich als *ev* auszeichnen.

#### **deType=„cluster“**

Der Wert *cluster* steht für Satzsequenzen unterschiedlicher Länge, wobei alle Sätze der Sequenz zu einem Topic gehören.

### **3.3 Annotation unter Serengeti**

Um sowohl an Qualität wie an Quantität bestmögliche Ergebnisse zu erzielen, sind neben einem Annotationsschema mit strikter Taxonomie und eindeutiger Interpretation einfach handhabbare Werkzeuge zur Annotation

und Organisation der Korpora nötig (vgl. z. B. MMAX; Müller u. Strube (2001)). Zu diesem Zweck wurde im Projekt A2 das Programm SERENGETI zur semi-automatischen Annotation semantischer Relationen entwickelt (<http://coli.lili.uni-bielefeld.de/serengeti/>). Das Annotationswerkzeug wird in Stührenberg u. a. (2007) detailliert beschrieben.

## 4 Annotation der (diskurs-)semantischen Beziehungen

Zur Markierung der (diskurs-)semantischen Beziehungen wird das Element `semRel` eingeführt, wodurch analog zu TEI und MATE die Annotation der (diskurs-)semantischen Beziehungen getrennt von der Annotation der Diskursentitäten vorgenommen wird. Dies hat den Vorteil, dass eine Diskursentität zu mehr als einer anderen Diskursentität gelinkt werden kann, wodurch einerseits Ambiguitäten bzgl. des Antezedens' auf der textuellen Repräsentationsebene gut gekennzeichnet werden können und es andererseits möglich wird, eine Diskursentität sowohl zu einer anderen Diskursentität im Text als auch zu einem Topic in der TopicMap in Beziehung zu setzen.<sup>6</sup>

Im Projekt A2 wird das Element `semRel` hinter dem Text eingefügt (siehe DTD in Anhang 6 auf Seite 31). Das Element `semRel` weist zur Annotation von Koreferenz bzw. zur Annotation von Kospezifikation oder indirekter Anaphorik die drei eingebetteten Tags `corefLink`, `cospecLink` und `bridgingLink` auf.

Listing 6: DTD-Ausschnitt für `semRel`

```
<!ELEMENT semRel (corefLink | cospecLink | bridgingLink)+>
<!ATTLIST semRel semRelID ID #IMPLIED
  sentenceIDRef IDREF #IMPLIED>
```

Die beiden optionalen Attribute `semRelID` und `sentenceIDRef` sind zu Dokumentationszwecken eingeführt worden.

### 4.1 Annotation von Koreferenz

Koreferenz wird als Beziehung zwischen einer referentiellen Diskursentität und einem Topic der TopicMap, dem sog. Referenzanker, interpretiert. Dazu

<sup>6</sup> Dieses Vorgehen lässt außerdem Raum, um das Schema auszubauen und zum Beispiel Objekte der visuellen Situation als Bezugsgröße anzugeben.

wird im Projekt B1 das Element `corefLink` eingeführt. Für dieses Element werden die zwei Attribute `deIDRef` mit einer `deID` und `tmIDRef` mit einer `TopicMapID` als Wert vergeben:

Listing 7: DTD-Ausschnitt für `corefLink`

```
<!ELEMENT corefLink EMPTY>
<!ATTLIST corefLink deIDRef IDREF #REQUIRED
tmIDRef CDATA #REQUIRED>
```

Im Projekt B1 wird die Situation, dass ein Ausdruck im Text eindeutig referentiell gebraucht wird, aber (noch) kein Eintrag in der `TopicMap` vorhanden ist, so gelöst, dass für den Wert des Attributs `tmIDRef` neben `TopicMapID` auch `unknown` vorgesehen ist.

## 4.2 Annotation von Kospezifikation

Kospezifikation wird als Beziehung zwischen zwei Diskursentitäten interpretiert. Dazu wird ein Element `cospecLink` eingeführt. Für dieses Element wird zum ersten das Attribut `relType` eingeführt, dessen Wert einer der in Listing 8 aufgeführten Werte ist (siehe auch Abschnitt 5.2.1). Zum zweiten wird das Attribut `phorIDRef` (Typ: `IDREF`) mit einer `deID` als Wert vergeben. Der Wert dieses Attributs verweist auf das entsprechende phorische Element. Die Bezugsgröße des phorischen Ausdrucks wird ebenfalls markiert. Da eine Diskursentität mehrere Bezugsgrößen aufweisen kann, wird das Attribut `antecedentIDRefs` (Typ: `IDREFS`) mit einer oder mehreren `deID` als Wert vergeben.

Listing 8: DTD-Ausschnitt für `cospecLink`

```
<!ELEMENT cospecLink EMPTY>
<!ATTLIST cospecLink relType (ident | propName | namedEntity | synonym |
hyperonym | hyponym | paraphrase | addInfo | isA) #REQUIRED
phorIDRef IDREF #REQUIRED
antecedentIDRefs IDREFS #REQUIRED>
```

Bei ambigen Bezugsgrößen wird das Element `cospecLink` entsprechend vervielfacht. Als Antezedens wird jeweils der *letzte nicht-pronominale Ausdruck* in der Kospezifikationskette gewählt.

### Listing 9: Annotationsbeispiel für Kospezifikation

```
<de deID="de21" deType="nom">Maik</de> und  
<de deID="de22" deType="nom">Marie</de> sind nicht obdachlos.  
<de deID="de23" deType="nom">Sie</de> sehen nicht verlottert aus. Aber  
<de deID="de24" deType="nom">sie</de> sind benachteiligt - ganz einfach  
deshalb, weil <de deID="de25" deType="nom">sie</de> weniger haben als  
die meisten anderen Kinder in der Republik.  
  
<semRel>  
  <cospecLink relType="ident" phorIDRef="de23"  
    antecedentIDRefs="de21_de22"/>  
  <cospecLink relType="ident" phorIDRef="de24"  
    antecedentIDRefs="de21_de22"/>  
  <cospecLink relType="ident" phorIDRef="de25"  
    antecedentIDRefs="de21_de22"/>  
</semRel>
```

### Nicht annotierte anaphorische Beziehungen

Bindungsphänomene werden *nicht* annotiert, denn Bindung ist eine ausschließlich syntaktisch lizenzierte und somit satzgebundene anaphorische Beziehung. Im folgenden Satz wird beispielsweise das anaphorische Possessivpronomen *seinem* von der quantifizierten NP *jeder Autor* syntaktisch gebunden.

Jeder Autor sollte die Verlinkung in seinem Hypertext selbst vornehmen.

Ebenfalls nicht annotiert werden VP-Ellipsen, die ein anaphorisches Pronomen enthalten:

Otto<sub>i</sub> [<sub>VP</sub> stellte seinen<sub>i</sub> Hypertext fertig], Max<sub>j</sub> auch.

Die fehlende VP im rechten Teil (d.h. im zweiten Teilsatz) ist vom Wortlaut identisch zu der VP im linken Teil (d.h. im ersten Teilsatz). Bei der Rekonstruktion der VP sind in Abhängigkeit von der Interpretation des anaphorischen Pronomens (als Konstante oder gebundene Variable) zwei Lesarten möglich, die durch unterschiedliche Indizierungen verdeutlicht werden können:

- a. Otto<sub>i</sub> [<sub>VP</sub> stellte seinen<sub>i</sub> Hypertext fertig], Max<sub>j</sub> auch.  
[<sub>VP</sub> stellte seinen<sub>i</sub> Hypertext fertig]
- b. Otto<sub>i</sub> [<sub>VP</sub> stellte seinen<sub>i</sub> Hypertext fertig], Max<sub>j</sub> auch.  
[<sub>VP</sub> stellte seinen<sub>j</sub> Hypertext fertig]

Im Falle von a. liegt eine strikte Interpretation des Pronomens *seinen* vor. Die NP *Otto* und das Pronomen *seinen* sind koindiziert. Der Satz wird so verstanden, dass Max den Hypertext von Otto fertigstellt. Das Pronomen wird daher semantisch als Konstante gedeutet. Im Falle von b. ist das Pronomen *seinen* mit der NP *Max* koindiziert und wird als gebundene Variable interpretiert. In dieser sog. *Sloppy-Identity*-Lesart stellt Max seinen eigenen Text fertig. (Vgl. hierzu auch Hardt, 1996).

### 4.3 Annotation von indirekter Anaphorik

Indirekte Anaphorik wird ebenfalls als Beziehung zwischen zwei Diskursentitäten interpretiert. Indirekte Anaphorik besteht dann, wenn das Antezedens zu einem anaphorischen Element nicht explizit im Text genannt ist, sondern aus dem Kontext erschlossen werden kann. Dazu wird ein Element `bridgingLink` eingeführt. Ein `bridgingLink` wird außerdem annotiert, wenn zwischen zwei Objekten eine explizite Possessionsrelation besteht (markiert durch Possessivpronomen oder Genitiv-NP).

Für dieses Element wird zum Ersten das Attribut `relType` eingeführt, dessen Wert einer der in Listing 10 aufgeführten Werte ist (siehe auch Abschnitt 5.2.3). Zum Zweiten wird das Attribut `phorIDRef` (Typ: IDREF) mit einer `deID` als Wert vergeben. Der Wert dieses Attributs verweist auf das entsprechende phorische Element. Die Bezugsgröße des phorischen Ausdrucks wird ebenfalls markiert. Da eine Diskursentität mehrere Bezugsgrößen aufweisen kann, wird das Attribut `antecedentIDRefs` (Typ: IDREFS) mit einer oder mehrerer `deID` als Wert vergeben.

Listing 10: DTD-Ausschnitt für `bridgingLink`

```
<!ELEMENT bridgingLink EMPTY>
<!ATTLIST bridgingLink relType (poss | setMember | hasMember | meronym |
  holonym | bridging) #REQUIRED
  phorIDRef IDREF #REQUIRED
  antecedentIDRefs IDREFS #REQUIRED>
```

Bei ambigen Bezugsgrößen wird das Element `bridgingLink` entsprechend vervielfacht. Als Antezedens wird jeweils der *letzte nicht-pronominale Ausdruck* gewählt.

Listing 11: Annotationsbeispiel für indirekte Anaphorik

```
Hier lebt <de deID="de1" deType="nom">Maik Celik</de> [...] mit
<de deID="de2" deType="nom">seinen zwei Brüdern</de> und
<de deID="de2" deType="nom">den Eltern</de> [...]
<de deID="de3" deType="nom">Seine Mutter</de> lebt seit 19 Jahren
von Sozialhilfe.

<semRel>
  <bridgingLink relType="poss" phorIDRef="de3" antecedentIDRefs="de1"/>
  <bridgingLink relType="setMember" phorIDRef="de3"
    antecedentIDRefs="de2"/>
</semRel>
```

## 5 Übersicht über semantische Relationen

Wie in den vorangegangenen Abschnitten beschrieben, umfasst die Annotation der sprachlichen Phänomene zwei Aspekte: die Annotation der relevanten Diskursentitäten und die Annotation der diskursemantischen Beziehungen zwischen einem Ausdruck und seiner Bezugsgröße. In den nachfolgenden Abschnitten werden anhand von Beispielen die jeweils vorgesehenen Werte eingeführt.<sup>7</sup>

### 5.1 Relationen der Koreferenzbeziehung

Es wird ein enger Koreferenzbegriff zugrunde gelegt: Zwei Diskursreferenten koreferieren, wenn sie mit demselben Referenzanker in der TopicMap verbunden sind. Anaphorik ist damit keine notwendige Bedingung für Koreferenz. Die Besonderheit des hier verwendeten Koreferenzbegriffes besteht darin, dass Referenz als „Vorhanden in der TopicMap“ interpretiert wird. Dies setzt natürlich voraus, dass alle Individuen der Welt auch im Wissensnetz

<sup>7</sup> In den Annotationsausschnitten wurde auf Wurzelemente sowie auf umschließende Elemente wie z. B. `semRel` verzichtet. Es sind nur die relevanten Diskursentitäten markiert.

repräsentiert sind. Eigennamen und Literaturangaben sind typischerweise verankert:

#### Listing 12: Annotationsbeispiel für Koreferenz

```
Das von <de deID="de1" deType="nom">Kuhlen 1991</de> skizzierte  
Grundmodell eines Hypertextsystems orientiert sich am Vorbild von  
Datenbankmanagementsystemen.
```

```
<corefLink refID="de1" tmIDRef="tm_1"/>
```

Das hier beschriebene Vorgehen erlaubt es u. a., dass ein referentieller Ausdruck mit zwei verschiedenen Ankern verbunden ist, wie im folgenden Beispiel illustriert. Der Name „Austin“ wird einmal für die Stadt und einmal für die Automarke verwendet. Entsprechend dürfen die beiden Ausdrücke nicht als koreferentiell markiert werden. Dies kann daraus abgeleitet werden, dass Austin in der TopicMap auf jeweils ein anderes Topic verweist.

#### Listing 13: Annotationsbeispiel für Koreferenz

```
<de deID="de2" deType="nom">Austin</de> liegt in Texas.
```

```
<corefLink refID="de2" tmRefAnchor="tm_2"/>
```

```
<de deID="de3" deType="nom">Der Austin</de> ist beliebt.
```

```
<corefLink refID="de3" tmRefAnchor="tm_3"/>
```

## 5.2 Relationen der Kospezifikation und indirekter Anaphorik

Kospezifikation beruht auf einer phorischen Beziehung zwischen sprachlichen Ausdrücken, wobei der phorische Ausdruck ein Pronomen, eine NP oder ein Eigenname sein kann.

**N. B.:** Der Einfachheit halber soll hier zunächst nur der anaphorische Fall betrachtet werden. Das vorgeschlagene Annotationschema kann aber leicht auf kataphorische Verhältnisse übertragen werden.

Indirekte Anaphorik beruht ebenfalls auf einer phorischen Beziehung zwischen sprachlichen Ausdrücken, wobei der phorische Ausdruck hier eine definite NP ist. „Bei den indirekten Anaphern handelt es sich um definite Ausdrücke, die in interpretativer Abhängigkeit zu bestimmten Ausdrücken der

vorausgehenden Textstruktur stehen und zwei textreferentielle Funktionen haben: die Einführung neuer (bisher noch nicht erwähnter) Textreferenten und die Weiterführung des globalen referentiellen Bezugs“ (Schwarz, 2000, S. 49).

### 5.2.1 Direkte Anaphorik (Kospezifikation)

#### **relType=„ident“**

Der Wert *ident* wird erstens vergeben, wenn sich ein Pronomen auf eine NP bezieht.

Listing 14: Annotationsbeispiel: ident – Pronomen

```
<de deID="de4" deType="nom">Ein Link</de> ist im Text meist farbig  
markiert. <de deID="de5" deType="nom">Er</de> ist dadurch gut sichtbar.  
<cospecLink relType="ident" phorIDRef="de5" antecedentIDRefs="de4"/>
```

Dieser Wert wird zweitens vergeben, wenn sich eine NP auf eine rekurrente NP (d. h. mit identischem Kopfnomen) bezieht.

Listing 15: Annotationsbeispiel: ident – identisches Kopfnomen

```
<de deID="de6" deType="nom">Ein Link</de> ist im Text meist farbig  
markiert. <de deID="de7" deType="nom">Der Link</de> ist dadurch gut  
sichtbar.  
<cospecLink relType="ident" phorIDRef="de7" antecedentIDRefs="de6"/>
```

Weitere Beispiele:

im türkischen Parlament – im Parlament  
Hauswand – Hauswand  
das Atomgesetz – das Atomgesetz  
einen Lizenzentzug – der Lizenzentzug

#### **relType=„namedEntity“**

Dieser Wert wird vergeben, wenn sich eine Nominalphrase, die nicht vom Typ *namedEntity* ist, auf eine Nominalphrase vom Typ *namedEntity* bezieht, also solche, die beispielsweise den Klassen *Person*, *Organisation*,

*Ortsbeschreibungen* (Länder, Regionen, Städte, Flüsse) zugeordnet werden können.

**N. B.:** Nur Personennamen werden explizit als Diskursentitäten vom Typ *namedEntity* annotiert. Bei der manuellen Annotation soll der Relationstyp *namedEntity* jedoch auch für solche Nominalphrasen vergeben werden, die durch eine Named-Entity-Erkennung als solche markiert würden.

#### Listing 16: Annotationsbeispiel: *namedEntity*

```
In einem Bericht für den Sozialausschuss des Landtages kritisierte
<de deID="de028" deType="namedEntity">Trauernicht</de> Vattenfall erneut
scharf und warf dem Unternehmen falsche Angaben nach den Störfällen vor.
<de deID="de033" deType="nom">Die Ministerin</de> schlug eine
Beweislastumkehr im Atomgesetz vor.

<cospecLink id="sr11" relType="namedEntity" phorIDRef="de033"
antecedentIDRefs="de028" />
```

Weitere Beispiele:

- Petra Kaiser – die Frau
- Ali Klç – der Unternehmer
- Ali Klç – der Deutschländer
- Meyer – der entmachtete Richter
- Vattenfall – das Unternehmen
- die Niederlande – das Land

#### **relType=„propName“**

Die Anapher ist ein Eigenname, der sich auf eine nominale Bezugsgröße bezieht. Die kospezifizierte NP kann sowohl eine Diskursentität vom Typ *nom* als auch eine Diskursentität vom Typ *namedEntity* (für Personennamen) sein.

**N. B.:** Wie für den Relationsstyp *namedEntity* gilt auch hier, dass nur Personennamen explizit als Diskursentitäten vom Typ *namedEntity* annotiert sind. Bei der manuellen Annotation soll der Relationstyp *propName* jedoch auch für solche Nominalphrasen vergeben werden, die durch eine Named-Entity-Erkennung als solche markiert würden.

Listing 17: Annotationsbeispiel: propName (1)

```
<de deID="de1" deType="namedEntity">Sabine Rolfs</de> ist erst seit  
kurzem arm. &gt;Ich war froh, dass noch ein Aufziehhauto für 3,50 Euro zu  
vergeben war&lt;;, berichtet  
<de deID="de2" deType="namedEntity">Rolfs</de>  
  
<cospecLink relType="propName" phorIDRef="de2" antecedentIDRefs="de1"/>
```

Listing 18: Annotationsbeispiel: propName (2)

```
Mit <de deID="de1" deType="nom">diesem XML-Editor</de> lässt es sich  
sehr gut arbeiten. <de deID="de2" deType="nom">XML Spy</de> ist daher  
trotz seines Preises sehr beliebt.  
  
<cospecLink relType="propName" phorIDRef="de2" antecedentIDRefs ="de1"/>
```

Weitere Beispiele:

Ali Klç – Ali Klç  
der Geschäftsmann Ali Klç – Ali Klç  
Petra Kaiser – Kaiser  
Petra – Petra  
der Zusteller UPS – UPS  
das ferne Istanbul – Istanbul

**relType=„synonym“**

Dieser Wert wird vergeben, wenn die *Kopfnomen* von Anapher und Antezedens in einer Synonymierelation zueinander stehen: „Semantische Relation der Bedeutungsgleichheit (bzw. Bedeutungsähnlichkeit) zwischen zwei [...] sprachlichen Ausdrücken [...] S. entsteht durch ein Nebeneinander dialektaler und hochsprachlicher, umgangssprachlicher und fachsprachlicher Varianten, durch euphemistische Beschreibungstendenzen (*sterben* vs. *entschlafen*), durch Sprachlenkung (*Ostzone* vs. *DDR*) und durch Übernahme von Fremdwörtern (*Stockwerk* vs. *Etage*)“. (Bußmann, 1990, S. 763)

Für die Annotation wird ein weiter Begriff der Synonymie angenommen, d. h. die kospezifizierte NP ist in dem Kontext des Textes synonym zu ihrer Bezugsgröße (partielle Synonymie, vgl. Bußmann, 1990). Beispiel für partielle Synonymie: Im Kontext „Hypertext“ besteht eine Synonymierelation

zwischen den Begriffen *Link* und *Verweis*. Diese Synonymie beschränkt sich jedoch auf diese Lesart von *Verweis*, im Kontext „Fußball“ gilt sie nicht (\*Ballack bekam einen Link).

**N. B.:** Im Gegensatz zur Paraphrase findet Synonymie nur auf Ebene der Lexeme statt:

Pilz – Schwammerl (Synonym)  
Junggeselle – unverheirateter Mann (Paraphrase)

Listing 19: Annotationsbeispiel: synonym (1)

```
Gestern besuchte Marie <de deID="de1" deType="nom">ihre Oma</de>.
<de deID="de2" deType="nom">Die Großmutter</de> lebt auf einem
Bauernhof.

<cospecLink relType="synonym" phorIDRef="de2" antecedentIDRefs="de1"/>
```

Listing 20: Annotationsbeispiel: synonym (2)

```
<de deID="de8" deType="nom">Ein Link</de> ist im Text meist farbig
markiert. <de deID="de9" deType="nom">Der Verweis</de> ist dadurch gut
sichtbar.

<cospecLink relType="synonym" phorIDRef="de9" antecedentIDRefs="de8"/>
```

**N. B.:** Abkürzungen werden ebenfalls als Synonyme annotiert. Falls in Anschluss an die Langform die Abkürzung in Klammern genannt ist, wird die Langform als Antezedens gewählt; ansonsten wird die letzte nicht-pronominale Nennung als Antezedens markiert (siehe folgendes Listing).

Listing 21: Annotationsbeispiel: synonym (3)

```
Bundesweite Streiks bei der Bahn hat <de deID="de14" deType="nom">die
Gewerkschaft Deutscher Lokomotivführer</de> (<de deID="de15"
deType="nom">GDL</de>) für die erste Juli-Woche angekündigt.
<de deID="de24" deType="nom">Die GDL</de> fordert drastische Lohn-
erhöhungen und einen eigenständigen Tarifvertrag [...]
```

```
Neben <de deID="de41" deType="nom">der Gewerkschaft Deutscher
Lokomotivführer</de> existieren noch die Gewerkschaft Deutscher
Bundesbahnbeamten und Anwärter (GDBA) sowie Transnet.
```

```
<cospecLink relType="synonym" phorIDRef="de24" antecedentIDRefs="de14"/>
<cospecLink relType="synonym" phorIDRef="de41" antecedentIDRefs="de24"/>
```

Weitere Beispiele:

Tragespuren – Gebrauchsspuren  
Trümmer – Überreste  
Gehirn – Denkapparat  
das Max-Planck-Institut – MPI

### **relType=„hyperonym“**

Das Kopfnomen der Anapher ist ein Hyperonym zum Kopfnomen des Antezedens: „Semantische Relation der lexikalischen Überordnung [...] zur Kennzeichnung hierarchieähnlicher Gliederungen des Wortschatzes: *Obst* ist ein Hyperonym von *Apfel*, *Birne*, *Pflaume*, denn der Übergang von z. B. *Apfel* zu *Obst* bringt eine Verallgemeinerung der Bedeutung mit sich.“ (Bußmann, 1990, S. 317)

Listing 22: Annotationsbeispiel: hyperonym (1)

```
<de deID="de1" deType="nom">HTML-Editoren</de> erleichtern die
Erstellung von Webseiten wesentlich. Deswegen sind
<de deID="de2" deType="nom">diese Programme</de> bei den Anwendern auch
sehr beliebt.

<cospecLink relType="hyperonym" phorIDRef="de2" antecedentIDRefs="de1"/>
```

Listing 23: Annotationsbeispiel: hyperonym (2)

```
<de deID="de1" deType="nom">Lurup</de> ist ein sozialer Brennpunkt
<de deID="de2" deType="nom">der Hansestadt</de>, ein Vorort mit
Einzelhäusern, aber auch vielen Wohnblocks im Westen
<de deID="de3" deType="nom">der Stadt</de>

<cospecLink relType="hyperonym" phorIDRef="de3" antecedentIDRefs="de2"/>
```

Weitere Beispiele:

Studium – Ausbildung  
Psychologen – Wissenschaftler

### **relType=„hyponym“**

Das Kopfnomen der Anapher ist ein Hyponym zum Kopfnomen des Antezedens: „die Relation der Unterordnung im Sinne einer inhaltsmäßigen Spezifizierung. Z. B. ist *Apfel* hyponym zu *Obst*, da *Apfel* eine spezifischere Bedeutung als *Obst* hat.“ (Bußmann, 1990, S. 317)

Listing 24: Annotationsbeispiel: hyponym

```
<de deID="de12" deType="nom">Die passende Software</de> für ein zu  
lösendes Problem zu finden ist nicht immer leicht, zumal die Programme  
oft nicht halten, was sie versprechen. Vorsicht vor  
<de deID="de13" deType="nom">Spracherkennungssoftware </de> oder  
<de deID="de14" deType="nom">Maschinellen-Übersetzungsprogrammen</de> aus  
dem Supermarkt. Leistungsfähige Programme haben nämlich ihren Preis.  
  
<cospecLink relType="hyponym" phorIDRef="de13" antecedentIDRefs="de12"/>  
<cospecLink relType="hyponym" phorIDRef="de14" antecedentIDRefs="de12"/>
```

Weiteres Beispiel:

Überblick – Preisübersicht

### **relType=„addInfo“**

Die kospezifizierte NP führt neue oder zusätzliche Informationen ein.

Listing 25: Annotationsbeispiel: addInfo (1)

```
<de deID="de19" deType="nom">Der HTML-Editor</de> ist ständig  
weiterentwickelt worden. <de deID="de20" deType="nom">Die Version  
5.0</de> kommt gerade auf den Markt.  
  
<cospecLink relType="addInfo" phorIDRef="de20" antecedentIDRefs="de19"/>
```

Listing 26: Annotationsbeispiel: addInfo (2)

```
<de deID="de6">Ein gepanzertes Kettenfahrzeug</de> der britischen  
Streitkräfte ist gestern Mittag gegen <de deID="de9">ein
```

```

Einfamilienhaus</de> im Dorf Asseln bei Lichtenau gefahren. [...]
<de deID="de52">Ein Gutachter</de> bezifferte den Schaden
a<de deID="de54">m rund 70 Jahre alten Haus</de>, das aus
Bruch- und Backsteinen errichtet worden ist, auf rund 30.000 Euro.
<cospecLink relType="addInfo" phorIDRef="de54" antecedentIDRefs="de9" />

```

Weiteres Beispiel:

das ferne Istanbul – die 12-Millionen-Metropole

### **relType=„paraphrase“**

Die Anapher ist eine Paraphrase (Umschreibung) ihres Antezedens.

Listing 27: Annotationsbeispiel: paraphrase

```

<de deID="de1" deType="nom">Das Artefakt</de> ist im Ägyptischen
Museum von Kairo zu sehen: [...] Forscher datieren das Alter
<de deID="de2" deType="nom">der Ersatzzehe</de> aus Leder und Holz
auf 2600 bis 3000 Jahre.
<cospecLink id="sr8" relType="paraphrase" phorIDRef="de2"
antecedentIDRefs="de1" />

```

**N. B.:** Übersetzungen werden ebenfalls als Paraphrase annotiert:

Armoured Personal Carriers – die Fahrzeuge

### **relType=„isA“**

Wenn es aus praktischen Erwägungen nicht möglich ist, die Relationen *namedEntity*, *synonym*, *hyperonym*, *hyponym*, *addInfo* und *paraphrase* bei der Annotation zu unterscheiden, ist es denkbar, diese Relationen zusammenzufassen und die entsprechenden *relType*-Werte einheitlich durch den Wert *isA* zu ersetzen.

## **5.2.2 Kospezifikation mit Propositionen oder Ereignissen als Bezugsgröße**

Die folgenden Relationstypen beschreiben Vorkommen von Kospezifikation, bei denen die Bezugsgröße eine nichtnominale Diskursentität ist.

### **relType=„abstrProp“**

Der kospezifizierte Ausdruck bezieht sich auf eine Proposition.

Listing 28: Annotationsbeispiel: abstrProp

```
<de deID="de_p_1" deType="Prop">Hypertexte sind sprachtheoretisch  
interessante Texte</de>. <de deID="de_n_21" deType="nom">Dies</de>  
ist bereits mehrfach festgestellt worden.  
  
<cospecLink relType="abstrProp" phorIDRef="de_n_21"  
antecedentIDRefs="de_p_1"/>
```

### **relType=„abstrProjProp“**

Der kospezifizierte Ausdruck bezieht sich auf eine projektive Proposition.

Listing 29: Annotationsbeispiel: abstrProjProp

```
<de deID="de_pp_1" deType="projProp">Welche Suchmaschine die beste ist  
</de>, <de deID="de_n_22" deType="nom">diese Frage</de> ist nicht mehr  
eindeutig zu beantworten, denn Vivissimo strukturiert die Suchergebnisse  
bereits in Clustern und hat daher aufgeholt. Allerdings ist Google in  
Sachen Trefferquote nach wie vor ungeschlagen.  
  
<cospecLink relType="abstrProjProp" phorIDRef="de_n_22" antecedentIDRefs  
="de_pp_1"/>
```

### **relType=„abstrEvType“**

Der kospezifizierte Ausdruck bezieht sich auf einen Ereignistyp.

Listing 30: Annotationsbeispiel: abstrEvType

```
Google will mit der Produktsuchmaschine Froogle  
<de deID="de_et_1" deType="evType">neue Kunden gewinnen</de>.  
Ob <de deID="de_n_23" deType="nom">das</de> gelingen wird, bleibt  
abzuwarten.  
  
<cospecLink relType="abstrEvType" phorIDRef="de_n_23"  
antecedentIDRefs="de_et_1"/>
```

### **relType=„abstrEv“**

Der kospezifizierte Ausdruck bezieht sich auf ein Ereignis.

### Listing 31: Annotationsbeispiel: abstrEv

```
Eine Studentin hat den gesamten Korpus manuell
<de deID="de_e_1" deType="ev">annotiert</de>.
<de deID="de_n_24" deType="nom">Das</de> dauerte zwei Monate.

<cospecLink relType="abstrEv" phorIDRef="de_n_24"
  antecedentIDRefs="de_e_1"/>
```

#### **relType=„abstrCluster“**

Der kospezifizierte Ausdruck bezieht sich auf eine Summe von Propositionen bzw. auf einen Textabschnitt.

### Listing 32: Annotationsbeispiel: abstrCluster

```
<de deID="de_c_1" deType="cluster">Die Spezifikation des Ankerkonzepts
in Tochtermann (1995) [...] und Bereich-Bereich-Links</de>.
<de deID="de_n_25" deType="nom">Weiterhin</de> unterscheidet er [...].

<cospecLink relType="abstrCluster" phorIDRef="de_n_25"
  antecedentIDRefs="de_c_1"/>
```

#### **relType=„unknown“**

Falls keine der aufgeführten Beziehungen zutrifft, wird der Wert `unknown` vergeben.

### 5.2.3 Indirekte Anaphorik

In der ersten Version des Annotationsschemas für das Projekt B1 wurden die von Schwarz (2000) beschriebenen Ausprägungen indirekter Anaphorik (semantische Rollenzuweisung, nominalsemantische Relation Meronymie, nominalsemantische Relation Hyperonymie, Schema-basierte indirekte Anaphorik und Inferenz-basierte indirekte Anaphorik) auf den `relType`-Wert *bridging* abgebildet. In der Version für das Projekt A2 werden Bridgingrelationen genauer modelliert. Zu diesem Zweck wird eine Unterteilung vorgenommen, anhand derer verschiedene Bridgingrelationen annotiert werden können. Die einzelnen Relationstypen wurden so ausgewählt, dass sie durch linguistische Ressourcen beschreibbar sind. Dies ist der Fall für die semantischen Relationen Meronymie und Holonymie, die in GERMANET

modelliert sind, sowie für die Possessionsrelation, die durch Possessivpronomen bzw. Genitivkonstruktionen syntaktisch markiert ist. Neben diesen Relationen wurden weitere Relationen ausgewählt: *setMember* und *hasMember* zur Beschreibung von Klassenzugehörigkeit, sowie *bridging* für sonstige, nicht explizit berücksichtigte Relationstypen (z. B. Schema-basierte indirekte Anaphorik).

### relType=„poss“

Der Wert *poss* wird vergeben, wenn der phorische Ausdruck explizit durch ein Possessivpronomen oder eine Genitiv-NP besitzanzeigend markiert ist.

Listing 33: Annotationsbeispiel: poss (1)

```
<de deID="de3">Andere Berater</de> kümmern sich um Stipendien für
<de deID="de4">ihre Schützlinge</de>.

<bridgingLink relType="poss" phorIDRef="de4" antecedentIDRefs="de3"/>
```

Listing 34: Annotationsbeispiel: poss (2)

```
<de deID="de18">Ali Klç</de>, [...] ein Mann, der es in 25 Jahren in
Deutschland zu etwas gebracht hat [...] Darüber staunt so mancher aus
<de deID="de47">Klç&apos; deutschen Freundes- und Bekanntenkreis</de>.

<bridgingLink relType="poss" phorIDRef="de47" antecedentIDRefs="de18"/>
```

Weitere Beispiele:

Peter – Peters Text  
Peter – sein Text

### relType=„meronym“

Die Kopfnomen von Anapher und Antezedens stehen in einer Meronymiere-lation (Teil-von-Beziehung) zueinander. Die Anapher beschreibt den Teil, das Antezedens das Ganze, z. B. ist ein *Lüfter* ein Teil eines *Computers*.

Listing 35: Annotationsbeispiel: meronym (1)

```
Hans hat <de deID="de1" deType="nom">einen neuen Computer</de> gekauft.
<de deID="de2" deType="nom">Der Lüfter</de> war unerträglich laut.
```

```
<bridgingLink relType="meronym" phorIDRef="de2" antecedentIDRefs="de1"/>
```

### Listing 36: Annotationsbeispiel: meronym (2)

```
Ein Gutachter bezifferte den Schaden  
a<de deID="de1" deType="nom">m rund 70 Jahre alten Haus</de>  
[...] auf rund 30.000 Euro. Vorsorglich wurden  
<de deID="de2" deType="nom">Wohnzimmer</de> und  
<de deID="de3" deType="nom">Dachboden</de> gesperrt.  
  
<bridgingLink relType="meronym" phorIDRef="de2" antecedentIDRefs="de1"/>  
<bridgingLink relType="meronym" phorIDRef="de3" antecedentIDRefs="de1"/>
```

Weitere Beispiele:

Tür – Klinke  
ein Kettenfahrzeug – die rechte Kette

**N. B.:** Städte werden als Meronym des entsprechenden Landes annotiert:

Deutschland – München

### **relType=„holonym“**

Das Kopfnomen der Anapher ist ein Holonym zum Kopfnomen des Antezedens', d. h. die Anapher beschreibt das Ganze, das Antezedens beschreibt einen Teil des Ganzen.

Beispiele:

Seite – Buch  
Henkel – Tasse

**N. B.:** Ein Land wird als Holonym zu einer Stadt annotiert:

München – Deutschland

### **relType=„hasMember“**

Die Anapher beschreibt eine Menge, das Antezedens beschreibt ein Element dieser Menge.

### Listing 37: Annotationsbeispiel: hasMember

```
<de deID="de6" deType="nom">Maik</de> hat kein eigenes Fahrrad und  
<de deID="de8" deType="nom">Marie</de> fährt nicht in den Urlaub.  
<de deID="de10" deType="nom">Zwei Kinder</de>, eine Gemeinsamkeit:  
Ihre Eltern haben nur halb so viel Geld im Monat wie ein deutscher  
Durchschnittsbürger.  
  
<cospecLink relType="hasMember" phorIDRef="de10"  
antecedentIDRefs="de6_de8"/>
```

Weiteres Beispiel:

#### Deutschland – Europa

Im Beispiel wird Europa in der Lesart als die Menge der dazugehörigen Länder (Europa als Zusammenschluss der Länder) gesehen.

#### **relType=„setMember“**

Der phorische Ausdruck ist Element einer durch die Bezugsgröße beschriebenen Menge.

### Listing 38: Annotationsbeispiel: setMember (1)

```
Gestern traf ich <de deID="de1" deType="nom">drei Frauen</de>,  
<de deID="de2" deType="nom">die Älteste</de> traf ich heute Morgen  
wieder.  
  
<bridgingLink relType="setMember" phorIDRef="de2"  
antecedentIDRefs="de1"/>
```

### Listing 39: Annotationsbeispiel: setMember (2)

```
Dabei wurden sie in <de deID="de18">Gruppen</de> aufgeteilt und durften  
entweder nur durch die Nase oder nur durch den Mund atmen. Es stellte  
sich heraus, dass <de deID="de27">die Mundatmer</de> [...] häufiger  
während des Films mitgähnen mussten als  
<de deID="de31">die Nasenatmer</de>.  
  
<bridgingLink relType="setMember" phorIDRef="de27"  
antecedentIDRefs="de18" />  
<bridgingLink relType="setMember" phorIDRef="de31"  
antecedentIDRefs="de18" />
```

Weiteres Beispiel:

Europa – Deutschland

### **relType=„bridging“**

Der phorische Ausdruck steht in einer Relation zu seiner Bezugsgröße, die nicht durch die bisher in diesem Abschnitt beschriebenen Werte abgedeckt ist (z. B. Schema-basierte indirekte Anaphorik oder Inferenz-basierte indirekte Anaphorik).

Listing 40: Annotationsbeispiel

```
Am Wochenende war ich zu <de deID="de1" deType="nom">einer Hochzeit</de>
eingeladen - <de deID="de2" deType="nom">die Torte</de> war köstlich.

<bridgingLink relType="bridging" phorIDRef="de2"
  antecedentIDRefs="de1"/>
```

Listing 41: Annotationsbeispiel

```
Durch die Unterstützung eines Kinder- und Familienzentrums wird sie mit
Obst oder <de deID="de224" deType="nom">getragenen Hosen</de> versorgt -
an Vitaminen und <de deID="de226" deType="nom">Markenklamotten</de>
fehlt es Maik also nicht.

<bridgingLink relType="bridging" phorIDRef="de226"
  antecedentIDRefs="de224"/>
```

Weitere Beispiele:

die türkische Politik – Parlamentswahlen  
Einfamilienhaus – die beiden Bewohner

## **6 Annotationsprozess**

Bei der Annotation anaphorischer Relationen soll zunächst entschieden werden, ob eine gegebene Diskursentität (DE) im Text anaphorisch verwendet wird: DEs indefiniter Nominalphrasen sind nicht anaphorisch, diejenigen definiter NPs sollen im Projekt A2 vollständig einem Antezedens zugeordnet werden (sofern möglich). Für jede anaphorisch verwendete Diskursentität

muss danach entschieden werden, welches das korrekte Antezedens ist. Innerhalb einer Kospezifikationskette soll immer *der letzte nicht-pronominale Kandidat* ausgewählt werden.

Das Vorgehen zur Auswahl der Relation zwischen Anapher und Antezedens ist in Abbildung 3 auf der nächsten Seite dargestellt: Zunächst soll entschieden werden, ob es sich um eine Kospezifikationsrelation (Element `cospecLink`) oder um eine Bridging-Relation (Element `bridgingLink`) handelt. Kospezifikation gilt dann, wenn Anapher und Antezedens das selbe Objekt beschreiben; Bridging gilt dann, wenn Anapher und Antezedens nicht das selbe Objekt beschreiben, jedoch miteinander durch die in Abschnitt 5.2.3 aufgeführten sekundären Relationen miteinander verbunden sind.

Im Fall von Kospezifikation soll wie in der Abbildung 3 dargestellt entschieden werden, welche Unterrelation zutrifft, zunächst wird *ident* überprüft, dann *propName* usw. Die Sekundärrelation *paraphrase* ist am wenigsten restringiert und sollte dann ausgewählt werden, wenn keine der anderen zutrifft.

Im Fall von Bridging wird ähnlich vorgegangen: Zunächst wird *poss* überprüft, dann *meronym/holonym*, danach *hasMember/setMember*. Die Sekundärrelation *bridging* ist am wenigsten restringiert und soll dann ausgewählt werden, wenn keine der anderen zutrifft.

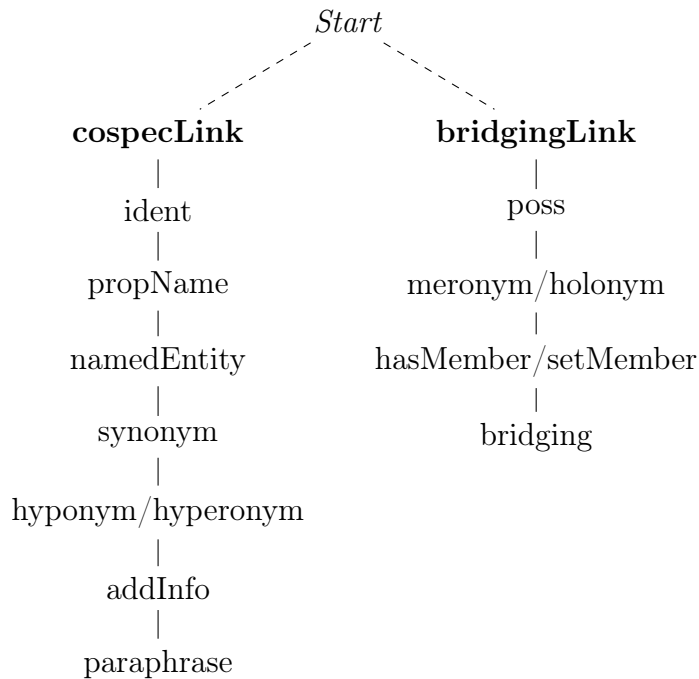


Abbildung 3: Entscheidungen im Annotationsverlauf

## Anhang A: DTD zum Annotationsschema

Listing 42: DTD zum Annotationsschema

```

1 <!ELEMENT txt_text (cnx-pi_analysis)>
2 <!ATTLIST txt_text lang (de | en | fr | es) #REQUIRED>
3
4 <!ELEMENT cnx-pi_analysis (txt_para | txt_list | txt_nontext |
5   standoff)*>
6
7 <!ELEMENT standoff ( semRel? , cnx-pi_token_ref+ ) >
8
9 <!ELEMENT semRel (cospecLink | bridgingLink)+>
10
11 <!ELEMENT cospecLink EMPTY>
12 <!ATTLIST cospecLink
13   id ID #IMPLIED
14   relType (ident | propName | synonym | hyperonym | hyponym |

```

```

15  paraphrase | addInfo | isA | identAdv | abstrProp |
16  abstrProjProp | abstrEvType | abstrEv | abstrCluster ) #REQUIRED
17  phorIDRef IDREF #REQUIRED
18  antecedentIDRefs IDREFS #REQUIRED
19  comment CDATA #IMPLIED>
20
21 <!ELEMENT bridgingLink EMPTY>
22 <!ATTLIST bridgingLink
23  id ID #IMPLIED
24  relType (poss | setMember | hasMember | meronym | holonym |
25  bridging) #REQUIRED
26  phorIDRef IDREF #REQUIRED
27  antecedentIDRefs IDREFS #REQUIRED
28  comment CDATA #IMPLIED>
29
30 <!ELEMENT txt_para (cnx-pi_token | de | cnx-pi_sentence | txt_inline)*>
31 <!ATTLIST txt_para
32  lang (de | en | fr | es) #IMPLIED
33  skip (yes | no) "no"
34  type (caption | ling_example | other) #IMPLIED>
35
36 <!ELEMENT cnx-pi_sentence (cnx-pi_token | de | txt_inline)* >
37 <!ATTLIST cnx-pi_sentence id ID #REQUIRED
38  auto (yes | no) "yes">
39
40 <!ELEMENT cnx-pi_token (#PCDATA) >
41 <!ATTLIST cnx-pi_token
42  ref IDREF #REQUIRED>
43
44 <!ELEMENT cnx-pi_token_ref (#PCDATA) >
45 <!ATTLIST cnx-pi_token_ref
46  auto (yes|no) "yes"
47  cnx-output (correct|incorrect) "correct"
48  id ID #REQUIRED
49  text CDATA #REQUIRED
50  lemma CDATA #IMPLIED
51  dependHead IDREF #IMPLIED
52  dependValue CDATA #IMPLIED
53  syntax CDATA #IMPLIED
54  pos CDATA #IMPLIED
55  heur (yes|no) "no"
56  morpho CDATA #IMPLIED >
57
58 <!ELEMENT de (cnx-pi_token | de | cnx-pi_sentence)*>

```

```
59 <!ATTLIST de
60   deID ID #REQUIRED
61   headRef IDREF #REQUIRED
62   deType (nom | namedEntity | prop | projProp | evType | ev |
63     state | cluster | adv) "nom"
64   deAuto (1|0) "1"
65   deAnaphoric CDATA #IMPLIED>
66
67 <!ELEMENT txt_list (txt_item+)>
68 <!ATTLIST txt_list type (ordered | unordered) #REQUIRED>
69
70 <!ELEMENT txt_item (txt_list | txt_nontext | txt_para)+>
71
72 <!ELEMENT txt_nontext (cnx-pi_token | de | txt_para | txt_nontext)*>
73 <!ATTLIST txt_nontext
74   type (image | table | other) #REQUIRED>
75
76 <!ELEMENT txt_inline (cnx-pi_token+)>
77 <!ATTLIST txt_inline
78   lang CDATA #IMPLIED>
```

## Literatur

- [Bußmann 1990] BUSSMANN, Hadumod: *Lexikon der Sprachwissenschaft*. 2. Auflage. Kröner, 1990
- [Clark 1977] CLARK, Herbert H.: Bridging. In: JOHNSON-LAIRD, P. N. (Hrsg.) ; WATSON, P. C. (Hrsg.): *Thinking: Readings in cognitive Sciences*. London ; New York : Cambridge University Press, 1977
- [Fligelstone 1992] FLIGELSTONE, S.: Developing a Scheme for Annotating Text to Show Anaphoric Relations. In: LEITNER, G. (Hrsg.): *New Directions in English Language Corpora: Methodology, Results, Software Developments*. Berlin : Mouton de Gruyter, 1992, S. 153–170
- [Garside u. a. 1997] GARSIDE, R. ; FLIGELSTONE, S. ; BOTLEY, S.: Discourse Annotation: Anaphoric Relations in Corpora. In: GARSIDE, R. (Hrsg.) ; LEECH, G. (Hrsg.) ; MCENERY, A. (Hrsg.): *Corpus Annotation: Linguistic Information from Computer Text Corpora*. London : Addison-Wesley Longman, 1997, S. 66–84
- [Hardt 1996] HARDT, D.: Centering in Dynamic Semantics. In: *Proceedings of COLING 1996* Bd. Volume 1: The 16th International Conference on Computational Linguistics, 1996
- [Hirschmann 1997] HIRSCHMANN, L.: MUC-7 Coreference Task Definition (version 3.0). In: HIRSCHMAN, L. (Hrsg.) ; CHINCHOR, N. (Hrsg.): *Proceedings of Message Understanding Conference (MUC-7)*, 1997
- [Holler u. a. 2004] HOLLER, A. ; MAAS, J.-F. ; STORRER, A.: Exploiting coreference annotations for text-to-hypertext conversion. In: *Proceeding of LREC* Bd. II. Lisboa, 2004, S. 651–654
- [Holler 2003] HOLLER, Anke: Ein Annotationsschema für Koreferenzphänomene / TEMIS Dtl. GmbH und Universität Dortmund. 2003. – Forschungsbericht
- [Holler-Feldhaus 2004] HOLLER-FELDHAUS, Anke: Koreferenz in Hypertexten: Anforderungen an die Annotation. In: *Osnabrücker Beiträge zur Sprachtheorie (OBST)* 68 (2004), S. 9–29

- [Kamp u. Reyle 1993] KAMP, Hans ; REYLE, Uwe: *From Discourse to Logic*. Kluwer: Dordrecht, 1993
- [Krasavina u. Chiarcos 2007] KRASAVINA, Olga ; CHIARCOS, Christian: PoCoS - Potsdam Coreference Scheme. In: *Proceedings of the Linguistic Annotation Workshop*. Prague, Czech Republic : Association for Computational Linguistics, June 2007, 156–163
- [Müller u. Strube 2001] MÜLLER, Christoph ; STRUBE, Michael: Annotating anaphoric and bridging relations with MMAX. In: *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue*. Morristown, NJ, USA, 2001, S. 1–6
- [Poesio 2004] POESIO, M.: The MATE/GNOME Scheme for Anaphoric Annotation, Revisited. In: *Proceedings of SIGDIAL*. Boston, April 2004
- [Poesio 1998] POESIO, Massimo: *Coreference*. MATE Deliverable D1.1., 1998
- [Schwarz 2000] SCHWARZ, M.: *Indirekte Anaphern in Texten*. Tübingen : Niemeyer, 2000
- [Sidner 1979] SIDNER, Candace L.: *Towards a computational theory of definite anaphora comprehension in English discourse*, MIT, Diss., 1979
- [Stührenberg u. a. 2007] STÜHRENBERG, Maik ; GOECKE, Daniela ; DIEWALD, Nils ; CRAMER, Irene ; MEHLER, Alexander: Web-based Annotation of Anaphoric Relations and Lexical Chains. In: *Proceedings of The Linguistic Annotation Workshop (LAW)*. Prague, Czech Republic : Association for Computational Linguistics, Juni 2007, 140–147
- [Thompson u. McKelvie 1997] THOMPSON, Henry S. ; MCKELVIE, David: Hyperlink semantics for standoff markup of read-only documents. In: *Proceedings of SGML Europe '97: The next decade – Pushing the Envelope*. Barcelona, 1997, 227–229
- [Webber 1988] WEBBER, Bonnie: Discourse deixis: Reference to discourse segments. In: *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistic (ACL-88)*. State University of New York at Buffalo, June 27–30 1988, S. 113–122