

# Spezifikation für ein Annotationsschema für Koreferenzphänomene im Hinblick auf Hypertextualisierungsstrategien

Projekt  
**Hypertextualisierung auf  
textgrammatischer Grundlage**  
([www.hytex.info](http://www.hytex.info))

Anke Holler  
(Kooperationspartnerin TEMIS GmbH, Heidelberg)

2003

## 1. Einführung

Im Folgenden wird ein Vorschlag für ein Schema zur Annotation von Koreferenzphänomenen unterbreitet. Die Annotation soll als eine von verschiedenen textgrammatischen Annotationsschichten im Projekt „Hytex (Hypertextualisierung auf textgrammatischer Grundlage)“ verwendet werden, um die angestrebten textgrammatischen Hypertextualisierungsstrategien zu unterstützen. Da das im Projekt verwendete Korpus in einer seiner Annotationsschichten mit dem Werkzeug *KaRoPars* (v.0.36) um morphosyntaktische Annotationen (POS, Chunks, topologische Felder) angereichert wurde<sup>1</sup>, geht diese Spezifikation davon aus, dass auf Ebene der sprachlichen Ausdrücke bereits eine Annotation vorhanden ist, auf welcher die Koreferenz-Annotation als weitere Schicht aufsetzen kann. Inwieweit die bereits vorhandenen Annotationen tatsächlich alle notwendigen Markierungen liefern, muss noch im Detail überprüft werden; ggf. müssen die Dokumente entsprechend nachannotiert werden.

*Morphologische Ebene:* Kasus, Genus, Numerus

*Wortebene:*

- Phorisches Pronomen
  - Personalpronomen, 3.Prs.
- Deiktisches Pronomen
  - Demonstrativpronomen
  - Relativpronomen im nichtrestr.RS
  - Personalpron., 1./2.Prs.
- Possessivpronomen
- 0-Pronomen (Kontrolle)

*Phrasale Ebene:*

- Einfache Nominalphrasen
  - Eigenname/Lit.-Angabe
  - Demonstrative NP
  - Definite NP
  - Indefinite NP
  - Quantifizierte NP
  - Plurale NP
  - Possessive NP
  - Koordinierte NP
  - NP mit Apposition
  - NP mit Genitivattribut
- Komplexe NP mit restr. Relativsatz

Im vorliegenden Dokument wird die Spezifikation der (diskurs-) semantischen Relationen im Vordergrund stehen. Grundlage dafür ist das von (Poesio 1998) im Rahmen von MATE vorgeschlagene Annotationsschema, das jedoch angepasst und präzisiert wird.

Die hier verfolgte Grundidee besteht darin, die Unterscheidung zwischen Kospezifikation (Sidner 1979) und Koreferenz in der Annotation abzubilden. Während zwei Ausdrücke nur

<sup>1</sup> Vgl. Ule / Müller (2004) sowie das „Stylebook for the Tübingen Partially Parsed Corpus of Written German (TüPP-D/Z)“ (Müller 2004).

dann koreferieren, wenn sie auf dieselbe Entität in der Welt verweisen, genügt für Kospezifikation, dass ein Ausdruck einen vorangegangenen sprachlich wieder aufgreift.

In der Annotation wird das dadurch umgesetzt, dass Kospezifikation als Beziehung zwischen Diskursentitäten markiert wird, während Koreferenz als Beziehung zwischen einer Diskursentität und einem Topic in der TopicMap ausgezeichnet wird.

Eine TopicMap ist ihrerseits eine Repräsentation der Welt, denn durch die TopicMap wird die Welt bzw. ein Ausschnitt der Welt modelliert.

Diese zusätzliche Ebene erlaubt es, den Bezug zu Entitäten in der Welt auch tatsächlich zu annotieren. Man kann somit sagen, dass Kospezifikation eine horizontale Annotationsbeziehung ist, denn sie wird auf 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).

In Abb. 1 wird das zugrundegelegte Text-Welt-Modell veranschaulicht:

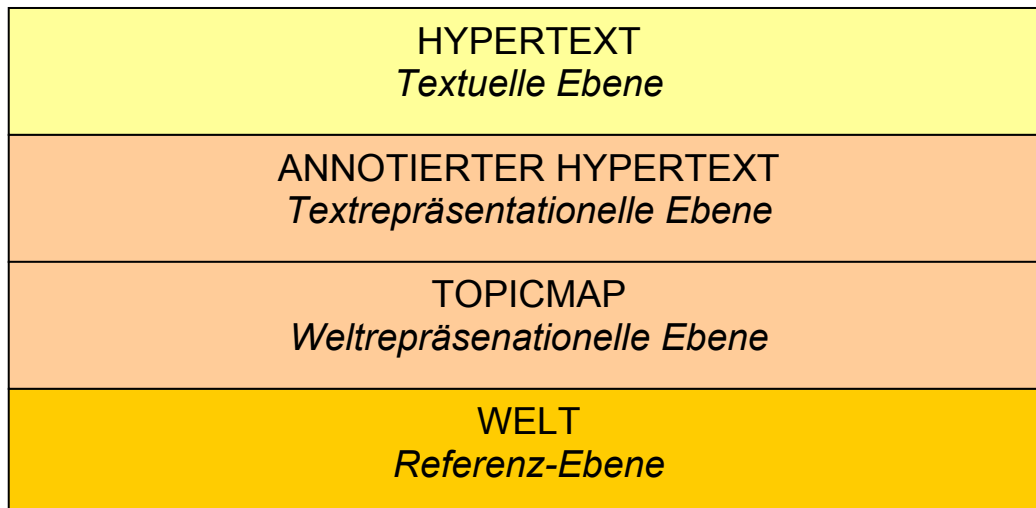


Abb. 1

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 oder der vorgestellten, konzeptuellen Welt Bezug zu nehmen. Dass hier Koreferenz als Beziehung zu einem Modell der Welt, dem als TopicMap realisiertem 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.

Eine vergleichbare Unterscheidung ist bei MATE nicht vorhanden. Unter MATE wird letztlich nur Kospezifikation annotiert.<sup>2</sup>

<sup>2</sup> Dadurch, dass die Tags unter MATE <coref:x> heißen, wird dieser Punkt verdeckt. Solange nur horizontal auf Textebene annotiert wird, kann Koreferenz nicht adäquat markiert werden, denn

## 2. Generelles Vorgehen

Die Annotation soll mehrstufig durchgeführt werden. In einem ersten Schritt werden alle vorkommenden relevanten Diskursentitäten mit einer Diskurs-ID versehen. In einem zweiten Schritt werden die für die Hypertextualisierung interessanten semantischen Beziehungen annotiert, die entweder zwischen Diskursentitäten (Kospezifikation) oder zwischen Diskursentitäten und Topics (Koreferenz) etabliert sein können.

### Schritt 1: Markierung der relevanten Diskursentitäten

Es wird ein Element `<de>` definiert, um relevante Diskursentitäten zu markieren. `<de>` trägt das obligatorische Attribut „deID“, dessen Wert vom Typ natürliche Zahl ist, und ein ebenfalls obligatorisches Attribut „deType“, das die Sorte der jeweiligen Diskursentität angibt. Als Werte sind `nom`, `prop`, `projprop`, `cluster`, `state`, `ev(ent)`, `ev(ent)-type` vorgesehen:

```
<de deID = WERT deType = WERT> SPRACHLICHER AUSDRUCK </de>
```

Für die ID der Diskursentität wird als Format „de\_X\_n“ 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<sup>3</sup>, und n eine natürliche Zahl. Man erhält dann z. B. als mögliche IDs „de\_n\_1“ und „de\_e\_1“.

Alternativ kann die ID natürlich auch für alle Diskursentitäten unabhängig von ihrer Sorte fortlaufend generiert werden.

Als relevante Diskursentität gelten alle sprachlichen Ausdrücke, die einen Diskursreferenten im Sinne von (Kamp&Reyle 1993) in die Diskurs- bzw. Textrepräsentation einführen können.

Es wird empfohlen, in einem ersten Schritt die nominalen Diskursentitäten zu markieren, weil diese im Text leicht zu identifizieren sind und danach die übrigen Entitäten.

### Schritt 2: Annotation der (diskurs-)semantischen Beziehungen

Zur Markierung der (diskurs-)semantischen Beziehungen wird das Element `<semRel>` eingeführt, wodurch analog zu TEI und MATE die Annotation von Kospezifikation bzw. Koreferenz getrennt von der Annotation der Diskursentitäten vorgenommen wird. Dies hat den Vorteil, dass eine Diskursentität zu mehr als einer anderen Entität 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>4</sup>

`<semRel>` kann am Ende eines `<segment-cluster>` eingeführt werden.

---

Koreferenz ist eine vertikale Beziehung. Tags zur Markierung einer vertikalen Beziehung sind bei MATE aber nicht vorgesehen.

- 3 Wenn man ein solches Format für die ID annimmt, kann X aus dem Wert des deType-Attributs abgeleitet werden.
- 4 Dieses Vorgehen lässt außerdem Raum, um das Schema auszubauen und zum Beispiel Objekte der visuellen Situation als Bezugsgröße anzugeben.

<semRel> weist zur Annotation von Koreferenz bzw. zur Annotation von Kospezifikation die zwei eingebetteten Tags <corefLink> und <cospecLink> auf.<sup>5</sup>

#### a) Annotation von Koreferenz

Koreferenz wird als Beziehung zwischen einer referentiellen Diskursentität und einem Topic der TopicMap, dem sog. Referenzanker, interpretiert. Dazu wird ein Element <corefLink> eingeführt.

Für dieses Element werden die zwei Attribute „refID“ mit einer *deID*<sup>6</sup> und „tmRefAnchor“ mit einer *TopicMapID* als Wert vergeben:

```
<corefLink refID = WERT tmRefAnchor = WERT </corefLink>
```

Da die Situation auftreten kann, dass ein Ausdruck im Text eindeutig referentiell gebraucht wird, aber (noch) kein Eintrag in der TopicMap vorhanden ist, muss für den Wert des Attributs „tmRefAnchor“ neben *TopicMapID* auch *unknown* vorgesehen werden.

#### b) 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 entweder aus der Menge der möglichen Relationen (vgl. 3.3) stammt oder als *unknown* angegeben wird, falls er nicht spezifiziert bzw. bekannt ist. Zum zweiten wird das Attribut „phorID“ mit einer *deID*<sup>7</sup> 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 ein weiteres Element eingebettet, das <linkTarget> heißen soll. Für dieses Element wird das Attribut „href“ mit einer *deID*<sup>8</sup> als Wert definiert.

```
<cospecLink relType = WERT phorID = WERT>
<linkTarget href = WERT/>
</cospecLink>
```

Bei ambigen Bezugsgrößen wird das Element <linkTarget> entsprechend vervielfacht.

#### Anmerkung:

Bindung wird NICHT annotiert, denn dies ist eine syntaktisch lizenzierte und daher immer satzgebundene Relation. Sie wird zwischen einer quantifizierten NP und einem anaphorischen Pronomen etabliert:

5 <semRel> wurde der Übersichtlichkeit halber eingeführt, kann also wenn nötig auch entfallen. In diesem Fall wären <corefLink> und <cospecLink> nicht mehr eingebettet.

6 Unter MATE ist der Wert des „href“-Attributs der gesamte Tag der Diskursentität, also <de>. Wenn gewünscht, kann das hier vorgestellte Schema entsprechend abgewandelt werden.

7 Siehe Fußnote 6. Wenn der Wert für das href-Attribut abgewandelt wird, sollte dies parallel sowohl für das <corefLink>-Element als auch das <cospecLink>-Element geschehen.

8 Siehe Fußnoten 6 und 7.

Jeder Autor sollte die Verlinkung in seinem Hypertext selbst vornehmen.

Es gibt weitere Phänomenbereiche, in denen die sog. Bound-Variable-Interpretation möglich ist, wie z.B. das *Sloppy Identity*-Phänomen. Diese werden ebenfalls nicht annotiert.

Otto stellte seinen Hypertext fertig, was Max auch tat.

### 3. Annotation der sprachlichen Phänomene

Wie bereits beschrieben, umfasst die Annotation der sprachlichen Phänomene zwei Aspekte: die Annotation der relevanten Diskursentitäten und die Annotation der diskurssemantischen Beziehungen zwischen einem Ausdruck und seinem Bezugsausdruck. In den nachfolgenden Abschnitten werden – zum Teil anhand von Beispielen – die jeweils vorgesehenen Werte eingeführt.

#### 3.1. Annotation der relevanten Diskursentitäten

##### 3.1.1 deType = „nom“

Die nominalen Diskursentitäten können automatisch identifiziert werden. Es wird vorgeschlagen, alle im Zuge der *KaRoPars*-Annotation als NP markierten Einheiten mit einer ID zu versehen. Damit bedeutet „relevante nominale Diskursentität“ zunächst „als NP unter *KaRoPars* ausgezeichnet“. Ein Konsequenz davon ist, dass wenn *KaRoPars* verschachtelte NP-Strukturen erzeugt, auch verschachtelte Diskursentitäten mit entsprechenden IDs annotiert werden müssen. Falls sich herausstellen sollte, dass mit diesem Verfahren übergeneriert wird oder dass wichtige Diskursentitäten unmarkiert bleiben, kann das Ergebnis manuell nachbearbeitet werden, um Markierungen zu löschen oder weitere Diskurs-IDs an entsprechende Ausdrücke zu vergeben.

##### 3.1.2 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.

##### 3.1.3 deType = „projProp“

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

##### 3.1.4 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.

#### 3.1.4.1 deType = „ev“

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

#### 3.1.4.2 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.

#### 3.1.4.3 deType = „cluster“

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

### 3.1.5 Annotation 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 repräsentiert sind.

Eigennamen und Literaturangaben sind typischerweise verankert:

```
Das von <de deID="de_n_1" deType="nom"> Kühlen 1991 </de/> skizzierte
Grundmodell eines Hypertextsystems orientiert sich am Vorbild von Da-
tenbankmanagementsystemen.
<corefLink refID="de_n_1" tmRefAnchor="tm_1" </corefLink>
```

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.

```
<de deID="de_n_2" deType="nom"> Austin </de/> liegt in Texas.
<corefLink refID="de_n_2" tmRefAnchor="tm_2" </corefLink>

<de deID="de_n_3" deType="nom"> Der Austin </de/> ist beliebt.
<corefLink refID="de_n_3" tmRefAnchor="tm_3" </corefLink>
```

### 3.1.6 Annotation der unter Kospezifikation vorkommenden (diskurs-)semantischen Relationen

Der Großteil der in den Fachtexten vorkommenden Beziehungen ist vom Typ „Kospezifikation“. Kospezifikation beruht auf einer phorischen Beziehung zwischen sprachlichen Ausdrücken, wobei der phorische Ausdruck ein Pronomen, eine NP oder ein Eigenname sein kann.

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

Kospezifikation kann mit direkter Anaphorik oder indirekter Anaphorik einhergehen.

„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 2000:49)

### 3.1.6.1 Direkte Anaphorik

#### 3.1.6.1.1 relType=„ident“

Dieser Wert wird erstens vergeben, wenn sich ein Pronomen auf eine NP bezieht.

```
<de deID="de_n_4" deType="nom"> Ein Link <de/> ist im Text meist farbig
markiert. <de deID="de_n_5" deType="nom"> Er <de/> ist dadurch
gut sichtbar.
<corefSpec relType="ident", phorID="de_n_5" <linkTarget
href="de_n_4"/> </corefSpec>
```

Dieser Wert wird zweitens vergeben, wenn sich eine NP auf eine rekurrente NP bezieht.

```
<de deID="de_n_6" deType="nom"> Ein Link <de/> ist im Text meist farbig
markiert. <de deID="de_n_7" deType="nom"> Der Link <de/> ist da-
durch gut sichtbar.
<corefSpec relType="ident", phorID="de_n_7" <linkTarget
href="de_n_6"/> </corefSpec>
```

#### 3.1.6.1.2 relType=„synonym“

Die kospezifizierte NP ist synonym zu ihrer Bezugsgröße.

```
<de deID="de_n_8" deType="nom"> Ein Link <de/> ist im Text meist farbig
markiert. <de deID="de_n_9" deType="nom"> Der Verweis <de/> ist
dadurch gut sichtbar.
<corefSpec relType="synonym", phorID="de_n_9" <linkTarget
href="de_n_8"/> </corefSpec>
```

#### 3.1.6.1.3 relType=„hyperonym“

Die kospezifizierte NP ist ein Hyperonym ihrer Bezugsgröße.

```
<de deID="de_n_10" deType="nom"> HTML-Editoren <de/> erleichtern die
Erstellung von Webseiten wesentlich. Deswegen sind <de deID="de_n_11"
deType="nom"> diese Programme <de/> bei den Anwendern auch sehr be-
liebt.
<corefSpec relType="hyperonym", phorID="de_n_11" <linkTarget
href="de_n_10"/> </corefSpec>
```

### 3.1.6.1.4 relType=„hyponym“

Die kospezifizierte NP ist ein Hyponym ihrer Bezugsgröße.

Die passende `<de deID="de_n_12" deType="nom"> 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="de_n_13" deType="nom"> Spracherkennungssoftware </de>` oder `<de deID="de_n_14" deType="nom"> Maschinellen-Übersetzungsprogrammen </de>` aus dem Supermarkt. Leistungsfähige Programme haben nämlich ihren Preis.

```
<corefSpec relType="hyponym", phorID="de_n_13" <linkTarget
href="de_n_12"/> </corefSpec>
<corefSpec relType="hyponym", phorID="de_n_14" <linkTarget
href="de_n_12"/> </corefSpec>
```

### 3.1.6.1.5 relType=„propName“

Die kospezifizierte NP ist ein Eigenname, der sich auf eine nominale Bezugsgröße bezieht.<sup>9</sup>

Mit `<de deID="de_n_15" deType="nom"> diesem XML-Editor </de>` lässt es sich sehr gut arbeiten. `<de deID="de_n_16" deType="nom"> XML Spy </de>` ist daher trotz seines Preises sehr beliebt.

```
<corefSpec relType="propName", phorID="de_n_16" <linkTarget
href="de_n_15"/> </corefSpec>
```

### 3.1.6.1.6 relType=„paraphrase“

Die kospezifizierte NP ist eine Paraphrase ihrer Bezugsgröße.

`<de deID="de_n_17" deType="nom"> Der HTML-Editor </de>` wird ständig weiterentwickelt, denn mit `<de deID="de_n_18" deType="nom"> Werkzeugen zum Erstellen von Webseiten </de>` lässt sich immer noch Geld verdienen.

```
<corefSpec relType="paraphrase", phorID="de_n_18" <linkTarget
href="de_n_17"/> </corefSpec>
```

### 3.1.6.1.7 relType=„addInfo“

Die kospezifizierte NP führt neue oder zusätzliche Information ein.

`<de deID="de_n_19" deType="nom"> Der HTML-Editor </de>` ist ständig weiterentwickelt worden. `<de deID="de_n_20" deType="nom"> Die Version 5.0 </de>` kommt gerade auf den Markt.

```
<corefSpec relType="addInfo", phorID="de_n_20" <linkTarget
href="de_n_19"/> </corefSpec>
```

### 3.1.6.1.8 relType=„is-a“

<sup>9</sup> Der Eigenname selbst ist mit der TopicMap verlinkt.

Wenn es aus praktischen Erwägungen nicht möglich ist, die Relationen Synonymie, Hyperonymie, Hyponymie, Paraphrase und AddInfo bei der Annotation zu unterscheiden, ist denkbar, die Relationen zusammenzufassen und die entsprechenden relType-Werte einheitlich durch den Wert „is-a“ zu ersetzen.

### 3.1.6.1.9 relType=„abstrProp“

Der kospezifizierte Ausdruck bezieht sich auf eine Proposition.

```
<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.
<corefSpec relType="abstrProp", phorID="de_n_21" <linkTarget
href="de_p_1"/> </corefSpec>
```

### 3.1.6.1.10 relType=„abstrProjProp“

Der kospezifizierte Ausdruck bezieht sich auf eine projektive Proposition.

```
<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. Aller-
dings ist Google in Sachen Trefferquote nach wie vor ungeschlagen.
<corefSpec relType="abstrProjProp", phorID="de_n_22" <linkTarget
href="de_pp_1"/> </corefSpec>
```

### 3.1.6.1.11 relType=„abstrEvType“

Der kospezifizierte Ausdruck bezieht sich auf einen Ereignistyp.

```
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.
<corefSpec relType="abstrEvType", phorID="de_n_23" <linkTarget
href="de_et_1"/> </corefSpec>
```

### 3.1.6.1.12 relType=„abstrEv“

Der kospezifizierte Ausdruck bezieht sich auf ein Ereignis.

```
Eine Studentin hat den gesamten Korpus manuell <de deID="de_e_1" de-
Type= "ev"> annotiert </de/>. <de deID="de_n_24" deType= "nom"> Das
</de/> dauerte zwei Monate.
<corefSpec relType="abstrEv", phorID="de_n_24" <linkTarget
href="de_e_1"/> </corefSpec>
```

Dass in Fachtexten tatsächlich Bezüge auf Ereignisse vorkommen, ist allerdings eher unwahrscheinlich.

### 3.1.6.1.13 relType=„abstrCluster“

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

```
<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 nach der Anzahl der [...].
<corefSpec relType="abstrCluster", phorID="de_n_25" <linkTarget href="de_c_1"/> </corefSpec>
```

### 3.1.6.1.14 relType=„unknown“

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

### 3.1.6.1.15 Indirekte Anaphorik

Mit Schwarz (2000) lassen sich fünf verschiedene Ausprägungen indirekter Anaphorik annehmen: Semantische Rollenzuweisung, Nominalsemantische Relation Meronymie, Nominalsemantische Relation Hyperonymie, Schema-basiert indirekte Anaphorik und Inferenzbasierte indirekte Anaphorik. Es scheint aber unrealistisch, im vorliegenden Korpus diese Unterscheidung treffen zu können. Deswegen wird vorgeschlagen, alle fünf Relationen trotz ihrer inhaltlichen Unterschiede auf den „relType“-Wert *bridging* abzubilden.<sup>10</sup>

```
Hans hat <de deID="de_n_26" deType="nom"> einen neuen Computer </de> gekauft. <de deID="de_n_27" deType="nom"> Der Lüfter </de> war unerträglich laut.
<corefSpec relType="bridging", phorID="de_n_27" <linkTarget href="de_n_26"/> </corefSpec>
```

### 3.1.6.2 Possession

Ein dritter Fall von Kospezifikation ist die Possession. Um diese Relation zu beschreiben wird der „relType“-Wert *poss* eingeführt.

```
<de deID="de_n_28" deType="nom"> Hans </de> hat sich einen neuen Browser installiert. <de deID="de_n_29" deType="nom"> Sein Bruder </de> hat ihm nämlich geraten, Mozilla statt dem Internet Explorer von Microsoft zu verwenden.
<corefSpec relType="poss", phorID="de_n_29" <linkTarget href="de_n_28"/> </corefSpec>
```

Das vorliegende Dokument soll als Vorschlag für ein Anotationsschema verstanden werden. Erfahrungen aus der praktisch durchgeführten Annotation können selbstredend zu entsprechenden Änderungen oder Erweiterungen des Schemas führen.

<sup>10</sup> Der Begriff *Bridging Reference* geht auf Clark (1977) zurück.

**Referenzen:**

- Clark, H.H. (1977): Bridging. In P.N. Johnson-Laird und P.C. Watson (Hrsg.) Thinking: Readings in cognitive Sciences. Cambridge University Press, London und New York.
- Kamp, H. und U. Reyle (1993): From Discourse to Logic. Kluwer: Dordrecht.
- Müller, F. H. (2004): Stylebook for the Tübingen Partially Parsed Corpus of Written German (TüPP-D/Z). Annotations-Manual (PDF). Online:  
<http://www.sfs.uni-tuebingen.de/~fhm/Biblio/stylebook-04.pdf>
- Poesio, M. (1998): Coreference. MATE Deliverable D1.1.
- Schwarz, M. (2000): Indirekte Anaphern in Texten. Niemeyer: Tübingen.
- Sidner, C.L. (1979): Towards a computational theory of definite anaphora comprehension in English discourse. Ph.D. thesis. MIT.
- Ule, T. / Müller F.H. (2004): KaRoPars: Ein System zur linguistischen Annotation großer Textcorpora des Deutschen. In Mehler / Lobin (Hg.): Automatische Textanalyse. Systeme und Methoden zur Annotation und Analyse natürlichsprachiger Texte. VS: Wiesbaden, 185-202.