

INTERNAL REPORT OF THE DFG-RESEARCH GROUP 437
“TEXT-TECHNOLOGICAL MODELLING OF INFORMATION”

Project A2 “Sekimo”

Coreference, Cospecification and Bridging: Annotation Scheme

Daniela Goecke
Maik Stührenberg
Fakultät für Linguistik
und Literaturwissenschaft
Universität Bielefeld

daniela.goecke@uni-bielefeld.de
maik.stuehrenberg@uni-bielefeld.de

Anke Holler
Seminar für Deutsche Philologie
Georg-August-Universität Göttingen
anke.holler@phil.uni-goettingen.de

August 27, 2008

Contents

1	Introduction	3
2	Presuppositions on the Annotation Scheme	3
2.1	Coreference and Cospecification	3
2.2	Comparison with Existing Annotation Schemes	5
3	General Procedure	6
3.1	Marking the Relevant Discourse Entities	7
3.2	Types of Discourse Entities	8
3.3	Annotation in Serengeti	10
4	Annotation of (Discourse-)semantic Relations	10
4.1	Annotation of Coreference	11
4.2	Annotation of Cospecification	11
4.3	Annotation of Indirect Anaphora	13
5	Overview of Semantic Relations	14
5.1	Relations of Coreference	15
5.2	Relations of Cospecification and Indirect Anaphora	16
5.2.1	Direct Anaphora (Cospecification)	16
5.2.2	Cospecification with Propositions or Events as Reference Values	24
5.2.3	Indirect Anaphora	26
6	Annotation process	30
	References	34

1 Introduction

These guidelines describe the annotation of anaphoric relations in the project A2 „Sekimo“ of the research group 437 “Text-Technological Modelling of Information”. This annotation scheme is based on suggestions for the annotation of coreference phenomena, which was developed for the annotation of coreference structures in the project B1 „HyTex“ (cf. Holler, 2003, Holler et al., 2004, Holler-Feldhaus, 2004).¹

This scheme was developed further for the project A2 in order to enable annotation of indirect anaphora (bridging relations, cf. Clark, 1977). In what follows the main focus will be on the annotation in the project A2. The document includes the following sections: firstly, basic ideas about the annotation scheme will be expressed. At this point we will look in particular at the differentiation between the relations *coreference*, *cospecification* and *bridging*. The annotation of semantic relations will be described in Section 4, and an overview of semantic relations will be given in Section 5.

2 Presuppositions on the Annotation Scheme

2.1 Coreference and Cospecification

The basic idea of this annotation scheme is to explore the differences between cospecification (Sidner, 1979) and coreference in the annotation. While two coreferent expressions have to refer to the same entity in the world, the criterion for cospecification is already met when one expression picks up another linguistically. This differentiation already occurred in the first version of this scheme in the project B1. The scheme extension for the project A2 considers additionally indirect anaphora²; here the antecedent to an anaphor is not explicitly realized, but has to be inferred from context.

In the annotation this differentiation is made clear through the marking of cospecification and indirect anaphoric relations between entities (projects

¹ Basis for the design of the project B1 was the annotation schema suggested by Poesio (1998) within the framework of MATE, which was extended and specified for the project B1.

² In the first version of the scheme for the project B1 the possibility of annotating indirect anaphora was already given by the introduction of the relation type bridging. This relation type was not further elaborated on, as bridging relations were not annotated in the project B1.

A2 and B1), while coreference is described as relationship between an entity and a topic using a TopicMap (project B1). A TopicMap itself is a representation of the world, since it models the world, or rather a part of it. As discussed in (Holler-Feldhaus, 2004), this extra level de facto enables the annotation of relations concerning entities in the world. Using the just outlined account, cospecification can be understood as a horizontal annotation relation, because it is marked on the same level (the level of text representation); coreference however represents a vertical relation, a relation between different levels, here, between the text and the world representational level (TopicMap). Figure 1 shows the underlying text-world-model.

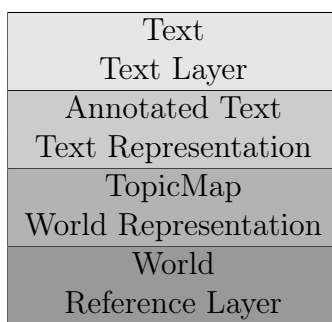


Figure 1: Text-World-Model

It is important to note at this point that coreference is formally semantically a sign-entity-relation and it is interpreted as reference identity in the world. Coreference exists, if two or more expressions are used by discourse participants to refer to *the same referent in the real world*. In this case, for purely practical reasons, coreference is represented as a relation to a model of the world, realized as a TopicMap. This enables coreference to be annotated, and the differentiation of cospecification and coreference to be shown directly in the annotation.

The annotation of cospecification as well as indirect anaphora are based on the following interpretation: cospecification and indirect anaphora are annotated within the level of text representation: one expression takes up a preceding expression linguistically. In cospecification as well as – even more clearly – in indirect anaphora, the semantic interpretation exists besides the textual level: linguistic expressions introduce discourse referents (in

this case *discourse entities*) or they refer to discourse referents that were already introduced (cf. Webber, 1988); semantic relations can exist between discourse referents. Figure 2 shows this connection.

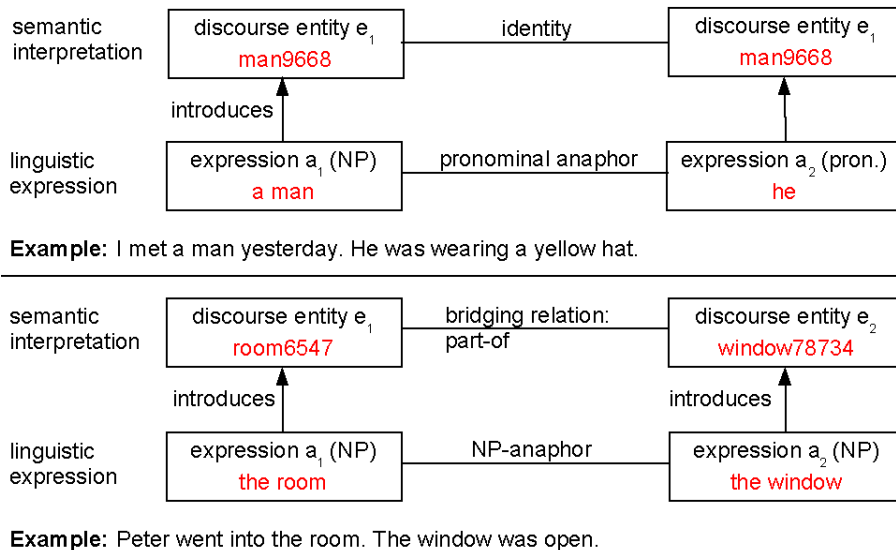


Figure 2: Differentiation between the text level and the level of semantic interpretation

The anaphoric relation between linguistic expressions is annotated as *cospecLink* or *bridgingLink* in a XML-representation (cf. Sections 4.2 on page 11 and 4.3 on page 13) and the (discourse-)semantic interpretation is represented by the relation type. The type of anaphoric relation (pronominal anaphora vs NP-anaphora) is not explicitly annotated, but rather coded implicitly in the morpho-syntactic information of linguistic expressions, which introduces discourse entities.

2.2 Comparison with Existing Annotation Schemes

In recent years there were several different annotation schemes developed for anaphoric relations, among other things the UCREL-Scheme (Fligelstone, 1992, Garside et al., 1997), the SGML-based MUC-Annotation scheme

(Hirschmann, 1997), as well as the XML-based MATE/GNOME-Scheme (Poesio, 1998, 2004).

Similar to the MATE/GNOME-Scheme this annotation scheme is XML-based and uses standoff annotations (Thompson and McKelvie, 1997). An explicit differentiation of coreference and cospecification does not exist in Poesio (1998); in MATE, only cospecification is annotated.³

In Poesio (2004) the difference between coreference and anaphoric relations is pointed out, however the distinction is not realized using different elements in the annotation scheme (cf. Poesio, 2004, Section 2.1). These aspects are discussed more detailed in (Holler-Feldhaus, 2004).

Another scheme is presented by Krasavina and Chiarcos (2007). That scheme focuses on the annotation of multilingual corpora (*cross-linguistic anaphoric annotation*, cf. Krasavina and Chiarcos, 2007, p. 156); an explicit distinction of coreference and cospecification is not assumed. In this respect the annotation scheme presented here, is a further development.

3 General Procedure

The basic corpus building data are the German corpus data of the project C1 (47 scientific articles on linguistics) as well as 21 German newspaper articles that were collected in the project A2. The corpus is administered by means of the version control system SUBVERSION (SVN). If not otherwise stated, all of the following descriptions are related to the project A2; for preparation of the corpus for the project B1 see Holler (2003).

For the annotation in the project A2 a preprocessing using the functional dependent parser-tagger MACHINESE SYNTAX by the company Connexor Oy is carried out. The morpho-syntactic information of the MACHINESE SYNTAX is annotated on word level and XML is a possible output format.

The annotation is a multilevel annotation. Firstly, all occurring relevant discourse entities are detected and provided with a discourse ID (cf. Section 3.1 on the next page). Secondly, the relevant semantic relations are annotated. Those are either established between discourse entities

³ Thus the tags in MATE are called `coref:x`, this point is not apparent. As long as annotation is solemnly done endophorically on text level, (*horizontal* in respect to Figure 1 on page 4), coreference cannot be adequately marked, because it is a text-world-relation (*vertical* in relation to Figure 1). Tags for marking a vertical relation are not intended in MATE.

(cospecification/indirect anaphora) or between discourse entities and topics (coreference) (cf. Section 4 on page 10).

3.1 Marking the Relevant Discourse Entities

The annotation scheme is XML-based, this means that the annotation can be done with any XML-editor (i.e. Oxygen⁴). The annotation scheme is formalized in form of a XML-DTD (cf. appendix 6 on page 31) as well as in form of an XSD. An element `de` is defined in order to mark relevant discourse entities. The element `de` has an obligatory attribute `deID` assigned to it, whose value has to be a *qualified name* (*QName*) and also an obligatory attribute `deType`, which indicates the type of each discourse entity. Possible values for `deType` are given in Listing 1. The obligatory attribute `headRef` has as its value the reference to the head noun of the NP it is based on. The attribute `deAuto` was introduced in order to mark manually corrected discourse entities; the default value `1` is automatically assigned to all detected discourse entities. The optional attribute `deAnaphoric` specifies whether its use is anaphoric for those discourse entities that are introduced by the pronoun *es(it)*.

Listing 1: DTD excerpt: discourse-entities

```
<!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>
```

The format `de_X_Y` is suggested for the ID of a discourse entity, in which *X* is a variable, that depending on the type of the current discourse entity can be specified as *n*, *p*, *pp*, *c*, *s*, *e* or *et*; and *Y* is a natural number. Possible IDs would then be `de_n_1` or `de_e_1`. In the project A2 *n* is used as default value for the `deType` attribute, since only nominal discourse entities are to

⁴ <http://www.oxygenxml.com>

be annotated. Therefore the infix variable X is omitted together with its surrounding underscores.

Relevant discourse entities are all lexical expressions, that introduce a discourse referent in the sense of Kamp and Reyle (1993) into the discourse or text representation. It is advisable to mark the nominal (as well as *named entity*) discourse entities in a first step, as these are easily identifiable in a text and thereafter all other entities.

In the project A2 *nominal* discourse entities are automatically detected on the base of MACHINESE SYNTAX annotated word forms. Here, firstly simple discourse entities are marked; those are discourse entities that are realized as a simple NP. Based on simple discourse entities also complex entities can be annotated. Complex NPs are NPs with a prepositional phrase (i.e. *der Mann auf dem Bahnsteig*) or NPs with an NP as premodifier (i.e. *der dem Bahnsteig stehende Mann*). NPs with relative clauses are *not* marked as complex discourse entities. Restrictive and non-restrictive relative clauses are treated equally. The cospecification is marked by a `cospecLink` (cf. Section 4.2 on page 11) between the discourse entities of the NP and the relative pronoun. In the following section the different types of discourse entities will be illustrated.

3.2 Types of Discourse Entities

In the project A2 only discourse entities of the type *nom* or *namedEntity* are annotated.

deType="nom"

Discourse entities of the type *nom* are introduced into the discourse by nominal phrases. Nominal discourse entities can be automatically identified using syntactic analysis. Thus the project A2 uses the parser-tagger output by MACHINESE SYNTAX.

Listing 2: Annotation example: 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”

The value *namedEntity* describes a subset of nominal discourse entities, namely those that can be detected automatically using *Named Entity Recognition*; for example those that can be assigned to the classes *Person*, *Organization*, *Location* (countries, regions, cities, rivers).⁵ In a first step only those entities that belong to the class *Person* were to be marked as **de** element with the **deType** *namedEntity* in the project A2.

Listing 3: Annotation example: deType=“namedEntity”

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

N. B.: Bibliographical references are annotated as complex discourse entities of the type *nom* (cf. following listing).

Listing 4: Annotation example: Bibliographical References (1)

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

Listing 5: Annotation example: Bibliographical References (2)

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

deType=“prop”

The value *prop* marks propositions. All declarative sentences are marked this way, as well as all embedded *dass*-sentences, the non-restrictive relative clauses and adverbial sentences. Interrogative sentences (although they are embedded) must *not* be marked as **prop**.

deType=“projProp”

The value *projProp* marks the so-called projective propositions. Among those are mostly interrogative sentences and the complements of modal operators.

⁵ cf. http://www.itl.nist.gov/iaui/894.02/related_projects/muc/

deType=“evType”

The value *evType* is used to mark (*event-types*). The full verb and the arguments belonging to each event-type have to be annotated.

deType=“ev”

The value *ev* stands for (*events*) and is used for event verbs. The corresponding event verb is annotated.

deType=“state”

The value *state* is supposed to mark static verbs. It is possible out of practical reasons not to make such a subtle differentiation and to leave out this value; then all verbs are consistently annotated as *ev*.

deType=“cluster”

The value *cluster* stands for sequences of sentences with varying length, where all sentences in a sequence belong to the same topic.

3.3 Annotation in Serengeti

In order to achieve optimum results in quality as well as quantity, it is necessary to have not only annotation guidelines with strict taxonomy and an unambiguous interpretation, but also easy to handle tools for annotation and organisation of corpora (cf. i.e. MMAX; Müller and Strube (2001)). For this purpose the programme SERENGETI was developed in the project A2. It was developed for the semi-automatic annotation of semantic relations (<http://coli.lili.uni-bielefeld.de/serengeti/>). This annotation tool is described in more detail in Stührenberg et al. (2007) and in Diewald (2008).

4 Annotation of (Discourse-)semantic Relations

To mark the (discourse-)semantic relations the element `semRel` is introduced, whereby analogous to TEI and MATE the annotation of (discourse-)semantic relations is done separately from the annotation of the discourse entities. This has the advantage of linking one discourse entity to more than one other discourse entity, by means of which on the one hand ambiguities concerning the antecedent on the textual representational level can be labeled well. On the other hand it is possible to link one discourse entity to others in the

text as well as with a topic in a TopicMap⁶.

In the project A2 the `semRel` is inserted behind the text (cf. DTD in Appendix 6 on page 31). The element `semRel` provides the three nested tags `corefLink`, `cospecLink` and `bridgingLink` for the annotation of coreference or the annotation of cospecification and indirect anaphora.

Listing 6: DTD excerpt for `semRel`

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

The two optional attributes `semRelID` and `sentenceIDRef` were introduced for documentation purposes.

4.1 Annotation of Coreference

Coreference is interpreted as a relation between a referential discourse entity and a topic of the Topic Map, the so-called reference anchor. For that purpose the project B1 created the element `corefLink`. This element has two attributes: `deIDRef` with a (deID) and `tmIDRef` with a TopicMapID as values:

Listing 7: DTD excerpt for `corefLink`

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

In the project B1 the situation that an expression in a text is clearly referential, but there is not yet an entry in the TopicMap, is resolved by assigning the value *unknown* to `tmIDRef`.

4.2 Annotation of Cospecification

Cospecification is interpreted as a relation between two discourse entities. For this purpose the element `cospecLink` is introduced. Firstly, for this element the attribute `relType` is created, whose value is one of the values

⁶ This procedure gives way to further development of the scheme; and to state e.g.

given in Listing 8 (cf. also Section 5.2.1). Secondly, the attribute `phorIDRef` (Typ: *IDREF*) with a `deID` as its value is created. The value of this attribute refers to the corresponding phoric element. The reference value of the phoric expression is marked as well. Since one discourse entity can have several reference values, the attribute `antecedentIDRefs` (Typ: *IDREFS*) with one or more `deID` is assigned as its value.

Listing 8: DTD excerpt for `cospecLink`

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

As antecedent the *last non-pronominal expression* in each cospecification chain is chosen. In case of ambiguous reference values the element `cospecLink` is duplicated accordingly.

Listing 9: Annotation example for cospecification

```
<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>
```

Non-annotated Anaphoric Relations

Binding phenomena are *not* annotated, because binding is solely syntactical licensed and therefore a sentence bound anaphoric relation. In the following

objects of the visual situation as reference value.

sentence for example, the anaphoric possessive pronoun *seinem*(his) is syntactically bound by the quantified NP *jeder Autor*(every author).

Jeder Autor sollte die Verlinkung in seinem Hypertext selbst vornehmen.

Likewise VP Ellipses that include an anaphoric pronoun are not annotated:

Otto_i [_{VP} stellte seinen_i Hypertext fertig], Max_j auch.

The missing VP on the right site (i.e. in the second part of the sentence) has the identical wording to the VP on the left site (i.e. in the first part of the sentence). In the reconstruction of the VP, there are two possible versions, depending on the interpretation of the anaphoric pronoun (as a constant or a bound variable). Those can be made clear by different indications:

- a. Otto_i [_{VP} stellte seinen_i Hypertext fertig], Max_j auch.
[_{VP} stellte seinen_i Hypertext fertig]
- b. Otto_i [_{VP} stellte seinen_i Hypertext fertig], Max_j auch.
[_{VP} stellte seinen_j Hypertext fertig]

In case a. there is a strict interpretation of the pronoun *seinen*(his). The NP *Otto* and the pronoun *seinen*(his) are coindexed. The sentence is understood the way that Max finishes Otto's Hypertext. Therefore the pronoun is read semantically as a constant. In case b. the pronoun *seinen*(his) coindexes with the NP *Max* and is interpreted as a bound variable. In this so-called *Sloppy-Identity*-reading Max finishes his own text. (cf. for this Hardt, 1996).

4.3 Annotation of Indirect Anaphora

Indirect anaphora is interpreted as relation between two discourse entities as well. An indirect anaphoric relation holds if the antecedent to an anaphoric element is not explicitly mentioned in the text, but has to be inferred from context. For this an element `bridgingLink` is introduced. A `bridgingLink` is annotated too, if there is an explicit possessive relation between two objects (marked by a possessive pronoun or genitive-NP).

Firstly, an attribute `relType` is introduced. The attribute value is one of the values given in Listing 10 (cf. also Section 5.2.3). Secondly, the attribute `phorIDRef` (Typ: IDREF) with a `deID` as value is assigned. The value of this attribute points to the corresponding phoric element. The reference value of the phoric expression is marked as well. Since one discourse entity can have several reference values, the attribute `antecedentIDRefs` (Typ: IDREFS) with one or more `deID` as a value is selected.

Listing 10: DTD excerpt for `bridgingLink`

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

As antecedent the *last non-pronominal expression* in each cospecification chain is chosen. In case of ambiguous reference values the element `bridgingLink` is duplicated accordingly.

Listing 11: Annotation example for indirect anaphora

```
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 Overview of Semantic Relations

As described in the previous sections, the annotation of linguistic phenomena under consideration comprises two aspects: the annotation of the relevant discourse entities and the annotation of the discourse semantic relations between one expression and its reference value. In the following sections

the expected values are introduced with the help of examples.⁷

5.1 Relations of Coreference

A narrow concept of coreference is applied: Two discourse referents are coreferent, if they are linked with the same reference anchor in a TopicMap. The existence of an anaphoric relation is not a necessary condition for coreference. The specific feature of the coreference concept used here is that reference is interpreted as “Exists in the TopicMap”. This certainly requires that all entities of the world are represented in a knowledge network, which typically holds for proper nouns and bibliographical references:

Listing 12: Annotation example for coreference

```
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"/>
```

The procedure explained here enables among other things that a referential expression can be linked to two different anchors, as illustrated in the following example. The Name “Austin” is used once for the city and once for the car brand. Accordingly both expressions cannot be marked as coreferential. This can be derived from the fact that Austin refers to different topics in the TopicMap.

Listing 13: Annotation example for coreference

```
<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"/>
```

⁷ In the annotation excerpts root elements and enclosing elements as for example `semRel` were renounced. Only relevant discourse entities are marked.

5.2 Relations of Cospecification and Indirect Anaphora

Cospecification is based on a phoric relation between linguistic expressions. A phoric expression could be a pronoun, an NP, or a proper noun.

N. B.: To simplify matters first of all only the anaphoric case is dealt with. However, the presented annotation scheme could easily be transferred to cataphoric contexts.

Indirect anaphora is based on the phoric relation between linguistic expressions, too. In this case the phoric expression is a definite NP. Schwarz defines indirect anaphoric relations as follows: “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, p. 49).

5.2.1 Direct Anaphora (Cospecification)

relType=„ident“

Firstly the value *ident* is assigned, if a pronoun refers to an NP.

Listing 14: Annotation example: ident – Pronoun

```
<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"/>
```

Secondly this value is assigned, if an NP refers to a recurrent NP (i.e. with identical head noun).

Listing 15: Annotation example: ident – identical head noun

```
<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"/>
```

Additional Examples:

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

relType="namedEntity"

This value is assigned, if a noun phrase that is not of the type *namedEntity*, refers to a noun phrase of the type *namedEntity*. Examples are noun phrases that belong to the classes *Person*, *Organization* and *Location* (countries, regions, cities, rivers).

N. B.: Only names of persons are annotated explicitly as discourse entities of the type *namedEntity*. During the manual annotation the relation type *namedEntity* is to be used as well for those noun phrases that would have been marked in an automatic named-entity recognition.

Listing 16: Annotation example: 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" />
```

Additional Examples:

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

relType="propName"

The anaphora is a proper noun that refers to a nominal reference value. The cospecified NP can be a discourse entity of the type *nom* or *namedEntity* (for names of persons).

N. B.: The same as for *namedEntity* is applied here as well; only names of persons are explicitly annotated as discourse entities of the type *namedEntity*. During manual annotation the relation type *propName* is to be used as well for those noun phrases that would have been marked in an automatic named-entity recognition.

Listing 17: Annotation example: propName (1)

```
<de deID="de1" deType="namedEntity">Sabine Rolfs</de> ist erst seit  
kurzem arm. &gt;Ich war froh, dass noch ein Aufziehauto 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: Annotation example: 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"/>
```

Additional Examples:

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"

This value is assigned if synonymy holds between the *head nouns* of the anaphor and its antecedent. Bussmann defines synonymy as follows: "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, p. 763)

For the annotation a broader definition of synonymy is applied, for example in the context of the text the cospecified NP is a synonym to its reference value (partial synonymy, cf. Bußmann, 1990). Example for partial synonymy: In the context "Hypertext" there is a synonymous relation between *Link* (engl. link) and *Verweis* (engl. reference or sending-off). This synonymy is limited only to this reading of *Verweis* (engl. reference), in the context of "Fußball" it is not valid (*Ballack bekam einen Link).

N. B.: In contrast to a paraphrase, synonymy holds between lexemes only (not between noun phrases):

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

Listing 19: Annotation example: 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: Annotation example: 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.: Abbreviations are annotated as synonyms, too. If the abbreviation is given in brackets after the full form, the full form is to be chosen as antecedent; otherwise the last non-pronominal naming is marked as antecedent (cf. following listing).

Listing 21: Annotation example: 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"/>
```

Additional Examples:

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

relType=„hyperonym“

The head noun of the anaphor is a hyperonym to the head noun of the antecedent. Bussmann defines hyperonymy as follows: „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, p. 317)

Listing 22: Annotation example: 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: Annotation example: 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"/>
```

Additional examples:

Studium – Ausbildung
Psychologen – Wissenschaftler

relType=„hyponym“

The head noun of the anaphor is a hyponym to the head noun of the antecedent. Bussmann defines hyponymy as follows: „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, p. 317)

Listing 24: Annotation example: 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"/>
```

Additional Example:

Überblick – Preisübersicht

relType=„addInfo“

The cospecified NP introduces new or additional information.

Listing 25: Annotation example: 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: Annotation example: 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" />
```

Additional Example:

das ferne Istanbul – die 12-Millionen-Metropole

relType=„paraphrase“

The anaphor is a paraphrase of its antecedent.

Listing 27: Annotation example: 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.: Translations are also annotated as a paraphrase:

Armoured Personal Carriers – die Fahrzeuge

relType=„isA“

If out of practical matters it is impossible to distinguish the relations *namedEntity*, *synonym*, *hyperonym*, *hyponym*, *addInfo* and *paraphrase* in the annotation, it is possible to subsume these relations and replace all *relType*-values with the value *isA*.

5.2.2 Cospecification with Propositions or Events as Reference Values

The following relation types describe the occurrence of cospecification, in which the reference value is a non-nominal discourse entity.

relType=„abstrProp“

This cospecified expression relates to a proposition.

Listing 28: Annotation example: *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“

This cospecified expression relates to a projective proposition.

Listing 29: Annotation example: *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“

This cospecified expression relates to an event-type.

Listing 30: Annotation example: 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“

This cospecified expression relates to an event.

Listing 31: Annotation example: 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“

This cospecified expression relates to a sum of propositions or a to part of the text.

Listing 32: Annotation example: 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“

In case none of the above relations can be applied, the value `unknown` is assigned.

5.2.3 Indirect Anaphora

In the first version of this annotation scheme that has been developed for the project B1 those types of indirect anaphora described by Schwarz (2000) have been subsumed by the `relType`-value *bridging* (semantic role assignment, nominal-semantic relation meronymy, nominal-semantic relation hyperonymy, schema-based indirect anaphora and inference-based indirect anaphora). In the version for the project A2 bridging relations are modelled in more detail. For this purpose a division regarding different bridging relations is made. The individual relation types were chosen so that they were describable with linguistic resources. This is the case for the semantic relations meronymy and holonymy, which are modelled in GERMANET; as well as for possession relations that are syntactically marked by possessive pronouns or genitive constructions. In addition to these relations further relations were chosen: *setMember* and *hasMember* for the description for class affiliation, as well as *bridging* for other, not explicitly considered relation types (i.e. scheme-based indirect anaphora).

`relType=„poss“`

The value *poss* is assigned, if the phoric expression is explicitly marked by a possessive pronoun or a genitive-NP.

Listing 33: Annotation example: *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: Annotation example: *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"/>
```

Additional Examples:

Peter – Peters Text

Peter – sein Text

relType=„meronym“

The head nouns of the anaphor and its antecedent are in a meronymy relation (part-of-relation). The anaphor describes a part and the antecedent the whole, for example a *fan* is part of a *computer*.

Listing 35: Annotation example: 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: Annotation example: 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"/>
```

Additional Examples:

Tür – Klinke
ein Kettenfahrzeug – die rechte Kette

N. B.: Cities are annotated as meronym of the country they belong to:

Deutschland – München

relType=„holonym“

The head noun of the anaphor is a holonym to the head noun of the antecedent, for example the anaphor describes the whole and the antecedent describes a part.

Examples:

Seite – Buch
Henkel – Tasse

N. B.: A country is annotated as holonym to a city:

München – Deutschland

relType=„hasMember“

The anaphor outlines a set and the antecedent is an element of that set.

Listing 37: Annotation example: 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"/>
```

Additional example:

Deutschland – Europa

In the example Europe is read as collection of countries that belong to it (Europe as union of countries).

relType=„setMember“

The phoric expression is an element of a set given by its reference value.

Listing 38: Annotation example: 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: Annotation example: 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" />
```

Additional Example:

Europa – Deutschland

relType=„bridging“

This value is chosen if none of the above mentioned values apply (for example scheme-based indirect anaphora or inference-based indirect anaphora).

Listing 40: Annotation example

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: Annotation example

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"/>
```

Additional Examples:

die türkische Politik – Parlamentswahlen
Einfamilienhaus – die beiden Bewohner

6 Annotation process

When annotating anaphoric relations, it has to be decided first whether a given discourse entity (DE) in a text is used anaphorically: DEs of indefinite noun phrases are not anaphoric whereas those of definite NPs are supposed to be used anaphorically and thus to be related completely to an antecedent (if possible) in the project A2. In a second step, a correct antecedent for each anaphoric discourse entity has to be found. Within a cospecification chain always *the last non-pronominal candidate* has to be chosen.

The procedure of selecting the appropriate relation type between anaphor and antecedent is shown in Figure 3 on the facing page: At first it has to be decided on whether it is a cospecification relation (element `cospecLink`) or a bridging relation (element `bridgingLink`). Cospecification is valid if anaphor and antecedent refer to the same object; bridging is valid if anaphor and antecedent do not refer to the same object, but are connected with each other through one of the secondary relations described in Subsection 5.2.3.

In the case of cospecification it has to be proceeded as shown in Figure 3. To determine which sub-relation applies, *ident* is checked first, then *propName* etc. The secondary relation type *paraphrase* is the least restricted one and should only be selected, if none of the others can be applied.

In case of bridging there is a similar procedure: Firstly *poss* is checked, then *meronym/holonym*, after that *hasMember/setMember*. The secondary relation type *bridging* is the least restricted one and should only be selected, if none of the other relation types can be applied.

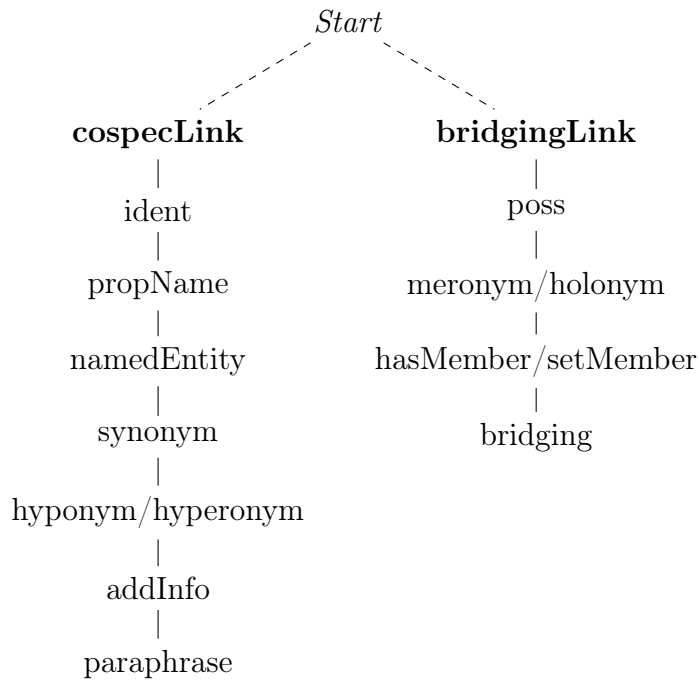


Figure 3: Decisions in the Annotation Process

Appendix A: DTD for the Annotation Scheme

Listing 42: DTD for the Annotation Scheme

```

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>
```

References

- Bußmann, H. (1990). *Lexikon der Sprachwissenschaft*. Kröner, 2nd edition.
- Clark, H. (1977). Bridging. In Johnson-Laird, P. and Watson, P., editors, *Thinking: Readings in cognitive Sciences*. Cambridge University Press, London ; New York.
- Diewald, N. (2008). Serengeti - a brief starting guide. Interne Reports der DFG-Forschergruppe 437 "Texttechnologische Informationsmodellierung".
- Fligelstone, S. (1992). Developing a Scheme for Annotating Text to Show Anaphoric Relations. In Leitner, G., editor, *New Directions in English Language Corpora: Methodology, Results, Software Developments*, pages 153–170. Mouton de Gruyter, Berlin.
- Garside, R., Fligelstone, S., and Botley, S. (1997). Discourse Annotation: Anaphoric Relations in Corpora. In Garside, R., Leech, G., and McEnery, A., editors, *Corpus Annotation: Linguistic Information from Computer Text Corpora*, pages 66–84. Addison-Wesley Longman, London.
- Hardt, D. (1996). Centering in dynamic semantics. In *Proceedings of COLING 1996*, volume Volume 1: The 16th International Conference on Computational Linguistics.
- Hirschmann, L. (1997). MUC-7 Coreference Task Definition (version 3.0). In Hirschman, L. and Chinchor, N., editors, *Proceedings of Message Understanding Conference (MUC-7)*.
- Holler, A. (2003). Ein Annotationsschema für Koreferenzphänomene. Technical report, TEMIS Dtl. GmbH und Universität Dortmund.
- Holler, A., Maas, J.-F., and Storrer, A. (2004). Exploiting coreference annotations for text-to-hypertext conversion. In *Proceeding of LREC*, volume II, pages 651–654, Lisboa.
- Holler-Feldhaus, A. (2004). Koreferenz in Hypertexten: Anforderungen an die Annotation. *Osnabrücker Beiträge zur Sprachtheorie (OBST)*, 68:9–29.

- Kamp, H. and Reyle, U. (1993). *From Discourse to Logic*. Kluwer: Dordrecht.
- Krasavina, O. and Chiarcos, C. (2007). PoCoS - Potsdam Coreference Scheme. In *Proceedings of the Linguistic Annotation Workshop*, pages 156–163, Prague, Czech Republic. Association for Computational Linguistics.
- Müller, C. and Strube, M. (2001). Annotating anaphoric and bridging relations with MMAX. In *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue*, pages 1–6, Morristown, NJ, USA. Association for Computational Linguistics.
- Poesio, M. (1998). Coreference. MATE Deliverable D1.1.
- Poesio, M. (2004). The MATE/GNOME Scheme for Anaphoric Annotation, Revisited. In *Proceedings of SIGDIAL*, Boston.
- Schwarz, M. (2000). *Indirekte Anaphern in Texten*. Niemeyer, Tübingen.
- Sidner, C. (1979). *Towards a computational theory of definite anaphora comprehension in English discourse*. PhD thesis, MIT.
- Stührenberg, M., Goecke, D., Diewald, N., Cramer, I., and Mehler, A. (2007). Web-based annotation of anaphoric relations and lexical chains. In *Proceedings of The Linguistic Annotation Workshop (LAW)*, pages 140–147, Prague, Czech Republic. Association for Computational Linguistics.
- Thompson, H. S. and McKelvie, D. (1997). Hyperlink semantics for standoff markup of read-only documents. In *Proceedings of SGML Europe '97: The next decade – Pushing the Envelope*, pages 227–229, Barcelona.
- Webber, B. (1988). Discourse deixis: Reference to discourse segments. In *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistic (ACL-88)*, pages 113–122, State University of New York at Buffalo.