

# Towards a standard for annotating abstract anaphora

Stefanie Dipper, Heike Zinsmeister

Institute of Linguistics, Institute of Linguistics  
Bochum University, Konstanz University  
dipper@linguistics.rub.de, Heike.Zinsmeister@uni-konstanz.de

## Abstract

This paper presents a survey on the annotation of abstract (= discourse-deictic) anaphora, i.e. anaphora that involves reference to abstract entities such as events and propositions. The survey identifies features that are common to the majority of relevant annotation efforts. Based on these, we propose a small set of recommendations, which can be viewed as a first, small step towards a standard for the annotation of abstract anaphora. As the overview shows, English is the language that most of the resources have been created for. However, many of them contain only few instances of abstract anaphora. Hence, the currently available evidence only supports preliminary conclusions.

## 1. Introduction

The paper presents a survey on the annotation of abstract (= discourse-deictic) anaphora, i.e. anaphora that involves reference to abstract entities such as events and states (Asher, 1993). The survey tries to identify the features that are common to the majority of relevant annotation efforts, and can be viewed as a first, small step towards a standard for the annotation of abstract anaphora.

In the last couple of years, several related but nevertheless different approaches have been proposed for both the inventory of annotation tags as well as the coding schemes of the relations and the markables in the texts. The differences can be traced back to the following reasons:

- the “theory” behind abstract entities: e.g. whether abstract entities are defined by reference to syntactic, semantic, and/or pragmatic properties;
- the kind of data that is analyzed: e.g. dialogues/spoken language or written text;
- the language under consideration: e.g. languages with zero pronouns or clitics require annotation schemes different from schemes for English data.

The contribution of this paper is a survey of the state of the art of abstract anaphora annotation which highlights categories that are generally agreed on and takes different points of view on anaphoric encoding into account. The paper first addresses proposals that have been made with regard to representational issues (Sec. 2.). In Sec. 3. we describe relevant annotation efforts, followed by our recommendations in Sec. 4. The Appendix contains a synoptic table of the studies considered in this paper.

## 2. Standards

Annotation of abstract anaphora is not an easy task. Hence, just like all precious resources, corpora annotated with abstract anaphora relations should be maximally reusable and exploitable for further applications. Maximal reusability can be achieved by adherence to standards, which regulate both *content* and *form* of annotation. This paper deals with the first aspect.

*Content* is standardized by means of tagsets, specifying obligatory and/or optional tags (“data categories”), along

with annotation guidelines. We do not know of any proposal to standardize content of abstract anaphora annotation, and propose a small set of recommendations in Sec. 4. *Form* is standardized by reference to data models and physical data structures, which are used, e.g., for data interchange. Data structures are specified, e.g., in the form of DTDs or XML schemata, which define an XML representation format. We would like to point out two proposals, MATE/GNOME (Poesio 2000a/b)<sup>1</sup> and RAF (Reference Annotation Framework, Salmon-Alt and Romary (2004)). Both proposals agree in that they do *not* encode anaphoric relations by pointers that are attached to the anaphor and point to the antecedent. Instead, they define extra, autonomous elements that represent the anaphoric relation. This opens up the possibility of easily annotating a discourse entity with multiple anaphoric links, as well as recursively defining complex markables, or annotating empty strings, such as zero pronouns.

A MATE-style XML example of an anaphoric “identity” link would look as follows (the outer element refers to the anaphor, the embedded element to its “anchor”, i.e. its antecedent):

```
<coref:link type="ident" href="...">  
  <coref:anchor href="..." />  
</coref:link>
```

## 3. State of the art

In this section, we present a series of relevant work on abstract anaphora.<sup>2</sup> The main features are summarized in Table 1 in the Appendix. This overview will lead us to a comparative assessment of the features used in the annotations. Column 1 of Table 1 lists the authors of the study. Columns “Data” (2–4) inform about the data used in the research:

<sup>1</sup>The original GNOME scheme is restricted to concrete anaphora annotation.

<sup>2</sup>See Recasens (2008, ch.2) for a similar overview of coreference annotation in general (MUC, ACE, MATE, and AnCora schemes), including a comparison of annotated English and non-English corpora. Müller (2008, ch.2 and ch.5) contains a discussion of different projects of abstract anaphora annotation.

column 2 displays the codes of the language(s) that the papers deal with; columns 3 and 4 presents general and statistical information about the corpus.<sup>3</sup> Columns “Anaphor” and “Antecedent” focus on the syntactic and semantic properties that are taken into account by these studies.<sup>4</sup> “Reliability” columns report whether inter-annotator agreement has been computed. For anaphors, agreement is computed for semantic annotation; for antecedents, agreement usually concern the marking of segment boundaries. Column “Criteria” indicates whether the study provides tests (e.g., in the form of annotation guidelines) that can be applied by the annotators.<sup>5, 6</sup>

### 3.1. The anaphor

The majority of research considered here restricts their investigations to pronominal anaphors. Exceptions are Vieira et al. (2002), Poesio and Modjeska (2005), and Botley (2006), who consider *this*- and *that*-NPs, i.e. “full” NPs which start with the respective (translated) demonstrative determiner, and Recasens and Martí (2010), who take all kinds of NPs and pronouns into account.

Identifying pronouns in general is considered a trivial task. However, identifying *abstract* (also called: discourse-deictic, indirect) anaphors and distinguishing them from *concrete* (also called: individual) anaphors is a relevant issue. Hence, reliability studies for this task provide important information. However, not all studies distinguish between abstract and concrete anaphora but define other basic classes. In addition, more fine-grained labels are sometimes introduced. In this paper, we only consider labels that apply to abstract anaphora. Labels for concrete anaphora and pronouns are subsumed under “others” (see Table 1), and reported along with the total number of such labels.

Eckert and Strube (2000), Navarretta and Olsen (2008), and Dipper and Zinsmeister (2009a) define *vague* anaphora, which refers to some general discourse topic which is not overtly expressed.<sup>7</sup> Müller (2008) uses the label *vague* in a more general sense, to mark pronouns with no clearly-defined textual antecedent.

<sup>3</sup>Abbreviations used: *T*: total number of tokens; *C*: anaphora candidates (e.g., number of NPs); *AA*: number of abstract anaphors.

<sup>4</sup>Abbreviations used: *Dem*: demonstrative pronouns, *Pers*: personal pronouns, *Poss*: possessive pronouns, *Rel*: relative pronouns, *Zero*: zero pronouns, *Cl*: clitics, *Expl*: expletives/idioms; *Dem-NP*: NP with a demonstrative determiner; *AA*: abstract anaphors, *concr*: concrete, *abstr*: abstract, *non-ref*: non-referring, *indir*: indirect.

*Clauses* means that antecedents are syntactically defined, e.g. as sentences, infinitives, gerunds; *V-head* means that only the verbal head is marked.

<sup>5</sup>Several annotation guidelines make use of GNOME, e.g., Poesio and Modjeska (2005) and Navarretta and Olsen (2008), but only for the annotation of concrete anaphora.

<sup>6</sup>Goecke et al. (2008) present an annotation scheme for anaphoric relations in German, which includes specifications for abstract entities. Abstract types are defined syntactically (propositions and projective propositions) or semantically (events, event-types, states). In their project, however, only concrete anaphora has been annotated.

<sup>7</sup>This label is called *deict* in Dipper and Zinsmeister (2009a).

Botley (2006), who considers *this*-NPs, investigates the semantics of the (abstract) anaphoric head nouns in detail. He distinguishes three main types of abstract anaphora: (i) “Label” anaphora, which serves to encapsulate (or to label) stretches of text (following Francis (1994)). Label anaphora is further classified as *general* or as *metalinguistic*, with subtypes *illocutionary*, *language activity*, *mental process*, *text*. (ii) “Situation” anaphora, with subtypes *eventuality* (e.g. events, processes, states) and *factuality* (e.g. fact, proposition) (following Fraurud (1992)). (iii) “Text deixis”.

Distinctions similar to Botley’s “situation” anaphora subtypes are made by Hedberg et al. (2007), Navarretta and Olsen (2008), and Dipper and Zinsmeister (2009a). Recasens and Martí (2010) define subtypes *token*, *type*, *proposition*. In contrast to most other work, Dipper and Zinsmeister (2009a) annotate these subtypes both to the abstract anaphors and their antecedents (see Sec. 3.2.).

Poesio and Artstein (2008) annotate the reference status of NPs and pronouns: *anaphoric*, *discourse-new*, *non-referring*. In addition, they classify them semantically, e.g. as *person*, *animate*, *concrete*, *space*, *time* etc.

Sometimes, abstract anaphora is subsumed under the more general label *indirect*, see Botley (2006), Vieira et al. (2002), and, with a slightly different classification, Hedberg et al. (2007).<sup>8</sup> Other members of these classes are bridging relations, occurring with concrete anaphora. Bridging relations are akin to abstract anaphora in that antecedents are not readily available but require additional interpretational efforts.

Kučová and Hajičová (2004) define the label *text* for inter-sentential general coreference relations, and the label *segm* which is used for anaphors with multi-node/multi-rooted antecedents (in the dependency framework).

Usually, *all* referring pronouns are annotated, and reliability results are reported that measure inter-annotator agreement on the entire set of referring pronouns (and their antecedents). Whenever appropriate information is available, we distinguish between agreement on personal and demonstrative pronouns. Personal pronouns (at least in English) predominantly refer to concrete entities, demonstrative pronouns often refer to abstract entities. The results listed in Table 1 indicate that—as is expected—anaphora resolution is considerably easier with concrete entities than abstract entities.

If no identification criteria and/or reliability results are listed in Table 1, this means that none are mentioned in the respective papers.

### 3.2. The antecedent

Antecedents of abstract anaphora are abstract objects, such as actions and events. Accordingly, they correspond to linguistic entities which include at least a verb: partial clauses, clauses, sequences of sentences, or even discontinuous strings, as illustrated by the following example (the antecedent of the anaphor *it* is underlined):<sup>9</sup>

<sup>8</sup>This label is called *other* in Vieira et al. (2002).

<sup>9</sup>Example taken from file ep-04-03-31.txt of the Europarl corpus.

- (1) I would like to draw particular attention to the fact that people who have made their lives here in the European Union still do not have the right to vote, even though the European Parliament has called for **it** on many occasions.

The approaches differ as to whether they restrict the marking of the antecedent to the verbal head, as in Müller (2007) or Pradhan et al. (2007), or approximate it by predefined constituents, e.g. clauses (Byron, 2003), or whether the annotators are allowed to mark free spans of text, e.g. Vieira et al. (2002), or Dipper and Zinsmeister (2009a). Dipper and Zinsmeister aim at determining the exact scope of the anaphor, i.e., the exact extension of the antecedent's string, including examples as (1). For this task, they propose a paraphrase test. Other annotation efforts deliberately do not aim at identifying exact boundaries, some even do not mark antecedents at all (Poesio and Modjeska, 2005).

The identification of antecedents is easier in monologue or written texts than in dialogues (Poesio, 2004). In general, identifying antecedents is easier in monologues or written texts than in dialogues (Poesio, 2004). For instance, different speakers may have different assumptions about the situation. In addition, incomplete or ungrammatical sentences often occur in spoken language, due to hesitations or disfluencies. Therefore, annotations of dialogues often recur to independently-defined units, such as dialogue acts (e.g., Eckert and Strube (2000)).

Kučová and Hajičová (2004), just like other approaches, mark the verbal head of the antecedent. However, in their dependency framework, the verbal head is the root node of the clause, and, hence, the marking specifies the extension of the antecedent.

Further properties of the antecedent are investigated only in a subset of the studies. Hedberg et al. (2007) consider the saliency of the antecedent to specify the cognitive status of the anaphor. Dipper and Zinsmeister (2009a) determine the semantic subtype of the antecedent by a replacement test, deliberately ignoring the anaphor.

### 3.3. Summary

As can be seen from Table 1, semantic annotation is considered more relevant for anaphors than for antecedents. Annotation efforts that consider both concrete and abstract anaphors often annotate the distinction concrete–abstract for anaphors. With NP anaphors, the head noun determines its class, e.g. *this situation*. With pronominal anaphors, people apply two strategies: (i) The clausal context of the pronoun, e.g., its governing verb, impose selectional constraints on the semantic type of the anaphor. (ii) The semantics of the antecedent is used to determine the anaphor's type.

However, it is often assumed that interpreting abstract anaphora involves an additional interpretational step (Weber, 1988), and the resolution process can involve a kind of type-raising operation (coercion, Hegarty et al. (2001), Consten et al. (2007)). This has to be taken into account in the design of the annotation process.

In general, annotation efforts tend to focus on anaphors rather than antecedents, partly because it is anaphors that

have to be resolved and partly because antecedents are sometimes difficult to determine. Antecedents that are made up by arbitrary sequences are usually restricted to written texts. In contrast, dialogue annotation tends to recur to “syntactically”-defined antecedents, i.e., antecedents that correspond to segments of dialog acts (or the respective verbal heads).

## 4. Towards a standard

Based on the observations made in the previous sections, we propose that “reference” corpora with abstract anaphora, which aim at sustainability and reusability, should adhere to the following principles.

### The anaphor

*Form:* Many languages distinguish between pronouns that are prototypical realizations of abstract anaphora (e.g. demonstrative pronouns), and non-typical ones (e.g. personal pronouns). In addition, NPs, depending on the semantics of their head noun, can refer to abstract entities.

*Proposal:* We propose that reference corpora minimally should annotate prototypical pronominal realizations.

*Semantics:* The distinction between concrete and abstract can be made rather easily and reliably. For finer-grained labels, the situation is more complex: no commonly-used set of labels has been yet proposed, and people annotate considerably different types of information, such as speech acts, eventualities and factualities, or type-token distinction.

*Proposal:* Minimal annotation should include the distinction concrete–abstract.

### The antecedent

*Form:* In most cases, marking the verbal head vs. the entire clause are equivalent solutions. It only makes a difference if the antecedent does not contain a verb, or if it consists of multiple clauses (in this case, a necessarily discontinuous string of multiple verbal heads would have to be marked). In both cases, clause marking seems more suitable than verbal-head marking. However, marking of verbal heads does not require any preprocessing.

*Proposal:* Antecedents are to be marked. Minimally, (sequences of) clauses or verbal heads should be annotated.

*Semantics:* Only very few projects have annotated semantic properties of abstract antecedents so far, and the issue still waits further investigations to be better understood.

*Proposal:* Currently none.

As we have seen, many annotation studies deal with anaphora in general, and—since concrete anaphors occur considerably more frequently than abstract anaphors—are restricted by an extremely low number of abstract anaphors. Hence, current results achieved so far are of limited significance, and considerably more data has to be produced to allow for serious investigations. We therefore call for annotation efforts focusing on the annotation of abstract anaphora.

## 5. References

Ron Artstein and Massimo Poesio. 2006. Identifying reference to abstract objects in dialogue. In *Proceedings of Brandial 2006*.

- Nicholas Asher. 1993. *Reference to Abstract Objects in Discourse*. Kluwer Academic Publishers, Boston MA.
- Simon Botley and Tony McEnery. 2001. Demonstratives in English: A corpus-based study. *Journal of English Linguistics*, 29:7–33.
- Simon Botley. 2006. Indirect anaphora: Testing the limits of corpus-based linguistics. *International Journal of Corpus Linguistics*, 11(1):73–112.
- Donna K. Byron. 2002. Resolving pronominal reference to abstract entities. In *Proceedings of the ACL-02 conference*, pages 80–87.
- Donna K. Byron. 2003. Annotation of pronouns and their antecedents: A comparison of two domains. Technical Report, University of Rochester.
- Manfred Consten, Mareile Knees, and Monika Schwarz-Friesel. 2007. The function of complex anaphors in texts: Evidence from corpus studies and ontological considerations. In *Anaphors in Text*, pages 81–102. John Benjamins, Amsterdam/Philadelphia.
- Stefanie Dipper and Heike Zinsmeister. 2009a. Annotating discourse anaphora. In *Proceedings of LAW III*, pages 166–169.
- Stefanie Dipper and Heike Zinsmeister. 2009b. Annotation guidelines “Discourse-Deictic Anaphora”. Draft. Universities of Bochum and Konstanz.
- Miriam Eckert and Michael Strube. 2000. Dialogue acts, synchronising units and anaphora resolution. *Journal of Semantics*, 17(1):51–89.
- Gill Francis. 1994. Labelling discourse: an aspect of nominal group lexical cohesion. In Malcolm Coulthard, editor, *Advances in Written Text Analysis*, pages 83–101. London: Routledge.
- Kari Fraurud. 1992. Situation reference: What does ‘it’ refer to? GAP Working Paper No 24, Fachbereich Informatik, Universität Hamburg.
- Daniela Goecke, Anke Holler, and Maik Stührenberg. 2008. Coreference, cospecification and bridging: Annotation scheme. Technical Report, University of Bielefeld.
- Jeanette K. Gundel, Nancy Hedberg, and Ron Zacharski. 2006. Coding protocol for statuses on the Givenness Hierarchy. Unpublished manuscript (1993/2006). [http://www.sfu.ca/~hedberg/Coding\\_for\\_Cognitive\\_Status.pdf](http://www.sfu.ca/~hedberg/Coding_for_Cognitive_Status.pdf).
- Nancy Hedberg, Jeanette K. Gundel, and Ron Zacharski. 2007. Directly and indirectly anaphoric demonstrative and personal pronouns in newspaper articles. In *Proceedings of DAARC-2007*, pages 31–36.
- Michael Hegarty, Jeanette K. Gundel, and Kaja Borthen. 2001. Information structure and the accessibility of clausally introduced referents. *Theoretical Linguistics*, 27(2-3):163–186.
- Lucie Kučová and Eva Hajičová. 2004. Coreferential Relations in the Prague Dependency Treebank. In *Proceedings of DAARC-2004*, pages 97–102.
- Marie Mikulová et al. 2005. Annotation on the tectogrammatical level in the Prague Dependency Treebank. reference book. Technical report, ÚFAL MFF UK, Prague, Czech Republic.
- Christoph Müller. 2007. Resolving *it*, *this*, and *that* in unrestricted multi-party dialog. In *Proceedings of ACL-07 conference*, pages 816–823.
- Christoph Müller. 2008. *Fully Automatic Resolution of ‘it’, ‘this’, and ‘that’ in Unrestricted Multi-Party Dialog*. Ph.D. thesis, University of Tübingen.
- Costanza Navarretta and Sussi Olsen. 2008. Annotating abstract pronominal anaphora in the DAD project. In *Proceedings of LREC-08*.
- Costanza Navarretta and Sussi Olsen. 2009. The annotation of pronominal abstract anaphora in Danish texts and dialogues. DAD report 1. Centre for Language Technology, University of Copenhagen.
- Costanza Navarretta. 2008. Pronominal types and abstract reference in the Danish and Italian DAD corpora. In *Proceedings of the Second Workshop on Anaphora Resolution*, pages 63–71.
- Massimo Poesio and Ron Artstein. 2008. Anaphoric annotation in the ARRAU corpus. In *Proceedings of LREC-08*.
- Massimo Poesio and Natalia N. Modjeska. 2005. Focus, activation, and *this*-noun phrases: An empirical study. In António Branco, Tony McEnery, and Ruslan Mitkov, editors, *Anaphora Processing*, volume 263, pages 429–442. John Benjamins.
- Massimo Poesio. 2000a. Coreference. In Andreas Mengel et al., editor, *MATE Dialogue Annotation Guidelines*. <http://www.ims.uni-stuttgart.de/projekte/mate/mdag>.
- Massimo Poesio. 2000b. The GNOME annotation scheme manual. 4th version. [http://cswwww.essex.ac.uk/Research/nle/corpora/GNOME/anno\\_manual\\_4.htm](http://cswwww.essex.ac.uk/Research/nle/corpora/GNOME/anno_manual_4.htm).
- Massimo Poesio. 2004. The MATE/GNOME proposals for anaphoric annotation, revisited. In *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue*, pages 154–162.
- Sameer Pradhan, Lance Ramshaw, Ralph Weischedel, Jessica MacBride, and Linnea Micciulla. 2007. Unrestricted coreference: Identifying entities and events in OntoNotes. In *Proceedings of the IEEE-ICSC*.
- Marta Recasens and M. Antònia Martí. 2010. AnCorACO: Coreferentially annotated corpora for Spanish and Catalan. *Language Resources & Evaluation*. DOI 10.1007/s10579-009-9108-x.
- Marta Recasens, M. Antònia Martí, and Mariona Taulé. 2007. Where anaphora and coreference meet. Annotation in the Spanish CESS-ECE Corpus. In *Proceedings of RANLP-07*, pages 504–509.
- Marta Recasens. 2008. Towards coreference resolution for Catalan and Spanish. Master’s thesis, University of Barcelona.
- Susanne Salmon-Alt and Laurent Romary. 2004. RAF: Towards a reference annotation framework. In *Proceedings of LREC-04*.
- Renata Vieira, Susanne Salmon-Alt, and Caroline Gasperin. 2002. Coreference and anaphoric relations of demonstrative noun phrases in a multilingual corpus. In *Proceedings of DAARC-2002*.
- Bonnie L. Webber. 1988. Discourse deixis: Reference to discourse segments. In *Proceedings of the ACL-88 conference*, pages 113–122.

# Appendix

Table 1: Survey of studies involving abstract anaphora

Study	Data		Anaphor		Antecedent		Criteria	Misc			
	Lang	Corpus	Statistics	Form	Semantics	Reliability (Semantics)			Form (AA)	Semantics (AA)	Reliability (Form)
Eckert and Strube (2000)	EN	Switchboard (telephone conversations)	T: ?, C: 678, AA: <b>154</b>	Dem, Pers (no Expl)	<i>concr, abstr, vague, others</i> (1)	$\kappa = .80$ (Dem), $\kappa = .81$ (Pers)	Clauses	–	96.1–98.4% ( <i>concr</i> ), 85.7–94.3% ( <i>abstr</i> )	Compatibility constraints, in algorithm	Algorithm (not implemented)
Vieira et al. (2002)	FR, PT	MLCC Corpus (written EU inquiries)	FR: T: 50K, C: 291, AA: <b>136</b> ( <i>other</i> ); PT: T: 50K, C: 243, AA: <b>116</b> ( <i>other</i> )	Dem-NPs	<i>concr, abstr, other</i> (incl. AA), others (2)	FR: $\kappa = .79$ PT: $\kappa = .65$	Arbitrary sequences	–	Total/partial: FR: 69.8/79.2% PT: 51/55.1%	Some form- and meaning-based criteria	Annotation tool MMAX
Byron (2003)	EN	TRAINS93 (problem-solving dialogues), BUR (radio news stories)	TR/BUR: T: 10K/13K, C: 346/380, AA: <b>22/47</b>	Dem, Pers (no Expl)	<i>concr, abstr</i>	Replacement test (TR/BUR): $\kappa = .56/.53$ (Dem), $\kappa = .71/.82$ (Pers)	Clauses	–	TR/BUR: $\kappa = .37/.62$ (Dem), $\kappa = .77/.95$ (Pers)	Replacement test; compatibility constraints, in algorithm	Implementation (Byron, 2002)
Kučová and Hajičová (2004)	CZ	PDT (mostly newspaper texts)	T: ? (35K sentences), C: 15K, AA: <b>274</b>	Pers, Poss, Dem, Zero	<i>text, segm, others</i> (3)	–	V-head (dependency structure)	–	–	Guidelines (Marie Mikulová et al., 2005)	Annotation tool TRED
Poesio and Modjeska (2005)	EN	GNOME (museum descriptions)	T: ? (500 sentences), C: 112, AA: <b>19</b>	Dem ( <i>this, these, this-NPs</i> )	<i>abstr, type</i> (generic), others (5)	$\kappa = .82$	–	–	na	Decision tree	–
Botley (2006), Botley and McEnergy (2001)	EN	AP (newswire), Hansard (parliament proc.), APHB (literary texts)	T: 300K, C: 648, AA: <b>10</b> ( <i>indir</i> )	Dem, <i>this/that-NPs</i>	<i>indir</i> (incl. AA, with multiple subtypes; Sec. 3.1.), others (3)	–	–	–	na	Guidelines (cf. Botley and McEnergy (2001))	–

Table 1 (cont'd)

Study	Data			Anaphor		Antecedent			Criteria	Misc	
	Lang	Corpus	Statistics	Form	Semantics	Reliability (Semantics)	Form (AA)	Semantics (AA)			Reliability (Form)
Hedberg et al. (2007)	EN	Newspaper texts	T: ? C: 321 (Dem), 43 (it), <b>AA: ?</b>	<i>it, this, that</i>	<i>dir, indir</i> (both incl. AA, with multiple subtypes (Sec. 3.1.))	<i>dir</i> vs. <i>indir</i> : $\kappa = .7$ (Dem), $\kappa = .16$ (it)	Arbitrary sequences	cognitive status: <i>in focus, activated</i>	Labels (cognitive status): $\kappa = .46$ (Dem), $\kappa = .58$ (it)	Replacement test; guidelines (Gundel et al., 2006)	
Müller (2007), Müller (2008)	EN	ICSI Meeting Corpus (dialogues)	T: ? (5 dialogues), C: 343 (anaphoric chains), <b>AA: 59</b> chains	<i>it, this, that</i>	<i>referring, vague, non-ref</i>	–	V-head	–	$\alpha = .43-.52$	“Simple instructions”	Annotation tool MMAX; implementation
Pradhan et al. (2007)	EN	Wall Street Journal (newswire)	T: 300K, C: 11,400 (coref chains), <b>AA: ?</b>	Dem, Pers (no Expl), Poss, NPs (specific)	–	na	V-head	(Only events are annotated)	–	Annotated examples	Implementation
Navarretta and Olsen (2008), Navarretta (2008)	DA, IT	DAD Corpus (mixed)	DA: T: 125K, C: 1,612, <b>AA: 455</b> ; IT: T: 135K, C: 890, <b>AA: 114</b>	Dem, Pers, Zero	Multiple labels (Sec. 3.1.)	AA: DA: $\kappa = .71-.89$ IT: $\kappa = .78-.89$	Clauses and larger seq.	–	Segments: DA: $\kappa = .79-.81$ IT: $\kappa = .87-.89$	Guidelines (Navarretta and Olsen, 2009)	Annotation tool PALinkA
Poesio and Artstein (2008), Artstein and Poesio (2006, Experiment 1)	EN	ARRAU (mixed: TRAINS dialogue, narrative, WSJ, ...)	T: 95K, C: 24,321, <b>AA: 455</b>	all NPs/ Prons	Multiple labels (Sec. 3.1.)	–	Arbitrary sequences/ clauses	–	Exp.1, TRAINS (T: 1,421, C: 181, AA: 35): $\alpha = .55$ (for best 16 AA)	MATE/ GNOME manuals (modified)	Annotation tool MMAX
Dipper and Zinsmeister (2009a)	DE	Europarl (parliament proc.)	T: ? (32 texts), C: 48, <b>AA: 48</b>	Dem ( <i>dites</i> ‘this’)	Multiple labels (Sec. 3.1.)	$\alpha = .37-.66$	Arbitrary sequences	Multiple labels (Sec. 3.1.)	Segments: 85%, labels: $\alpha = .52-.60$	Guidelines (Dipper and Zinsmeister, 2009b)	
Recasens and Martí (2010), Recasens (2008)	CA, ES	AnCora (newsper texts)	CA: T: 385K, C: 120K, <b>AA: 643</b> ; ES: T: 420K, C: 135K, <b>AA: 748</b>	all NPs/ Prons (incl. Cl, Zero; no Expl)	AA with subtypes (Sec. 3.1.), others (4)	ES: <i>coref</i> vs. <i>non-coref</i> : $\alpha = .85-.90$	Verbs, clauses and larger seq.	–	ES: $\alpha = .85-.89$	Guidelines (Recasens et al., 2007)	PALinkA and AnCoraPipe annotation tools