Coreference Annotation Guidelines[1]

Erwin R. Komen
Radboud University, Nijmegen, Netherlands
June 2009

Summary

These guidelines describe the principle of enriching texts from syntactically annotated corpora with coreference information. This information can be added manually or with a program like CESAC. This paper explains the different types of relationships used for coreference links. It sets out general guiding principles that can be used to determine which NPs should be linked to which other NPs or IPs. Furthermore, it illustrates which types of relationships are appropriate in which circumstances.

These guidelines contain a decision tree that should help prospective annotators in their task.

1. Introduction

Syntactically annotated corpora like the YCOE and the PPCME2 are currently being enriched with coreference information. Such information is added as a basis for research into information structure phenomena. These guidelines establish a common practice for the coreference annotation. If these guidelines are adhered to, then coreference annotation can be provided in a consistent enough way, so that texts, enriched with coreference information, can confidently be used for research purposes.

While care has been taken to address all kinds of situations where choices have to be made by the coreference annotator, it is quite likely that unforeseen situations come up. When the guidelines provided in this document do not suffice, I would kindly request people to supply enough information, so that such situations can be addressed too.

2. Coreference annotating

The process of providing coreference annotation in the program CESAC is described in the manual of Cesac (Komen, 2009). This section, however, discusses some important points within the process of coreference annotation. Consider the following small story:

Figure 1 The invitation

As the sun was setting, Jim stared out of his window. He could see the shadows grow larger. The doorbell unexpectedly interrupted his thoughts. He went downstairs and opened the door. "Hello", said his neighbour, "I thought I'd invite you for a cup of coffee. That is, if you can spare a moment from your busy life." His "busy life"? What was his neighbour talking about? Didn't he know, that Jim had been fired a week ago now? Would he accept the invitation?

In linguistics it is common practice to use indices in order to keep participants apart. So consider how the story looks like with indices on the main participants.

Figure 2 An indexed version of "The invitation"

As the sun was setting, Jim$_i$ stared out of his$_i$ window. He$_i$ could see the shadows grow larger. The doorbell unexpectedly interrupted his$_i$ thoughts. He$_i$ went downstairs and opened the door. "Hello", said his$_i$ neighbour$_j$, "I$_j$ thought I$_j$'d invite you$_i$ for a cup of coffee. That is, if you$_i$ can spare a moment from your$_i$ busy life." His$_i$ "busy life"? What was his$_i$ neighbour$_j$ talking about? Didn't she$_j$ know, that Jim$_i$ had been fired a week ago now? Would he$_i$ accept the invitation?

While the indices *i* and *j* here do show who is who, they don't really help in establishing how things are related to one another. Besides, there are several definite NPs that don't have an index, but they are clearly in some kind of relationship with others. For instance the *shadows* result from the *setting sun*. So they don't come completely out of the blue, but are inferred. In order to provide the basic instruments to derive such coreference relations, we need to provide all necessary NPs (and perhaps some IPs too) with unique labels. Let us have a look at the story once more, but now with the labels.

Figure 3 A labelled version of "The invitation"

As [$_1$ the sun] was setting, [$_2$ Jim] stared out of [$_3$ [$_4$ his] window].
[$_5$ He] could see [$_6$ the shadows] grow larger.
[$_7$ The doorbell] unexpectedly interrupted [$_8$ [$_9$ his] thoughts].
[$_{10}$ He] went downstairs and opened [$_{11}$ the door].
    "Hello", said [$_{12}$ [$_{13}$ his] neighbour],
    "[$_{14}$ [$_{15}$ I] thought [$_{16}$ I]'d invite [$_{17}$ you] for [$_{18}$ a cup of coffee]].
    That is, if [$_{19}$ you] can spare [$_{20}$ a moment from [$_{21}$ [$_{22}$ your] busy life]."
[$_{23}$ [$_{24}$ His] "busy life"]?
What was [$_{25}$ [$_{26}$ his] neighbour] talking about?
Didn't [$_{27}$ he] know, that [$_{28}$ Jim] had been fired [$_{29}$ a week ago] now?
Would [$_{30}$ he] accept [$_{31}$ the invitation]?

Pronouns like [$_9$ his], [$_5$ he] and [$_4$ his] all point back to [$_2$ Jim]. For our coreference annotation task we should provide relationships of the type "Identity" between these referring expressions. Notice that the relationships point backwards. That is, 9 points back to 5, 5 to 4, and 4 to 2. Each labelled NP can have exactly one link with a previous occurrence of a referent. This link should give the following information:
a) The identifying label of the source NP.
b) The identifying label of the NP or IP to which it is referring back (or forward).
c) The kind of relationship established between the two referents.
Our final version of the story has the coreference information added to each source NP where it is applicable.

Figure 4 An annotated version of "The invitation"

As [$_1$ the sun] was setting, [$_2$ Jim] stared out of [$_3$ [$_{4\text{-Identity-}2}$ his] window].
[$_{5\text{-Identity-}4}$ He] could see [$_{6\text{-Inferred-}1}$ the shadows] grow larger.
[$_{7\text{-Inferred-}3}$ The doorbell] unexpectedly interrupted [$_8$ [$_{9\text{-Identity-}5}$ his] thoughts].
[$_{10\text{-Identity-}9}$ He] went downstairs and opened [$_{11\text{-Inferred-}7}$ the door].
    "Hello", said [$_{12}$ [$_{13\text{-Identity-}10}$ his] neighbour],

"[14 [15-CrossSpeech-12 I] thought [16-Identity-15 I]'d invite [17-CrossSpeech-13 you] for [18 a cup of coffee]].
That is, if [19-Identity-17 you] can spare [20 a moment from [21 [22-Identity-19 your] busy life]."
[23-Identity-21 [24-CrossSpeech-22 His] "busy life"]?
What was [25-CrossSpeech-16 [26-Identity-24 his] neighbour] talking about?
Didn't [27-Identity-25 he] know, that [28-Identity-26 Jim] had been fired [29 a week ago] now?
Would [30-Identity-28 he] accept [31-Identity-14 the invitation]?

The colour coding shows how different coreference *chains* run through the story. The blue coloured NPs all refer to Jim. Most relationships in this chain are of the type "Identity". However, while NP number 17 refers to Jim, it does so across a direct speech border. Therefore this relationship is of type "CrossSpeech". The same kind of relationship occurs for NP number 24, which refers from the narrator back into the direct speech, and thereby crosses a direct speech border.

There are several shorter chains, where one NP is linked with another one in an "Inferred" relationship. For instance, NP 6 *the shadows* are inferred from NP 1 *the sun*. The *door* of 11 is inferred from *the doorbell* in 7. Actually, *window* in 3 triggers "house" in the minds of the reader, and both *doorbell* in 7 as well as *door* in 11 are more closely related to this unmentioned "house" (they are in a PartOfWhole relationship). But for practical purposes it suffices to relate 11 to 7 with "Inferred", and 7 to 3 with "Inferred".

We should pay attention to the "Identity" relationship between the NP *the invitation* in 31, and the IP *I thought I'd invite you for a cup of coffee* in 14. Sometimes coreference relations exist between an NP and a sentence, or even a larger stretch of text. Since the referents do coincide in such situations, an "Identity" relation is in order. When there is a set of IPs to choose from, all of which are equally applicable, then NEARFIRST in section 3says that we establish the link with the nearest IP.

3. Common rules

Before addressing specific situations, I would like to set a handful of common rules that should provide a clear starting point for the whole idea of coreference annotation. These rules might be perceived as trivial at first glance, but some of them will prove to come in quite handy, when more complex situations are addressed.

(1) **DOCP**: do not overlook coreference possibilities.
Any NP that *could* receive coreference information *should* receive it. All pronouns are by definition referring expressions and therefore *must* receive coreference information.

(2) **NEARFIRST**: choose the nearest antecedent.
An NP should establish a coreference relation with the nearest antecedent possible.

(3) **IDENTITYFIRST**: choose an "Identity" type relation.
If it is possible to refer back to an antecedent with an Identity relationship, then this should be done.

The first rule is an important one. Whenever there are situations when a coreference relation should be established, an annotator must establish it. A program like Cesac allows the user to determine which syntactical situations *must* receive a coreference link. One rule in Cesac could for instance stipulate that any NP immediately dominating a Pronoun must have a

coreference link. The appendix of this document contains a recommendation for phrase types that should receive coreference information. Cesac allows for an easy way to check whether there are any phrases left that must still be checked.[2]

The second and third rules, NEARFIRST and IDENTITYFIRST, could be in conflict with one another. There might be an antecedent, with which a relation that is not of type *Identity* could be established. Let us first have a look at the different types of coreference relations that could be used by annotators. For information structure purposes it seems best to limit the number of different types to a bare minimum. Such an approach should reduce potential interrater disagreement. Table 1 presents the coreference relation types I propose to use.

Table 1 Coreference relation types

Type	Meaning
Identity	Referring expressions point to the same entity.
CrossSpeech	Referring expressions point to the same entity (so an *Identity* relation is implied), but the link crosses a direct speech boundary (from speech to narration or from narration to speech—but not *between* speech parts!).
Inferred	The entity referred to by the source of the link is loosely connected to the entity referred to by the link's destination. Such a connection could be something like part of a whole (a body's feet, hands, nose etc), or subset (one member from a group), or something else.

Let us now return to the possibility of a tie between NEARFIRST and IDENTITYFIRST. The order of the coreference relation types is of importance. A *CrossSpeech* relation should be preferred over an *Identity* one, and an *Identity* relation outweighs an *Inferred* one. The hierarchy of relations is as shown in (4).

(4) CrossSpeech >> Identity >> Inferred

So when there seems to be competition between NEARFIRST and IDENTITYFIRST, it is IDENTITYFIRST that wins, unless the coreference link crosses a direct speech boundary. In such a situation *CrossSpeech* is chosen.

4. Specific situations

This section addresses several specific situations, where different choices are in principle possible. We propose several rules, in order to come to a unified coreference annotating approach.

4.1. Identity

Coreference relations of the type *Identity* must adhere to strict rules. This means, for instance, that they must agree in number. In principle, an *Identity* relation can only be established between two referents that are identical. This is made explicit by the following rule.

(5) **IDENTITY**: Only establish *Identity* relations between identical referents.
    The referent of a link's source should completely coincide with a link's destination, when a link of type *Identity* is established.

---

[2] Selecting **M**ust/**F**irst (or F7) under Cesac takes the user to the first phrase that must receive coreference information, and has not received it yet.

Care must be taken to not establish a link with an antecedent that is *related* to the referent we would like to refer back to, but not *identical* to it. Such a situation is illustrated in Figure 5. There should be a coreference relation between *an expert in applied physics* and its antecedent *you*. The type of such a relation can safely be regarded as *Identity*. However, this relation is unidirectional. It is true that *you* is an *expert*, but not every *expert* is identical to *you*. So the next pronoun *your* can not refer back to this *expert*. It must refer back to *you* instead.

Figure 5 Avoid identity mismatches

> "Good afternoon, **sir**. I am calling **you**, since **you** are **an expert in applied physics**. I would very much like to hear **your** opinion in a situation, which I shall now describe to **you**."

The situation sketched in Figure 5 leads to a *forking* situation. This is illustrated in Figure 6. The chain of references *sir, you, you, your, you* forks at the second *you*. Forking results in this and analogous situations like appositions (section 4.3) and bound anaphors (section 4.4). The common reason for forking is the appearance of a link that is not symmetrically *Identity*.

Figure 6 Forking
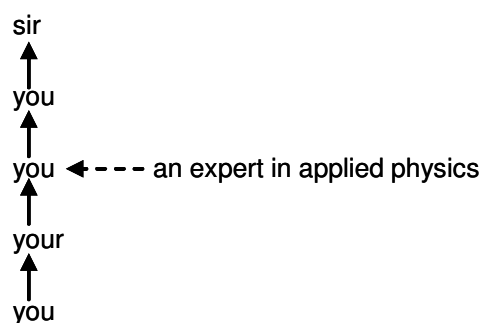
One final word about the *Identity* type coreference relation concerns question words. Since a *wh*-question word is an open variable, it cannot refer back to an entity. So no coreference relation should be made from any *wh*-question word.

4.2.   Direct speech

Direct speech has been touched upon in the introduction to the whole process of coreference annotating in chapter 2. This section adds some specific guidelines for direct speech situations.[3] The fundamental idea is that any link across a direct speech boundary should be labelled as *CrossSpeech*.

First, consider the situation where direct speech follows a narrative section. Suppose a participant is mentioned both in the direct speech as well as in the subsequent narrative section. Such a situation is shown in Figure 7. The references to *Jim* are third person pronouns in the narrative part, and then the direct speech contains a second person pronoun referring to *Jim*. It is true that *you* refers back to the same person *Jim* as do the third person pronouns. However, a direct speech border is crossed, and for that reason I suggest that the link between *you* in the direct speech and *his* in the clause *said his neighbour* should be of type *CrossSpeech*.

---

[3] These guidelines are mainly derived from discussions in the ETC group of the Radboud University (Dreschler, 2009).

Figure 7 Linking back from direct speech into a narrative

> As the sun was setting, **Jim** stared out of **his** window. **He** could see the shadows grow larger. The doorbell unexpectedly interrupted **his** thoughts. **He** went downstairs and opened the door.
> > "Hello", said **his** neighbour, "I thought I'd invite **you** for a cup of coffee".

Only the first mention of *Jim* in direct speech has a link crossing a direct speech boundary, so subsequent mentions of Jim within the direct speech are linked together with *Identity*.

Next, let us consider the situation where a participant is mentioned in direct speech, and then picked up again in a subsequent narrative section, as illustrated in Figure 8.

Figure 8 Linking back from narrative into direct speech

> As the sun was setting, **Jim** stared out of **his** window. **He** could see the shadows grow larger. The doorbell unexpectedly interrupted **his** thoughts. **He** went downstairs and opened the door.
> > "Hello", said **his** neighbour, "I thought I'd invite **you** for a cup of coffee. That is, if **you** can spare a moment from **your** busy life."
> **His** "busy life"? What was **his** neighbour talking about?

Jim is referred to with a third person pronoun in the narrative before and after the direct speech interruption. However, the nearest possible antecedent of *his* in *His "busy life"?* is the second person pronoun *your* in the direct speech. I suggest that a coreference relation be established between *his* and *your* of type *CrossSpeech*. This is completely in line with the prominent place of the NEARFIRST principle, as discussed in chapter 3.

Finally, let us consider a situation where a participant in two direct speech sections, while it does not occur in a preceding narrative section, as illustrated in Figure 9. Jim's neighbour mentions Jim's *daughter* in his direct speech. Subsequently, Jim answers and also refers to his daughter, only now he does so by her name *Sheila*. Both instances refer to the same person and exist within the same "world"—the direct speech realm. I suggest therefore, that *Sheila* be linked back to *your daughter* with an *Identity* type relation.

Figure 9 Linking between direct speech sections

> > "Hello", said his neighbour, "I thought I'd invite you for a cup of coffee. That is, if you can spare a moment from your busy life. Do bring **your daughter**, by the way."
> His "busy life"? What was his neighbour talking about?
> > "Thanks for the invitation, Jack. I think I'll accept it. But **Sheila** is not at home right now."

Some final words are necessary at this points, in order to clarify the relationship between the *CrossSpeech* type relation and other types like *Identity* and *Inferred*. Basically, any *CrossSpeech* relation implies that the referents of the source and target are identical. So by default a relation of type *CrossSpeech* is also of type *Identity*.

When a relation is made between a source and target that crosses a direct speech boundary, yet the referents do *not* completely coincide, it is *not* possible to use *CrossSpeech*, because using *Crossspeech* implies *Identity*. In such case, when it is indeed necessary to make an explicit link, the type *Inferred* should be used.

4.3.   Apposition

Consider a situation such as illustrated in Figure 10, where an NP is specified further by another NP appositive to it. Clearly the NP *Jim's daughter* uniquely specifies *Sheila*, so there should be an *Identity* relation between them. The question is whether *she* should point back to *Sheila* or to *Jim's daughter*. I propose that *she* should point back to the head of the apposition relation, which is *Sheila*. Let us see why this makes sense, by looking at one more apposition relation in Figure 10.

The NP *a very observant person* is in an apposition relation to *their next-door neighbour*. While the *next-door neighbour* has a uniquely identifiable referent, the NP *a very observant person* does not. In fact, it is impossible to refer back to *a very observant person* with a pronoun in subsequent sentences, due to semantic scope restrictions. Since it is clear in this situation that the apposition's head should be the NP referred back to subsequently, the same principle should better apply to all apposition situations.

Figure 10 Apposition relations

Shortly after his neighbour had left, **Sheila**, Jim's **daughter**, came home from school. **She** called: "Dad, where are you?"
Their next-door neighbour, a very observant person, heard Sheila shouting.

The explanations above lead to the guideline spelled out in (6).

(6) **APPOSITIONHEAD**: Refer back to apposition's head.
An NP should establish a coreference relation with the head of an apposition NP, and not with the daughter of such an NP.

In the Old and Middle English text corpora appositive NPs have the extension –PRN (which is also used for "parentheticals"). Rule (6) implies, therefore, that no NP with extension –PRN can be referred back to. Linking should always be made to the head of an apposition relation (or, for that matter, of a parenthetical relation).

4.4.   Bound anaphors

Reflexives like *myself, himself* and *yourself* are locally bound to the subject of the sentence. Such NPs are pronouns, and therefore must receive coreference information (see the DOCP rule in chapter 3). Since they are locally bound, just like appositive NPs, no further reference to them is possible. I suggest the following guideline for bound anaphors:

(7) **BOUNDANAPHOR**: Locally bound anaphors should be linked to their binder.
A coreference relation of type *Identity* must be made for any locally bound anaphor.

4.5.   Empty NPs

Texts taken from the YCOE (York-Toronto Corpus of Old English texts) and PPCME2 (the second edition of the Penn-Helsinki Parsed Corpus of Middle English) code several instances if elided (null) subjects and objects (Kroch & Taylor, 2000; Taylor, Warner, Pintzuk, & Beths, 2003). I suggest that at least one category of empty NPs should be included in the process of coreference annotation. This is the category of "empty subjects elided under conjunction". Such NPs are labelled as *con* (Taylor, 2003). Since it could be important for the work on Information Structure to keep track of the usage of empty subjects, I suggest a guideline spelled out in (8).

(8) **EMPTYSUBJECTS**: Empty subjects elided under conjunction should be linked.
An empty subject elided under conjunction is labelled as *con*. Such an NP should be linked with an antecedent.

Appendix 7.3 explains how Cesac can be configured such, that empty subjects are categorized as phrases that must be supplied with coreference information.

4.6. Inferred relations

The default relation "when all else fails" is called *Inferred*. Several of these situations are illustrated in Figure 11.

First, there is a larger group of *three daughters*. One member, a subset, of this group is *his youngest daughter Sheila*. The relationship between the one member and the whole group is not one of *Identity*. A relationship of *Inferred* is chosen instead.

Next, in the second paragraph of the story there is a larger whole, which is the *body* of Jim. In subsequent clauses mention is made of parts of this whole: feet, ankles, muscles, head. In this particular situation there would be four *Inferred* relations established, all of which have *his whole body* as target.

Figure 11 Inferred relations

Jim had been running through the woods. **His whole body** hurt. His **feet** were sore, his **ankles** were hurting, his **muscles** tired, and his **head** was bursting.

Jim had **three daughters**. Shortly after his neighbour had left **his youngest daughter Sheila** came home from school. **She** called: "Dad, where are you?"
Their next-door neighbour, a very observant person, heard **Sheila** shouting.
A car arrived. The **driver** came out.

Other possible *Inferred* relations could be established between *a car* and *the driver* in the final paragraph of the story. If the driver had been mentioned before, then the hierarchy from (4) says that an *Identity* relation must be established if there is one (no matter whether such a relation has to be established over a greater distance). If this is the first mention of the driver's referent, then it would really be wise to establish an *Inferred* type link with *a car*, since *the driver* is a definite NP.

4.7. IP-selection

NPs do not only refer back to other NPs, but coreference relations can also be established between an NP and an IP. Such an NP should point to some kind of action, and the IP it refers back to should be chosen with care. It should denote the core idea referred to be the NP. This is captured in the following guideline.

(9) **IP-SELECTION**: Choose an IP target with care.
An IP referred to by an NP must express the core idea of the NP source.

Many times there is more than one IP being involved with the idea to which the NP is referring. In such cases NEARFIRST applies: select the IP nearest to the source NP.

4.8. Identity versus Inferred

There are situations where it seems to be difficult to decide upon the type of coreference relation. First there is a possibility to refer back to a nearer by NP, using a relationship of the type *Inferred*. Alternatively, there is the possibility to refer back to an NP that is further away, but with a relationship of type *Identity*. Such a situation is exemplified in Figure 12.

There seem to be two options for the antecedent of the NP *his daughter*. First, an *Inferred* relation could be established with the nearer by pronoun *he*, since *daughter* is inferred from the fact that *he = Sheila's father*. Alternatively, an *Identity* relation could be established with the pronoun *her*, since this is the nearest by entity pointing to the same referent.

Figure 12 Competition between Identity and Inferred

Sheila's father had been talking to **her**. **He** wasn't quite sure whether **his daughter** finally understood the core of the matter. Would **daughters** ever?

In these and other cases rule (4) provides a clear answer. It states that there is a preference in relationships. Whenever there exists an antecedent with which a relation of type *Identity* can be established, one should not resort to establishing a relation of type *Inferred* with a closer-by antecedent.

4.9.   Anchored NPs

Anchored NPs contain a possessive pronoun like *his* or a genitive like *Jim's*. Such an anchor (the possessive pronoun or the genitive) is already linked to an antecedent. Therefore, care must be taken when linking the whole NP to a possible antecedent. If there is a clear antecedent for the whole NP, then an *Identity* relation must be made with it, since rule DoCP applies. However, if there is no antecedent with which an *Identity* relation can be made, then there is no need to make a link of type *Inferred* from the whole NP to any appropriate antecedent, since there already is an "anchor" running from within the NP (the possessive pronoun or the genitive) with type *Identity* to an appropriate antecedent.

Figure 13 Anchored NPs

As the sun was setting, Jim stared out of **his window**. He could see the shadows grow larger. The doorbell unexpectedly interrupted **Jim's thoughts**. He went downstairs and opened the door. "Hello", said **his neighbour**, "I thought I'd invite you for a cup of coffee. That is, if you can spare a moment from **your busy life**." **His "busy life"**? What was **his neighbour** talking about? Didn't he know, that Jim had been fired a week ago now? Would he accept the invitation?

Rule (10) captures this coding practice:

(10) **ANCHOREDNP**: Do not link an anchored NP unless there is a clear antecedent. An anchored NP already has an identity link from one of its elements to an antecedent. If the whole NP can be linked with *Identity* to an antecedent, this must be done. Otherwise, the whole NP should not be linked further.

5. Conclusions

This paper describes guidelines for enriching syntactically annotated corpora with coreference information. A small set of three core principles is provided, while a preference ranking is determined between them. Next, a set of supplementary rules is proposed. When all guidelines are adhered to, it is hoped that the process of enriching corpora with coreference information can be done with a high interrater agreement.

6. References

Dreschler, G. (2009). *Codeerafspraken over (direct) speech in een narrative*. Nijmegen, Netherlands: Radboud University.

Komen, E. R. (2009). *CESAC: Coreference Editor for Syntactically Annotated Corpora.* Nijmegen: Radboud University.

Kroch, A., & Taylor, A. (2000). *Penn-Helsinki Parsed Corpus of Middle English, second edition.* Retrieved 19/Jun/2009, from http://www.ling.upenn.edu/hist-corpora/PPCME2-RELEASE-2/

Taylor, A. (2003). *The York-Toronto-Helsinki Parsed Corpus of Old English Prose*. Retrieved 18/Jun/2009, from http://www-users.york.ac.uk/~lang22/YCOE/doc/annotation/YcoeRef.htm#elision_nps.

Taylor, A., Warner, A., Pintzuk, S., & Beths, F. (2003). The York-Toronto-Helsinki Parsed Corpus of Old English Prose. *Electronic texts and manuals available from the Oxford Text Archive*, from http://www-users.york.ac.uk/~lang22/YCOE/YcoeHome.htm

7. Appendix: essentials

This appendix gives a summary of the essential information provided by this coding manual.

7.1. Coreference relation types

Table 1 is repeated here for convenience.

Type	Meaning
CrossSpeech	Referring expressions point to the same entity.
Identity	The source is an anaphor locally bound to the destination.
BoundAnaphor	Referring expressions point to the same entity (so an _Identity_ relation is implied), but the link crosses a direct speech boundary (from speech to narration or from narration to speech—but not *between* speech parts!).
Inferred	The entity referred to by the source of the link is loosely connected to the entity referred to by the link's destination. Such a connection could be something like part of a whole (a body's feet, hands, nose etc), or subset (one member from a group), or something else.

7.2. Coding rules

The rules introduced in this document are repeated here for convenience.

(1) **DOCP**: do not overlook coreference possibilities.
   Any NP that *could* receive coreference information *should* receive it.

(2) **NEARFIRST**: choose the neares antecedent.
   An NP should establish a coreference relation with the nearest antecedent possible.

(3) **IDENTITYFIRST**: choose an "Identity" type relation.
   If it is possible to refer back to an antecedent with an Identity relationship, then this should be done.

(4) CrossSpeech >> Identity >> Inferred

(5) **IDENTITY**: Only establish _Identity_ relations between identical referents.
   The referent of a link's source should completely coincide with a link's destination, when a link of type _Identity_ is established.

(6) **APPOSITIONHEAD**: refer back to apposition's head.
   An NP should establish a coreference relation with the head of an apposition NP, and not with the daughter of such an NP.

(7) **BOUNDANAPHOR**: Locally bound anaphors should be linked to their binder.
   A coreference relation of type _Identity_ must be made for any locally bound anaphor.

(8) **EMPTYSUBJECTS**: Do link empty subjects elided under conjunction.
   An empty subject elided under conjunction is labelled as *con*. Such an NP should be linked with an antecedent.

(9) **IP-SELECTION**: Choose an IP target with care.
   An IP referred to by an NP must express the core idea of the NP source.

(10) **ANCHOREDNP**: Do not link an anchored NP unless there is a clear antecedent.
An anchored NP already has an identity link from one of its elements to an antecedent. If the whole NP can be linked with _Identity_ to an antecedent, this must be done. Otherwise, the whole NP should not be linked further.

7.3. Cesac settings

Table 2 presents Cesac settings that I propose as standard ones. First the situations are defined where a coreference annotation _must_ be made. Line #1 states that all NPs that have a determiner (defined as D^*) as a first child must receive coreference information. Especially for Old English this is an important one (for Middle English and early Modern English indefinite NPs are, unfortunately, also selected as "must"). In general such NPs are definite ones, and are referring back to an antecedent.

Lines 2 and 3 state that NPs having a DPRO (demonstrative pronoun) or a PRO (pronoun) as first child must be linked to an antecedent.

Line #4 continues by stating that all NPs having an empty subject elided under conjunction must receive coreference information.

Line #5 singles out possessive pronouns, which must also be linked to an antecedent.

Next, line #6 states that all appositive NPs must be visited, and line #7 states that all reflexive NPs must be addressed.

In general _all_ NPs _could_ refer back to something, which is what line #8 says. For the moment line #9 also includes ADVPs. On top of that the genitive forms of names (NPR$) form their own category, as stated by line #10.

IPs and PPs can serve as destinations for coreference relations, but not as sources, as stated by line #11 and #12.

Table 2 Standard settings for Cesac

#	Node	Type	Target	Child
1	NP	Must	Any	D
2	NP	Must	Any	DPRO
3	NP	Must	Any	PRO
4	NP	Must	Any	*con*
5	PRO$	Must	Any	*
6	NP*PRN	Must	Any	*
7	NP*RFL	Must	Any	*
8	NP	Can	Any	*
9	ADVP	Can	Any	*
10	NPR$	Can	Any	*
11	IP	Can	Dst	*
12	PP	Can	Dst	*

7.4. Decision tree

Figure 14 is a practical help for deciding if a coreference relation should be made, and of what type it should be made.

Figure 14 Decision tree