13. Appendix

13.1. Bridging in WSJ documents: annotation guidelines

13.1.1. Markables

Markables (and thus candidates for bridging anaphors) are all NPs that have been gold annotated in the OntoNotes corpus. These markables are highlighted in green in the annotation web interface.

Pre-marked NPs in OntoNotes include

- nominal phrases: *the president*

- proper names: *Mr. Bush*

- pronouns: personal, possessive, demonstrative, reflexive

- quantifier phrases: *all the products*

They do not include

- nominal premodification: the *US* president

- interrogative or relative pronouns

Choose the longest span of the NP that refers to the entity, including determiners and adjectives, dependent PPs and relative clauses.

(1)   There have been concerns that the Big Board's basket could attract investors with a short term perspective who whould rapidly turn over the product, thus increasing volatility.

13.1.2. Antecedents

As a general principle, one antecedent has to be chosen. In special cases, e.g. comparative cases where two antecedents are needed, the annotator may create two or several links.

(2) <u>President Bush</u>, <u>the Canadian prime minister</u> and **14 other members of the Committee**.

We include nominal and abstract antecedents. In case of abstract anaphora, the abstract antecedent (VP or clausal) has to be added as a markable in order to create a link, as links can only be added between two markables in the annotation tool.

(3) <u>What is the meaning of life?</u> **The answer** cannot be expressed in one sentence.

13.1.3. Link types

As there are different types of links covered under the term " bridging" in previous annotation efforts, we distinguish a number of bridging types, for pragmatic reasons. The phenomena can then be studied separately, if needed. Bridging-contained, as described in Riester and Baumann 2012 is not annotated as bridging because it is not an anaphoric phenomenon and as such a special case where the antecedent modifies the bridging anaphor.

(4)   *the windows in the room*

(5)   *the mother's room* or *her room*

13.1.4. Bridging

Bridging anaphora, or sometimes called associative anaphora, are discourse-new, anaphoric expressions which are dependent on the previous context, and for which the text presents an antecedent NP which does not stand in the relation of identity, but in some other form of relation to the associative phrase. The antecedent may be an associate in a typical relation such as part-of, is-a, or any kind of associate as long as there is a clear relation between the two phrases.

(6)   We use <u>a classifer</u> to distinguish between the two categories.
        The training data consists of ...

(7)   I went to <u>a wedding</u> last weekend. **The bride** was a friend of mine. She baked
        the cake herself.

(8)   Our correspondent in <u>Egypt</u> is reporting that **the opposition** is holding a rally
        against **the constitutional referendum**.

(9)   <u>What is the meaning of life?</u> **The answer** cannot be expressed in one sentence.

The anaphor is lacking an implicit argument (the antecedent) which enables the interpretation of the expression.

(10)   **the opposition** (in Egypt)

(11)   **the answer** (to this question)

Definite use

Most bridging anaphors are definite NPs. In this case, just link the markable with an antecedent of your choice.

(12)   I went into <u>the room</u>. **The windows** were broken.

Note that bare singulars can sometimes also count as definite, in case when the insertion of the definite article is more plausible than the insertion of an indefinite article. Bare plurals always count as indefinites.

(13)   We performed <u>the experiments</u> using ... . **Evaluation** is done by means of
         10-fold cross validation.

Indefinite use

Some bridging anaphors are indefinite expressions. In this case, add an indefinite label to the markable and link it to the preferred antecedent.

(14)   I bought <u>a bicycle</u>. **A tire** was already flat.

(15)   <u>Health insurance</u> ... **Costs** have exploded over the last ten year.

Comparative anaphora

Comparative anaphora have been excluded from the bridging category and treated as a separate category in the ISNotes corpus. We label these cases as comparative and link the comparative markable to the antecedent.

(16)   About 200,000 East Germans marched in Leipzig and thousands more staged protests in **three other cities**

(17)   President Bush, the Canadian prime minister and **14 other members of the Committee**.

13.1.5. Annotation process

The annotation is done using the annotation tool Slate (Kaplan citation). The markables are presented in green. Coreferent entities shown in red are already marked and can thus not be marked as bridging. Exceptions are the first mention in a coreference chain, which can of course be of the category bridging. In case of definite bridging anaphora, the annotator should link the green markable to the best-suited antecedent. In case the antecedent is not a pre-selected markable, a markable should be created for the antecedent in order to place the link. In other cases (indefinite anaphors, comparative anaphors), the NP should be labeled with the respective category and then this newly created entity should be linked to the antecedent. We refrain from annotating attributes in order not to complicate the annotation process.