same ilocutive value but different meaning (b), cases of subjectivity (c), cases of potential coreference (d), (e) and (f), etc.

a) The two had argued that <u>you shouldn't go there</u> .
   He and Zilkha believed that <u>this is unfair</u> .

b) I want some fresh air.
   Could you open the window?

c) The U.S.-led <u>invasion</u> of Iraq .
   The U.S.-led <u>liberation</u> of Iraq.

d) They got married <u>last year</u> .
   They got married <u>in 2004</u> .

e) I live <u>here</u> . I live <u>in Barcelona</u> .

f) They will come <u>later</u> .
   They will come <u>this afternoon</u>

N.B.: Paraphrase and coreference overlap considerably. Those cases that may corefer, but at the same time are paraphrases, should be annotated as paraphrases.

In cases (d), (e) and (f), the linguistic information is not enough to link the two members of the pair, we need to know which point in the time or in the space are we taking as reference. Thus, they are annotated as nonparaphrases. Cases in (g), (h) and (i) can be linked only through linguistic information (a year in the past, a 'city' type of entity, a masculine singular entity, respectively). Thus, they are annotated as paraphrases.

g) They got married <u>last year</u> .
   They got married <u>a year ago</u> .

h) I live in <u>Barcelona</u> .
   I live <u>in a city</u> .

i) I love <u>John</u> .
   I love <u>him</u> .

N.B.: Although sometimes a non-paraphrase fragment may actually affect the meaning of the full sentence, only the fragment in question will be tagged as NON-PARAPHRASE (j) and the rest of the sentence will be annotated independently of this fact.

j) <u>Mike and Lucy</u> decided to leave .
   <u>Mark</u> decided to leave .

N.B.: When two linguistic units having a different meaning are not aligned formally nor informatively, they should be tagged as two different ADDITION/ DELETION cases (1 and 2 in k), not as NON-PARAPHRASES.

k) Yesterday,$_1$ Google failed .
Google failed because of the crisis$_2$.

A.3.7.3   Entailment

Definition: Fragments having an entailment relation. **N.B.:** It should be noted that entailment relations are present in many paraphrase types (e.g. general/specific in SAME-POLARITY or ADDITION/DELETION). We will only use the EN-TAILMENT tag when there is a substantial difference in the information content. Entailment is always **negative sense preserving**.

- Google was in talks to buy Youtube .
Google bought Youtube

A.4   Annotating non-paraphrases

Annotating non-paraphrases (negative examples of paraphrasing in the MRPC corpus) is a non-trivial task that has not been carried out for other paraphrase typology corpora. The non-paraphrases in the MRPC corpus have many of the properties of paraphrases, they have a very high degree of lexical and syntactic similarity. In a) we can see an example of a non-paraphrase pair. The two sentences talk about the same NEs (Yucaipa and Dominick) in the same syntactic-semantic roles of the same actions (buying, selling, owning). At the same time, there are key differences between the two sentences – the price of the sale in the first sentence is $2.5 billion, while in the second it is $1.8 billion.

a) Yucaipa owned Dominick's before selling the chain to Safeway in 1998 for $2.5 billion.
Yucaipa bought Dominick's in 1995 for $693 million and sold it to Safeway for $1.8 billion in 1998.

Due to the complex nature of the non-paraphrasing, the annotation of these pairs goes in three steps

1) (Re)evaluation of the paraphrasing or non-paraphrasing relation between the two sentences as a whole (this is the first step for both paraphrases and non-paraphrases).

2) (After the pair has been annotated as non-paraphrases) Annotation of the non-sense-preserving phenomena, responsible for the non-paraphrasing label of the pair.

3) Annotation of the sense-preserving phenomena, responsible for the high degree of similarity between the two sentences.

An example annotation of the pair in a) follows:

1) The relation between the two sentences is non-paraphrases

2) The non-sense-preserving phenomena responsible for the "non-paraphrase" label of the pair is "Lexical Substitution (Named Entities)":
Yucaipa owned Dominick's before selling the chain to Safeway in 1998 for <u>$2.5 billion</u> .
Yucaipa bought Dominick's in 1995 for $693 million and sold it to Safeway for <u>$1.8 billion</u> in 1998.

3) The sense-preserving phenomena, responsible for the high degree of similarity are:

    a. Same polarity substitution (contextual)
    Yucaipa owned Dominick's before selling <u>the chain</u> to Safeway in 1998 for $2.5 billion.
    Yucaipa bought Dominick's in 1995 for $693 million and sold <u>it</u> to Safeway for $1.8 billion in 1998.

    b. Entailment
    Yucaipa <u>owned</u> Dominick's before selling the chain to Safeway in 1998 for $2.5 billion.
    Yucaipa <u>bought</u> Dominick's in 1995 for $693 million and sold it to Safeway for $1.8 billion in 1998.

    c. Inflectional changes
    Yucaipa owned Dominick's before <u>selling</u> the chain to Safeway in 1998 for $2.5 billion.
    Yucaipa bought Dominick's in 1995 for $693 million and <u>sold</u> it to Safeway for $1.8 billion in 1998.

    d. Order
    Yucaipa owned Dominick's before selling the chain to Safeway <u>in 1998</u> for $2.5 billion.
    Yucaipa bought Dominick's in 1995 for $693 million and sold it to Safeway for $1.8 billion <u>in 1998</u>.