"*" element taken from the Treebank parse which serves as a placeholder for the missing pronoun.

In order to keep the annotation feasible at high agreement levels, only intra-document anaphoric coreference is being marked. Furthermore, while annotation is not limited to any fixed list of target entity types, noun phrases that are generic, underspecified, or abstract are not annotated.

Attributive NPs are not annotated as coreference because the meaning in such cases can be more appropriately taken from other elements in the text. For example, in "New York is a large city", the connection between New York and the attributive NP "a large city" comes from the meaning of the copula "is". Similarly, in "Mary calls New York heaven", the connection comes from the meaning of the verb "call". Thus these cases are not marked as IDENT coreference.

Appositive constructions are marked with special labels. For example, in "Washington, the capital city, is on the East coast", we annotate an appositive link between Washington (marked as HEAD) and "the capital city" (marked as ATTRIBUTE). The intended semantic connection can then be filled in by supplying the implicit copula.

While annotating the broadcast conversation data, we realized that the length of these documents, typically recordings of entire shows covering various topics, was prohibitive for full-document coreference annotation. We therefore chose to break the documents into multiple parts, breaking along story boundaries as much as possible, and to annotate coreference within those parts independently. The different parts of each document thus currently behave as independent documents, and the coreference chains do not carry any information across parts. This required some changes to the document format, as described in a later section. In the future, we hope to be able to fill in the coreference links that cross part boundaries, so as to create fully-coherent document-level annotation.

2.6 Entity Names Annotation

Names (often referred to as "Named Entities") are annotated according to the following set of types:

PERSON	People, including fictional
NORP	Nationalities or religious or political groups
FACILITY	Buildings, airports, highways, bridges, etc.
ORGANIZATION	Companies, agencies, institutions, etc.
GPE	Countries, cities, states
LOCATION	Non-GPE locations, mountain ranges, bodies of water
PRODUCT	Vehicles, weapons, foods, etc. (Not services)
EVENT	Named hurricanes, battles, wars, sports events, etc.
WORK OF ART	Titles of books, songs, etc.
LAW	Named documents made into laws

LANGUAGE	Any named language

The following values are also annotated in a style similar to names:

DATE	Absolute or relative dates or periods
TIME	Times smaller than a day
PERCENT	Percentage (including “%”)
MONEY	Monetary values, including unit
QUANTITY	Measurements, as of weight or distance
ORDINAL	“first”, “second”
CARDINAL	Numerals that do not fall under another type