

**Co-reference Guidelines
for English OntoNotes**

Version 6.0

© COPYRIGHT BBN TECHNOLOGIES 2004-2007

1	Overview	2
1.1	Mentions	2
1.1.1	Noun phrases.....	2
1.1.2	Premodifiers.....	2
1.1.3	Verbs.....	2
1.2	Co-reference link types	2
1.2.1	Identical (IDENT).....	2
1.2.2	Appositives (APPOS)	3
2	IDENT	3
2.1	Pronouns And Demonstratives.....	3
2.2	Specificity / Generic mentions.....	3
2.3	Pre-modifiers	4
2.4	Nested Proper Names.....	5
2.5	Copular verbs.....	5
2.6	Determining which entity to add.....	6
2.7	Small clauses.....	6
2.8	Temporal expressions	6
3	APPOS	6
3.1	Marking appositive heads	7
3.2	Linking appositive spans to other referents	7
3.3	Special cases	8
4	Special Issues.....	8
4.1	Organization and members	8
4.2	Gender and Number.....	9
4.3	Indefinite uses of proper nouns.....	9
4.4	GPEs and governments	9
4.5	Quantifiers, partitives and other “of” expressions	9
4.6	Possessive extents	10
4.7	Formulaic mentions	11
4.8	Sentence fragments	11

1 Overview

The purpose of this task is to co-reference all specific entities and events, and distinguish between types of co-reference as needed to improve accuracy and scope. Co-reference is limited to noun phrases, proper noun pre-modifiers and verbs. This initial overview briefly describes the types of mentions and the types of co-reference applied.

1.1 Mentions

1.1.1 Noun phrases

All noun phrases with distinct headwords are extracted from previously treebanked data and presented to annotators prior to annotation. Annotators do not need to identify any additional NP spans. Whenever head-sharing NPs are nested, the largest logical span is used for IDENT links.

1.1.2 Premodifiers

Proper premodifiers that can be co-referenced to other mentions should be added to the list of mentions. Adjectival premodifiers are not marked. GPE abbreviations, such as U.S. in the example below, are not marked in premodifier position, since they are adjectival in nature.

(1) U.S. economy = American economy

1.1.3 Verbs

Verbs are added as single-word spans if they can be co-referenced with a noun phrase. This includes morphologically related nominalizations (2) and noun phrases that refer to the same event but are lexically distinct (3).

(2) Sales of passenger cars [grew]_x 22%. [The strong growth]_x followed year-to-year increases.

(3) Japan's domestic sales of cars, trucks and buses in October [rose]_x 18% from a year earlier to 500,004 units, a record for the month, the Japan Automobile Dealers' Association said. [The strong growth]_x followed year-to-year increases of 21% in August and 12% in September.

Only the single-word head of the verb phrase is included in the span, even in cases where the entire verb phrase is the logical co-referent.

1.2 Co-reference link types

Two types of co-reference chains are marked: Identical (IDENT) and Appositive (APPOS).

1.2.1 Identical (IDENT)

Names, nominal mentions, pronominal mentions and verbal mentions of the same entity, concept or event are co-referenced as IDENT. There is no restriction on what semantic types can be considered for co-reference; in particular, co-reference is not limited to ACE types.

- (4) She had [a good suggestion]_x and [it]_x was unanimously accepted by all.

1.2.2 Appositives (APPOS)

Appositives consist of a *head*, or *referent* (noun phrase that points to a specific object/concept in the world), and one or more *attributes* of that referent. An appositive construction contains a noun phrase that modifies an immediately-adjacent noun phrase (which may be separated only by a comma, colon, dash, or parenthesis). An attribute often serves to rename or further define the first mention. Appositive constructions are marked because there is no explicit copula providing a word sense relation.

2 IDENT (anaphoric co-reference)

The IDENT type is used for anaphoric co-reference. By *anaphoric co-reference*, we mean links between pronominal, nominal and named mentions of specific referents. It does not include entities that are only mentioned as generic, underspecified or abstract. In order to mark IDENT co-reference, there must be a specific mention, usually pronominal, named, or definite nominal:

- (5) [Elco Industries Inc.]_x said [it]_x expects net income in the year ending June 30, 1990, to fall below a recent analyst's estimate of \$ 1.65 a share. [The Rockford, Ill. maker of fasteners]_x also said [it]_x expects to post sales in the current fiscal year that are "slightly above" fiscal 1989 sales of \$ 155 million.

2.1 Pronouns And Demonstratives

All pronouns and demonstratives are linked to their referents. Pronouns in quoted speech are also marked. Expletive pronouns (it, there) and generic 'you' are not linked. In the following example, the pronoun 'you' would not be marked:

- (6) Senate majority leader Bill Frist likes to tell a story from his days as a pioneering heart surgeon back in Tennessee. A lot of times, Frist recalls, [you]'d have a critical patient lying there waiting for a new heart, and [you]'d want to cut, but [you] couldn't start unless [you] knew that the replacement heart would make it to the operating room.

2.2 Specificity / Generic mentions

Generic nominal mentions are linked to referring pronouns and other definite mentions, but not to generic nominal mentions. This allows linking of the bracketed mentions in (7) and (8), but not (9).

- (7) [Officials]_x said [they]_x are tired of making the same statements.
- (8) [Meetings]_x are most productive when [they]_x are held in the morning. [Those meetings]_x, however, generally have the worst attendance.

- (9) Allergan Inc. said it received approval to sell the PhacoFlex intraocular lens, the first foldable silicone lens available for *[cataract surgery]. The lens' foldability enables it to be inserted in smaller incisions than are now possible for *[cataract surgery].

Bare plurals, as in (7) and (8), are always generic.

In example (10) below, there are three generic instances of 'parents.' These are marked as three distinct Ident chains, each containing a generic and the referring pronouns.

- (10) [Parents]_x should be involved with their children's education at home, not in school. [They]_x should see to it that [their]_x kids don't play truant; [they]_x should make certain that the children spend enough time doing homework; [they]_x should scrutinize the report card. [Parents]_y are too likely to blame schools for the educational limitations of [their]_y children. If [parents]_z are dissatisfied with a school, [they]_z should have the option of switching to another.

Below, the verb cannot be linked to 'a reduction of 50%,' since 'a reduction' is indefinite.

- (11) Argentina said it will ask creditor banks to [halve] its foreign debt of \$64 billion -- the third-highest in the developing world . Argentina aspires to reach [a reduction of 50%] in the value of its external debt.

2.3 Pre-modifiers

Premodifiers that are proper nouns are linkable, unless they are in a morphologically adjectival form. Adjectival forms of GPEs such as 'Chinese' in 'the Chinese leader,' are not eligible for co-reference. Thus, in 'the [United States] policy' the proper noun is linked to other references, but not in 'the American policy'. Nationality acronyms are considered adjectival as well; i.e., [U.S.S.R.] or [U.S.].

- (12) But [the Army Corps of Engineers]_x expects the river level to continue falling this month (...) the flow of the Missouri River is slowed, an [Army Corps]_x spokesman said.

Acronymic premodifiers are co-referenced unless they refer to nationality. In the examples below, 'FBI' is co-referenced to other mentions, but 'U.S.' is not.

- (13) [FBI] spokesman

- (14) [U.S.] spokesman

Even when acronymic nationality-premods act like their non-acronymic counterparts, they are not considered proper premodifiers. Here 'Japan' can be co-referenced, but 'U.S.' cannot:

- (15) On U.S.-[Japan] relations: "I'm encouraged."

Proper pre-modifiers that include acronyms in the span, however, are eligible for co-reference:

(16) A [U.S. Treasury] spokesman

Pre-modifying dates and monetary amount are also co-referenced.

(17) The current account deficit on France's balance of payments narrowed to 1.48 billion French francs (\$236.8 million) in August from a revised 2.1 billion francs in [July]_x, the Finance Ministry said. Previously, the [July]_x figure was estimated at a deficit of 613 million francs.

(18) The company's [\$150]_x offer was unexpected. The firm balked at [the price]_x.

2.4 Nested Proper Names

Proper names are considered to be atomic, and nested mentions inside proper names are not annotated separately. In the following examples, the location names that form part of the organization names are not eligible for co-reference.

(19) [Massachusetts Institute of Technology]

(20) [Bank of America]

(21) [the Chicago Board of Trade]

2.5 Copular verbs

Attributes signaled by copular structures; these are attributes of the referent they modify, and their relationship to that referent will be captured through word sense tagging.

(22) [John]_x is a [linguist]. [People]_y are nervous around [John]_x, because [he]_x always corrects [their]_y grammar.

Copular (or ‘linking’) verbs are those verbs that function as a copula, and are followed by a subject complement. Some common copular verbs are: *be*, *appear*, *feel*, *look*, *seem*, *remain*, *stay*, *become*, *end up*, *get*. Subject complements following copular verbs will be considered attributes, and will not be linked in this task. In the following example, no link is marked between mentions, since “called” is copular.

(23) Called [Otto's Original Oat Bran Beer], [the brew] costs about \$12.75 a case.

The copular sense of “or” is captured by word sense; therefore neither an IDENT nor an APPOS relation is marked in the example below.

(24) Among other things, Mr. Bologna said that the sale will facilitate Gen-Probe's marketing of a diagnostic test for [acquired immune deficiency syndrome], or [AIDS].

2.6 Determining which entity to add

In a copular structure, the subject is added to the co-reference chain. In the example below, John Smith is the attribute of the NP “the president of the bank”.

(25) [The president of the bank]_x is [John Smith]. [He]_x loves [his]_x job.

2.7 Small clauses

(26) John considers Fred an idiot.

‘Fred’ and ‘an idiot’ are not linked. This small clause construction is interpreted as missing the copula (“John considers Fred *to be* an idiot”).

2.8 Temporal expressions

Temporal expressions are eligible for co-reference, including deictic expressions such as: *now*, *then*, *today*, *tomorrow*, *yesterday*, etc. or other temporal expressions that are relative to the time of the writing of the article.

(27) John spent[three years]_x in jail. In [that time]_x...

In the example below, ‘the end of this period’ and ‘that time’ are co-referenced, as are ‘this period’ and ‘from three years to seven years.’

(28) The limit could range [from three years to seven years]_x, depending on the composition of the management team and the nature of its strategic plan. At [the end of [this period]_x]_y, the poison pill would be eliminated automatically, unless a new poison pill were approved by the then-current shareholders, who would have an opportunity to evaluate the corporation's strategy and management team at [that time]_y.

In multi-date temporal expressions, embedded dates are not linked to other mentions of that date. For example, in [Nov. 2, 1999], the sub-span ‘Nov.’ is not linked to other instances of [November].

3 APPOS (appositives)

Appositives consist of a *head*, or *referent* (a noun phrase that points to an object or concept in the world), and one or more *attributes* of that referent. An appositive construction contains a noun phrase that modifies an immediately-adjacent noun phrase (these may be separated by only a comma, colon, or parenthesis). The attribute often serves to rename or further define the first mention.

(29) [[the PhacoFlex intraocular lens]_x<HEAD>, [the first foldable silicone lens available for cataract surgery]_x<ATTRIB>]

As an additional test for apposition, either part of the appositive by itself should make sense with the rest of the sentence.

- (30) a. [[John]_x, [a linguist]_x], is coming to dinner.
b. John is coming to dinner.
c. A linguist is coming to dinner.

3.1 Marking appositive heads

For each appositive construction, we will mark the head/referent according to the specificity scale below:

- (31) Proper noun > Pronoun > Def. NP > Indef. spec. NP > Non-spec. NP
John > He > the man > a man I know > man

In the following cases, the underlined element is the head (referent):

- (32) [John]_x, [a linguist]_x
(33) [A famous linguist]_x, [he]_x studied at...
(34) [a principal of the firm]_x, [J.S]_x

In cases where the two members of the appositive are equivalent, the left-most member of the appositive will be considered the head/referent. Definite NPs include NPs with a definite marker (the) as well as NPs with a possessive adjective (his).

- (35) [The chairman]_x, [the man who never gives up]_x...
(36) [The sheriff]_x, [his friend]_x...
(37) [His friend]_x, [the sheriff]_x...

For the purpose of determining relative specificity, specific names of diseases and technologies are classified as proper names, whether they are capitalized or not.

- (38) [A dangerous bacterial]_x, [bacillium]_x, is found...

3.2 Linking appositive spans to other referents

Only the single span containing the entire appositive construction is in turn linked in an IDENT chain. In the example below, the entire span can be linked to later mentions of 'Richard Godown.' The sub-spans: [Richard Godown] and [president of the Industrial Biotechnology Association] are not included in the IDENT chain.

- (39) [Richard Godown, president of the Industrial Biotechnology Association]

3.3 Special cases

Adjacent spans containing equivalent amounts in different currencies are marked as appositives.

(40) [50 million Canadian dollars]`<head>` ([US\$ 42.5 million]
`<attrib>`)

Appositives that contain adverbs are marked, providing the adverb does not affect the “size” of the entity

(41) The problem has been particularly damaging to [the OTC market]`<head>`, **traditionally** [a base for the small investor]`<attrib>`

In the following example, the adverb “primarily” serves to further delimit the referent, ‘outside vendors.’ No appositive is marked.

(42) Mr. Barr said yesterday's plea by Gulf Power, which came after months of negotiations, was based on evidence that Gulf Power had set up an elaborate payment system through which it reimbursed [outside vendors] -- **primarily** [three Florida advertising agencies] -- for making illegal political contributions on its behalf.

Numeric ages are interpreted as ellipses of a full noun phrase, for example, ‘a 42-year-old’. These are marked as appositives.

(43) [Mr. Smith]`<head>`, [42]`<attrib>`, ...

Multiple-attribute appositives include only one head, and multiple attributes.

(44) [Robert V. Van Fossan]`<head>`, [63]`<attrib>`, [chairman of Mutual Benefit Life Insurance Co.]`<attrib>`

4 Special Issues

4.1 Organization and members

No co-reference is marked between an organization and a subset of its members.

(45) It was an ideal place for [the Orange Workers]_x to start [their]_x new nation, unencumbered by the demographics that have undermined apartheid elsewhere in South Africa. So far, [about 150 Orange Workers] have moved here, spending nearly \$ 1 million buying up property over the past three years.

(46) IDENT ([the Orange Workers],[their])

4.2 Gender and Number

If there is a disagreement in number or gender, yet both noun phrases clearly refer to the same entity, singular and plural entities are linked.

(47) And lawmakers are putting the finishing touches on a compromise that would give the Air Force nearly all of the \$ 2.4 billion it wants for production of [Northrop Corp.'s radar-eluding B-2 bombers, which cost \$ 530 million apiece]_x. The final [B-2]_x agreement is certain to require detailed testing and verification of [the bomber's]_x capabilities.

4.3 Indefinite uses of proper nouns

When a proper noun usually used to refer to a specific entity is used as an indefinite reference, it is not marked. In (48), 'a Hungary' is not linked to either the other indefinite reference or to the definite reference to the country.

(48) Nor is it [a Hungary], where yesterday the parliament approved constitutional changes meant to help turn the Communist nation into a multiparty democracy... Erasing the differences still dividing Europe, and the vast international reordering that implies, won't endanger the statehood of [a Poland] or [a Hungary] ... With this year's dislocations in China and the Soviet Union, and the drive to democracy in [Poland] and [Hungary], the East German leadership grew still more defensive.

4.4 GPEs and governments

GPEs are linked to references to their governments, even when the references are nested NPs, or the modifier and head of a single NP.

(49) Christian legislators are insisting on a Syrian troop pullout from [Lebanon]_x before agreeing to political changes giving [the nation's]_x Moslems a greater role in [[Beirut's]_x government].

(50) IDENT([Lebanon], [Beirut's government], [the nation's], [Beirut's])

4.5 Quantifiers, partitives and other “of” expressions

Quantifiers (of the form “X of Y”) should **not** be co-referenced with the entities they modify:

(51) [all of [the scientists]_{y-no coref}]_x

(52) [the total of [the loan]_{y-no coref}]_x

(53) [the amount of [\$10 billion]_{y-no coref}]_x

(54) [a mile of [highway]_{y-no coref}]_x

In the above cases, the larger extent should be co-referenced with other mentions of the same entity. For partitives referring to a set of “entities of type Y” the entire span is co-referenced with other mentions:

- (55) [a group of [doctors] _{y-no coref}] _x
- (56) [a bunch of [flowers] _{y-no coref}] _x
- (57) [a number of [American citizens] _{y-no coref}]
- (58) [a pinch of [salt] _{y-no coref}] _x

“Of” expressions representing approximate quantities should be co-referenced as a single entity, as if the expression consisted of a cardinal number plus the rest of the expression.

- (59) [Thousands of [protesters] _{y-no coref}] _x = several thousand protesters
- (60) [dozens of [friends] _{y-no coref}] _x = several dozen friends
- (61) [hundreds of [years] _{y-no coref}] _x = several hundred years

Quantifiers or partitives that represent a subset of a non-generic group can be co-referenced independently with other mentions.

- (62) [Half of [the Palestinian population] _y] _x
- (63) [A group of [the doctors] _y] _x
- (64) [Most of [the attendees] _y] _x

However, if the group is generic (including all bare plurals) or underspecified, it is not eligible for co-reference.

- (65) [A handful of [Palestinians] _{y-no coref}] _x
- (66) [A group of [doctors] _{y-no coref}] _x
- (67) [A busload of [attendees] _{y-no coref}] _x

4.6 Possessive extents

Noun phrases extracted from the treebank may include the possessive 's in the NP. The 's ending should be included in the extent of the noun phrases that are co-referenced.

- (68) [Iowa 's]_x governor ...Postville, [Iowa]_x

4.7 Formulaic mentions

In broadcast news, the reporter's introduction and sign off often follows the following formula:

(69) Introduction: As [ABC 's]_x Jim Sciutto reports

Sign off: Jim Sciutto , [ABC News]_x , Postville Iowa

In these cases, [ABC's] and [ABC News] are interpreted as the same entity and are co-referenced.

4.8 Sentence fragments

When appositive-like mentions appear in adjacent sentence fragments, these should be annotated as IDENT co-reference.

(70) [The price]_x ? [\$ 300 *U*]_x . [A lot]_x by current standards.