



Arabic Co-reference Guidelines

For OntoNotes

V 0.4

Last update: 5/29/08

© Copyright BBN Technologies 2004-2008

1	Overview	2
1.1	Mentions	2
1.1.1	NPs.....	2
1.1.2	Premodifiers.....	2
1.1.3	Verbs.....	2
1.2	Co-reference links.....	3
1.2.1	Identical (IDENT).....	3
1.2.2	Appositives (APPOS)	3
2	Annotation Tool.....	4
3	IDENT (anaphoric co-reference)	4
3.1	Head-sharing NPs	5
3.2	Pronouns And Demonstratives	5
3.2.1	Detached pronouns	5
	Nominative personal pronouns (subject) and demonstrative pronouns are detached in Arabic.	5
3.2.2	Attached pronouns	5
3.3	Specificity / Generic mentions.....	6
3.4	Pre-modifiers	7
3.5	Copular verbs.....	8
3.6	Small clauses	8
3.7	Temporal expressions	8
4	APPOS (appositives)	9
4.1	Marking appositive heads	9
4.2	Extents of heads and attributes	10
4.3	Linking appositive spans to other referents	10
4.4	Specific appositive guidelines	10
5	Special Issues.....	11
5.1	Organization and members	11
5.2	Number	12
5.3	Indefinite uses of proper nouns.....	12
5.4	GPEs and governments.....	13
5.5	Determining which entity to add.....	13
5.6	Nested Proper Names	13
5.7	Quantifiers, partitives and other “of” expressions	14
5.8	Possessive extents	15
5.9	Formulaic mentions	15

5.10	Sentence fragments	15
6	Addendum: Callisto Tool	16
6.1	Overview	16
6.2	Appositives with nesting.....	17

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. Co-reference has been marked using RelTag, one of BBN's in-house annotation tools, and Callisto, developed at Mitre. This initial overview briefly describes the types of mentions and the types of co-reference applied.

1.1 Mentions

1.1.1 NPs

NP mentions are extracted from Treebank prior to annotation. Annotators should not add additional NP spans to the mentions list, unless a pre-removed NP is needed to create an appositive.

1.1.2 Premodifiers

Proper Premodifiers in Arabic are generally adjectival, and *ARE NOT* co-referenced.

❖ المبعوث (الأميركي) الخاص الى السودان جون دانفورت

The(American) Commissary to Sudan, John Danforth

❖ قرار (أميركي) في شأن العراق

An(American)decision about Iraq

1.1.3 Verbs

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

لقد (نما) الإقتصاد الأوروبي بسرعة خلال السنوات الماضية، (هذا النمو) ساهم في رفع...

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

(2) Japan's domestic sales of cars, trucks and buses in October [rose] 18% from a year earlier to 500,004 units, a record for the month, the Japan Automobile Dealers' Association said. [The strong growth] 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 links

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

1.2.1 Identical (IDENT)

Names, nominal mentions, and pronominal mentions of the same entity are co-referenced as IDENT. There is no limit to what semantic types of NP entities can be considered for co-reference; in particular, co-reference is not limited to ACE types.

❖ بدء محادثات ل وقف النار في (جبال النوبة) بين الخرطوم و المتمردين في
سويسرا. بدأ ممثلون ل الحكومة السودانية و " الجيش الشعبي ل تحرير السودان
" في سويسرا محادثات ل وقف النار في (جبال النوبة) ب وسط السودان ...

[Negotiations for ceasefire in (Alnawba Mountains) had been started between Khartoum and the rubbles in Switzerland.

The representatives of the Sudan government and "Sudan People's Liberation Army", started negotiations in Switzerland to ceasefire in the (Alnawba Mountains) in the middle if Sudan...]

نحو (650 جندياً اميركياً) سينضمون تباعاً الى قوات فيليبينية. الا ان (الجنود الاميركيين)
حذروا من المخاطر.

1.2.1* Note: In an NP containing two or more entities conjoined by a "و" ("and"), the entities should be separately extracted to form nested mentions. They can then be coreferenced to other relevant mentions in the text. In some cases, annotators will have to manually extract the entities. Note that this rule does **not** apply to proper names, which are considered atomic (see section 5.6).

التقى شارون وبوش في واشنطن اليوم وبحثا في إمكانية إقامة دولة فلسطينية مستقل
وكان هذا بعد لقاء بين عرفات وبوش دام أكثر من ساعتين...

The mentions of "Bush" should be extracted from both phrases and linked together.

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 (separated only by a comma, colon, dash, or parenthesis). It often serves to rename or further define the first mention. We need to mark appositive constructions because there is no explicit copula.

دانیلا ستوفل
Head

الناطق ب أسم وزارة الخارجية السويسرية
Attrib

(3) [[Daniela Stovell]<HEAD>, [The Swiss Foreign Ministry's Spokeswoman]<ATTRIB>]

Test for apposition: Either part of the appositive by itself should make sense with the rest of the sentence.

- (4) a. [[Daniela Stovell], [The Swiss Foreign Ministry's Spokeswoman], is coming to dinner.
- b. Daniela is coming to dinner.
- c. The Swiss Foreign Ministry's Spokeswoman is coming to dinner.

المسؤول في " كتائب شهداء الاقصى " التابعة ل حركة فتح

محمد رائد الكرمي
Head

2 Annotation Tool

[Add notes on Callisto here.]

3 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.] said [it] 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] also said [it] expects to post sales in the current fiscal year that are "slightly above" fiscal 1989 sales of \$ 155 million.

IDENT ([Elco Industries Inc.], [it], [The Rockford, Ill. maker of fasteners], [it])

3.1 Head-sharing NPs

Head-sharing NPs are two (or more) extracted entities, the shorter one(s) contained within the span of the longer, sharing the same content word as their headword. In such cases, the longest logical span should be used in co-reference with other mentions.

(6) There's already word of [[[a possible Israeli-Palestinian summit] in Egypt] in the next several days]_x. [This summit]_x ...

- IDENT chain: [a possible Israeli-Palestinian summit in Egypt in the next several days], [This summit]

3.2 Pronouns And Demonstratives

Subject, object, demonstrative and possessive (including adjectival) pronouns are eligible for coreference. Null subjects and objects (represented by *) are also eligible for coreference.

3.2.1 Detached pronouns

Nominative personal pronouns (subject) and demonstrative pronouns are detached in Arabic.

The subject pronouns in Arabic include:

انتن / انتم / انتما / نحن / هن / هم / هما	We, you , they
هي / هو / انت / انا	I, you, he, she

Subject pronouns are often null in Arabic; overt subject pronouns are rare, but do occur.

3.2.2 Attached pronouns

Object pronouns are attached to the verb (direct objects) or preposition (indirect objects). The following object pronouns can be coreferenced:

- ها / هـ / هـ / كـ / يـ	Me, you, him, her
- هُنـ / هُمـ / كُما / كُنـ / كُمـ / نا	Us, you them

Possessive (adjectival) pronouns are identical to object pronouns, but are attached to nouns. The following object pronouns can be coreferenced:

- ها / هـ / هـ / كـ / يـ	My, your, his, her
- هُنـ / هُمـ / كُما / كُنـ / كُمـ / نا	Our, your their

Subject, object and possessive pronouns (as listed above) are all eligible for coreference. Verbal inflections are *not* pronominal and *cannot* be coreferenced:

(6) Do **NOT** coreference the final **-ت** in the example below. **This is an inflection, not a pronoun.**

صرحت الناطقة ب أسم وزارة الخارجية السويسرية دانييلا ستوفل : " إن **ها** ليست في برن و لا في جنيف

The Swiss foreign ministry's spokeswoman announced the **she** is neither in Burne nor in Geneva
Pronouns in quoted speech are also marked.

*Do not link expletive pronouns (it, there). Also, do not mark generic *you*. In the following example, the pronoun **نا** would **NOT** be marked:

(9) قالت ان " ما يثير أكثر من تساؤل لدى أبناء شعب (**نا**) هو المبررات التي تدفع....

In the following example:

جماعة أبو سياف التي تربط (**ها**) صلات ب تنظيم القاعدة...

The pronoun (**ها**) is co-referenced with (**جماعة أبو سياف**)

3.3 Specificity / Generic mentions

Generic nominal mentions should be linked to referring pronouns and other definite mentions, but not to generic nominal mentions. This would allow linking of the bracketed mentions in **Error! Reference source not found.** And 0, but not **Error! Reference source not found.**

(مسؤولون) قالوا* أن (**هم**) تعبوا من تكرار نفس العبارات

(الاجتماعات) تكون اكثر فعالية عندما تعقد في الصباح. لكن (**هذه الاجتماعات**) لدي **ها** اسوأ حضور.

Bare plurals, as in **Error! Reference source not found.** and 0, are always generic.

In example (7) below, there are two generic instances of *parents*. These are marked as two distinct Indent chains, each containing a generic and the referring pronouns.

والدين يجب أن يكون لهما دور في تعليم اولادهم في المنزل, ليس في المدرسة. يجب أن يتأكدوا أن الاولاد يقضون وقتا كافيا في كتابة الوظائف. يجب عليهم التدقيق في سجلاتهم المدرسية. ليس غريبا على والدين ان يلوموا المدرسة على تقصير اولادهم في الدراسة. اذا كان لوالدين اعتراض على المدرسة, لهم الحق في النقل لأخرى.

(7) [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] 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 are dissatisfied with a school, [they]_y should have the option of switching to another.

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

- (8) 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.

3.4 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*, should not be linked. So we can co-reference the [United States] policy with another referent, but not the [American] policy. Nationality acronyms are considered adjectival as well; i.e., [U.S.S.R.] or [U.S.].

بعض الأشخاص من (واشنطن) يعملون على مكافحة الارهاب في العالم (linkable)

(Linkable) وفد من (الأمم المتحدة) زار اليوم مصر.
(linkable) موظفوا (سيرياتل) تظاهروا ضد ارتفاع الضرائب

(Unlikable) القائد (الصيني)

Acronymic premodifiers can be coref-ed unless they refer to nationality. In the examples below, FBI can be co-referenced to other mentions, but U.S. cannot.

- (9) [FBI] spokesman
(10) [U.S.] spokesman

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

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

Do mark pre-modifying dates.

خطف سبعة أشخاص على يد جماعة ارهابية في (حزيران). وقد تبين ان هذه الجماعة هي جماعة " ابو سيف " التي ما زالت تحتجز الرهائن , بينها أميركيان منذ (حزيران) الماضي.

Do mark pre-modifying monetary amounts.

- (12) عرض الشركة لمبلغ (150) دولار لم يكن متوقعا, صدمت باقي الشركات (بالسعر)

Do extract proper pre-modifiers that include acronyms in the span

قال المتحدث بإسم وزارة المالية عز الدين بدر الذي (linkable)

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

3.5 Copular verbs

Do not mark 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.

(14) 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.

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

No link between mentions. “Called” is copular, so neither IDENT nor APPOS is marked.

3.6 Small clauses

شادي يعتقد أن فادي متسرع في إتخاذ قراراته

<شادي> and <متسرع> are **not** linked together. Think of this small clause construction as missing the copular.

3.7 Temporal expressions

Link temporal expressions such as:

(16) عبد الله بن مروان قضى (عشرة سنوات) في السجن, في (ذاك الوقت) ...

Do link 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, and may therefore require knowledge of the time of the writing to resolve the co-reference. Annotators can use this knowledge from outside the text to resolve co-reference.

In the following example:

(17) The limit could range [from three years to seven years], depending on the composition of the management team and the nature of its strategic plan. At [the end of [this period]], 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].

[the end of this period] and [that time] can be co-referenced, also [this period] and [from three years to seven years].

In a multi-date temporal expression, do not link embedded dates to other mentions of that date. For example, in “Nov. 2, 1999”, do not link “Nov.” to another instance of “November”)

4 APPOS (appositives)

Appositives consist of a head, or referent (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 (separated only by a comma, colon, or parenthesis). It often serves to rename or further define the first mention.

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

4.1 Marking appositive heads

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

(19) Proper noun > Pronoun > Def. NP > Indef. spec. NP > Non-spec. NP
سعيد > هو > الرجل > رجل أعرفه > رجل

So in these cases, the underlined element is the head (referent):

(20) (سعيد), (المترجم), يتكلم عدة لغات

(21) قانون (الشركة), (سيرياتل), لا يسمح بذلك

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).

(22) The chairman, the man who never gives up...

(23) The sheriff, his friend...

(24) His friend, the sheriff...

In the specificity scale, specific names of diseases and technologies are classified as Proper Name, whether they are capitalized or not.

(25) A dangerous bacteria, bacillium, is found...

4.2 Extents of heads and attributes

If the most appropriate span is not pre-extracted, don't create it. Instead, for appositives, add the span that includes the head of the span that would ideally be used. In the following example, only [the creators] was extracted by the system and not the desired [the creators of the money fund concept]. Thus, [the creators] is the attribute:

(26) But the Reserve Fund, America's first money fund, was not named, nor were {[the creators] of the money-fund concept}, [Harry Brown and myself]

If the proper span is not extracted, but rather a span including more than the desired span, do not create the proper span and do not add the larger span to the appositive (material between {} is not added, in the example below):

(27) *{two \$ 200 million [Real Estate Mortgage Investment Conduits]}, [a kind of collateralized mortgage obligation]

No APPOS relation.

اعلن (نائب حاكم مصرف لبنان) (هاروتيون صامونيليان)

Head

Attribute

نائب حاكم مصرف لبنان was extracted by the system, but you will have to extract هاروتيون صامونيليان manually to be the attribute.

4.3 Linking appositive spans to other referents

Only the single span containing the entire appositive construction is in turn linked in an IDENT chain. None of the nested NP spans should be added. However, if some nested spans are already included since they were pre-linked as sharing a head, leave the already-included nested spans in. In the example below, the entire span can be linked to later mentions to Richard Godown. The sub-spans: [Richard Godown], [president of the Industrial Biotechnology Association] are not included in the IDENT chain.

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

اعلن (نائب حاكم مصرف لبنان هاروتيون صامونيليان) عن انعقاد مؤتمر.....

و عقد (صامونيليان) مؤتمرا صحافيا امس في مقر الحاكمية ب المصرف المركز

4.4 Specific appositive guidelines

Do mark money-appositives:

(29) [50 million Canadian dollars] ([US\$ 42.5 million])

تم تحويل مبلغ (خمسون مليون ليرة سورية)، مع يعادل (مليون دولار أمريكي) إلى الدول المنكوبه

Do mark appositives that contain adverbs, providing the adverb does not affect the “size” of the entity:

- (30) The problem has been particularly damaging to [the OTC market]<head>, **traditionally** [a base for the small investor]<attrib>
- (31) 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.

APPOS: No relation.

Do mark ages as attributes (as if they were ellipses of, for example, ‘a 42-year-old’):

تحدث (الرئيس الفرنسي)، (56 عاما) ، إلى مندوب ...
Attrib Head

Do mark multiple-attribute appositives:

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

Do NOT mark appositives with “or”, as the copular sense of “or” will be captured by word sense:

- (33) 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].

Appositive **is marked in** the following example:

- (34) Japan has found [another safe outlet for its money]: [U.S. home mortgages] .

- (35) وجدت اليابان (مخرجا آمنا لأموالها) : (رهن البيوت)

5 Special Issues

5.1 Organization and members

No coreference is marked between an organization and a subset of its members.

(36) It was an ideal place for [the Orange Workers] to start [their] 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.

[The Orange Workers] and [about 150 Orange Workers] are **NOT** linkable. While ([the Orange Workers], [their]) are linked together as IDENT.

وكان قد تظاهر في مدينة لوس انجلس الأمريكية (حشد كبير من الناس) احتجاجا على إرتفاع أسعار الوقود. (معظم المتظاهرين) هم من سائقين السيارات العمومية

Do NOT link (معظم المتظاهرين) with (حشد كبير من الناس)

5.2 Number

If there is a disagreement in number (or gender) yet both NPs clearly refer to the same entity, singular and plural entities can be linked.

(37) 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]. The final [B-2] agreement is certain to require detailed testing and verification of [the bomber's] capabilities.

(38) كان مراد قد إستقبل وفدا منهم وإستمع لمطلبهم وشدد على ضرورة إبرام (٢٨٨ من عقودهم). وطالب الأساتذة بإعادة النظر (بعقودهم المرفوعة) إلى مجلس الجامعة وإلغاء قرار الإبرام

5.3 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 **Error! Reference source not found.**, [a Hungary] is NOT linked to either the other indefinite reference or to the definite reference to the country.

(39) Nor is it **a Hungary**, where yesterday the parliament approved constitutional changes meant to help turn the Communist nation into a multi party 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.

5.4 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.

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

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

المسؤولون يصرون على انسحاب القوات السورية من **(لبنان)** ، قبل الموافقة على التغيرات السياسية التي من شأنها إعطاء المسلمون في **(البلاد)** دورا اكبر في **(حكومة بيروت)**

5.5 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”.

[The president of the bank] is [John Smith].

When there is no clear syntactic hierarchy, the most specific element should be added, the others are attributes. When two proper nouns are used, both are added.

(41) he added a thousand bureaucrats to [the RDF in Tampa] and renamed [it] [Central Command]

5.6 Nested Proper Names

Proper names are considered to be atomic, and nested mentions inside proper names are not annotated separately.

Nested mentions not linked:]

(42) [Massachusetts Institute of Technology]

(43) [Bank of America]

(44) [the Chicago Board of Trade]

(45) In the examples above, do NOT extract Massachusetts, America or Chicago.

(46) “معهد القاهرة الموسيقي”

(47) “مصرف لبنان المركزي”

Don't extract القاهرة or لبنان to link them to other entities.

5.7 Quantifiers, partitives and other “of” expressions

Quantifiers (of the form “X of Y”) should **NOT** be coreferenced with the entities they modify:

- (48) [all of [the scientists]_{y-no coref}]_x
- (49) [the total of [the loan]_{y-no coref}]_x
- (50) [the amount of [\$10 billion]_{y-no coref}]_x
- (51) [a mile of [highway]_{y-no coref}]_x

In the above cases, the larger extent should be coreferenced with other mentions of the same entity.

For partitives referring to a set of “entities of type Y” the entire span should be coreferenced with other mentions:

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

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

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

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

- (59) [Half of [the Palestinian population]_y]_x
- (60) [A group of [the doctors]_y]_x
- (61) [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.

(62) [A handful of [Palestinians] _{y-no coref}] _x

(63) [A group of [doctors] _{y-no coref}] _x

(64) [A busload of [attendees] _{y-no coref}] _x

5.8 Possessive extents

The syntax trees for Year 2 data include the 's in the NP. The 's ending should be included in NPs that are coreffed.

(65) [Iowa 's] governor ...Postville, [Iowa]

5.9 Formulaic mentions

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

(66) **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 to be treated as the same entity are co-referenced.

5.10 Sentence fragments

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

(67) [The price] _{x?} [\$ 300 *U*] _x.

[A lot] _x by current standards

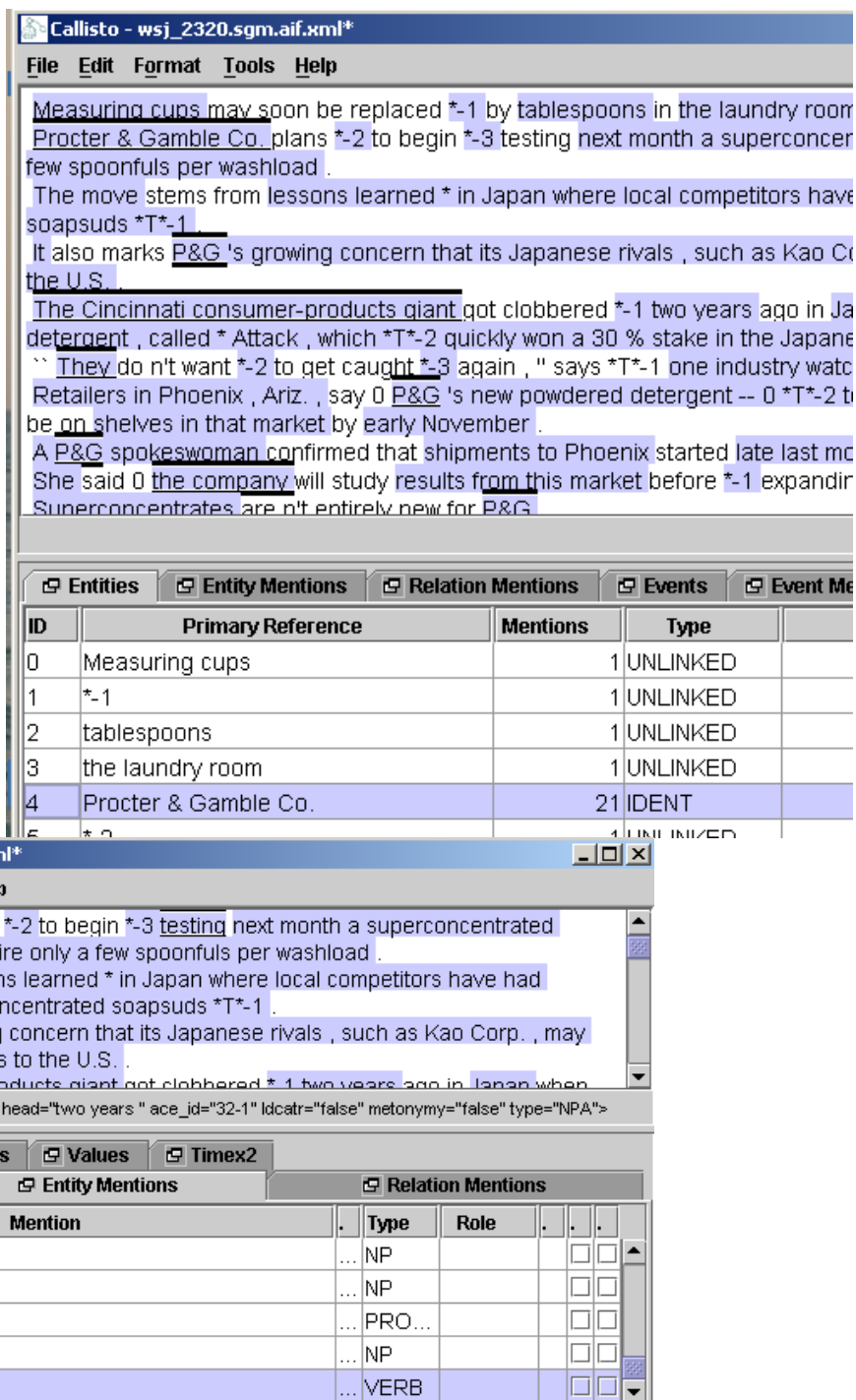
6 Addendum: Callisto Tool

6.1 Overview

Annotation performed in 2006 is being done with the Callisto annotation tool.

Files are prepared, as they were for RelTag, with NPs pre-selected based on Treebank. They are then imported into the Callisto tool using a version of the ACE Event Task, and ACE2004 APF v. 4.0.1 or higher.

Nested extents that share a head are not co-referenced.



The screenshot displays the Callisto annotation tool interface. The top window shows a text document with several sentences. The first sentence is "Measuring cups may soon be replaced *-1 by tablespoons in the laundry room". The second sentence is "Procter & Gamble Co. plans *-2 to begin *-3 testing next month a superconcentrated detergent that *-1 will require only a few spoonfuls per washload". The third sentence is "The move stems from lessons learned * in Japan where local competitors have had phenomenal success with concentrated soapsuds *-1". The fourth sentence is "It also marks P&G 's growing concern that its Japanese rivals , such as Kao Corp. , may bring their superconcentrates to the U.S.". The fifth sentence is "The Cincinnati consumer-products giant got clobbered *-1 two years ago in Japan when".

The bottom window shows a table of entities. The table has columns for ID, Primary Reference, Mentions, and Type. The data is as follows:

ID	Primary Reference	Mentions	Type
0	Measuring cups	1	UNLINKED
1	*-1	1	UNLINKED
2	tablespoons	1	UNLINKED
3	the laundry room	1	UNLINKED
4	Procter & Gamble Co.	21	IDENT
5	*-2	1	UNLINKED

The bottom window also shows a table of events. The table has columns for ID, Mention, Type, and Role. The data is as follows:

ID	Mention	Type	Role
E2-2	its best-selling Tide	NP	
E3-2	Hugh Zurkuhlen	NP	
E3-4	he	PRO...	
E3-6	Mr. Zurkuhlen	NP	
14-4	testing	VERB	

Verbs, pronouns and proper pre-modifiers are added when they are co-referent with a NP.

They are labeled as such on the Entity Mentions tab.

6.2 Appositives with nesting

If there are redundant nested spans within an appositive, assign the <HEAD> and <ATTRIB> spans to the APPOS, assign the entire appositive span to the relevant IDENT chain (if any), and ignore the redundant nested spans. For example:

(68) John, the guy from Boston, said he loves annotation.

[[John] 1, [[the guy] 2 from Boston] 3]4, said [he] 5 loves annotation.

APPOS ([John] 1, [the guy from Boston] 3)

IDENT ([John, the guy from Boston] 4, [he] 5)

Do not coreference [the guy] 2 with anything.