



OntoNotes Chinese Co-reference Guidelines

Version 4.0

© COPYRIGHT BBN TECHNOLOGIES 2008

1	OVERVIEW	3
2	BASIC CONCEPTS.....	3
2.1	MENTIONS	3
2.1.1	<i>Mentions and Entities</i>	3
2.1.2	<i>Noun Phrases (NPs) and Nested NPs</i>	3
2.1.3	<i>Premodifiers</i>	4
2.1.4	<i>Verbs.....</i>	4
2.2	IDENT (IDENTICAL)	4
2.3	APPOS (APPOSITIVE).....	5
3	IDENT (CO-REFERENCE)	5
3.1	SPECIFIC VS. GENERIC MENTIONS.....	5
3.1.1	<i>Specific vs. generic nouns.....</i>	5
3.1.2	<i>Tests for generic NPs.....</i>	6
3.1.3	<i>Annotation Guidelines for specific vs. generic nouns.....</i>	7
3.1.4	<i>Some Special Cases of Generic NPs.....</i>	8
3.2	PRE-MODIFIERS.....	9
3.3	NESTED NPS	9
3.3.1	<i>Proper nouns</i>	9
3.4	PRONOUNS, REFLEXIVES AND DEMONSTRATIVES	9
3.5	TIME EXPRESSIONS	10
3.5.1	<i>Period of time</i>	10
3.5.2	<i>Point in time</i>	10
3.5.3	<i>Deictic time expressions</i>	10
3.5.4	<i>Approximate time expressions</i>	11
3.6	MONETARY EXPRESSIONS	12
3.7	GPEs AND GOVERNMENTS.....	12
3.8	ORGANIZATIONS AND MEMBERS	15
3.9	COPULAR VERBS	15
3.10	CHANGE OVER TIME	16
3.11	*PRO*.....	17
3.11.1	<i>Recoverable *pro*s</i>	17
3.11.2	<i>Irrecoverable *pro*s</i>	18
4	APPOS (APPPOSITIVES)	20
4.1	MARKING APPOSITIVE HEADS	20
4.2	EXTENTS OF HEADS AND ATTRIBUTES.....	21
4.3	ATTRIBUTES MISSING COMMAS OR LINKED BY CONJUNCTIONS	21
4.4	LINKING APPOSITIVE SPANS TO OTHER REFERENTS	21
4.5	SPECIFIC APPOSITIVE GUIDELINES	22
5	PENDING ISSUES	22
5.1	FUNCTIONS	22
5.2	EQUIVALENT EXPRESSIONS	23

1 Overview

The purpose of OntoNotes project is to co-refer linguistic expressions representing the same specific entity or event in a discourse. Co-reference is limited to noun phrases (including proper nouns, nominals, and pronominals), verb phrases and, sometimes, numerals, time expressions, and monetary amounts.

2 Basic Concepts

In this section, we will briefly describe the basic concepts that will be used in the guidelines.

2.1 Mentions

2.1.1 Mentions and Entities

“An entity is an object or set of objects in the world. A mention is a reference to an entity. Entities may be referenced by their name, indicated by a common noun or noun phrase, or represented by a pronoun.” (Huang, S. 2003: 3).

In the current task, all the mentions have been extracted from previously treebanked data and presented to annotator prior to annotation. Annotators do not need to add new mentions themselves. The only exception to this rule will be discussed in section 4 on appositive constructions (specifically, in clause 4.2, “Extents of heads and attributes.”

2.1.2 Noun Phrases (NPs) and Nested NPs

Noun phrases (NPs) in Chinese can be complex noun phrases or bare nouns (nouns that lack a determiner such as “the” or “this”). Complex noun phrases contain structures modifying the head noun, as in the following examples:

- (1) [他担任 总统 任内 最后 一次 的 [亚太 经济 合作 会议 [高峰会]]].
[[His last APEC [summit meeting]] as the President]
- (2) [越南 统一 后 [第一 位 前往 当地 访问 的 [美国 总统]]]
[[The first [U.S. president]] who went to visit Vietnam after its unification]

In these examples, the smallest phrase in square brackets is the bare noun. The longer phrase in square brackets includes modifying structures. All the expressions in square brackets, however, share the same head noun, i.e., “高峰会 (summit meeting)”, and “美国总统 (U.S. president)” respectively.

Nested noun phrases, or nested NPs, are contained within longer noun phrases. In the above example, “summit meeting” and “U.S. president” are nested NPs. **Wherever NPs are nested, the largest logical span is used in co-reference** (also see 3.1.10 for examples).

2.1.3 Premodifiers

A premodifier is a word that precedes and modifies a noun. Proper noun premodifiers can be co-referenced to existing noun phrases **and/or** other proper premodifiers (see section 3.1.9 for examples).

2.1.4 Verbs

Verbs can be manually extracted by the annotator and added as single-word spans **if** (and only if) they can be co-referenced with an existing noun phrase. This includes morphologically related nominalizations, as in (3), and noun phrases that refer to the same event but are lexically distinct, as in (4).

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

(4) 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.**

2.2 IDENT (*Identical*)

Names, nominal mentions, and pronominal mentions of the same entity are co-referenced as IDENT. For example:

(5) [美国 总统 克林顿] 1 3 号 起程 前往 亚洲 ， 准备 出席 [他]担任 总统 任内 最后 一次 的 亚太 经济 合作 会议 高峰会 ， 同时 [他] 也 将 成为 越南 统一 后 第一 前往 当地 访问 的 美国 总统 。

On the 13th, [U.S. President Clinton] started [his] trip to Asia to attend [his] last APEC summit meeting. Meanwhile, [he] is also going to be the first US president to visit Vietnam after its unification.

In this example, the expressions in square brackets refer to the same person, i.e., US President Clinton. Therefore, we need to co-refer these expressions.

2.3 APPOS (Appositive)

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 (i.e., verbs such as “to be, to become”, and in Chinese, verbs such as “是 (to be), 成为 (to become)”) providing a word sense relation. (For a detailed explanation of copula, see section 3.9.)

- (6) *[[John] <HEAD>, [a linguist] <ATTRIB>]*

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

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

3 IDENT (co-reference)

If proper names, nouns, pronouns and empty categories (*pro*) refer to the same specific entity in the discourse, they should be co-referenced and given the label IDENT. (For a detailed explanation of the type *pro*, see section 3.11.) The entities can be objects, states, relationships, actions, and so on. IDENT 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.

3.1 Specific vs. Generic Mentions

3.1.1 Specific vs. generic nouns

NPs can either be generic or specific. Specific NPs are co-referenced. Generic NPs are linked to referring pronouns and other definite mentions, but **not** to each other. An NP is specific if it is used to refer to a particular object in the world. For example,

- (8) *[My friend Sara] came to Boston last week. I have known [Sara] ever since we were in first grade. [She] and I are good friends.*

In this example, the NPs in square brackets—i.e., *My friend Sara*, *Sara*, and *she*—are used to identify a specific individual in the world who is the speaker or writer’s friend.

An NP is generic when it is used to refer to any or all members of the type which the noun denotes. For example,

- (9) *[Birds] can fly.*
(10) *He saw a flock of [birds] at the park.*

In both of the above examples, the NP *[birds]* does not refer to any particular entity in the world but instead denotes any member of the set *birds*. Therefore, the NPs are generic and cannot be co-referenced to each other.

3.1.2 Tests for generic NPs

Sometimes, the distinction between specific and generic NPs is difficult to ascertain. The following examples draw from Huang's tests for generic NPs (for more details, see Huang, 2003).

3.1.2.1 Bare Nouns and Bare Plurals

Singular bare nouns—that is, nouns that lack a determiner such as “the” or “this”—are often generic. Bare plurals are always generic. For example,

- (11) 很多小朋友喜欢[童话故事]。
Many kids like [fairy tales].
(12) 有一间房子失[火]了。
There is a house on [fire] now.
(13) *He does not enjoy riding [horses].*

3.1.2.2 Underspecified NPs

In the following examples, the noun phrases do not refer to any specific person or entity in the world and are therefore generic:

- (14) 目前[没有人]来这里玩。
At present, [nobody] comes here to play.
(15) 有[人]来这里玩吗?
Is there [anyone] who comes here to play?

3.1.2.3 Words that are Commonly Generic

Noun phrases such as “任何+noun (*any* +noun)”, “(绝) 大多数+noun (*most* + noun)”, “少数+noun (*few* +noun)”, “有些+noun (*some* +noun)” tend to be generic. For example,

- (16) 这儿 [大多数学校] 都有升旗仪式。

- (17) *There is a flag-raising ceremony in [most schools] here.*
今天 [有些学生] 没有交作业。
Today, [some students] haven't turned in their homework yet.
- (18) *[任何人] 都应该遵守交通规则。*
[Everyone] should obey traffic rules.

When the subject is a bare noun and has the reading of “any x”, it is generally generic:

- (19) *[鸟] 会飞。*
[Birds] can fly.

3.1.2.4 “Empty Shell” Mentions

“Empty shell” mentions refer to NPs that convey a hypothetical setting (Huang, 2003: 45). For instance,

- (20) *我们每年会选[一位主席]和[几位副主席]。*
Each year, we select [one chair] and [several vice chairs].

The NPs in brackets should be considered generic because in this context they do not refer to any particular “chair” or “vice chairs.” (These job titles, yet unfilled by real people, remain hypothetical.)

3.1.3 Annotation Guidelines for specific vs. generic nouns

In order to mark IDENT co-reference, there must be a specific mention, usually pronouns, proper names or definite nouns.

Generic nouns, i.e., non-specific nouns, should not be linked to other generic nominal mentions. **Generic NPs are not co-referenced to each other.** For example:

- (21) *卫生部已确定在未来几年内实施一项艾滋病的[公众]教育计划，以进一步提高人们的认识，增强*
[公众]自我保护能力。
*The Ministry of Public Health has decided to implement a *[public] education program in the prevention and cure of AIDS within the next few years, so as to further enhance people's awareness of AIDS and to strengthen the *[public]'s ability to protect themselves .*

In the above example, the “*” before the square bracket indicates that the two NPs in square brackets should not be co-referenced with each other. The first occurrence of 公众 (public) is not eligible for co-reference since it is only part of a noun phrase. The last 公众 is a noun phrase, but cannot be co-referenced because it is generic.

In the following English example, the NPs in brackets are generic and therefore cannot be co-referenced with each other:

(22) *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].*

However, **generic nominal mentions can be linked to pronouns and other definite nouns**. For example,

(23) *[Officials] said [they] are tired of making the same statements.*

(24) *[Meetings] are most productive when [they] are held in the morning. [Those meetings], however, generally have the worst attendance.*

In example below, there are three generic instances of *parents*. These are marked as three distinct IDENT chains, each containing a generic and the referring pronouns.

(25) *[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.*

3.1.4 Some Special Cases of Generic NPs

死刑(capital punishment), 世界(world—as distinct from “the world” meaning the planet Earth), 社会(society) are considered generic nouns. For example,

(26) *美国总统当选人小布什 22 号则是排除了联邦死刑案暂缓执行可能性，他并且重申 *PRO* 支持[死刑]，小布什在德州指出，如果[死刑]在执行上是公平、公正和没有疑问的，他支持[死刑]，因为他相信此项制度能够保障大众的生命安全。*

US President-Elect George Bush, Jr. excluded the possibility of a stay of execution on the twenty-second and reiterated his support for [capital punishment]. While in Texas, Bush pointed out that if the process of [capital punishment] is carried out fairly and beyond any reasonable doubt, he supports [capital punishment] because this policy will guarantee the safety of the public.

In this case, “capital punishment” refers to a general policy and lacks the specificity to be co-referenced.

3.2 Pre-modifiers

If pre-modifiers (the linguistic structure that modifies the head noun) are proper nouns, they can be linked to other mentions of the proper nouns in the same text. For example,

(27) *[[[中][美]]两国]关系这几年不断得到改善。*

The [[US]-[China]] relationship has been improving continuously these years.

In this example, if 中(China), 美(US) or 中美(China-US) or 中美两国(China-US) appear elsewhere in the same discourse, they can be linked together.

3.3 Nested NPs

3.3.1 Proper nouns

Proper names, including the titles of songs and other works of art 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.

(28) *[Massachusetts Institute of Technology]*

(29) *[Bank of America]*

(30) *[the Chicago Board of Trade]*

(31) *["I Left My Heart in San Francisco"]*

3.4 Pronouns, reflexives and demonstratives

Link all pronouns 你, 我, 他, 她, 它, 你们, 我们, 他们, 它们, 我, 您, 咱们 (you, me, he, she, and so on), and demonstrative pronouns: 这个, 那个, 这些, 那些 (this, that, these, those) in singular, plural or possessive forms, to anything they refer to.

(32) *[这位大使]为论证[他]的结论...*

To prove [his] conclusion, [the ambassador]...

IDENT: [这位大使 (the ambassador), 他 (his)]

Pronouns from classical Chinese such as 其中 (among which), 其 (he/she/it), 之 (he/she/it) should also be linked with other mentions to which they refer.

Pronouns such as 你, 您, 你们, 大家, 各位 can be considered generic. In this case, they are not linked to other generic mentions in the same discourse. For example,

(33) 请*[大家]带好自已的随身物品。*[大家]请下车。

*Please take your belongings with *[you]. Please get off the train, *[everyone].*

3.5 Time expressions

In OntoNotes annotation, time and monetary expressions can be co-referenced if they refer to the exact same time, period of time, or monetary entity, as shown in the following.

3.5.1 Period of time

(34) *John spent [three years] in jail. In [that time]...*

(35) 过去[五年]是外商投资增长最快的[五年]。

The past [five-year period] is also the [five-year period] when foreign investments grow the fastest.

3.5.2 Point in time

(36) *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], the Finance Ministry said. Previously, the [July] figure was estimated at a deficit of 613 million francs.*

In the above example, both mentions of the term “July” can be coreferenced to each other because they refer to the very same July. (In other instances, two mentions of the same month name may merely refer to the month in general or in different years—for example: “Paul loves [July]. Every [July] he goes fishing.” These two mentions of July **cannot** be co-referenced.)

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”)

3.5.3 Deictic time expressions

Link deictic expressions, such as *now, then, today, tomorrow, yesterday*, etc. These expressions are relative to the time of the writing of the article and may therefore require annotators to use knowledge from outside the text to determine when the article was written and resolve co-reference.

Note the following example:

(37) *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.*

IDENT:

[the end of this period]_y and [that time]_y
[this period]_x and [from three years to seven years]_x

3.5.4 Approximate time expressions

Approximate time expressions, such as 现在(now), and 目前(at present), can only be co-referenced when they refer to the exact same event or indicate the same time frame. For example,

(38) [目前] 正在越南访问的美国总统克林顿，18号前往越战美军阵亡现场亲自凭吊。

US President Clinton, who is [now] visiting Vietnam, went to pay his condolences at the Vietnam Veterans Memorial site on the 18th.

(39) 根据美国国防部的统计，到[目前]为止仍然有1992名美军下落不明。

According to the statistics from the US Department of Defense, there are still 1992 American soldiers missing up to [now].

The above two examples appear in the same article, and both mentions of “目前(now)” refer to the time when this piece of news was written. Hence, these two mentions can be co-referenced with each other.

But in this example, the two mentions of “现在(now)” **cannot** be co-referenced:

(40) 但是在金正日上台之后情况出现了变化，他特别在北韓的《劳动日报》上表示：现在已经和60年代不同，而应该用新的思考领导国家。

But the situation changed after Kim Jong-il came to power. He emphatically stated on North Korean Labor Day, “[Now] is different from the sixties, and we therefore need to lead the country with new thinking.”

(41) 金正日去年 5 月到北京时参观了北京的电子工业，现在又以上海跟深圳为参观重点，金正日进行经济改革的迫切性不言可喻。

Kim Jong-il visited the electronics industry during his trip to Beijing last May, and [now], this trip focuses on Shanghai and Shenzhen. It is therefore quite obvious that Kim Jong-il is views economic reform as an urgent matter.

In the two examples above, the two mentions of “现在(now)” cannot be co-referenced with each other because the first mention refers to “the current age vs. the sixties,” while the second mention refers to the time the article was written.

现在(now) and 目前(at present) sometimes can co-refer with each other, depending on the contexts. For example, (*pending*)

3.6 Monetary expressions

Monetary expressions can co-refer with each other only when they refer to the exact same entity:

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

If the expressions refer to the same monetary amount but **not** the exact same entity, they **cannot** be co-referenced. For example,

(43) *Last year, he earned *[500 dollars]. I also earned *[500 dollars].*

3.7 GPEs and governments

Geo-political entities (GPEs) are linked to references to their governments, even when the references are nested NPs, or the modifier and head of a single NP (for example, “China’s government”). Capital cities can sometimes be used to refer to the governments. In this case, they should also be linked to the governments.

(44) *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], [the nation], [Beirut]

(45) 一名政府官员在 22 号表示美国总统克林顿可能在下个星期宣布是否决定前往 [北韩] 访问，设法说服 [平壤当局] 冻结飞弹研发计划。

A government official said on the 22nd that US President Clinton may announce his decision whether to visit [North Korea] or not next week. The purpose of the visit will be to persuade [Pyongyang authorities] to freeze their missile research and development program.

In the example above, “北韩 (North Korea)” is co-referenced with “平壤当局 (Pyongyang authorities)”.

In the following example, “中共当局 (Chinese communist authorities)”, “中国 (China)”, “北京当局 (Beijing authorities)” are all co-referenced with each other:

(46) 过去一年尽管 [北京当局] 镇压未向政府登记的地下基督教会和天主教会，但是基督教在 [中国] 大陆已经越来越普及，[中共当局] 已经加强提防宗教团体发展为象法轮功等反抗组织。

Although [Beijing authorities] have been suppressing the underground Christian and Catholic churches in the past year, Christianity has become more and more popular in Mainland [China]. [Chinese communist authorities] have already enhanced their measures to prevent religious groups from developing into rebellious organizations like Falungong.

Note that mentions clearly referring to the location and nothing more **cannot** be co-referenced to mentions of the government; they should only be linked to other locative mentions. (These mentions are often, though not always, preceded by a locative preposition, such as *in*, *at*, *to*, *from*, etc.) For instance, country names should be co-referenced in separate chains when they act as a GPE in some cases and a location in others (or have other distinctly different senses). For example:

(47) *Today [China] sent a message to Russia.*

(48) *In [China] yesterday, a man won a million dollars.*

In the first sentence, the word "China" has the narrow meaning of the Chinese government. In the second, it denotes the actual physical territory. Therefore, these terms represent different entities that belong in different chains.

In the same vein, mentions of capital cities, such as “北京” (Beijing), sometimes refer to “北京当局” (Beijing authorities), which should be co-referenced with the Chinese government. “中国政府” (“the Chinese government”) sometimes refers to the City of Beijing, “北京市”. As a result, different mentions of “北京” (Beijing) do not always co-refer with each other.

Note that country and city mentions should be co-referenced in separate chains **only** when the two different senses (i.e., GPE and location) are **very clear**. If there is any ambiguity, all mentions of the country or city name should be linked in the same chain.

Along similar lines, GPEs generally cannot be co-referenced to the building where the president's office is situated. For example, "Washington D.C." cannot co-refer with "the White House" because there are different parts of the government in Washington. "The White House" only refers to the executive branch, while "Washington" can refer to the whole country. "Moscow" should not co-refer with "the Kremlin." (In a similar way, "the Bush administration" **cannot** be co-referenced with "the U.S. government" but **can** be linked to "the executive branch.")

The following example illustrates these points. Here, "Washington" cannot be co-referenced with "the country," and "the White House" cannot be co-referenced with "The Bush administration."

(49) *[The Bush administration] maintains that the President has [the country's] best interests at heart. This is Kelli Arena in [Washington], reporting live from [the White House].*

Note one exception: in extremely rare cases, the seat of a government's executive (e.g., the White House, 10 Downing Street) can be co-referenced with mentions of the government as a whole. This occurs **only** when the two mentions **clearly and without a doubt** refer to the same entity. For example,

(50) *In a statement released this afternoon, [10 Downing Street] called the bombings in Casablanca "a strike against all peace-loving people."*

(51) *In a statement, [Britain] called the Casablanca bombings "a strike against all peace-loving people."*

In this case, it is obvious that "10 Downing Street" and "Britain" are being used interchangeably in the text. Again, if there is any ambiguity, however, these terms should be co-referenced separately.

An ousted government or government-in-exile cannot refer to the country or area. For example,

(52) *尼泊尔政府也已经加强跟印度交界地区的安全管制，并且逮捕意图从*[西藏流亡政府]所在地的印度北部返回*[西藏]的西藏人。*

*The Nepalese government has already strengthened its control over the Nepal-India border and intends to arrest the Tibetans returning to *[Tibet] from the Northern part of India, where *[the Tibetan government-in-exile] is located.*

In the above example, “西藏 流亡 政府 (the Tibetan government-in-exile)” does not co-refer with “西藏(Tibet)”.

3.8 Organizations and members

No coreference is marked between an organization and its members.

(53) *[学生会]一共有*[500 个会员]。

*There are *[500 members] in *[the Student Union].*

3.9 Copular verbs

A copular structure consists of a referent (usually the subject), an attribute of that referent (usually the predicate), and a copula (most often, though not always, a 'linking' verb). The copula serves to equate (or connect) the referent with the attribute. Some examples of copular verbs in Chinese are 是(to be) and 为(to be, to serve as). In addition, the EDT guidelines include in this category other verbs (particularly so-called “light verbs”) that trigger an attributive reading on the following NP: “成为” (become), “(当)选为” (is elected), “称为” (is called), “(好)像” (looks like), “叫做” (is called), etc.

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. Subject complements following copular verbs will be considered attributes, and will not be linked in this task. For example,

(54) [上海]是*[中国最大的城市]。[上海]发展得很快。

[Shanghai] is[the largest city in China]. [Shanghai] develops fast.*

In the above example, the two mentions of “上海(Shanghai)” co-refer with each other, but the entity does not co-refer with “中国最大的城市(the largest city in China)”.

Taken together, the two examples below illustrate the rules for co-reference in sentences containing copular verbs. First,

(55) [长江]是*[中国第一大河]。

*[The Yangtze River] is *[the biggest river in China].*

In the same text, there may be other mentions of “长江(the Yangtze River)”:

(56) [长江]流经中国很多的省市。

[The Yangtze River] passes through many provinces and cities in China.

In that case, “长江(the Yangtze River)” in the first sentence should co-refer with “长江(the Yangtze River)” in the second. However, the entity should **not** co-refer with “中国第一大河(the biggest river in China),” since this noun phrase follows the copular verb “是(to be).”

3.10 Change over time

Entities that change over time (such as a country or organization) should be co-referenced **unless** there is some clear and concrete factor that distinguishes them (for instance, a dramatic transformation in a country's political regime or a gap in time between distinct historical periods). In the following phrases, for example, the bracketed terms would be co-referenced:

(57) *In [the old China], no one had cell phones.*

(58) *In [the new China], everyone has an I-Pod.*

Here, “the old China” and “the new China” do not represent particularly distinct entities and so can be linked.

However, in the examples below, the terms would form separate chains:

(59) *In [the old China] before the Communist Party took over...*

(60) *In [the new China] since the rise of the Communist Party...*

Here, “the old China” and “the new China” describe dramatically different eras in the country's history and political landscape and so should **not** be co-referenced.

Sometimes, an organization or facility or a person can grow larger or take on a new form (a neighborhood shop can become a large corporation, a politician has an ideological conversion). In most cases, the relevant terms should be co-referenced, depending on how the entity is discussed in the text. Note the following set of examples.

(61) *[The small factory]_x grew into a [large corporation]_y.*

In the above sentence, the bracketed terms would **not** be co-referenced because of the copular construction ("grew into"; see section 3.9 on copular verbs). However, there may be later references in the same text, as in the following:

(62) *[The small factory]_x has been a remarkable success story. [It]_x has tripled its size in one year.*

(63) *[The large corporation]_x now employs thousands of the town's workers.*

Here, the three entities marked with an _x clearly refer to the same entity in the text. All of these terms should therefore be linked.

3.11 *pro*

In Chinese, if the subject or object can be recovered from the context, or it is of little interest for the reader/listener to know, it can be omitted. In Chinese Treebank, the position where subject or object is omitted is annotated with small *pro*. (Please refer to *Bracketing Guidelines for the Chinese Treebank*, at <http://www.cis.upenn.edu/~chinese/ctb.html>.)

In OntoNotes annotation, we need to recover these *pro*s by co-referring them with other mentions of the same entity or event. (Note that in some cases, the *pro* cannot be recovered; see section 3.11.2 on irrecoverable *pro*s.)

In Chinese, *pro* can be replaced by overt NPs if they refer to the same entity or event. And *pro* and overt NPs do not have to be in the same sentence. Exactly what *pro* stands for is determined by the linguistic context in which it appears.

(64) 吉林省主管经贸工作的副省长全哲洙说：“[*pro*] 欢迎国际社会同[我们]一道，共同推进图们江开发事业，促进区域经济发展，造福东北亚人民。

*Quan Zhezhu, Vice Governor of Jinlin Province who is in charge of economics and trade, said: “[*pro*] Welcome international societies to join [us] in the development of Tumen Jiang, so as to promote regional economic development and benefit people in Northeast Asia.*

(Here, the *pro* stands for the omitted pronoun “we.”)

3.11.1 Recoverable *pro*s

In imperative sentences (such as requests and commands), subjects are usually omitted. In this situation, the omitted subject is usually the listener of the utterance. For example,

(65) *pro* 关上门。

Close the door.

Here, *pro* refers to the listener of this command.

Sometimes, the omitted subject/object can be recovered by an overt NP in the same sentence. For example,

(66) 已经有[13 名阿拉伯裔民众] 在[*pro*] 支持巴勒斯坦的示威事件中
丧生。

*[13 Arabs] have died in [*pro*] support-Palestine demonstration.*

(67) 为了[*pro*] 带女儿去看医生, [她] 上午没来上班。

*In order to [*pro*] take her daughter to the doctor's, [she] didn't come to work
this morning.*

In the above two examples, expressions in square brackets co-refer with each other. The overt NP appears in the same sentence with the *pro*s.

However, the overt NP does not have to be in the same sentence as the *pro*. For example,

(68) [俄罗斯总统 普京 和 外蒙古总统 巴嘎班迪] 今天 14 号在外蒙首府 乌兰
巴托举行 高峰会, 并且 签署了一项 军事条约, 保证 两国未来将会 维持和
平关系。

*Today the 14th, [Russian President Putin and Mongolian President Bagabandy] had a
summit meeting in the Mongolian capital of Ulan Bator, where they signed a military
treaty to guarantee peace-relations between the two countries in the future.*

(69) 今天[*pro*] 签署的 双边 协议 指出, 外蒙古的 非核区 有助支持禁止 核武
扩散 政策。

*The bilateral treaty [*pro*] signed today pointed out that a non-nuclear Mongolia will
help support the nuclear non-proliferation policy.*

The above two examples are from the same article. From the context, we see that *pro* can be co-referred with the overt NP in the text.

3.11.2 Irrecoverable *pro*s

Sometimes, *pro*s cannot be recovered in the text—i.e., they cannot be replaced by an overt NP in the text. There are different cases where *pro*s cannot be recovered.

Case One: *pro* in existential sentences usually cannot be recovered or linked in the annotation, as shown in this example:

(70) [*pro*] 有二十三顶高新技术项目进区开发。

There are 23 high-tech projects under development in the zone.

(71) 目前还[*pro*]没有学生来上学。

At present, no students come to school yet.

Case Two: *pro* does not refer to a specific entity or event. This kind of generic *pro* should not be linked. For example,

(72) 肯德基、麦当劳等速食店全大陆都推出了[*pro*]买套餐赠送布质或棉质圣诞老人玩具的促销。

*In Mainland China, fast food restaurants such as Kentucky Fried Chicken and McDonald's have launched their promotional packages by providing free cotton Santa toys for each combo [*pro*] purchased.*

(73) [*pro*]过圣诞节已经成为大陆城市居民生活重要的一部分。

*[*pro*] celebrating Christmas has become an important part of life for mainland city residents.*

Case Three: *pro*s in idiomatic expressions do not need to be linked. For example,

(74) [*pro*] 大约是公元1644年的时候...

*when [*pro*] was about 1644 AD, ...*

(75) [*pro*] 到了清朝末期...

*When [*pro*] reaches the end of Qing Dynasty, ...*

(76) 张启悦说中共和越南最近几年的关系发展顺利，去年2月确定双方发展关系的指导方针是“[*pro*]长期稳定、面向未来、睦邻友好、全面合作”。

*Qiyue Zhang said that in recent years, the China-Vietnam relationship has been developing smoothly. And in last February, both sides agreed that the guiding principle for the bilateral relationship is [*pro*] long-term stability, future-oriented, good-neighborly friendship and comprehensive cooperation.*

(77) 所以[*pro*] 整体而言，美国基于国家利益的考量对于两岸关系的基本处理架构不至于有太大的改变。

*Therefore, [*pro*] generally speaking, there won't be any major changes in the US government's framework for dealing with cross-strait relations between China and Taiwan, out of consideration for its own benefit.*

4 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 (separated only by a comma, colon, or parenthesis). It often serves to rename or further define the first mention. In OntoNotes annotation, we should link the head with its attribute.

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

(79) *[中国第三条南北大动脉]<ATTRIB> -[京九铁路]<HEAD> 在一九九五年顺利建成。*

[The third biggest north-south artery]<ATTRIB>, - [Jingjiu Railroad]<HEAD> was successfully established in 1995.

Sometimes, the HEAD can also appear before the ATTRIB:

(80) *[香港]<HEAD>[这个融合了中西的国际城市]<ATTRIB>*

[HongKong]<HEAD>, [the international city integrating Chinese and Western cultures]<ATTRIB>

4.1 Marking appositive heads

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

John > He > a man I know > man

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

John<HEAD>, a linguist

A famous linguist, he<HEAD> studied at...

a principal of the firm, J.S<HEAD>

(81) *[“九五”期间]<HEAD> (一九九六年 至 二 0 0 0 年)*

[“Nine-Five” period]<HEAD> (1996 to 2000)

If a proper name is followed by a title (such as Mr. or Dr.), the head is the proper name plus the title.

(82) [大陆著名的作家]<ATTRIB>[王朔先生]<HEAD>

[Famous writer in Mainland China]<ATTRIB>, [Mr. Shuo Wang]<HEAD>

4.2 Extents of heads and attributes

If the head is not pre-extracted, we need to add the appropriate head as a new mention and link the head with its attribute. For example, if we see the phrase,

(83) Michael Jordan, [the star athlete]...

and “Michael Jordan” is not already extracted, we will need to add it ourselves.

4.3 Attributes missing commas or linked by conjunctions

If the attributes in an appositive construction are linked by the conjunction "and," or if (as is sometimes the case in Chinese) they lack any commas in between, then they can all form one long attribute NP. In the following, the “and” connects the various features into a single attribute:

(84) ... [the world's greatest player and the defending champion and the most dedicated promoter of processed-sugar drinks] <ATTRIB>, [John Kashin] <HEAD>

Obviously, if the attributes are divided by commas, they can be tagged separately in the usual format, as in the following:

(85) ... [the world's greatest player] <ATTRIB>, [the defending champion] <ATTRIB>, [the most dedicated promoter of processed-sugar drinks] <ATTRIB>, [John Kashin] <HEAD>

4.4 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. For example,

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

In the above example, 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.

4.5 Specific appositive guidelines

Do mark multiple-attribute appositives:

(87) *[Robert V. Van Fossan]<HEAD>, [[founder], [chairman of Mutual Benefit Life Insurance Co.]]<ATTRIB>*

(88) *[[一百米短跑冠军], [我的朋友]]<ATTRIB> [王丽丽]<HEAD>
[[One-hundred meter dash champion], [my friend]]<ATTRIB> [Lili Wang]<HEAD>*

Appositives are marked in each of the following examples:

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

(90) *“You can't say it's a “[tsunami]” ([tidal wave]), but we're making some headway.*

(91) *Mr. Blair said his libel suit seeks [10 million Canadian dollars] ([US\$ 8.5 million])- from Canadian Express and Hees executives Manfred Walt and Willard L'Heureux.*

(92) *Moreover, as [a member of the Mitsubishi group, which is headed by one of Japan's largest banks], [it] is sure to win a favorable loan.*

(93) *As a result, marketers of faux gems steadily lost space in department stores to more [fashionable rivals] – [cosmetics makers].*

Expressions ending in “...等 等” (“etc.”) and tagged with “APPOS”, should be tagged as “UNLINKED” instead. For example,

(94) *小朋友们喜欢[苹果， 香蕉等]*<APPOS> 水果。*

Kids like fruits [such as apples, bananas and so on]<APPOS>.*

In the above example, please change the “APPOS” tag into “UNLINKED.” (Note that this rule only applies to data sets where APPOS mentions were pre-labeled.)

5 Pending Issues

5.1 Functions

Stock price this morning \$4.5, stock price now is \$6.

5.2 Equivalent expressions

Equivalent expressions linked by 也就是 (“that is”) are co-referenced. For example,

(95) [美国东部时间 上午九点], 也 就是[北京 时间下午十 点]...

<i>9:00am US Eastern Time—that is, 10:00pm Beijing Time.</i>
--

IDENT: ([美国东部时间上午九点], [北京时间下午十点])

(This treatment to equivalent expressions is subject for further argument. We can either co-reference the terms in the sentence or treat the construction as copular (in which case, it would not be co-referenced).