Co-reference Guidelines

for Chinese OntoNotes

Version 2.0

1 Overview

The purpose of this task is to co-reference all specific entities and events, and identify co-reference phenomena in Chinese requiring special consideration as needed to improve accuracy and scope. Co-reference is limited to noun phrases (including proper nouns, nominals, pronouns and null arguments).

This initial overview briefly describes the types of mentions and the types of co-reference applied.

1.1. Mentions

1.1.1. Noun phrases

Noun phrases are extracted from the Chinese Treebank prior to annotation. The example below shows five noun phrases that have been extracted from a sentence fragment.

(1) [建筑]NP 　是 [[*pro*]NP 开发 　[浦东]NP 的 　一 项 主要 　[经济 　活]NP]NP

CONSTRUCTION IS *PRO* 　DEVELOPS PUDONG-DE 　ONE-CL MAIN ECONOMIC ACTIVITY

Construction is a principle economic activity in developing Pudong.

Null NPs are extracted and added as *pro*, as shown above.

1.1.2. Proper Nouns

There is no distinction made between adjectival and nominal premodifiders. These are extracted as proper nouns when appropriate, and are eligible for co-reference.

(2) [[中国]NP [经济]NP]NP

CHINA 　　ECONOMY

Chinese economy

(3) [中国]NP 的 　[关贸 [总协定] NP] NP

CHINA 　DE 　TRADE AGREEMENT

China's General Agreement on Tariffs and Trade (GATT)

1.1.3. Null subjects and pronouns

In cases where small *pro* represents a null argument, the *pro* placeholder is eligible for co-reference.

> (4) [吉林省]x 主管 经贸 工作 的 副省长 全哲洙 说："[*pro*]x 欢迎 国际 社会 同 [我们]x 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."
>
> *The economics and trade work assistant deputy governor 全哲洙 in charge of **Jilin Province** said, "(**We**) welcome the international news service to join with **us** together, together advancing Tumen Jiang to develop enterprise, promote regional economic development, benefit the Northeast Asian people".*

In the example above, ***pro*** is linked to 我们 (we) and 吉林省 (Jilin Province). First-person plural pronouns used by government or corporate spokespersons will be assumed to refer to the entities (organizations or GPEs) they represent.

1.2. Co-reference links

Two types of co-reference chains are marked: 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 restriction on what semantic types of NP entities can be considered for co-reference; in particular, co-reference is not limited to ACE types.

> (5) [一次 荟萃 世界 华人 物理学家 的 "第一届 全球 华人 物理学 大会]x ,
>
> ONCE COMBINE WORLD ASIAN PHYSICIAN DE FIRST GLOBAL ASIAN PHYSICS CONFERENCE
>
> 将 于 一九九五年 八月 五日 至 九日 在 汕头 大学 举行。
>
> WILL 1995 YEAR AUGUST 5TH THROUGH 9TH AT SHANTOU UNIVERSITY HELD
>
> **[这次 规模 大、 规格 高 的 国际 会议]x,** 由
>
> THIS SCALE LARGE STANDARD HIGH DE INT'L MEETING, BY
>
> 中国 国家 自然 科学 基金 委员会 和 中国 科学院 联合 主办。
>
> CHINA NAT'L NATURAL SCIENCE FOUNDATION AND CHINA SCIENCE ACADEMY JOINTLY HELD
>
> *The first session of the Global Chinese Physics Meeting which will bring together Chinese physicists from all over the world will be held at Shantou University, from August 5 through 10 1995. This large scale high level international convention will be jointly held by the Chinese National Natural Science Foundation and the Chinese Academy of Science.*

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 the word sense relation.

```
(6) [[中共 中央 办公   厅 副     主任]<ATTRIB>  [陈福] <HEAD>] <APPOS SPAN>
    CCP CENTER OFFICE ASSISTANT DIRECTOR     CHEN FU
```

```
(7) [[中国   小将] <ATTRIB>-[罗莉]  <HEAD>] <APPOS SPAN>
    CHINA    YOUNG PLAYER  - LUO LI
```

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:

```
(8) [这 位 大使]x       为  论证   [他]x 的 结论...
    THIS CL AMBASSADOR FOR PROVE  HE   DE CONCLUSION
    This ambassador, to prove his conclusion...
```

2.1. Pronouns And Demonstratives

Link all pronouns: 他, 她, 我, 您, 他们, and demonstrative pronouns: 这些, 那些, 这(一)个 in singular, plural or possessive forms, to anything they refer to. Corresponding demonstrative adjectives are not marked.

```
(9) [17 岁  的 罗莉]x  在 预赛       中      以 9.900 分       名列    榜首
    17 AGE DE LUO LI IN PRELIMINARY CONTEST BY 9.900 POINTS NAMED 1ST PLACE
    今天    [她]x    第五    个 出场。
    TODAY SHE      FIFTH  CL ENTERS-STAGE
```
The 17 year old Luo scored 9.900 in the preliminary competition to take the lead. She was the fifth on the apparatus.

2.1.1. Genericity

Generic nominal mentions should be linked to referring pronouns and other definite mentions, but not to generic nominal mentions. Generic nominals are not co-referenced to each other.

In (10) below, the first occurrence of 公众 (public) is not eligible for coreference 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.

(10)	卫 生	部	已	确定	在	未来	几 年	内	实施
	PUBLIC HEALTH	MIN.	ALREADY	CONFIRM	AT	FUTURE	FEW YEARS	WITHIN	IMPLEMENT

一　项 艾滋病 防治]NP　　的 [公众] 教育　　　计划，以 进一步　　　　提高
ONE CL AIDS-PREVENTION　DE PUBLIC EDUCATION PLAN, THUS A STEP FUTHER ENHANCE
人们认识，　　　　　　　增强　[公众]NP 自我保护　　　　能力。
PEOPLE RECOGNITION IMPROVE PUBLIC　　SELF-PROTECTION　ABILITY.

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.

2.2. Pre-modifiers

Premodifiers that are proper nouns are eligible for co-reference.

(11)	[[海协]NP	[负责人]NP]NP
	SEA ASSOCIATION	DIRECTOR

(12)	[[台湾]NP	[军方]NP]NP
	TAIWAN	MILITARY

2.3. Copular verbs

Attributes signaled by copular structures are not marked; 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.

Chinese expresses copular structure through the use of copular verbs: 是 and 为, and light verbs such as 成为 (become), (当)选为 (is elected), 称为 (is called), (好)像 (looks like), etc. which indicate attributes.

In the example below, 崇明 (ChongMing) and 中国 第三 大 岛 (China's third largest island) are not co-referenced, because the copular verb 是 indicates that the second noun phrase is an attribute of the first.

(13)	[崇明]	是	[中国 第三 大　　岛]
	CHONGMING	IS	CHINA THIRD LARGE ISLAND

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".

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

3 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.

---

(15) [中国 著名　风景旅游城　　　] \<ATTRIB\> - [杭州]\<HEAD\>

     CHINA FAMOUS LANDSCAPE TOURIST CITY — HANGZOU

Hangzou, a famous scenic Chinese sightseeing city.

3.1. Marking appositive heads

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

    Name　 > pronoun > definite　 > indefinite

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

(14)   <u>John</u>, a linguist

(15)   A famous linguist, <u>he</u> studied at…

(16)   a principal of the firm, <u>J.S</u>

---

(17)   ["九五"]NP [期间]\<HEAD\>（[一九九六年 至 二０００年]\<ATTRIB\>）

    "NINTH FIVE" PERIOD (1996 YEAR TO 2000 YEAR)

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. None of the nested NP spans should be added.

3.3. Specific appositive guidelines

Appositives may have multiple attributes.

---

(18)   澳网官方网站预测 [亚洲一姐]\<ATTRIB\>[中国头号女单]\<ATTRIB\>

                     NO.1 OF ASIA　　CHINESE LEADING FEMALE SINGLES PLAYER

[本次大赛的 19 号种子]\<ATTRIB\>　 [李娜] \<HEAD\> 将会在今年的比赛中创造佳绩。

THE 19TH SEED OF THIS COMPETITION NA LI

The official website of the Australian Open predicts that NO.1 of Asia, Chinese leading female singles player, the 19th seed of this competition, Na Li will have a successful performance at this year's competition.