

# **ACE (Automatic Content Extraction) Chinese Annotation Guidelines for Entities**

Version 5.5 20050505

Linguistic Data Consortium

<http://www ldc upenn edu/Projects/ACE/>

1	Introduction .....	4
	Basic Concepts .....	4
2	Text to Annotate.....	5
3	Entity Types and Subtypes .....	5
	3.1 Persons .....	5
	3.1.1. Subtypes for Person .....	5
	3.1.2 Titles, Honorifics, and Positions.....	6
	3.1.3 Fictional characters, names of animals, and names of fictional animals....	7
	3.1.4 Groups of people .....	7
	3.2 Organizations .....	8
	3.2.1 Subtypes for Organizations.....	8
	3.2.2 Organization Subtyping Trumping Rules .....	14
	3.2.3 Organization Entities used in Person Contexts.....	15
	3.2.4 First Person Pronouns Referring to Organizations .....	16
	3.3 Geographical/Social/Political Entities (GPE) .....	16
	3.3.1 Subtypes for GPEs .....	17
	3.3.2 GPE-like Locations and Organizations .....	19
	3.3.3 Formulaic GPE Constructions: Nested Region Names.....	19
	3.3.4 GPE Mention Roles .....	19
	3. 4 Locations .....	23
	3.4.1 Subtypes for Locations .....	24
	3.4.2 Sub-parts of Locations and GPEs.....	25
	3.4.3 Non-Locations.....	25
	3.5 Facilities .....	26
	3.5.1 Subtypes for Facilities.....	26
	3.6 Vehicle .....	27
	3.6.1 Subtypes for Vehicles .....	27
	3.7 Weapon.....	29
	3.7.1 Subtypes for Weapons.....	29
4	Entity Class.....	30
	4.1 Attributive (ATR).....	31
	4.2 Generic Referential (GEN) .....	32
	4.3 Non-generic Specific Referential (SPC) .....	34
	4.4 Underspecified (USP).....	34
	4.5 Some fine distinctions and pronominals .....	35
5	Mention Types .....	35
	5.1 Mention Extent .....	36
	5.2 Mention Head.....	36
	5.3 Mention Types.....	38
	5.3.1 Simple mention headed by a proper name (NAM).....	39
	5.3.2 Simple mention headed by a common noun (NOM) .....	40
	5.3.3 Pronouns (PRO) .....	41
	5.3.4 Headless (HLS) .....	44
	5.3.5 Mentions in Conjunctions.....	44

5.4 Complex Constructions .....	46
5.4.1 Appositive Constructions (APP).....	46
5.4.2 Extended Appositive Construction (EAP).....	47
6 Nickname Metonymy .....	48
6.1 Capital City or Government Seat (FAC) Names standing in for Country's Government .....	48
6.2 City name for Sports Team.....	49
6.3 Cross-Type Metonymy Relations .....	49
7. Appendix:.....	49

# 1 Introduction

The Entity Detection task requires that selected types of entities mentioned in the source data be detected, their sense disambiguated, and that selected attributes of these entities be extracted and merged into a unified representation for each entity.

## **Basic Concepts**

An entity is an object or set of objects in the world. A mention is a reference to an entity. Entities may be referenced in a text by their name, indicated by a common noun or noun phrase, or represented by a pronoun. For example, the following are several mentions of a single entity:

**Name Mention:** *Joe Smith*

**Nominal Mention:** *the guy wearing a blue shirt*

**Pronoun Mentions:** *he, him*

Some examples in Chinese:

*北京奥申委常务副主席刘敬民在接受记者采访时表示，此行达到了他预想效果*

**Name Mention:** 刘敬民

**Nominal Mention:** 北京奥申委常务副主席

**Pronoun Mentions:** 他

For Phase 4 of ACE, entities are limited to the following seven types:

- Person - Person entities are limited to humans. A person may be a single individual or a group.
- Organization - Organization entities are limited to corporations, agencies, and other groups of people defined by an established organizational structure.
- GPE (Geo-political Entity) - GPE entities are geographical regions defined by political and/or social groups. A GPE entity subsumes and does not distinguish between a nation, its region, its government, or its people.
- Location - Location entities are limited to geographical entities such as geographical areas and landmasses, bodies of water, and geological formations.
- Facility - Facility entities are limited to buildings and other permanent man-made structures and real estate improvements.
- Vehicle – A vehicle entity is a physical device primarily designed to move an object from one location to another, by (for example) carrying, pulling, or pushing the transported object. Vehicle entities may or may not have their own power source.
- Weapon – Weapon entities are limited to physical devices primarily used as instruments for physically harming or destroying other entities.

For each entity, the annotation records the type of the entity (PER, ORG, GPE, LOC, FAC, VEH, and WEA), subtype, class, and all the textual mentions of that entity.

## 2 Text to Annotate

Only material between <TEXT> and </TEXT> tags is to be annotated. In newswire documents, material in headlines and slug sections is not to be tagged. In broadcast news, only the transcribed speech is to be tagged; added information, such as that within <TURN> tags or speaker identification tags, is not to be tagged.

## 3 Entity Types and Subtypes

### 3.1 Persons

Each distinct person or set of people mentioned in a document refers to an entity of type Person. For example, people may be specified by name (“John Smith”), occupation (“the butcher”), family relation (“dad”), pronoun (“he”), etc., or by some combination of these. Dead people and human remains are to be recorded as entities of type Person. So are fictional human characters appearing in movies, TV, books, plays, etc.

There are a number of words that are ambiguous as to their referent. For example, nouns, which normally refer to animals or non-humans, can be used to describe people. If it is clear to the annotator that the noun refers to a person in a given context, it should be marked as a Person entity.

*He is [a real turkey]*

*[The political cat of the year]*

*She's known as [the brain of the family]*

*他是一只[铁公鸡]。*

*他是[今年政坛上的一匹黑马]。*

*胡锦涛是[中国政府首脑]。*

#### 3.1.1. Subtypes for Person

We will further classify Person entities with the following subtypes.

##### **PER.Individual**

If the Person entity refers to a single person, tag it as PER.Individual.

*[胡锦涛]Per.Individual*

*[一个学生]Per.Individual*

##### **PER.Group**

If the Person entity refers to more than one person, tag it as PER.Group.

[访华代表团]Per.Group  
[一群恐怖分子]Per.Group

**PER.Indefinite**

If from the context you can't judge whether the Person entity refers to one or more than one person, tag it as PER.Indefinite.

有[人] PER.Indefinite 在爆炸中受伤。

**3.1.2 Titles, Honorifics, and Positions**

In English, titles and most honorifics precede the name. In Chinese, however, titles and positions follow the name, which also should not be treated as part of the head of the name of a Person. The parts of titles are taggable mentions if they refer to entities. Examples of English and Chinese are given: in the string “US Secretary of State Madeline Albright”, there would be four mentions of three distinct entities; in the string “朱容基总理 or 朱总理 ” there would be two mentions and they are coreferenced.

[Secretary of State <u>Madeline Albright</u> ]	PER
[Secretary of State]	
[US]	GPE
[State]	ORG
[朱容基总理 ]	PER
[朱容基]	

**3.1.2.1 Saints and other religious figures**

Religious titles such as saint, prophet, imam, or archangel are to be treated as titles.

[St. Christopher, the patron of transportation]

[Pope John Paul II]

[[阿安扎西]活佛]

[[丹增嘉措]达赖喇嘛]

[[马宝刚]阿訇]

References to “God” will be taken to be the name of this entity for tagging purposes. If it is used as a descriptor rather than a name, it will be considered a nominal mention.

If you believe in [god], you must... name mention

*Although he felt like he was [a god], he...*      nominal mention

毛泽东被一些中国人看作[一位神]。      Nominal mention

### 3.1.3 Fictional characters, names of animals, and names of fictional animals

Names of fictional characters are to be tagged; however, character names used as TV show titles will not be tagged when they refer to the show rather than the character name.

*[Batman] has become a popular icon*

*[Adam West]'s costume from Batman the TV series*

她像[鲁迅小说中的祥林嫂]一样唠叨。

[林黛玉]与[贾宝玉]是两个家喻户晓的小说人物。

Names of animals are not to be tagged, as they do not refer to person entities. The same is true for fictional animals and non-human characters. These two examples do not yield mentions.

*Morris the cat*

*Snuggle, the fabric softener bear*

我爱加菲猫。

她养了一只猫，叫咪咪。

### 3.1.4 Groups of people

Groups of people are to be considered an entity of type Person unless the group meets the requirements of an organization or a GPE described below.

*[The family]*

*[The house painters]*

*[The linguists under the table]*

*[中国申奥代表团]*

*[记者们]*

*[中华人民共和国归侨]*

*[库尔斯克号遇难官兵]*

#### 3.1.4.1 Family Names

Family names are to be tagged as Person.

*[The Kennedy]'s*

*[[布什]家族]*

*[[宋]氏家族]*

### 3.1.4.2 Ethnic and Religious Groups

Ethnic groups of people and religious groups that do not have a formal organization unifying them will be considered entities of type Person.

*[The Arabs]*

*[The Christians]*

*[华人]*

*[阿拉伯人]*

*[基督教徒]*

*[巴勒斯坦人]*

*[维吾尔族人]*

*[阿尔巴尼亚族]*

## 3.2 Organizations

Each organization or set of organizations mentioned in a document gives rise to an entity of type Organization. An Organization must have some formally established association. Typical examples are businesses, government units, sports teams, and formally organized music groups. Industrial sectors are also treated as Organizations.

### 3.2.1 Subtypes for Organizations

We will further classify Organization entities with the following subtypes. Organizations which do not fit into the subtypes defined below will not be tagged.

#### Government (GOV)

Government organizations are those that are of, relating to, or dealing with the structure or affairs of government, politics, or the state. The entire government of a GPE is excluded from this subtype and should be tagged GPE.ORG. Military organizations that are connected to the government of a GPE will be tagged as Government.

*[The Salzburg prosecutor's office] is investigating the disaster to determine if criminal charges could be filed.*

*Putin, a former [KGB] agent, defended [the court] [that] convicted Pope and [the security services],...*



The Financial Accounting Standards Board will take no conclusive action on its current project on business combinations until [Congress] has reconvened in 2001...

[The US navy] now says the USS Cole was being refueled when an explosion ripped through it in Yemen last week, killing 17.

[韩国统一部] 长官朴在归 12 号率领韩国代表团乘专机抵达平壤

[中国信息产业部] 电子信息产品管理司长张琪, 昨天在台北透露.....

[中国国务院]的工作规则首次公布, 领导一般不得为地方题词。

[美国军队]

[中国人民解放军]

## Commercial (COM)

A commercial organization is an entire organization or taggable portion of an organization that is focused primarily upon providing ideas, products, or services for profit.

Last February, he failed a drug test required by [Major League Baseball] and Commissioner Bud Selig suspended him for the 2000 season.

Pope, who owns [TechSource Marine Industries in State College, PA.], and an associate were...

Fifteen people, all with minor injuries, were rescued, according to [the Austria Press Agency].

中国大陆也将逐步对[诺基亚] ([Nokia]) 或[易利信] ([Ericsson]) 等外国厂商开放。

从传言到现实: [盛大] 联姻 [新浪], 震惊业界。

## Educational (EDU)

An educational organization is an entire institution or taggable portion of an institution that is focused primarily upon the furthering or promulgation of learning/education.

Pope, 54, a retired U.S. Navy officer from State College, PA., had worked for [the Applied Research Laboratory at Pennsylvania State University].

[NDSU] and [University of Minnesota] weeds specialist Alan Dexter says 98% of the plant survived.

[台湾大学] [国家发展研究所] 教授洪镰德指出.....

这台类人形机器人与[国防科技大学] 1990 年研制的我国首台两足步型机器人相比

[北京大学] [中国经济研究中心] 成立五周年庆典

## Entertainment (ENT)

Entertainment organizations are those whose primary activity is entertainment. This includes organizations such as Barnum and Bailey's Circus and HBO, but excludes provider giants such as Comcast and media conglomerates such as Disney and Time-Warner. These exceptions are all best annotated as Commercial Organizations.

*[the Roundabout Theater Company] is calling its new facility in Times Square the American Airlines Theater*

*Like [the famous Irish group] [the Chieftains] and [Alton], [Solace] frequently headlines in Celtic festivals.*

*[中北英皇]昨日正式启动国内首条“造星生产线”，实施AAA准艺员造星计划。*

*香港廉政公署去年中反转乐坛，经过半年调查，昨日正式落案起诉前[环球唱片有限公司]总裁陈少宝。*

*[中国东方歌舞团]是[在周恩来总理和陈毅副总理倡导和关怀下，于1962年1月13日成立的国家级歌舞团]。*

## Non-Governmental Organizations (NonGov)

NonGov Organizations are those Organizations that are not a part of a Government or Commercial Organization and whose main role is advocacy, charity or politics (in a broad sense). This subtype will include such diverse organizations as:

### 1. (Para-) Military Organizations:

*KKK*

*Al Aqsa Martyr's Brigade*

*Tamil Tigers*

*泰米尔猛虎组织*

*[尼泊尔游击队]“封锁”首都陆路交通*

*[联合国维和部队]*

### 2. Political Parties:

*Republican Party*

*Labour Party*

*共产党*

*国民党*

*新党*

*民主党*

### 3. Political Advocacy Groups and Think Tanks:

*ACLU*

*The Cato Institute*

卡托研究所  
东方人文思想研究所  
[中国民间保钓协会]

4. Professional Regulatory and Advocacy Groups:

*The American Bar Association*  
*The American Medical Association*  
美国律师协会  
中国作家协会  
美国语言协会

5. Charitable Organizations:

*The Red Cross*  
*The United Way*  
*Doctors Without Borders*  
红十字会  
中国慈善基金会  
宋庆龄儿童发展中心

6. International Regulatory and Political Bodies:

UN  
NATO  
*The World Bank*  
联合国  
北约组织  
世界卫生组织  
泛美健康组织

Under the above definition, all of the following would be annotated as Organizations of Subtype NonGov:

*three of the [U.N]. workers stationed in East Timor*  
*[International Monetary Fund] aid*  
*[Hizbollah]*  
*[Islamic Resistance]*  
*[Rally for Congolese Democracy]*  
*[the ``Caravan of Death],'' [a military party [that] killed 73 political prisoners]*  
*the leading deputy of [the Rally for Congolese Democracy], [one of [the biggest rebel movements supported by Uganda]]*  
*[Institutional Revolutionary Party]*  
*[The GOP]*  
*[the Democratic Party]*

*[the Socialist People's Party]*

*[Republican National Committee]*

*[Rock the Vote]*

*[American Medical Association]*

*[American Council on Education]*

*[College Board]*

*[National Rifle Association]*

*[American Diabetic Association]*

*[NAACP]*

*[National Center for Public Policy and High Education]*

*[American Bar Association]*

*[The Red Cross] said about 15 people managed to escape, but at least three people were dead.*

*The musings of [the Financial Accounting Standards Board], [the private-sector body based in Norwalk, Conn., [that] sets the nation's accounting standards], ...*

*美国、俄罗斯、英国、法国四个[[联合国]安理会]的常任理事国外长。*

*[中国共产党]历次全国代表大会*

*[世界卫生组织]赞扬中国援助印度洋海啸受灾国。*

*中国社会科学院*

## **Media (MED)**

Media organizations are those whose primary interest is the distribution of news or publications, regardless of whether the organization is privately or publicly owned. This will include media companies such as Time Magazine, but will exclude media conglomerates such as Time-Warner which should be annotated as a Commercial organization.

*Agence France Presse*

*Reuters international news agency*

*Time magazine*

*abc news*

*Associated Press*

*Chicago Sun-Times*

*Road & Track magazine*

*National Geographic*

人民日报  
新华社  
美国之音  
中国地产杂志社  
香港中国通讯社  
中国新闻社  
中国建筑工业出版社

## Religious (REL)

Religious organizations are those that are primarily devoted to issues of religious worship.

*...said Archbishop Khajag Barasamian, head of [the Diocese of [the Armenian Church in America, [whose] headquarters are in Manhattan]].*

*[German Bishops Conference]*

*[The Vatican]*

*[罗马教廷]*

*[中国基督教协会]*

*[中国天主教爱国会]*

*[中国基督教“三自”爱国运动委员会]*

*[中国道教学院]*

## Medical-Science (SCI)

Medical-Science organizations are those whose primary activity is the application of medical care or the pursuit of scientific research, regardless of whether that organization is publicly or privately owned.

*Overseas Chinese Physics Institute*

*Gulf Coast Research Laboratory*

*Massachusetts General Hospital*

*medical and health groups*

*[[上海中科院]神经所]*

*[上海中国遗传医学研究中心]*

*[德国分子医学研究中心]*

*[国内首家纳米技术生物医学实验室]*

## Sports (SPO)

Sports organizations are those that are primarily concerned with participating in or governing organized sporting events, whether professional, amateur, or scholastic. We will include groups whose sports are board games, card games, and games of chance in this category.

*A group of survivors belonging to [a German ski club in Vilseck, Germany], ...*

*Saudi Soccer Federation*

*Taekwondo Association*

*Philippines Olympic Committee*

*national hockey league*

*San Francisco 49ers*

*[费城 76 人队]*

*[中国队]*

*[中国奥林匹克委员会]*

*[中华全国体育总会]*

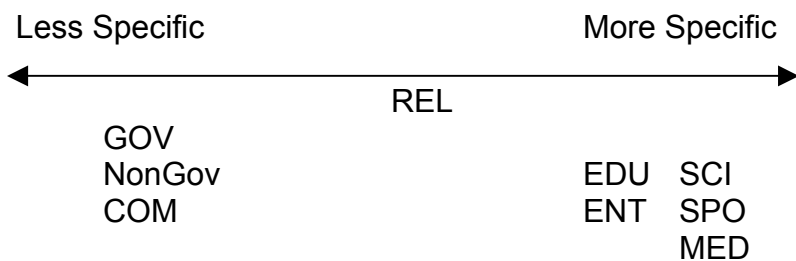
*[中国残疾人体育协会]*

*[中国羽毛球协会]*

*[武汉红金龙]击败[四川全兴]。*

### 3.2.2 Organization Subtyping Trumping Rules

The collection of organization subtypes is hierarchical in nature. Some organizations will naturally fit into more than one category. The following diagram displays the hierarchy of organization subtypes. As a rule, we will assign the most specific type possible.



### Exceptions to Trumping Rules

### ***GPE military***

The military organizations connected to a GPE's government will be tagged ORG-Government (GOV).

### ***Media Conglomerates***

Big media conglomerates such as Disney and Time-Warner will be tagged Commercial (COM). The subsidiary media organizations owned by these companies will be tagged Media (MED).

### ***Medical Schools and Research Labs***

Medical schools will be tagged Educational. Specific labs and research institutions which primarily devote their attention to medical or scientific research will be tagged Medical/Science (SCI) even when they are attached to educational institutions. (Note that medical schools should be tagged EDU.)

[北京大学附属医院]

### ***Soft Science Research Institutions***

Institutions whose primary activity is the study of social sciences will be tagged Non-Governmental (NonGov).

[中国社会科学院]

### ***Boy Scouts*** (童子军)

The Boy Scouts of America and similar organizations will be tagged Educational (EDU).

### **3.2.3 Organization Entities used in Person Contexts**

Whenever an organization takes an action, there are people within or in charge of the organization that one presumes actually made the decision and then carried it out. Thus many organization mentions could be thought of as metonymically referring to people within the organization.

However, there seems to be little to be gained in the usual case by thus “reaching inside the organization” to posit a mention of a Person entity. It seems better to adopt the view that organizations can be agentive, and take action on their own. Only when something in the context draws particular attention to the people within the organization should a separate mention of a Person entity be marked.

Sets of people who are not formally organized into a unit are to be treated as a Person entity rather than an Organization entity. It is often difficult to tell the difference between Organization entities and collections of individuals tagged as Person entities. Example organization-like nouns which are *not* organizations are “employees,” and “crew.” In the latter two cases, although the members of a

company or crew may work together in an organized and even hierarchical fashion, the groups are not organizations by themselves.

### 3.2.4 First Person Pronouns Referring to Organizations

First person plural pronouns are often used by representatives of an organization to refer to that organization. Pronouns are often used in this way by reporters representing a broadcasting station and spokespeople representing organizations. For example, in *our top story*, *our* refers to the broadcasting organization. In these cases, annotators should mark first person plural pronouns as Organization mentions, and not as Person mentions.

### 3.3 Geographical/Social/Political Entities (GPE)

Geo-Political Entities are composite entities comprised of a population, a government, a physical location, and a nation (or province, state, county, city, etc.). All mentions of these four aspects of a GPEs will be marked GPE and coreferenced.

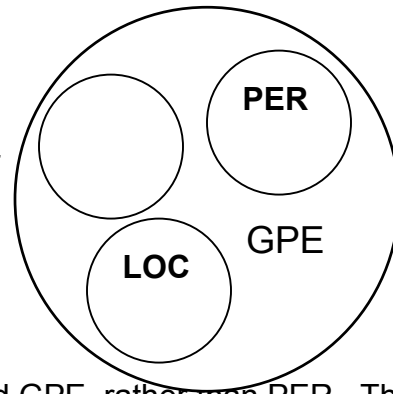
In this sentence,

*The people of France welcomed the agreement.*

there are two mentions

[The *people* of France] GPE

[France] GPE



The mention of the population of France is marked GPE, rather than PER. These mentions would be coreference as they refer to different aspects of a single GPE.

Explicit references to the government of a country (state, city, etc.) are to be treated as references to the same entity evoked by the name of the country. Thus "*the United States*" and "*the United States government*" are mentions of the same entity. On the other hand, references to a portion of the government ("*the Administration*" 当局, "*the Clinton Administration*") are to be treated as a separate entity (of type Organization), even if it may be used in some cases interchangeably with references to the entire government (compare "*the Clinton Administration signed a treaty*" and "*the United States signed a treaty*").

Sometimes the names of GPE entities may be used to refer to other things associated with a region besides the government, people, or aggregate contents of the region. The most common examples are sports teams:

[New York] defeated [Boston] 99-97 in overtime.



These are to be recorded as distinct entities, not as mentions of the GPE entity. Thus, in this example, both "New York" and "Boston" would evoke Organization-Sport entities.

### 3.3.1 Subtypes for GPEs

We will further classify GPE entities with the following subtypes. GPE entities which do not fit into the subtypes defined below will not be annotated.

#### Continent

Taggable mentions of the entireties of any of the seven continents: North America, South America, Antarctica, Europe, Asia, Africa, and Australia.

*Many people in [North America] will be able to experience a partial solar eclipse tomorrow.*

*该委员会认为，香港经济复苏的情况比 [亚洲]其他国家都要好*

#### Nation

Taggable mentions of the entireties of any nation.

*But the [U.S.] presence in Aden is unmistakable.*

*Hospital officials said all eight survivors were [German].*

*...the conversion to Christianity of the [Roman] emperor Constantine...*

*[中国] 国务院官员表示，为扩大外商在西部的投资领域.....*

#### State-or-Province

Taggable mentions of the entireties of any state, province, or canton of any nation.

*[Salzburg] governor Schausberger said...*

*在[黑龙江]、[陕西]、[河北]、[浙江]、[四川]等地纪念抗美援朝 50 周年的展览也相继开幕*

#### County-or-District

Taggable mentions of the entireties of any county, district, prefecture, or analogous body of any state/province/canton. In China, we reserve this term for entity 地区 such as 扬州地区, which is an administrative region between province and county.

Note in China, 地区, just like a province, also has a "capital city" where the government is located. The "capital" may have a different name from the county/prefecture. But when they have identical names, one must decide whether it's the city or the county that's being referred to. In most cases, it's probably the county. But when someone within the county says something like "I'm going to xxx",

where xxx is the county/city name, the speaker apparently is talking about the “capital” city and therefore the subtype of the entity should be “Population Center”.

Also note that 地区 is a widely used term in Chinese. Annotators should distinguish administrative 地区 from location 地区, such as 华中地区, 北京地区 etc. Only when X 地区 refers to administrative region can we tag it as GPE-County-or-District. Otherwise, treat it as LOC.

### Population-Center

Taggable mentions of the entirety of any GPE below the level of County-or-District. In China, this covers 市 city, 市辖区, 县 county, 乡 county, 镇 town, 村 village etc.

*John Martin, ABC news, [Washington]*

*The economic boom is providing new opportunities for women in [New Delhi].*

*...said Norbert Karlsboeck, mayor of Kaprun, [a town some 50 miles south of Salzburg in the central Austrian Alp.]*

*[哈尔滨市]一些参加过抗美援朝战争的老战士和民间收藏爱好者.....*

*作为[[中国最大的经济中心城市][上海]], 由于[它]独特的区位优势和管理、人文优势, 也必然会成为[跨国公司地区总部和外资研发中心的首选地]。*

### GPE-Cluster

Named groupings of GPEs that can function as political entities.

*东欧*

*[欧美国家]*

*[西方国家]*

*南亚*

*中东*

*欧盟*

### Special

This refers to a closed set of GPEs for which the conventional labels do not straightforwardly apply. Initially it restricts to Palestine, the Palestinian Authority, Taiwan, and mainland China (when it contrasts with Taiwan).and Native American reservations. (Note: Historical empires are tagged as GPE.Nation as long as they refer to it in the appropriate temporal context.)

*[[中国]GPE.NATION 大陆]GPE.SPC*

### 3.3.2 GPE-like Locations and Organizations

Other incidental, non-political clusters of GPEs are still considered Locations.

*the southern United States*

[美国南部]LOC

Coalitions of governments, as well as the UN, are organizational bodies and should be marked Organization.

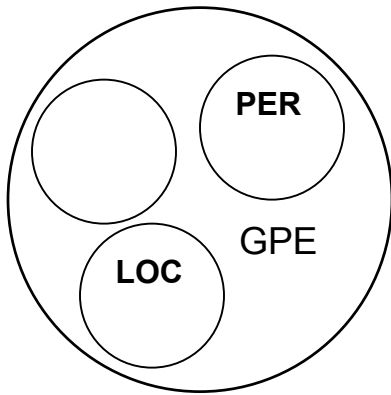
[NATO] peacekeepers arrived in the valley before nightfall.

### 3.3.3 Formulaic GPE Constructions: Nested Region Names

A series of nested region names, such as "*Provo, Utah*" evokes one entity for each region. Thus "*Provo, Utah*" evokes one entity for the population center (with mention "*Provo, Utah*") and a second one for the state (with mention "*Utah*").

### 3.3.4 GPE Mention Roles

Annotators need to decide for each entity mention in the text which role (Person, Organization, Location, GPE) the context of that mention invokes. The four roles are Person, Organization, Location and GPE. This judgment typically depends on the relations that the entity enters into.



- **GPE.ORG** - France signed a treaty with Germany last week.
- **GPE.PER** - France likes to eat cheese.
- **GPE.LOC** - The world leaders met in France yesterday.
- **GPE.GPE** - France produces better wine than New Jersey.

In the examples above, the name "France" refers to a range of concepts. Annotators must select the Role which matches the function of the GPE mention.

The GPE role may be used in contexts that highlight the nation (or state or province or city, etc.) aspect of the GPE entity, as distinct from the government, populace, and location, but it may also be used in contexts referring to an indistinct amalgam of more than one of the aspects of a GPE (government, population, location, and nation).

*France produces better wine than New Jersey.*      GPE Role (whole nation)

The following sections give particular guidelines for frequently encountered cases, with examples.

## GPE.ORG

GPE.ORG is used for GPE mentions that refer to the entire governing body of a GPE. It is important to differentiate between a part of the government (the executive branch, the courts) and the entire governing body. Below are some examples of contexts in which GPE.ORG should be used.

### ***Political Communication and Decision-making***

ORGs are responsible for decisions to take military actions. ORGs are also responsible for political communication events such as announcements, agreements, statements, denials, expressions of approval and disapproval, etc. So, if *China* agrees to something, *China* is a GPE.ORG.

*Ritter's return is seen as something of a test of that agreement, under which*  
**[[GPE.ORG]]** *Iraq* **[[GPE.ORG]]** *agreed to give inspectors full access to eight of Saddam Hussein's presidential palaces.*

**[[GPE.ORG]]** *美* **[[GPE.ORG]]** *俄* **[[GPE.ORG]]** *韩* *要求继续进行朝鲜核谈判。*

**[[GPE.ORG]]** *朝鲜* *警告* **[[GPE.ORG]]** *澳* *勿参与拦截否则遭核打击。*

**[[GPE.ORG]]** *朝鲜* *拒绝与* **[[GPE.ORG]]** *日本* *就横田惠假遗骨案进行会谈的要求。*

## **Governments**

While the entity type for governments is GPE, the role for governments should always be GPE.ORG.

*But* **[[GPE.ORG]]** *the Russian* *government* *and many politicians will be stridently critical of the United States if they believe they are being ignored.*

*维护和平、促进发展,是* **[[GPE.ORG]]** *中国政府* *外交方面的一贯立场*

## GPE.PER

As stated above, populations of a GPE are treated as GPE.PER. However, it is sometimes difficult to determine whether a reference to people is a reference to the population as a whole.

**[[GPE.PER]]** *The* *Japanese* *have a considerable responsibility for the wars of the first half of the century*

*[[GPE.PER]]* *中国* *人民* *热爱和平。*

**{{[GPE.PER]日本人}}**一直没有为上个世纪的战争道歉。

In this example, the phrase *the Japanese* may be interpreted as the population of Japan, or the government of Japan, or the Japanese military, or even some part of the Japanese population. If the annotator believes that the phrase in question refers to the population of the GPE, or most of the population of a GPE, then the annotation should be GPE.PER and the mention is a name mention. However, if the annotator believes the phrase refers to a group of people, then PER is the assigned annotation and the mention is nominal because it does not refer to the name of a person. Examples:

**{{[GPE.PER] Cubans}}** have been waiting for this day for a long time.

**{{[PER] A majority of [GPE.PER] Americans}}** believe the allegations against Mr. Clinton are true.

You and **{{[GPE.PER] the [GPE.GPE] American} people}}** have a right to get answers.

**{{[PER] A majority of [PER] Americans surveyed}}** believes allegations Mr. Clinton had an affair while he was President are not relevant.

Yet another cutting edge development by **{{[GPE.PER] the French}}** in their ongoing dealings with their enormous pet population.

**{{[PER] The rest of [GPE.PER] America}}**

**{{[PER] idealistic Europeans}}**

**{{[PER] Americans who want to come and, and learn, uh, from the communities how to live in a community, how to take decisions among the community}}**

I do think there is a danger that **{{[PER] some Chinese}}** may underestimate American will on the Taiwan issue.

当前, **{{[PER]一些中国人}}**热衷于地产投资。

## GPE.LOC

GPE.LOC is used when a mention of a GPE entity primarily references the territory or geographic position of the GPE.

**{{[GPE.LOC]France}}** has an area of 547,090 square kilometers

U.S. warplanes flew over **{{[GPE.LOC]Afghanistan}}**

Northern portions of **{{[GPE.LOC]Montana}}**

2008年奥运会将在[中国]举行.

[苏州]距离[上海]比[杭州]更近

Dateline mentions of GPEs are given a location role.

新华社记者{{GPE.LOC}纽约}报道

## **GPE.GPE**

GPE.GPE is used when more than one of the other GPE roles is being referenced at once or when no one role stands out in the context. Below are a few particular contexts in which GPE.GPE should always be used.

*[中国]近年来的蓬勃发展让世界瞩目.*

## **GPE Premodifiers**

Pre-modifiers are inherently vague and difficult to decompose. For this reason, all GPE pre-modifiers will be assigned the role GPE.GPE.

*{{GPE.GPE} Israeli} troops*  
*{{GPE.GPE} New York} policemen*  
*Prime Minister of {{GPE.GPE} Britain}*  
*{{GPE.GPE} New York} attorney*  
*{{GPE.GPE} U.S.} Commander-in-Chief*  
*{{GPE.GPE} U.S.} surveillance aircraft*  
*{{GPE.GPE} Iraqi} flag*  
*The {{GPE.GPE} California} company*  
*{{GPE.GPE} 美国}国会*  
*{{GPE.GPE} 以色列}官员*  
*{{GPE.GPE} 国家}主席*  
*{{GPE.GPE} 中国}的民用飞机*  
*{{GPE.GPE} 中国}北京*  
*{{GPE.GPE} 辽宁}副省长*  
*{{GPE.GPE} 辽宁}{{GPE.GPE}阜新}孙家湾煤矿*

## **Military Activity**

Similarly, military activities like invasions, military strikes, bombings, etc. are considered to be acts carried out by and directed at entire nations (not distinguishable from the government, people and location of that nation) and therefore are associated with GPEs. Both the aggressors and the victims in these cases are marked GPE.GPE.

*The city could have used some special protection in nineteen seventy-nine when the {{GPE.GPE} Soviet Union} invaded {{GPE.GPE} Afghanistan}.*

*{{[GPE.GPE]美国}}可能会在今年6月就对{{[GPE.GPE]伊朗}}实施军事打击*

### **Activities Associated with GPEs**

Certain activities are associated with GPEs and therefore invoke a GPE role. For example, in a *pro-Iraq rally*, *Iraq* is assigned a GPE.GPE annotation. A rally is generally concerned with a nation as a whole, rather than exclusively a location or government.

*The Palestinian Authority has banned rallies that are pro-{{[GPE.GPE] Iraq}}, but that ban has been widely ignored.*

*{{[GPE.GPE]中国}}近年来的蓬勃发展让世界瞩目.*

### **Athletes, Sports Teams, and GPEs**

Athletes and teams are associated with GPE.GPEs as in *Picabo Street of the United States* below. Please note that *Picabo Street* is a person who was a member of the United States Olympic team.

*Six days into the Nagano Games, one Alpine event \_ the women's super-G won on Wednesday by Picabo Street of the {{[GPE.GPE] United States}}\_ has been completed.*

However, when a GPE name is used as a team name (as in *Boston beat Philly*), the entity is ORG and marked as a Nickname Metonymy.

*{{[New York]}} had a shot to win but Chris Childs missed a three.*

### **Political associations**

Political associations hold between people and GPEs. So in *Hillary Clinton (D-NY)*, NY is marked GPE.GPE.

*"This is going to be a brutal fight," said Rep. Thomas C. Sawyer (D-{{[GPE.GPE] Ohio}}), who has been closely involved in the census and is among those who believe the ongoing debate played a role in Riche's departure.*

### **3. 4 Locations**

Places defined on a geographical or astronomical basis which are mentioned in a document and do not constitute a political entity give rise to Location entities. These include, for example, the solar system, Mars, the Hudson River, Mt. Everest, and Death Valley.

Places distinguished *only* by the occurrence of an event at that position ("the scene of the murder", "the site of the rocket launching") are not entities.

### 3.4.1 Subtypes for Locations

We will further classify Location entities with the following subtypes. Locations that do not fit into the subtypes defined below will not be tagged.

#### Address

A location denoted as a point such as in a postal system or abstract coordinates ("31° S, 22° W"). The name of a location in a postal system is also an address. (**Note:** Sometimes address can be a facility. In this case, make a judgement according to the primary function in the context.)

*[3600 Market Street]*

#### Boundary

A one-dimensional location such as a border between GPE's or other locations.

*Armenia occupies a politically challenging swath of the southern Caucasus, with [borders] shared by Turkey, Azerbaijan, and Georgia].*

#### Celestial

A location which is otherworldly or entire-world-inclusive.

*...as always, eclipse fans are being warned not to look directly at [the sun] because that can cause permanent eye damage.*

*In Armenia, the three of them will join other, similar delegations from around [the world]...*

#### Water-Body

Bodies of water, natural or artificial (man-made).

*[The Missouri River]*

#### Land-Region-natural

Geologically or ecosystemically designated, non-artificial locations.

*...a politically challenging swath of [the southern Caucasus]...*

*[山东半岛]*

*[钓鱼岛]*

*[喜马拉雅山]*

*[这座山]*



### Region-International

Taggable locations that cross national borders.

*...the main threats to [southern Africa] over the coming 15 years...*

[非洲南部]

### Region-General

Taggable locations that do not cross national borders.

*In [another part of the city],...*

*...said Hans Wallner, the director of [the tourist region of Kaprun].*

*...from neighboring Bavaria, in [southern Germany]...*

[北京城区]

[[华中]地区] (the two mentions are correferences)

[德国南部]

### 3.4.2 Sub-parts of Locations and GPEs

Portions of GPE entities or Location entities, such as "the center of the city", "the outskirts of the city", or "the southern half of New Jersey" constitute Location entities in their own right. When general locative phrases like "top," "bottom," "edge," "periphery," "center," and "middle" are used to pinpoint a portion of a markable location, they are markable locations.

*"They tend to live not in [the center of [the country]] but at [its periphery]"*

### 3.4.3 Non-Locations

It is easy to start interpreting all objects as locations. Every physical object implies a location because the space that each physical object occupies is the "location" of that object. In addition, our language is full of location modifiers (which are often prepositional phrases) that pinpoint objects and activities, and even abstract concepts:

*"Your coat is under the dog."*

*"The rabbit is hiding behind that rock."*

*"I have an idea in my head."*

Viewed from a certain angle, "the dog," "that rock" and "my head" become locations. Very "location-ish" nouns make such an interpretation even more tempting:

*"He dropped the logs on the ground."*

*“He put the lamp back in its place.”*

However, none of these are taggable location expressions. They do not fall within any of the classes defined above for taggable locations. The annotator must be careful not to fall down this slippery slope.

Do not tag compass points when they serve as adjectives or refer to directions, as in “the ants are heading north” and “they are found as far north as Maine.” Compass points should only be tagged when they refer to sections of a region, as in “the far west.”

### **3.5 Facilities**

A facility is a functional, primarily man-made structure. These include buildings and similar facilities designed for human habitation, such as houses, factories, stadiums, office buildings, gymnasiums, prisons, museums, and space stations; objects of similar size designed for storage, such as barns, parking garages and airplane hangars; elements of transportation infrastructure, including streets, highways, airports, ports, train stations, bridges, and tunnels. Roughly speaking, facilities are artifacts falling under the domains of architecture and civil engineering.

#### **3.5.1 Subtypes for Facilities**

We will further classify Facility entities with the following subtypes. Facility entities which do not fit into the subtypes defined below will not be tagged.

##### **Airport**

A facility whose primary use is as an airport.

*[new york's la guardia airport] has been a nightmare this year*

*该系统将会在 [115 个美国飞机场] 和 14 个海港 安装.....*

##### **Plant**

One or more buildings that are used and/or designed solely for industrial purposes: manufacturing, power generation, etc.

*...the train ran directly from [the oil refinery] to [the smelter]...*

*[大亚湾核电站]*

##### **Building-or-Grounds**

Man-made/-maintained buildings, outdoor spaces, and other such facilities. This includes anything from a tent to a hotel to a ranch to Disneyland.

*...at [the national archives].*

*[The Berlin Wall]*

*the parades at [Disneyland]*

按国际四星级标准设计建造了多功能现代化酒店--[广州大厦]

[这个位于加基武吉的工人宿舍]，占地共 2 万平方公尺，楼面近 3 万 4000 平方公尺，可容纳 6000 名住客

自 10 月 12 日美国一艘驱逐舰在[也门亚丁港]发生爆炸以来，没有一艘美国船只驶过苏伊士运河。

[柏林墙]

### **Subarea-Facility**

Taggable portions of facilities. The threshold of taggability of subarea-facility is the ability of the area to contain a normally proportioned person comfortably. Individual rooms of buildings are considered subarea-facility, but other portions of buildings, such as walls, windows, or doors, are not tagged.

*...two men who rented [an Aden apartment]...*

[位于直落古楼 3 楼的 房间] 内.....

设施包括[餐厅]、[医药中心（备有 X-光设施）]、[迷你市场]、[行政大厦]等

### **Path**

A facility that allows fluids, energies, persons or vehicles to pass from one location to another. For example: streets, canals, and bridges.

*...and undercover agents patrolling [Aden's streets].*

[Telephone lines] were knocked down...

[ 四川成绵高速公路 ] 是四川省高等级公路主骨架的重要组成部分

载满滑雪者的列车向基茨坦霍恩山顶爬升途中，在进入 [3 公里长的隧道] 行驶了 600 米后，突然起火

## **3.6 Vehicle**

A vehicle is a physical device primarily designed to move an object from one location to another, by (for example) carrying, pulling, or pushing the transported object. Vehicle entities may or may not have their own power source.

### **3.6.1 Subtypes for Vehicles**

We will further classify Vehicle entities with the following subtypes. Vehicles that do not fit into the subtypes defined below will not be tagged.

## Air

Vehicles designed to locomote primarily through the air, not touching water or land.

...[military helicopters] float overhead...

同样，旅客也可以乘坐 [美联航 飞机] 到北京、上海 ...

[[此次发射用的火箭][“宇宙-3M”]]属[轻型运载火箭]

## Land

Vehicles designed to locomote primarily upon land.

[The train] is reported to be about 600 yards inside the mountain.

In one of the worst train accidents in Indian history, [a passenger train bound for Amristar] was derailed, then struck moments later by [another train bound for Calcutta].

澳海关征收的进口豪华[汽车]税...

[载满滑雪者的列车]向基茨坦霍恩山顶爬升途中,在进入 3 公里长的隧道行驶了 600 米后,突然起火

## Water

Vehicles designed to locomote primarily on or submerged in water.

...the Oct. 12<sup>th</sup> bombing of [the U.S.S. Cole].

[俄罗斯核动力潜艇][库尔斯克号]9 月中在北冰洋的巴伦支海出事, 陷入海床。

一支国际潜水队开始打捞今年八月随同[[库尔斯克号]核潜艇]沉入海底的 118 名俄罗斯海军官兵的尸体

[[美国“库勒”号]驱逐舰]在也门亚丁港集装箱码头添加燃料时受到[一艘装有爆炸物的橡皮艇]的撞击而发生爆炸

## Subarea-Vehicle

A portion of a Vehicle entity that is of a size such that humans can fit inside with some degree of comfort. For example: train compartments, the cabin of an airplane, a car's interior, the deck of a boat. We will include the trunk of a car as a Subarea-Vehicle. We will not tag the glove compartment of cars and other compartments that are too small to contain a human.

...[the compartment of the train]...

将列车运行信息传送至车站调度室和[火车驾驶室]中的屏幕上...

乘客们纷纷从[车厢]里逃出来

## Underspecified

Vehicles whose subtype is not specified in the text, or sets of vehicles of different subtypes.

*We have [202,000 vehicles in our fleet].*

*[Israeli tanks and gunships]*

## 3.7 Weapon

A physical device that is primarily used as an instrument for physically harming or destroying entities (taggable or not).

### 3.7.1 Subtypes for Weapons

We will further classify Weapon entities with the following subtypes according to the primary method by which they cause harm. Weapon entities that do not fit into the subtypes defined below will not be tagged.

#### Blunt

Blunt weapons are those designed or used as bludgeoning instruments

*He was beaten repeatedly with [a baseball bat]....*

#### Exploding

Exploding weapons are those that are designed or used to accomplish damage through explosion

*...a small boat packed with [explosives]....*

*An international investigation led to charges against two alleged Libyan intelligence agents of putting [a bomb] on the plane.*

*美国“库勒”号驱逐舰在也门亚丁港集装箱码头添加燃料时受到一艘装有[爆炸物]的橡皮艇的撞击而发生爆炸*

#### Sharp

Sharp weapons are those designed or used to cut, slash, jab, & hack

*Knife*

#### Chemical

A chemical weapon is any device or substance that is designed or has been used for the purpose of causing death or serious injury through the release, dissemination or impact of toxic or poisonous chemicals or their immediate precursors. This is a ACE-oriented modification of the definition of chemical weapons as set out in the National Security Anti-Terror Law.

*Sarin gas*

## Biological

Biological weapons are bacteria, viruses, fungi, toxins, as well as the means of their dispersal, used for the spread of disease among humans, plants & animals. This definition is based on the definition proposed by basicint.org.

*Anthrax*

## Shooting

Shooting weapons are weapons designed or used to send projectile objects at great speed for the purpose of causing harm.

*He was caught attempting to carry [a semi-automatic pistol] through customs...*

## Projectile

Projectile weapons are weapons designed or used to be projected at great speed for the purpose of causing harm

*bullets*

## Nuclear

Nuclear weapons are those designed or used for the purpose of causing damage, death, and harm through the expenditure of radiological or nuclear energies.

*[Nuclear missiles], [hydrogen bombs], [dirty bombs], [deliberately sabotaged nuclear facilities leading to catastrophic meltdowns]*

## Underspecified

Underspecified weapons are weapons whose subtype is not specified in the text, or sets of weapons of different subtypes.

*They searched him for [a weapon].*

*U.N. [arms] inspectors*

*Hussein wanted to send [his tanks and Scuds] through Jordan.*

## 4 Entity Class

How mentions and entities are related in terms of referentiality are recorded according to the following types.

Positive	Referential	Generic	
		Non-Generic	Specific
			Underspecified

	Attributive
Negative	<i>Does not apply to Chinese</i>

We believe that Chinese does not have negative quantifiers over N or NP's. 有人 and 没有人, the approximate equivalents of English “somebody” and “nobody”, are actually existential structures, more like “there’s somebody who ...” and “there isn’t anybody who ...”. The primary reason is that these expressions can never occur in an object position, suggesting that they are not nouns or noun phrases. Therefore, we do not have negative quantified NP’s as a referential class for Chinese ACE.

The distinction between referential and attributive uses of an NP is given by the following definitions for ACE:

A mention is referential if the mention (a) introduces a new entity into the discourse, (b) is a definite descriptive term, a name, or an anaphoric expression for a referential mention previously occurred in the discourse, or (c) a repetition of (a) in subsequent sentences.

A mention is attributive if the mention (a) states a property or properties about an entity referenced by another mention within the same sentence – often as an appositive to or part of a predicate on the other mention, (b) anaphorically links to (a), or (c) qualifies an entity through immediate modification within the same phrase.

Referential mentions are further divided into generic and non-generic classes. A generic mention refers to a class/kind/species of objects or a typical representative of that class/kind/species and does not point to or pick out any specific individual object(s) of that class/kind/species. So if any property predicates of a generic mention, it means the entire class referred to by the mention has that property, or all/most/any members of that class have the property. (any member will satisfy the condition; in conditional clause; in object position)

A non-generic referential mention refers to one or more individual member entities of a particular class. The entity or entities can be accounted for by counting, pointing (specific) or cannot be precisely accounted for (underspecified).

#### **4.1 Attributive (ATR)**

For case (a), the first type of typical attributive mention-entity relationship is predicative NP’s, with or without the copular verb “是”.

*驱逐舰“科尔号”的两名男子的身份已经被证实，是[曾经在80年代协助阿富汗驱逐苏联军队的游击队老兵]。*

Other verbs (particularly so-called “light verbs”) triggering an attributive reading on the following NP include “成为”, “(当)选为”, “称为”, “(好)像”, “叫做”, etc.

小布什是否会当选[下一任美国总统], 还有待佛州的重新计票。  
美国作为[世界第一强国]

Sometimes there is no explicit triggers for an attributive, as the following example shows:

今日统一的德国正在从[经济大国]走向[政治大国]。

The second type of attributive class for (a) involves apposition (Refer to 5.4.1 for detailed discussion). The most common appositive structure is one where the first NP is head by a common noun while the second one is a proper name with no space or punctuation between them. For this kind of structure, we treat the first NP as attributive:

俄罗斯检方已经要求将[涉嫌间谍罪的[美国嫌犯]][波普]]判处 20 年徒刑。

(Note a typical appositive structure invokes three mentions, as shown in the example.) Under this new scheme, “记者” as in “记者张三” and “记者张三李四” is attributive. Similarly, “[俄罗斯总统]普京”, “[中共中央总书记]胡锦涛”.

A less common appositive structure reverses the order of the two NP’s, often with a hyphen (or comma). In this case, the second NP is considered attributive.

宁波, [中国发展最快的城市之一]

There’s also a construction that looks like the above but with a demonstrative and without a punctuation. We also consider it as appositive and it’s treated the same as above.

上海[这颗东方明珠]

For case (b), it includes relative pronouns (who, whom, where, which, that, etc. and 的 in Chinese) when they introduced a relative class modifying an attributive mention. In fact, relative pronouns always inherit their referentiality via anaphorically.

For case (c), only common nouns are tagged for Chinese. English annotation tags both common and proper nouns (or rather, proper adjectives).

## 4.2 Generic Referential (GEN)

The generic reading of an NP is not tied to any specific syntactic structure of the phrase. In fact, the subject NP in all the following examples has a generic reading:



- \* *A lion is a noble beast.*
- \* *The lion is a noble beast.*
- \* *Lions are noble beasts.*
- \* *The lions are noble beasts.*

What is common to these examples is the predicate: it's a kind-level predicate, meaning the predicate describes a property of a kind of entities. Individual- and species-level predicates are also triggers:

*Lions eat other animals.*  
*Dinosaurs are extinct.*

Non-subject NP's are harder to decide. Again their interpretation is dependent on their interaction with the predicate that contains them:

*I've seen lions. (non-generic)*  
*I like lions. (generic)*

Sometimes the NP may be ambiguous if the sentence is isolated. Here the context will play an important role.

*I like the lions. They are noble beasts. | But they're all sleep.*

An NP with an indefinite article can be hard to distinguish. Consider this example:

*He who beats his wife should be punished.*

Should "he" be generic? The answer is yes. Why? Because we cannot say:

\* *He who beat his wife has been punished.*

The first example is what we call "representative of a class", in this, "he" represents the class of "wife-beating husbands". Because it's a representative, it cannot take a stage-level predicate as the second example shows.

Examples from actual texts:

*建立[一支与打赢未来战争相适应的人才指挥 部队].*

*[高科技 部队] 如果没有[高素质 人才] 支撑, 再先进的装备也是一堆废铁.*

Personal pronouns – in particular, first and second person pronouns – may have a generic reading even when they are not used anaphorically. In Chinese the third

person pronouns can only have a generic reading when they refer back to a generic NP in the discourse.

一个共产党员无所畏惧。哪里有风险[他]就会出现在哪里。

Some hints for distinguishing between generic and non-generic:

1. If the NP is in the form of “a(n) + N”, can you replace the NP with a bare plural? If so, it’s generic.
2. For a bare plural NP, can you precede the NP with a demonstrative in subsequent references? If not, it’s generic.

Note that a universal quantifier does not necessarily trigger a generic reading even though the quantified NP refers to the entire set of entities.

Everyone likes fish. (GEN)  
Everyone went there. (SPC)

### **4.3 Non-generic Specific Referential (SPC)**

A non-generic specific referential NP refers to a particular, unique object (or set of objects), whether or not the author or reader is aware of the name of the entity or its anchor in the (local) real world. Criteria for verifiability include (but not exclusively): can be named; or can be pointed out; ...

### **4.4 Underspecified (USP)**

We reserve the term underspecified for non-generic non-specific reference. Underspecified entities include quantified NP’s in modal, future, conditional, hypothetical, negated, uncertain, question contexts (in all cases the entity/entities referenced cannot be verified, regardless of the amount of “effort”).

*[Many people] will participate in the parade.*

*I don’t know [how many people] came.*

*Do you know [how many people] came?*

*We will elect [five new officials].*

While we try to define the other categories as precisely as possible, annotators may still encounter NPs that cannot be classified. In these cases, annotators should make these NPs Underspecified. By partitioning these truly ambiguous cases into the USP category, annotators will be able to make clearer distinctions between the other four categories, thus improving consistency.

Note that some previously annotated “generic” references should now be tagged as USP.

对于长时间耽误请求外援的作法，[人们]尤其感到愤怒。

[您]现在收听的是美国之音的《时事经纬》节目。

有[人]猜测当初的爆炸发生以后，船上可能有[人]没有死亡。

他说 要把 [多数人]的利益，[多数人]的愿望，[多数人]的意见作为我们制定政策的出发点和归宿。

瑞士政府暂时将不向伊拉克派驻[大使]，但不久将派[两名外交官]赴 伊重开驻伊使馆。

#### **4.5 Some fine distinctions and pronominals**

In summary, an attributive mention does not directly refer to an entity or a class. It assigns a property to an already mentioned entity, group of entities, or class. In this sense, it’s more like an adjective than a noun in the conventional sense.

The distinction between generic and non-generic referenced entities is that the former are a class, a kind, or a species whereas the latter are individual instantiations of a class.

The distinction between specific and underspecific referenced entities is that the former exist or have existed in the real world and can be enumerated whereas the second may NOT exist in the real world and cannot be enumerated. Be careful about groups of people who will do something in the future but whose membership is known constantly. They should not be treated as USP.

Some NP’s may be ambiguous if the sentence is considered out of context. It is very important that for less clear cases, one has to look carefully at the context for disambiguation.

The reference relation between a pronominal and an entity can be direct or indirect (anaphoric). Direct reference of a pronominal points to an object/objects in the world directly – pronouns of all persons can be used this way. Thus “I” often refers to the addresser directly and whereas “you” often refers to the addressee directly. Anaphoric reference of a pronominal refers to an entity via co-reference with another mention of the entity in the same text.

### **5 Mention Types**

For each entity, we record and coreference all mentions of the entity. Mentions will frequently be nested; that is, they will contain mentions of other entities. For example, the phrase

*The president of Ford*

is a mention of an entity of type Person, and contains the name "Ford", a mention of an entity of type Organization. It is even possible for a noun phrase to contain an embedded mention of the same entity. For instance, the phrase

*The historian who taught herself COBOL*

evokes a Person entity with three mentions: the entire phrase, and the words "herself" and "who".

For each mention, we record its full extent and its head.

### **5.1 Mention Extent**

The extent of a mention consists of the entire noun phrase. In case of structures where there is some irresolvable ambiguity as to the attachment of modifiers, the extent annotated should be maximally inclusive. In the case of a discontinuous constituent, the extent goes to the end of the constituent, even if that means including tokens that are not part of the constituent. Thus, in

*I met some people yesterday who love chess.*

the extent of the mention is the entire phrase:

*[Some people yesterday who love chess]*

The extent includes all the modifiers of a noun phrase, including prepositional phrases and relative clauses.

Generally speaking, tokens are broken at white space, and each item of punctuation is treated as a separate character. As a rule, we do not include punctuation such as commas, periods, and quotation marks in the extent of a mention unless words included within the extent continue on after the punctuation mark. Possessive endings ('s) are treated as separate tokens, and contractions are split (so that "we're" becomes the two tokens "we" and "re"). Extents must begin at the beginning of a token and end at the end of a token.

### **5.2 Mention Head**

In addition to the extent of the nominal phrase, the head of the phrase must be marked. In

*The hurricane destroyed the new glass-clad skyscraper.*

the full extent of the mention is

*The new glass-clad skyscraper*

and the head is *skyscraper*. If the head is a proper name, the whole extent of the name is considered to be the head. In the following examples, the mention is enclosed in brackets and the head is underlined:

*[Fred Smith] became [the new prime minister].*

The job fell to [Abraham Abercrombie III].

[北京市] [黑龙江省] [香港特别行政区] [黄河] [青海湖]

Determining the head of a common noun (NOM) in Chinese is intrinsically hard and there are no agreed rules on word segmentation. The new task of annotating nominal pre-modifiers make it even harder because the consequence is either one or two mention-entity pairs. The annotator is encourage to follow the Chinese Tree Bank guidelines and their native intuitive. Here're some simple rules:

1. Never tag suffixes alone as the head (suffixes such as 家 in 作家, 者 in 记者, 人 in 候选人, 残疾人, 室 in 研究室 etc.
2. Bi-syllabicity is always preferred by the Chinese speaker. If possible, treat a bi-syllabic string as one word except the syntactic/morphological properties favor the other way around (such as in 5), for example:  
[各地] [各国] [各党];  
[本校] [全球] [全国] [当地] [每人]  
but [本单位]
3. For a tri-syllabic string “ABC”, do similarly as in 3:  
[饺子馆] [料理店] 石油[化工厂];
4. If a measure word can be inserted between the number without changing the meaning, separate the number from the head. Otherwise, treat it as a single head, for example:  
[三人] [数人] [三国] [四省] [两国] [多国]  
[三排] [一方] [三者] [一行]
5. Names of relatives such as [三叔][大姑父] is treated as one word
6. When the string is a pronoun and a noun, if both pronoun and noun are monosyllabic and noun is bound, treat it as a single head; otherwise, treat the noun only as a head, for example:  
[我校] [我国] [我单位] [我公司] [我们国家]
7. For the pattern X + N, where X modifies the N, treat X + N as a single head if X is a prefix. If x is a non-predicative adjective, treat X + N as a single head if both X and N are monosyllabic when the meaning of X + N is noncompositional. A simple test of compositionality is to insert 的 between X and N to see whether the meaning changes. If it doesn't, treat N alone as head. For unclear cases, if both X and N are monosyllabic, treat X + N as a single head. For example:  
[阿爸] [女人] [强队] [小媳妇] [大海]  
[小国] [大国]  
[前总统] [原在华老挝难民]
8. For localizer + N and N + localizer, if both localizer and N are monosyllabic, treat the string together as a head. For example:

[前院] [境外] [国内] [海外] [后花园]

9. Ask.

### 5.3 Mention Types

Mentions are categorized at several levels. At the top level there are two major types, simple and complex mentions. This top level distinction is motivated by the fact that some mentions have complex syntactic structures that cannot be easily annotated without breaking the syntactic analyses and/or information loss.

Subcategorizations of mentions types are syntactically motivated. The following table lists all the types we distinguish for Chinese.

Simple Mentions	Single head	Named ( <b>NAM</b> )
		Nominal ( <b>NOM</b> )
		Pronominal ( <b>PRO</b> )
	Headless nominal ( <b>HLS</b> )	
Complex Mentions	Apposition Constructions ( <b>APP</b> )	
	Extended Apposition Constructions ( <b>EAP</b> )	

Constructions that we DO NOT single out for Chinese:

1. Partitive: Although Chinese has partitive constructions, they are different from English partitive constructions. First of all, the “part” part may repeat the head of the whole part. Secondly, they’re much less common – their English translation is more like “x out of y”. The more common expression in Chinese – at least in the written language – is “其中...”. The only expression that looks like English “some of the ...”, “one of the ...” is “... 之一”. But we treat “—” as headless and the “...” is included in its extent.
2. Complex Constructions taking a Relative Clause (ARC): Replace it with EAP.

Furthermore, we do not distinguish between bare and quantified NP for Chinese, for the following reasons.

1. Chinese does not have articles – the so-called bare NP is extremely common in Chinese and they CAN be used for definiteness.
2. The possessive marker “的” may often be omitted and because of this, it’s often difficult to tell whether a pronominal modifier has a possessive meaning or simply qualifies the noun head. Furthermore, “的” is also used to mark an adjectival modifier.

### 5.3.1 Simple mention headed by a proper name (NAM)

Proper nouns and nicknames. Names are atomic. We will not tag portions of Names as mentions.

*[John]*

*[Defense Secretary William Cohen]*

*[The Jeluzoon Refugee Camp near [Ramallah]]*

*[North Dakota State University in [Fargo]]*

*[The house of representatives]*

*[多普]是在半年多之前因为被怀疑指控收集情报，从事间谍活动而被[俄国联邦安全局]拘捕的*

*[朱容基]*

*[[美国]佛罗里达州]*

Note that we do not look into the internal structure of a proper name, which may consist of a single proper noun or a group of common nouns or a mixture of both. We do however mark the head if there's any modifier. One must be careful when separating modifiers from the proper name itself.

Some specific details:

When a person proper name is followed by a title, the title is part of the extent, but not part of the head. In other words, titles are like post-modifiers.

*[[朱容基]<sub>NAM</sub> 总理]<sub>NOM</sub>*

*Zhu Rongji Prime Minister*

*[[张]<sub>NAM</sub> 教授]<sub>NOM</sub>*

*Zhang professor*

This is different from where a proper name is preceded by a position – such constructions are APP like so:

*[[[中国]总理]朱容基]*

*China PM Zhu Rongji*

Some vehicles, particularly aircraft carriers, submarines, famous airplanes, etc. may be named after something else. Frequently one will find the name followed by “号 NAME” and then a common noun for the type such as “航空母舰”, “潜艇”, “飞船”, etc. Both the common noun and the word “号” can be omitted from the head of the

Name if it is clear from the context what type of vehicle is being mentioned. If the common noun is present, we tag two mentions: a nominal mention with the common noun as the head and a NAM mention that extends the string up to “号” if it’s present. The two mentions are co-referenced.

[[库尔斯克号]<sub>NAM</sub>潜艇]<sub>NOM</sub>

*Kursk NAME Submarine*

[[空军一号]<sub>NAM</sub>飞机]<sub>NOM</sub>

*Air I NAME Plane*

[[长征三号]<sub>NAM</sub>火箭]<sub>NOM</sub>

*Changzheng III NAME Rocket*

[[神舟四号]<sub>NAM</sub>宇宙飞船]<sub>NOM</sub>

*Shenzhou IV NAME Spaceship*

[[东方红号]<sub>NAM</sub>列车]<sub>NOM</sub>

*Dongfanghong NAME Train*

If there’s any pre-modifier, it should be in the extent of the NOM mention.

[[俄罗斯][库尔斯克号]潜艇]

*Russian Kursk NAME Submarine*

A proper name is always taggable as long as it is NOT part of an atomic mention of an entity, regardless of whether or not the entity is taggable under the current specifications. For example:

[意大利]面 Italian pasta

[法国]红酒 French wine

[诺贝尔]奖 Noble Prize

### 5.3.2 Simple mention headed by a common noun (NOM)

A nominal is a noun phrase headed by a common noun.

[The lawyer]

[the Yemeni port of Aden]

[some American executives]

[the southern tip of Africa]

[a crowd of angry muslims]

[lawyers]



*[US officials]*  
*offices in [foreign countries]*  
*[representatives of the company's Tananger branch]*  
*[中国的互联网用户]*  
*China DE Internet users*  
*[俄罗斯和挪威潜水员]*  
*Russia and Norway divers*

Noun + noun compounding is the most frequent morph-syntactic process in Chinese. We annotate the nominal modifier as a NOM, and its entity class usually falls into Generic. We exclude proper names from this category, which is most often Specific. For example:

*[火车]司机*  
*南斯拉夫[总统]选举的计票结果是“骗人”和“作弊”*  
*国务院减轻[农民]负担联席会议决定*

### 5.3.3 Pronouns (PRO)

A pronominal is a word used as a substitute for a noun phrase. Pronominals refer to persons or things that are previously specified or understood from the context. Pronominals are marked whenever they reference a taggable entity.

*[他]目前在德国的兰斯图尔美国军队医院接受治疗*

The following are some additional rules that apply to pronominal mentions in Chinese.

#### 5.3.3.1 Personal Pronouns

The following table lists person pronouns in Chinese

	Singular	Plural
1 <sup>st</sup> Person	我/咱	我们/咱们
2 <sup>nd</sup> Person	你/您	你们
3 <sup>rd</sup> Person	他/她/它	他们/她们/它们

As shown above, the plural form simply consists of a singular pronoun and the plural morpheme 们. Also, there are a few dialectal variations such as 俺.

Personal pronouns in Chinese do not usually refer to animals and inanimate objects, not even the third person pronouns. The reason for this is that Chinese is a PRO-drop language.

With a few exceptions such as 可怜的我 (“Poor me!”), Chinese person pronouns are not modified.

The lack of possessive pronouns in Chinese is actually a benefit for ACE annotation. Consider:

*你的孩子来了, 我的没来.*

*You DE child come PFT, I DE not come*

*Your child has come. Mine hasn't.*

The English possessive pronoun “mine” here actually invokes two entities, “me – the speaker” and “the speaker’s child”. But there’s no easy way to mark the two entity mentions in English. For Chinese, however, we can simply mark “我的” as one mention with “的” as head (see “headless mentions” below) and “我” as another mention: [[我]<sub>2</sub>的]<sub>1</sub>.

### **5.3.3.2 Reflexives**

The reflexive morpheme 自己 (“self”) can be used in two ways, as a reflexive pronoun or as an adverbial to serve to contrast with oneself with others. Regardless of what function it serves in a sentence, it’s always marked as pronominal if the entity it refers to falls into one of the seven types.

However, the morpheme 自己 can appear immediately after a person pronoun (such as 他自己, 他们自己) and it’s sometimes not clear which function it serves. For simplicity, we mark it as a single pronominal mention.

### **5.3.3.3 Demonstratives**

Demonstratives (这, 那, 这些, 那些, 这儿, 那儿, 这里, 那里, 此) are marked as pronominal when appear alone as NP’s and stand for previously occurred NP’s (or their reference can be determined from the context). They are not markable when they server as determiners of NP’s.

### **5.3.3.4 Other pronouns and pronoun-like words/phrases:**

- Possessive pronoun 其: third person pronoun used in written/formal language. Mark 其 as a pronominal mention if the entity is one of the seven types. Additional expressions consisting of 其:
  - 其中: “lexicalized” phrase meaning “among them; of them; in it”. Mark 其 as a pronominal mention if the entity it refers to is one of the seven types.

- 其间: “lexicalized” phrase meaning “between them; in it”. As with 其中 above, mark 其 as a pronominal mention if the entity it refers to is one of the seven types.
- 其他/其它: highly lexicalized expression meaning “the rest of them; the others”. Syntactically, it behaves a like a determiner. It’s usually used a context where a subset of a set has previously been mentioned and this expression, with optional noun head and/or –de, is used to refer to the other subset. As with demonstratives, stand-alone 其他/其它 is marked as headless pronominal.
- 其余: Similar to 其他/其它.
- 彼此 (each other): meaning “each other”. Treat the entire unit as pronominal. However, 相互 and 互相 is not taggable because they are adverbials.
- 本 (self): this morpheme, originally meaning “the root of a plant”, can function as a “determiner” meaning “this” – often related to the speaker. Two words with this morpheme draws special attention.
  - 本人: can mean “I (myself, me)” or “oneself/in person”. In either case, treat it as pronominal.
  - 本身: meaning “itself/in itself”. Treat it as pronominal.

If 本 is only used as a determiner, do not treat it as pronominal.

*缅甸贸易部长吞基和泰国外交部长格森\*格森西分别代表{本国政府}在协定上签字.*

*Burma Trade Minister Tunki and Thai Foreign Minister Gesen respectively represent self country government at agreement sign*

*Burmese Minster of Trade Tunki and Thai Minister of Foreign Affair Gesen respectively represented their own government to sign on the agreement.*

- 自身(self): There are many words with 自 as their component morpheme. Among them, we consider 自身 as pronominal similar to 自己. We DO NOT mark 自 that’s part of a compound/complex verb (for example, 自重, 自爱, 自筹, like self-determination, self-esteem in English), where 自 is non-specific anyway.
- 前者/后者: meaning “the former/the latter”. Treat them as pronominal.
- Locatives with demonstrative determiners: 这里, 那里, 这儿, 那儿  
Strictly speaking, these expressions are full-fledged noun phrases and used as locative phrases. Morpho-syntactically, they’re different from “here” and “there” in English. But for ACE, we treat them just like “here” and “there” and tag them as pronominal mentions.

- 我+GPE entity (monosyllabic) – one mention; 我+ORG entity (monosyllabic) – one mention.
- 我们+GPE entity – two mentions of the same entity; 我们+ORG entity – depending on the context, either two mentions of the same entity (in which case 我们 would be an ORG mention) or of two distinct entities (in which case 我们 would be a PER mention).
- 大家 (everyone)– tag it as pronominal. It can be generic or specific depending on the context.
- 对方 (the other party)– tag it as pronominal. It may or may not be coreferenced with an antecedent just as like English “each other”. Note 相互 or 互相 is not taggable because it’s an adverbial.
- 双方 (both parties)– tag it as pronominal.
- 此间 –tag it as pronominal.

### 5.3.4 Headless (HLS)

A headless mention is an NP without an overt syntactic head, but the head (in Chinese) is recoverable syntactically from the context even though the version that includes the head may not be preferred for the context by the native speaker. Although these mentions are technically headless, we will mark the rightmost character or word of the headless mention as the head in annotation.

A: 你喜欢红色的小轿车吗?

B: 我喜欢白色的。

Here [白色的] is a headless mention and we tag the last word 的 as the head.

Other cases include “number + classifier” (三个没来), “bare quantifier” (很多没来), etc.

Again we do not have a separate type for “partitive constructions” in Chinese Entity.

### 5.3.5 Mentions in Conjunctions

Mentions of a similar type are frequently used together via conjunction, disjunction, enumeration, etc. and the entire expression appears a single syntactic phrase in a sentence. The following patterns have been observed on Chinese data:

- 1) NPs can be in conjunction without any common component such as 张三和李四 *Zhangsan and Lisi*;

- 2) NPs can be in conjunction which functions as premodifiers such as 张三和李四的书 *Zhangsan and Lisi De book*, 张三和李四的妻子 *Zhangsan and Lisi De wife*;
- 3) NPs can be in conjunction with common components that are (pre-) modifiers such as 该公司的张三和李四 *this company DE Zhangsan and Lisi*.

For 1) and 2), each mention in the conjunction should be tagged separately, but a simple conjunction like 张三和李四 as a whole will not be tagged (each conjoined mentions, however, are tagged individually as NAME-PER), neither do we tag the conjunction in a modifier position as in 张三和李四的书 or 张三和李四的妻子.

For construction as 3) where there's a modifier (regardless of whether or not the modifier has taggable mentions) that applies to two or more mentions that are conjoined, we treat the entire extent of conjunction including the modifier as the extent; each of the conjoined mentions are tagged as the head; For example, in 该公司的张三和李四, there are three mentions:

[该公司的张三和李四]<sub>NAM</sub>  
 [该公司的张三和李四]<sub>NAM</sub>  
 [该公司]<sub>NOM</sub>的张三和李四  
*this company DE Zhangsan and Lisi*

Here are some additional examples for different cases:

[俄罗斯的总统和官员]<sub>NOM</sub>  
 [俄罗斯的总统和官员]<sub>NOM</sub>  
 [俄罗斯]<sub>NAM</sub>的总统和官员  
 [报名参展的 国家和地区]<sub>NOM</sub>  
 [报名参展的 国家和地区]<sub>NOM</sub>  
 [[上海]<sub>NAM</sub>、 [江苏]<sub>NAM</sub>、 [广西]<sub>NAM</sub> 等省市自治区]<sub>NOM</sub>  
 [[上海]<sub>NAM</sub>、 [江苏]<sub>NAM</sub>、 [广西]<sub>NAM</sub> 等省市自治区]<sub>NOM</sub>  
 [[上海]<sub>NAM</sub>、 [江苏]<sub>NAM</sub>、 [广西]<sub>NAM</sub> 等省市自治区]<sub>NOM</sub>  
 [江西和湖南省]<sub>NAM</sub>  
 [江西和湖南省]<sub>NAM</sub>  
 [江西和湖南省]<sub>NOM</sub>  
 [[江西]<sub>NAM</sub> 和 [湖南]<sub>NAM</sub> 两省]<sub>NOM</sub>

[南北韩] NAM  
 [南北韩] NAM  
 [南北韩] NAM  
 [黄浦南浦大桥] NOM  
 [黄浦南浦大桥] NAM  
 [黄浦南浦大桥] NAM  
 [[中国] NAM 的神舟三、四号飞船] NOM  
 [中国的神舟三、四号飞船] NAM  
 [中国的神舟三、四号飞船] NAM  
 [[俄罗斯] NAM 的明斯克和库尔斯克号潜艇] NOM  
 [俄罗斯的明斯克和库尔斯克号潜艇] NAM  
 [俄罗斯的明斯克和库尔斯克号潜艇] NAM  
 [[中] NAM [日] NAM [韩] NAM 等三国] NOM  
 [[钱其琛] NAM 和 [温家宝] NAM 副总理] NOM  
  
 [[国务院] NAM 副总理] NOM 兼 [[卫生部] NAM 部长] NOM [吴仪] NAM] APP  
 [[[国务院] NAM 副总理] NOM [钱其琛] NAM 和 [温家宝] NAM] APP

But for expressions like “上海这颗东方明珠” and “张三这个无赖”, we treat them as APP. See below.

## 5.4 Complex Constructions

The purpose of complex constructions is to identify difficult regions where the simple mention extent rules do not apply. We do not identify heads for complex constructions. Within the extent of a complex construction, simple mentions will be annotated. Each of these complex constructions has rules for extent.

### 5.4.1 Appositive Constructions (APP)

Superficially apposition looks similar to noun-noun compound in which the first noun modifies the second. However, they are different in that in an appositive relationship, the first NP and the second NP roughly refers to the same entity, whereas this is not the case in a noun-noun compound. In an appositive construction, one mention is

always the attribution of the other. See 4.1 for details on which one is attributive and which one is referential.

[[人口调查局发言人]<sub>NOM</sub> [安格拉.布里提鄂姆]<sub>NAM</sub>]<sub>APP</sub> 说  
(Note [人口调查局发言人] is attributive)  
[[波斯湾]<sub>NAM</sub> [国家]<sub>NOM</sub> [巴林]<sub>NAM</sub> 和 [卡塔尔]<sub>NAM</sub>]<sub>APP</sub>  
[[此次发射用的火箭]<sub>NOM</sub> [“宇宙-3M”]<sub>NAM</sub>]<sub>APP</sub>

The entity type and reference class of an appositive construction are determined by the referential mention. In Chinese, this usually is the second element of an APP, although it's possible for the first element to be non-attributive when there's a hyphen between the two or the second element has a demonstrative determiner as follows.

上海——一个拥有“东方巴黎”之称的大都市

上海这颗东方明珠

张三这个无赖

#### 5.4. 2 Extended Appositive Construction (EAP)

Sometimes, an APP may have a modifier. For example,

正在柬埔寨访问的国家主席江泽民

The relative clause 正在柬埔寨访问的 actually modifies the entirety of 国家主席江泽民 (English guidelines may treat the modification between 正在柬埔寨访问的 and 江泽民, hence the notion of ARC). We use the notion of “extended appositive construction” to capture this structure, like so:

[正在柬埔寨访问的[[国家主席]<sub>NOM</sub> [江泽民]<sub>NAM</sub>]<sub>APP</sub>]<sub>EAP</sub>

Here are more examples:

[正在崛起的[[新兴城市]<sub>NOM</sub> [宁波]<sub>NAM</sub>]<sub>APP</sub>]<sub>EAP</sub>.

莫斯科的一个法庭否决了[涉嫌从事间谍活动的[[美国人]<sub>NOM</sub> [波普]<sub>NAM</sub>]<sub>APP</sub>]<sub>EAP</sub> 要求医疗检查的请求

[获得诺贝尔奖的[[中国作家]<sub>NOM</sub> [高行健]<sub>NAM</sub>]<sub>APP</sub>]<sub>EAP</sub>

Sometimes it is ambiguous to tell whether the extension is modifying the nominal head or the entire apposition. We simply uses “是” test for a quick decision:

**Given X A B, if AB is an appositive structure and if “AB 是 X” is good, tag it as EAP like [X[[A] [B] APP]EAP; otherwise, tag it as APP like [[XA][ B]]APP.**

So under this schema: we have [[此次发射用的火箭] NOM [“宇宙－3 M”] NAM]APP, because it is bad to say 火箭“宇宙－3 M” 是 此次发射用的, but good to say “宇宙－3 M” 是 此次发射用的火箭.

## 6 Nickname Metonymy

Metonymy occurs when a speaker uses a reference to one entity to refer to another entity (or entities) related to it. For example, in the sentence below *Beijing* is a capital city name that is used as a reference to the Chinese government:

*Beijing will not continue sales of anti-ship missiles to Iran.*

While this phenomenon occurs in many different contexts and to varied degrees, we are only interested in what we are calling “Nickname Metonymy” for the purposes of this stage of the annotation process. This kind of metonymy occurs when the name of one entity is used to refer to another entity. The sections below outline several common examples. This is not an exhaustive list. Any example of this kind of reference should be identified. We will coreference the mention with the entity to which the mention refers in the context and indicate that this is an example of Nickname Metonymy by selecting the check box.

### 6.1 Capital City or Government Seat (FAC) Names standing in for Country’s Government

Cases in which the capital city or the building that is the seat of government is used to refer to the nation's government are marked as classic metonyms.

*Secretary of Defense William S. Cohen said today that he is satisfied {[GPE.ORG]Beijing} will not continue sales of anti-ship missiles to Iran as he wrapped up a four-day visit here that underscored improving Sino-American military ties.*

In this example there are two senses of the word Beijing: the city Beijing and the government of China. We will tag the government of China (GPE.ORG) mention and coreference it with the China entity. If there is a later mention of the city of Beijing (for example, Cohen left {the city} this morning), it would be a GPE.LOC mention of the Beijing entity.



Common examples of government seats used to refer to the nation's government are "The White House" and "The Kremlin". We will tag these according to the entity to which they refer.

## 6.2 City name for Sports Team

When the name of a GPE refers to a sports team, the mention is marked for Nickname Metonymy. The mention is coreferenced with the sports team's entity.

*{{[ORG]America} brought home the gold.*

## 6.3 Cross-Type Metonymy Relations

At the Entity annotation stage, we will group entities of different types together when they refer to different aspects of the same underlying entity simply by clicking the button of Cross\_Type Met in the tool after choosing the two entities.

The most common Cross-Type Metonymy link occurs between Organizations and the Facilities they occupy. These two entities are often using the same strings of text.

### Examples

In this example, there are two mentions of a hospital. The first mention is referencing the physical, building or hospital facility which is a FAC. The second refers to the organization that runs or administrates the hospital which is an ORG.

Wouters, 42, died an hour later at **St. John Macomb Hospital**. The suspect died later the same night, **hospital** spokeswoman Rebecca O'Grady said Thursday. His name wasn't released.

We will then choose the two entities in the annotation window and then click Cross\_Type Met to mark that one is a metonymy of another.

## 7. Appendix:

Some notes on certain mentions:

1 What's the head for the following entities? AS SHOWN.

[各地] [各国] [各党][两国]

[饺子馆] [料理店] [石油化工厂] ?

2. These entities should be tagged? YES.

[公安][工商][纪检] [法院] 等部门

架[车]逃亡, [车]上乘客, 开[枪]射击

3 全国[各地], 全省[各地], the entity of [各地] is a LOC

4 [江西[全省]] or [江西][全省], we prefer the first one

6 What's the difference between GPE.ORG and ORG.Government?

GPE.ORG is the entire government while ORG.GOV is a subpart of the government. In Chinese, when 政府 is used, it can either refer to government or administration. Tag it as GPE.ORG when it refers to government, as ORG.GOV when it refers to administration.

7 How to distinguish GEN, SPC, USP:

--Always keep the hierarchical tree in mind

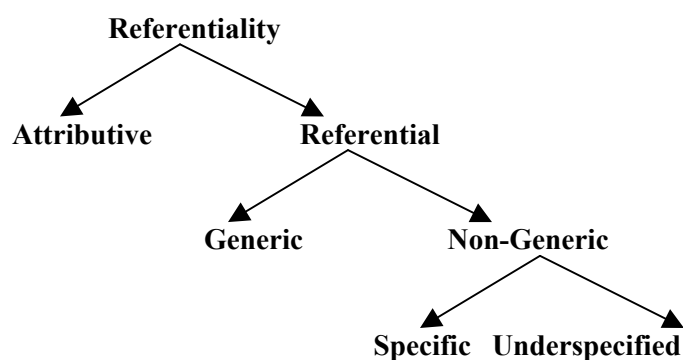
--If the entity is involved in a specific event, at a specific time or place, it is most likely SPC.

#### Entity Types and subtypes

<i>Person</i>	<i>Organization</i>	<i>GPE</i>	<i>Location</i>	<i>Facility</i>	<i>Weapon</i>	<i>Vehicle</i>
Group	Government	Continent	Address	Airport	Blunt	Air
Individual	Commercial	Nation	Boundary	Plant	Exploding	Land

indefinite	Educational	State or Province	Celestial	Building/ Grounds	Sharp	Water
	Entertainment	County or District	Water Body	Subarea Building	Chemical	Subarea Vehicle
	Media	Population Center	Land-Reg-Natural	Path	Biological	Underspecified
	Religious	GPE Cluster	Region International		Shooting	
	Medical-Science	Special	Region General		Projectile	
	Sports				Nuclear	
	Non-Govern				Underspecified	

**Entity Class**



**Mention Type**

Simple Mentions	Single head	Named ( <b>NAM</b> )
		Nominal ( <b>NOM</b> )
		Pronominal ( <b>PRO</b> )
	Headless nominal ( <b>HLS</b> )	
Complex Mentions	Apposition Constructions ( <b>APP</b> )	
	Extended Apposition Constructions ( <b>EAP</b> )	