

Chinese Annotation Guidelines for Entity Detection and Tracking (EDT)

Version 4.2.4 20040205¹

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*

For Phase 3 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.
- Facility - Facility entities are limited to buildings and other permanent man-made structures and real estate improvements.
- Location - Location entities are limited to geographical entities such as geographical areas and landmasses, bodies of water, and geological formations.
- 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.
- 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.

¹ Chapters 1~3 and chapter 6 are almost the same as the English version (Version 4.2.4 20040204). Readers familiar with the English version may skip these chapters.

- Weapon – Weapon entities are limited to physical devices primarily used as instruments for physically harming or destroying animals (often humans), buildings, or other constructions.

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]

3.1.1. Subtypes for Person

We will not tag subtypes for Person entities.

3.1.2 Titles, Honorifics, and Positions

In English, titles and most honorifics precede the name. We will not consider these to be part of the name of a Person. The parts of titles are taggable if they refer to entities.

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]

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. Note that capitalization information may not be available in speech transcripts.

If you believe in [god], you must... name mention
Although he felt like he was [a god], he... 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

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

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]

3.1.4.1 Family Names

Family names are to be tagged as Person.

[The Kennedy]'s

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]

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.

Government – Government organizations are those that are of, relating to, or dealing with the structure or affairs of government, politics, or the state.

Governing bodies of GPE's are included in this subtype.

[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...

Commercial – 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].

Educational – 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.

Non-profit – A non-profit organization is an organization or taggable portion of an organization that is focused primarily upon promoting, promulgating ideas/products/services and which retain capital solely for the purpose of the continuation of the organization, rather than for profit.

[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],

...

Other – Other will be used for all other organizations.

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

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

3.2.2 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.

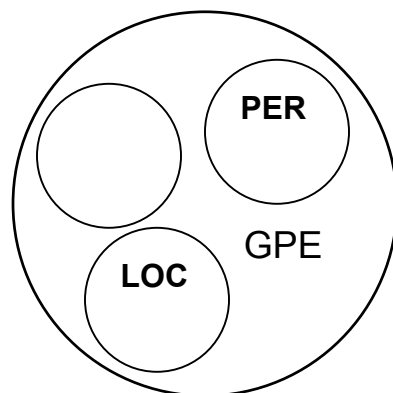
Sets of people who are not formally organized into a unit are to be treated as a person entity rather than an organization entity. It is often difficult to tell the difference between organization entities and collections of individuals tagged as person entities. Example organization-like nouns which 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.

3.2.3 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*", "*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 entities.

3.3.1 Subtypes for GPEs

We will further classify GPE entities with the following subtypes.

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.

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].

State-or-Province – Taggable mentions of the entireties of any state, province, or canton of any nation.

[Salzburg] governor Schausberger said...

County-or-Prefecture – Taggable mentions of the entireties of any county, district, prefecture (地区 such as 扬州地区) or analogous body of any state/province/canton.

Recounts are only just beginning in [Palm Beach] and [Volusia] counties.

Note in China, a county or prefecture, just like a province, also has a “capital” 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 “CITY”.

Population Center – Taggable mentions of the entireties of any GPE that can be called as city, town, or village.

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.]

As noted above, Chinese annotators should be careful to choose between “county/prefecture” and “city/town” when there are no distinguishable names.

Other – All other GPE’s, including Palestine, Taiwan, and mainland China (when it contrasts with Taiwan).

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

...and the government tried to redefine [the West].

[[中国]_{GPE.NATION} 大陆]_{GPE.OTHER}

3.3.2 Honorary GPEs

For this annotation task, named geographical entities that can also function as political entities that are commonly referred to by those names will be considered GPEs rather than Locations. Following is a non-exhaustive list of entities that should also be GPEs for this task.

Eastern Europe, Western Europe, EU, the Middle East, Palestine, Southeast Asia

Other incidental, non-political clusters of GPEs are still considered Locations. For example, *the southern United States* is a Location. On the other hand, coalitions of governments, as well as the UN, are organizational bodies and should be marked Organization.

[NATO] peacekeepers arrived in the valley before nightfall.

3.3.2 Honorary GPEs

For this annotation task, named geographical entities that can also function as political entities that are commonly referred to by those names will be considered GPEs with Other as the Subtype rather than Locations. Following is a non-exhaustive list of entities that should also be GPEs for this task.

Eastern Europe, Western Europe, EU, the Middle East, Palestine, Southeast Asia
 Other incidental, non-political clusters of GPEs are still considered Locations. For example, *the southern United States* is a Location. On the other hand, coalitions of governments, as well as the UN, are organizational bodies and should be marked Organization.

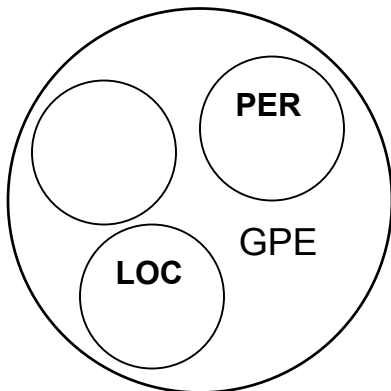
[NATO] peacekeepers arrived in the valley before nightfall.

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

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 that 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).

- | | |
|---|--------------------------------|
| <i>France produces better wine than New Jersey.</i> | GPE Role (whole nation) |
| <i>France's greatest national treasure</i> | GPE Role (indistinct referent) |

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 agreed to give inspectors full access to eight of Saddam Hussein's presidential palaces.*

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.*

GPEs and Government Organizations

GPEs modifying government organizations, like *New York police department* and *Kentucky fire marshal's office*, reflect a relationship between the organizations and the governmental aspect of the GPE, so they are assigned a GPE.ORG markup.

*The department said Sonabend can appeal to **[[GPE.ORG]]** Switzerland's supreme court.*

Note that the GPE mention is a possessive in this example and not a premodifier.

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*

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 th- **{{[GPE.PER] the {{[GPE.GPE] American} people}}** have a right to- 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.

GPE.LOC

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

*The coast of **{{[GPE.LOC]Britain}}***

***{{[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}}***

In nested mentions of the form [child],[parent], the parent GPE always takes a LOC role; the child's role depends on context.

*My brother was born in **{{[GPE.LOC]Tehran, {{[GPE.LOC]Iran}}***

***{{[GPE.ORG]Dallas, {{[GPE.LOC]Texas}}** announced its decision to build a subway system.*

Dateline mentions of GPEs are given a location role.

***{{[GPE.LOC]Sydney,{{[GPE.LOC]Australia}}** (AP) -*

GPE.GPE

GPE.GPE is used when more than one of the other GPE roles are 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.

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

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}.

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.

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 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

Locations 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.

Positions 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.

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 of a point in a postal system is also an address.

[Capital Hill]

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 other worldly 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]...

Region_City-or-Town – Taggable location portions of GPE.City-or-Towns.

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

Region_Provincial – Taggable location portions of GPE.State-or-Provinces.

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

Region_National – Taggable location portions of GPE.Nations.

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

Region_Global – Taggable locations that cross national borders.

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

Other – All other taggable locations.

But scientists say it won't block much Christmas sunlight in [most places].

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 large, 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.

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]...

Building - One or more buildings that form or can be referred to as a unit.

...at [the national archives].

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

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

Bounded Area – A man-made/-maintained outdoor space, often including buildings.

the parades at [Disneyland]

Conduit - A facility that allows fluids or energies to pass from one location to another.

[Telephone lines] were knocked down...

Path - A facility that allows persons or vehicles to move from one location to another.

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

Barrier - A facility that blocks the movement of fluid, energy, persons, and/or vehicles.

[The Berlin Wall]

Other – All other facilities.

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.

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].

3.6.1 Subtypes for Vehicles

We will further classify Vehicle entities with the following subtypes.

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

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

Land - Vehicles designed to locomote primarily upon land.

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

Water – Vehicles designed to locomote primarily on or submerged in water.

...the Oct. 12th bombing of [the U.S.S. Cole].

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]...

Other – all other vehicles, especially those that fall into more than one of the types above. For example: amphibious vehicles, sea planes.

3.7 Weapon

A physical device that is primarily used as an instrument for physically harming or destroying animals (often humans), buildings, or other constructions.

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

3.7.1 Subtypes for Weapons

We will further classify Weapon entities with the following subtypes.

Blunt - weapons designed or used as a bludgeoning instruments

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

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

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

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

Knife

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

Sarin gas

Biological - This definition is based on the definition proposed by basicint.org: 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.

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 – weapons 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]

Other – all other weapons

Nooses, garrotes

4 Referential Class (a.k.a. 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 seemingly 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]。(CBS20001117.1000.1027)

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。
(CBS20001108.1000.0511)

The second type of attributive class for (a) involves apposition. 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。
(CBS20001129.1000.0848)

(Note a typical appositive structure invokes three mentions, as shown in the example.)

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.

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.

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”.

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).

Since attributive mentions are non-referential by definition, two attributive mentions of identical strings in the same text do not necessarily ascribe to the same entity, and a single attributive mention may ascribe to two or more entities. Therefore, every attributive mention should occupy one row in the entity table (although this has not been followed strictly in the current annotation) and linking should allow an attributive mention linked to multiple entities.

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 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:

(to be filled in)

Personal pronouns – in particular, first and second person pronouns – may have a generic reading even if 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 an individual entity of a class or a group of individual entities of a class that must be verifiable. Criteria for verifiability include (but not exclusively): can be named; or can be pointed out; ...

While the generic reading of an NP is usually tied to the predicate, that's not the case for specific referential NP's. A specifically referenced entity can be predicated by an individual-level predicate. The difference between the two is that a generic referenced entity can ONLY be predicated by an individual-level predicate.

4.4 Underspecified (USP)

We reserve the term underspecified for non-generic non-specific reference. Underspecified references means that the (quantified) NP refers to individual objects in the world that cannot be verified. Contexts that often trigger underspecified reference modal, future, conditional, hypothetical, negated, uncertain, question, etc.

While we try to define the other three (four for English) categories as precisely as possible, the annotator will tag an NP that doesn't fit into one of the other categories. The rationale is that annotators will have fewer inconsistencies for the other categories.

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。(previously generic)
[您]现在收听的是美国之音的《时事经纬》节目。(previously generic)

有人猜测当初的爆炸发生以后，船上可能有人没有死亡。(modal context)

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.

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 Mentions and 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 two mentions: the entire phrase, and the word "herself".

Mentions can be fully-fledge noun phrases (1) with or without modifiers, and or (2) with or without a syntactic head. We call the entire mention as “mention extent” and the syntactic head, if present, “mention head”. Both mention head and mention extent are recorded except for (1) exact correspondence between head and mention or (2) special cases as noted in the following sections.

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 (NAM)
		Nominal (NOM)
		Pronominal (PRO)
	Headless nominal (HLS)	
	Nominal modifier in a compound (PCN)	

	Relative clause-De	Eliminated
Complex Mentions	Parallel Multiple Mentions (PMM)	
	Extended Parallel Mentions (EPM)	
	Apposition Constructions (APP)	
	Extended Apposition Constructions (EAP)	

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 “一” 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 prenominal modifier has a possessive meaning or simply qualifies the noun head. Furthermore, “的” is also used to mark an adjectival modifier.

5.1 Simple Mentions

A simple mention is one that has a “simple” syntactic structure such that either has a single head or the head is syntactically missing (headless). A simple mention may have taggable modifiers that are either simple or complex mentions.

5.1.0 Head and Extent

For simple mentions, we record both the extent and the head. The extent of a mention is the entire phrase including all the modifiers if any. The head of a

mention is the last word of the syntactic head. This invokes notions of “wordhood” and “syntactic head”, which are unfortunately not easy to define for Chinese.

A word must be “free”. That is, it can appear anywhere in a sentence so long as its part-of-speech is well suited for the position. See Appendix for details. A syntactic head is the word that determines its syntactic category. In Chinese, noun phrases are head-final, which means that we tag the last free word as the head for simple nominal mentions. Pronominals are atomic and don’t usually have modifiers. They constitute the syntactic head on their own. In the rare case where a pronominal has modifiers, the extent must include the modifiers.

A proper name is also atomic, but it may have modifiers that are not part of the name per se. It is therefore very important for the annotator to identify the atomic part of the name. Unfortunately, organization names, especially subsidiary organizations and GPE.ORG’s, are particularly hard in this respect. Here’s the rule:

If the modifier of an ORG is its parent ORG (or GPE.ORG), treat the modifier as a separate mention. And
If the mention is a mention of a single entity in the context, treat it as NAM.
If the mention is a mention of multiple entities in the context, treat it as a NOM.

Examples:

[[中国]_{NAM} 外交部]_{NAM}
[[[[中日]_{PMM} 两国]_{NOM} 外交部]_{NOM}

5.1.1 Simple mention headed by a proper name (NAM)

Single proper names with or without modifiers.

[John]

[Defense Secretary William Cohen]

[Israeli] and [Palestinian] leaders

[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

中国国务院总理[朱容基]

[[美国]佛罗里达州]

Note that we 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.

[朱容基总理]

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]

Some vehicles, particularly aircraft carriers, submarines, famous airplanes, etc. may be named after something else. Frequently one will find the name followed by “号” 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 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]潜艇]

[[空军一号]飞机]

[[长征三号]火箭]

[[神舟四号]宇宙飞船]

[[东方红号]_{NAM} 列车]_{NOM}

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]

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.

5.1.2 Simple mention headed by a common noun (NOM)

A NP mention with one common noun head:

[中国的互联网用户]

[俄罗斯和挪威潜水员]

Determining the head of an NP is intrinsically hard and there's no agreed up 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. While the annotator is encourage to follow the CTB guidelines and their native intuitive. Here're some simple rules:

1. If the string AB doesn't mean "A 的 B", never include A in the head;
2. Never tag suffixes alone as the head;
3. 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);
4. For a tri-syllabic string "ABC", do similarly as in 3;
5. Separate the number from the head if the number is not an inseparable component of the word;
6. Ask.

5.1.3 Single pronominals (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.

5.1.3.1 Person Pronouns

The following table lists person pronouns in Chinese

	Singular	Plural
1 st Person	我/咱	我们/咱们
2 nd Person	你/您	你们
3 rd 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.

Person pronouns in Chinese do not usually refer to animals and inanimate objects, not even the third person pronouns. The compensation 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 cannot usually be modified.

Lack of possessive pronouns in Chinese is actually a plus instead of a minus for ACE annotation. Consider:

你的孩子来了，我的没来。

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]₂的]₁.

5.1.3.2 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 functions it serves in a sentence, it’s always marked as pronominal if the entity it refers to falls into one of the five 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 decide that if the sequence is in an object position (verbal or prepositional), mark it as a single pronominal mention. Otherwise, mark the pronoun and the reflexive as two pronominal mentions (of the same entity).

5.1.3.3 Demonstratives

Demonstratives (这, 那, 这些) 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.

5.1.3.4 Other pronouns and pronoun-like words/phrases:

- 其: third person pronoun used in written/formal language. Mark 其 as a pronominal mention if the entity is one of the five 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 five 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 five 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: meaning “each other”. Treat the entire unit as pronominal. However, 相互 and 互相 is not taggable because they are adverbials.
- 本: 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。
(Chtb_027)

- 自身: 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 – two mentions of the same entity; 我+ORG entity – same ??
- 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 – tag it as pronominal. It can be generic or specific depending on the context.
- 对方 – 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.

5.1.4 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 head version may not be preferred for the context by the native speaker. Although these mentions are technically headless, we will assign as head the right most premodifier that falls directly before the spot where the head would be.

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 EDT.

Caution: we have abandoned annotating 的 as a relative clause marker. Therefore 的 should not be tagged twice.

5.1.5 Nominal modifier in compound nouns (PCN)

Noun + noun compounding is the most frequent morph-syntactic process in Chinese. We annotate the nominal modifier iff both the modifier and the head noun are free words and tag the modifier as PCN. We exclude proper names from this category. In other words, a PCN is always a common noun. In fact a PCN can only be attributive. For example “火车” as in [[火车]司机].

One has to be careful if the noun directly modifies the head noun (hence part of compound noun) **Or** if it's simply an NP modifier.

国务院减轻农民负担联席会议决定

In this example, 农民负担 is not a type of 负担, but rather it's 负担 that 农民 (in general, thus generic) has to endure.

One simple test is to insert the 的 marker. If it doesn't change the meaning, it cannot be a PCN, otherwise it is PCN. Another test is to insert another modifier between the nominal and the head. If insertion is allowed, the nominal is very

likely referential. Otherwise, the nominal should be attributive. For example, 好吃懒做的火车司机 vs. *火车好吃懒做的司机. The context also often helps. If a co-reference can be established between the modifier and some other non-attributive mention in the context, it's a good indication that the modifier is not a PCN.

The requirement that both the modifier and the head noun in a compound must be free is often related to the syllable length. In general, if the compound is more than 4 syllables long where each of the component nouns has more than two syllables, both the modifier and the head are likely free. On the other hand, if it has 3 syllables only, one of them is likely bound and therefore the compound is likely to be lexicalized and we do not look inside a lexicalized compound noun.

The following section has been suspended due to too many misses in actual annotation.

5.1.6 DE in relative clauses (DE)

~~A relative clause in Chinese always misses the element that it modifies and the element will always fill in the empty slot if the clause is a standalone sentence. With a few exceptions, Chinese relative clauses are usually marked by “DE”. We treat “DE” as if it was the missing element and tag it if the modified mention refers to a taggable entity. Since the modified mention can be omitted in Chinese — which means it's a headless mention, we do not tag the last element of a headless mention as the head to avoid tagging “DE” twice. (need modification)~~

~~悼念[几天前在也门亚丁港应军舰发生爆炸而死亡[的]17名美国水军]~~

~~If a relative clause is used directly as argument, 的 is tagged twice. Here's an example:~~

~~[肇事[的]]是小客车驾驶~~

~~的 is tagged by itself in the clause 肇事[的] where it acts as the subject of 肇事. Then the entire clause 肇事的 is tagged as the subject of the main clause with 的 as the head. We have yet to check if the tool allows this.~~

5.2 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.

5.2.1 Parallel Multiple Mentions (PMM) and Extended Parallel Multiple Mentions (EPM)

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) NP's can be in parallel structure without any common component;
- 2) NP's can be in parallel structure with common components that are either (pre-) modifiers or integral part of any of the NP's or N's if the NP or N appeared alone;

We use the notion of parallel multiple mention (PMM) to capture any portion of parallelism regardless where it happens in a phrase AND use that notion recursively to include any integral part of the NP's or N's. This recursive property can allow us to extract the entirety of a proper name.

A simple expression like 张三和李四 is annotated as

[[张三]_{NAM}和[李四]_{NAM}]_{PMM}

whether they appear together as argument in a clause or in a modifier position (as in 张三和李四的书 or 张三和李四的妻子).

On the other hand, an example like 黄埔和南浦大桥 is annotated as

[[[黄埔]_{NAM}和[南浦]_{NAM}]_{PMM}大桥]_{PMM}

Thus 黄埔大桥 is a proper name, so is 南浦大桥.

Note that there cannot be any intervening mention between two embedded PMM's.

If there's any modifier (regardless of whether or not it has taggable mentions) that applies to the entirety of PMM (the outmost PMM if there're embedded PMM's, we tag the entire mention as EPM as if the PMM is the head of the EPM (but we don't highlight a head because complex mentions are considered inherently headless), like so:

[[该公司]_{NOM}的[[张三]_{NAM}和[李四]_{NAM}]_{PMM}]_{EPM}

In general, only entities of the same type can be referred to in this way. Thus the strategy for deciding the entity type of an EPM or PPM mention is to keep just one mention by removing all the other mentions from the PPM and see what the entity should be. Recursively embedded PMM's and the embedding EPM (if any) should all be in the same row of the EDT table.

Here are some additional examples that we've discussed extensively

[国[[内]_{NOM}[外]_{NOM}]_{PMM}]_{PMM}

[[[江西]_{NAM}和[湖南]_{NAM}]_{PMM}省]_{PMM}

[[[江西]_{NAM}和[湖南]_{NAM}]_{PMM}两省]_{NOM}

[[中国]_{NAM}的[神舟[[三]_{NAM}、[四]_{NAM}]_{PPM}号]_{PMM}飞船]_{NOM}

[[俄罗斯]_{NAM}的[[[明斯克]_{NAM}和[库尔斯克]_{NAM}]_{PMM}号]_{PMM}潜艇]_{NOM}

[报名参展的 [[国家]_{NOM}和[地区]_{NOM}]_{PMM}]_{EPM}

[[[上海]_{NAM}、[江苏]_{NAM}、[广西]_{NAM}]_{PMM}等[[省]_{NOM}[市]_{NOM}[自治区]_{NOM}]_{PMM}]_{EPM}

[[[中]_{NAM}[日]_{NAM}[韩]_{NAM}]_{PMM}等三国]_{NOM}

[[[国务院]_{NAM}副总理]_{NOM}[[钱其琛]_{NAM}和[温家宝]_{NAM}]_{PMM}]_{APP}

[[[钱其琛]_{NAM}和[温家宝]_{NAM}]_{PMM}副总理]_{NOM}

Note how the last example is annotated. “副总理” is treated as a common noun, not a title. If it were treated as a “title”, the entire mention would be tagged as PMM. The reason for tagging it as a common noun is that titles in general cannot be used with multiple individuals. This is shown by the anomaly of *钱和温副总理.

The introduction of PMM and EPM puts a lot of workload on annotation and annotators may take some time to get used to them and do annotation without misses.

Note we are getting rid of “member set constructions” as a distinct complex mention type. A few examples have already shown how they should be annotated. In general, if it’s a many-to-one relationship, we simply treat it as modification and tag the first part as PPM and the second part as NOM. If the first has just one member with 等, the first part is simply a NAM mention.

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.

5.3.2 Appositive Constructions (APP) and Extended Appositive Construction (EAP)

An appositive construction refers to one where one mention is attributive 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][安格拉.布里提鄂姆]]说

(Note [人口调查局发言人] is attributive)

[[[波斯湾]_{NAM}国家]_{NOM}[[巴林]_{PPM}和[卡塔尔]_{NAM}]_{PMM}]_{APP} (6 mentions!)

The entity type and reference class of an appositive construction are determined by the non-attributive 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

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][江泽民]]_{APP}]_{EAP}

Sometimes, it can be ambiguous:

正在崛起的新兴城市宁波.

The ambiguity is primarily due to how “unique” the first element is if it’s used referentially. The following test may work in most cases:

江泽民 是 正在柬埔寨访问的国家主席 - bad

上海是中国最大的城市 - good

宁波是正在崛起的新兴城市 - good

We use this simple “是” test for a quick decision:

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

莫斯科的一个法庭否决了[[涉嫌从事间谍活动[的]美国人][波普]]_{APP} 要求医疗检查的请求

[[获得诺贝尔奖的中国作家]_{NOM}[高行健]_{NAM}]_{APP}

The notion of EAP vs. APP is in a sense parallel to the notion of EPM.

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. Below are 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

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.*