Annotation Guidelines for

Entity Detection and Tracking (EDT)

Version 4.2.6 200400401

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 textual 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.

- **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 included within the head of the name of a Person. The parts of titles are taggable as premodifiers (PRE).

Mr. Smith	_[[Mr] Smith]_
former-President Bill Clinton	_[[former-President] Bill Clinton]_
Justice Sandra Day O'Connor	_[[Justice] Sandra Day O'Connor]_
US Secretary of State Colin Powel	_[[[US] Secretary of State] Colin Powel]_

3.1.2.1 Saints and other religious figures

Religious titles such as saint, prophet, imam, or archangel are to be treated as premodifiers (PRE).

> *[[[St]. [Christopher], [the patron of transportation]]*

Please note that this phrase contains four mentions:

two specifically reference the saint

> *St. Christopher, the patron of transportation – APP*
>
> *St. Christopher - NAM*

and two ascribe properties to him.

> *St. - PRE*
>
> *the patron of transportation - NOM*

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 or reliable in the texts.

> *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*
>
> *Tony the tiger*

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 type Person.

> [The <u>Kennedy</u>]'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 <u>Arabs</u>]
> [The <u>Christians</u>]

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, and sports teams. 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. Political parties do not fall into the subtype of Government and should be tagged as ORG.Other.

> [The Salzburg prosecutor's <u>office</u>] is investigating the disaster to determine if criminal charges could be filed.
>
> Putin, a former [<u>KGB</u>] agent, defended [the <u>court</u>] [<u>that</u>] convicted Pope and [the security <u>services</u>],…
>
> The Financial Accounting Standards Board will take no conclusive action on its current project on business combinations until [<u>Congress</u>] 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. Industries and industrial sectors are of the subtype commercial.

> Last February, he failed a drug test required by [<u>Major League Baseball</u>] and Commissioner Bud Selig suspended him for the 2000 season.
>
> Pope, who owns [<u>TechSource Marine Industries</u> in State College, PA.], and an associate were…
>
> Fifteen people, all with minor injuries, were rescued, according to [the <u>Austria Press Agency</u>].

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. Common examples include political parties and the UN.

> *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 though 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.
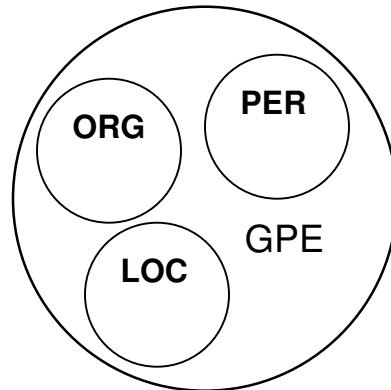
3.2.3 First Person Pronouns Referring to Organizations

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.

> *[Japan]'s intellectual elite*
> *In [Spain], the images were shown on most TV channels*

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

> *The governor of [Texas]*

County-or-District – Taggable mentions of the entireties of any county, district, prefecture, or analogous body of any state/province/canton.

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

City-or-Town – Taggable mentions of the entireties of any GPE below the level of County-or-Prefecture.

> *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.]*

Other – Named geographical entities that can also function as political entities that are commonly referred to by those names will be considered GPE.Other rather than Locations. Following is a non-exhaustive list of entities that should be GPE.Other 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.

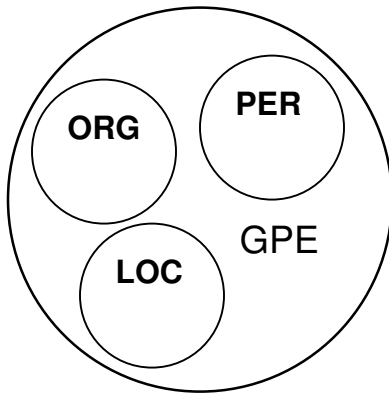> *[NATO] peacekeepers arrived in the valley before nightfall.*
> *[NATO] – NAM, ORG.Other*

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 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** - <u>France</u> signed a treaty with Germany last week.
- **GPE.PER** - <u>France</u> likes to eat cheese.
- **GPE.LOC** - The world leaders met in <u>France</u> yesterday.
- **GPE.GPE** - <u>France</u> 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 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).

> ***France*** *produces better wine than New Jersey.*   GPE Role (whole nation)
>
> ***France****'s greatest national treasure*   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.

> Ritte*r's return is seen as something of a test of that agreement, under which {[GPE.ORG]* <u>Iraq</u>} *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 <u>government</u>} and many politicians will be stridently critical of the United States if they believe they are being ignored.*

GPEs and Government Organizations

GPEs mentioned in close context with government organizations often 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] <u>Switzerland</u>}'s supreme court.*

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 <u>Japanese</u>} 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] <u>Cubans</u>} have been waiting for this day for a long time.*

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

> *You and th- {[GPE.PER] the {[GPE.GPE] <u>American</u>} <u>people</u>} have a right to- to get answers.*

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

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

> *{[PER] The <u>rest</u> of {[GPE.PER] <u>America</u>} }*

> *{[PER] idealistic <u>Europeans</u>}*

> *{[PER] <u>Americans</u> 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 <u>Chinese</u>} 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] <u>Britain</u>}
>
> {[GPE.LOC] <u>France</u>} has an area of 547,090 square kilometers
>
> U.S. warplanes flew over {[GPE.LOC] <u>Afghanistan</u>}
>
> Northern portions of {[GPE.LOC] <u>Montana</u>}

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] <u>Tehran</u>, {[GPE.LOC] <u>Iran</u>}}
>
> {[GPE.ORG] <u>Dallas</u>, {[GPE.LOC] <u>Texas</u>}} announced its decision to build a subway system.

Dateline mentions of GPEs are given a location role.

> {[GPE.LOC] <u>Sydney</u>,{[GPE.LOC] <u>Australia</u>}} (AP) -

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.

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] <u>Israeli</u>} troops
>
> {[GPE.GPE] <u>New York</u>} policemen
>
> Prime Minister of {[GPE.GPE] <u>Britain</u>}
>
> {[GPE.GPE] <u>New York</u>} attorney
>
> {[GPE.GPE] <u>U.S.</u>} Commander-in-Chief
>
> {[GPE.GPE] <u>U.S.</u>} surveillance aircraft
>
> {[GPE.GPE] <u>Iraqi</u>} flag
>
> The {[GPE.GPE] <u>California</u>} 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] <u>Soviet Union</u>} invaded {[GPE.GPE] <u>Afghanistan</u>}.

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 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]…*

Region_Local – Taggable location portions of GPE.City-or-Towns.

> *In [another part of the city],…*

Region_Subnational – 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_International – 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]"*

In this example,

> *[They] live west of [the city].*

we would not tag "*west of the city*" as a location mention. Constructions like this establish a physical relationship between two entities. They do not reference entities in their own right.

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.

> [The *Erie Canal*] had an enormous impact on New York

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 <u>compartment</u> of the train]…

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

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 <u>bomb</u>] 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 <u>bat</u>]…*

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

> *…a small boat packed with [<u>explosives</u>]…*

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

> *<u>Knife</u>*

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.

> *<u>Sarin gas</u>*

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.

> *<u>Anthrax</u>*

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 <u>pistol</u>] through customs…*

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

> *<u>bullets</u>*

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

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

Other – all other weapons

> *<u>Nooses</u>, <u>garrotes</u>*

4 Entity Class

Each taggable entity must be assigned a class that describes the kind of reference the entity makes to something in the world. The distinction between referential and attributive uses of an NP is given by the following definitions for ACE:

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

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 – or (b) 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 on 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.

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 pointing (specific) or cannot be precisely accounted for (underspecified).

Please see Appendix A for the Decision Tree for Entity Class. This tree steps through the process of assigning a class to an entity.

4.1 Negatively Quantified (NEG)

An entity is NEG when it has been quantified such that it refers to the empty set of the type of object mentioned.

> *[No sensible <u>lawyer</u>] would take that case.*
> *[No <u>one</u>] has claimed responsibility.*

There are [no confirmed suspects] yet, but officials say several Middle East groups are expected to be investigated.

Please note that we do not assign NEG for entities introduced by negated predicates.

They are not [lawyers].

The class of 'lawyers' in this example would be ATR.

4.2 Non-referential/Attributive/Ascriptive (ATR)

An entity is ATR when it is not being used to refer, but rather to attribute some property or attribute to some entity.

John is [a lawyer].

Richard Zolinger is [a weed specialist in North Dakota State University] and [part of an International Monitoring Network].

David Lule, [the head of light entertainment for Pearson Television, [one of the largest international distributors of programming]], said…

Dr. Alan Leshner is [the director of the National Institute on Drug Abuse].

No one should doubt our resolve to remain [a force for peace and for stability]…

4.3 Specific Referential (SPC)

An entity is SPC when the entity being referred to is 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.

[John's lawyer] won the case.

This afternoon, [a crowd of angry muslims] set fire to [a hotel].

[Lee Hawk Seder] is Jerusalem Bureau Chief for [the Washington Post]

[[Columbia University]'s Institute of War and Peace Studies]

4.4 Generic Referential (GEN)

An entity is GEN when the entity being referred to is not a particular, unique object (or set of objects). Instead GEN entities refer to a kind or type of entity. Notice that the mentions in question are still understood to be referential in that they point to actual things in the world rather than saying that an object 'has that property' or some similar notion.

[Lawyers] don't work for free.

About 231 feet to 264 feet of water is considered shallow for [submarines].

But the sense of urgency for this meeting matches the rage felt by both [Israelis] and [Palestinians] after yesterday's violence.

…[extremist groups] have a lot of support these days and a lot of power.

Japan's equivalent of [a naval force] is officially referred to as the Japan Maritime Self-Defense Force.

4.5 Under-specified Referential (USP)

We reserve the term underspecified for non-generic non-specific reference. Underspecified references 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 <u>people</u>] will participate in the parade.*
>
> *I don't know [how many <u>people</u>] came.*
>
> *Do you know [how many <u>people</u>] came?*
>
> *We will elect [five new <u>officials</u>].*

Imprecise quantifications often yield USP-tags; a few examples are as follows:

> *A few,*
>
> *some (plural),*
>
> *most,*
>
> *more than half*

Another example of an underspecified entity is a mention of a large number of entities where the actual members of the set are not necessarily identifiable and the number used is an estimate.

> *[15 thousand <u>people</u>] attended the rally.*

From the context in which they appear, the under-specificity of these quantifications may be ascertained; however, the appearance of these strings often indicate under-specificity.

While we try to define the other four categories as precisely as possible, annotators may still encounter NPs that cannot be classified. In these cases, annotators should make these NPs Underspecifed. 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.

5 Mention Types/Mention Levels

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*".

5.1 Simple Mentions

Simple mentions are full noun phrases. For each simple mention, we record its full extent and its head.

5.1.1 Mention Extent

The extent of a mention consists of the entire nominal phrase. In case of structures where there is some irresolvable ambiguity as to the attachment of modifiers, the extent annotated should be the maximal extent. 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 <u>people</u> yesterday who love chess]

The extent includes all the modifiers of a nominal 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.

5.1.2 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 mention is

The new glass-clad skyscraper

and the head is *skyscraper*. If the syntactic head of the phrase is a multi-token item, the last token is marked. If the head is a proper name, however, then 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:

[<u>Fred Smith</u>] became [the new prime <u>minister</u>].
The job fell to [<u>Abraham Abercrombie III</u>].

5.1.3 Names (NAM)

Proper nouns and nicknames.

[<u>John</u>]
[Defense Secretary <u>William Cohen</u>]
[The <u>Jeluzoon Refugee Camp</u> near [<u>Ramallah</u>]]
[<u>North Dakota State University</u> in [<u>Fargo</u>]]

[The <u>house of representatives</u>]

5.1.4 Quantified Nominal Constructions (NOM)

A noun quantified with a determiner, a quantifier, or a possessive.

[The <u>lawyer</u>]
[the Yemeni <u>port</u> of Aden]
[some American <u>executives</u>]
[the southern <u>tip</u> of Africa]
[a <u>crowd</u> of angry muslims]

5.1.5 Nominal Mentions with Multiple-word Heads (MWH)

A construction in which two or more heads that share the same premodifiers or postmodifiers. The conjoined heads form the head of a single mention.

[20 angry <u>men and women</u>]
[Palestinian <u>towns and villages</u>]
[Hollywood <u>actors and writers</u>]
[The <u>movers and shakers</u> in Washington]

5.1.6 Unquantified (Bare) Nominal Mentions (BAR)

An unquantified nominal construction. Both singular and plural constructions may be BAR.

[<u>lawyers</u>]
[US <u>officials</u>]
offices in [foreign <u>countries</u>]
[<u>representatives</u> of the company's Tananger branch]
[<u>Americans</u>] eagerly await the results of the election.

5.1.7 Pronouns (PRO)

Pronouns with the exception of wh-question words and the specifier 'that'.

[<u>he</u>]
[<u>they</u>]
[<u>her</u>]
[<u>I</u>]

5.1.8 WH-Question Words and Specifiers (WHQ)

WH-question words and the specifier 'that'.

[who]
the executive, [<u>who</u>] spoke on the condition of anonymity
the Russian navy, [<u>which</u>] waited several days before attempting to launch any rescue mission.
Germany, [<u>where</u>] they were taken yesterday

5.1.9 Pre-modifier Mentions

Premodifier mentions are those mentions, name or nominal, which occur in a modifying position before another word(s). It is immaterial whether or not the word being modified is a taggable entity.

In almost all cases, the construction of the premodifier mention must be identical to the construction of the mention as it would occur in a BAR, NOM, or NAM construction. The only exception to this rule is the transformations that occur on name-mentions of LOC's and GPE's in their premodifier positions. Transformations of names or nominal mentions for any other types of entities are not taggable.

Taggable	Not Taggable
the *[Russian] foreign minister*	*Stalinist nations*
[mountain] regions	*mountainous regions*
[government] grants	*federal grants*
[Tuscan] food	*Machiavellian strategy*
[Alpine] villages	
the *[Microsoft] [court] case*	
[bomb]-making materials	
[school] vouchers	
the *[state] budget this year*	
[[Democratic] [presidential] nominee] Al Gore	

5.1.9.1 Nested Premodifiers

Cases where premodifiers are nested occur frequently in the texts especially when people are modified with titles and their professions. When a premodifier is itself modified, its extent should reflect the nested constructions.

> *[[[White House] [press] secretary] Scott McClellan]*

This example illustrates both nested and non-nested premodifiers. 'White House' and 'press' modify 'secretary', but do not nest with each other. This phrase gives us four mentions.

> *White House* – ORG.Government, PRE, ATR
>
> *Press* – ORG.Commercial, PRE, ATR
>
> *White House press secretary* – PER, PRE, ATR
>
> *White House press secretary Scott McClellan* – PER, NAM, SPC

5.1.10 Headless Mentions (HLS)

Headless mentions are constructions in which the nominal head is not overtly expressed. 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.

>*[the <u>toughest</u>]*
>*[more than <u>30</u>]*
>*[<u>many</u>] on both sides*
>*[60<u>%</u>] said*
>*[sixty <u>percent</u>] said*
>*[<u>35</u>] were injured*

5.1.11 Partitive Constructions (PTV)

Partitive constructions have two elements: the part and the whole. The first element of a partitive construction lacks a head and quantifies over the second element. Just as in Headless mentions, we will tag the right most premodifier of the first element as the head of the partitive construction.

>*[<u>some</u> of the lawyers]*
>*[<u>one</u> of the houses]*
>*[<u>half</u> of the team]*
>*[<u>all</u> of them]*

There are some constructions with prepositional phrase that greatly resemble partitives, but should not be tagged as partitives. The first element of these constructions is a nominal that can function as a head.

Examples of non-partitives (two entity mentions):

>*two members of the team*
>*the bottom of the ocean*
>*the heart of the Alps*

5.3 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.

5.3.1 Conjoined Mention Constructions (CMC)

A construction which consists of two (or more) full entity mentions conjoined. The fashion in which an CMC-construction appears is: quantification, relative clauses, prepositional phrases, and/or modifiers distributing over two or more entity mentions. This distribution conjoins the entity mentions and affects the CMC-construction. When tagging CMC-mentions, the component entity mentions as well as the full CMC-mention are tagged. There is no head-assignment in CMC-mentions.

>*[[15 <u>men</u>] & [20 <u>women</u>] [<u>who</u>] stormed the barracks] were captured…*
>*[[<u>Kennedy</u>], [<u>Lieberman</u>], [<u>Gephardt</u>], & [<u>Kerry</u>] all] attended the fundraiser.*

[[Elderly women] and [young children] in the refugee camp]…

[Approximately 80 [immigrant teenagers] and [Swedish natives]] died in the blaze.

5.3.2 Appositive Constructions (APP)

A construction which consists of two or more full entity mentions which refer to (or predicate on) the same entity. As with MNH-mentions, the component mentions within the APP-constructions will be tagged with their heads as appropriate. However, the APP-mention itself has no head-assignment.

[[Bill], [John's lawyer]]

[[Mr. Black, 58], [a victim of the terrorist assault]], told the Associated Press…

[[the show's production company], [Celador]]

[[the heavy lift ship], [Blue Marlin]]

[[Sauache County], [home of the Watchtower]]

5.3.3 Complex Constructions taking a Relative Clause (ARC)

An ARC-construction is an appositional construction with an adjacent relative clause that refers to the initial, referent (SPC) mention of the entity, rather than the latter, attributive (ATR) mention(s) of the entity. In ARC-constructions, the component entity mentions and the WHQ mention all are tagged and assigned heads, after which the headless ARC-tag is applied.

[[The former company CEO], [John Fisher], [who] fired the board of directors]…

[[Dennis R. Beresford], [an accounting professor at the University of Georgia] [who] was then chairman of the accounting board]

[John Richards], [the party leader] [who] attended the DNC convention] said…

In the final example, it is unclear whether the relative clause refers to "John RIchards" or "the party leader". In cases of ambiguity like this, the ACE policy is to tag the string as an ARC-construction. Also, please note that the embedded APP-constructions within ARC's are not tagged. Because an embedded apposition is requisite to ARC's, it is redundant to tag the APP-mention within them.

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.*

Appendix A: Decision Tree for Entity Class