ACE (Automatic Content Extraction) Spanish Annotation Guidelines for Entities

Version 1.6 2006.11.01

Linguistic Data Consortium

http://www.ldc.upenn.edu/Projects/ACE/

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*
> ***Name Mention:*** *Gabriel Garcia Márquez*
> ***Nominal Mention:*** *el escritor colombiano*
> ***Pronoun Mention:*** *él*

Entities are limited to the following seven types:

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

For each entity, the annotation records the type of the entity (PER, ORG, GPE, LOC, FAC, VEH, 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 (PER)

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. Body parts are not to be annotated. 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 <u>turkey</u>] 1*
> *[The political <u>cat</u> of the year]*
> *Él fue [la <u>estrella</u>] de la noche.*
> *Ella es [la <u>columna</u> vertebral] de su hogar*
> *Eres [una <u>gallina</u>]*

3.1.1. Subtypes for Person
We will further classify Person entities with the following subtypes.

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

> *[<u>Vicente Fox</u>]*

---

[1] For the remainder of the document, square brackets [ ] will indicate the extent of an Entity Mention and <u>underlining</u> will be used to indicate its head. When the head is the same as the extent, the underlining is often omitted.

[la <u>madre</u>]

[el <u>Presidente</u> de los Estados Unidos.]

PER.Group

If the Person entity refers to more than one person, tag it as PER.Group unless the group meets the requirements of an Organization or a GPE described below. This will include family names and ethnic and religious groups that do not have a formal organization unifying them.

[estudiantes]

[los <u>profesores</u> de la universidad]

[la <u>familia</u>]

[Los <u>cristianos</u>]

PER.Indefinite

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

3.1.2 Titles, Honorifics, and Positions

In English, titles and most honorifics precede the name. In Spanish, this is usually also the case. Because of the nature of appositive constructions (APP) in Spanish, titles and honorifics that in English would be tagged as premodifiers would not be in Spanish, such that whereas in English, the "President" in [[President] Bush] is a premodifier, "presidente" in [[El <u>presidente</u>] [Bush]] is the head of a NOM mention. In Spanish, we tag all honorifics and titles that are preceded by a determiner (el, la) as appositions.   See section 5.3.1 for more about APPs.

[*APP [**NOM.REF** La <u>Secretaria</u> de Estado de EEUU] [**NAM.ATR** <u>Madeline Albright</u>]]*

[*APP [**NOM.REF** El <u>Presidente</u>] [**NAM.ATR** <u>George Bush</u>]]*

Saints and other religious figures

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

St. Christopher

[St. <u>Christopher</u>]	PER
[St].	

San Bartolomé

[San <u>Bartolomé</u>]	PER
[San]	PER

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.**

Si crees en [diós], debes…   name mention

Aunque siente que es [un <u>diós</u>], él…   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] ha sido un héroe popular.

El disfraz de [<u>Adam West</u>] del programa Batman.

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.

Snuggle, the fabric softener bear

El ratón Topo Gigio

3.2 Organizations

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

3.2.1 Subtypes for Organizations

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

Government (GOV)

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

[The Salzburg prosecutor's <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>] [that] 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 [Congress] has reconvened in 2001…

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

[El <u>Banco Popular de China</u>] estableció un nuevo embargo entre China y las naciones africanas.

[<u>La Dirección General de Asuntos Consulares</u> de la cancillería] puntualizó que [el nuevo <u>consulado</u>] fue establecido en el edificio Espigón Gorlero.

[El <u>Ministerio del Exterior</u> de Rusia] reiteró hoy que se opone a la "independencia de Taiwán"

Commercial (COM)

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

Pope, who owns [<u>TechSource Marine Industries</u>] in State College, PA.], and an associate were…

[<u>Yukos</u>] fue [la <u>empresa</u> petrolera más grande de Rusia].

Educational (EDU)

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

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

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

Fundada en el año 1701, [<u>Yale</u>] es [la segunda <u>escuela</u> más antigua del país].

Como todas las universidades estadounidenses, [<u>UCLA</u>] tiene una mascota y sus colores preferidos para distinguirla de las demás.

Entertainment (ENT)

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

[the <u>Roundabout Theater Company</u>] is calling its new facility in Times Square the American Airlines Theater

Like [the famous Irish <u>group</u>] [the <u>Chieftains</u>] and [Alton], [Solace] frequently headlines in Celtic festivals.

Popularizada por [los <u>Gipsy Kings</u>], la rumba flamenco es una mezcla de flamenco con ritmos afro-cubanos.

Non-Governmental Organizations (NonGov)

Non-governmental organizations are those organizations that are not a part of a government or commercial organization and whose main role is advocacy, charity or politics (in a broad sense). This subtype will include such diverse organizations as:

1. (Para-)Military Organizations:

> *[<u>KKK</u>]*
> *[<u>Fuerzas Armadas Revolucionarias de Colombia</u>] [(<u>FARC</u>)]*
> *[<u>Ejército Zapatista de Liberación Nacional</u>] [(<u>EZLN</u>)]*
> *[<u>ETA</u>]*

2. Political Parties:

> *[<u>Republican Party</u>]*
> *[<u>Labour Party</u>]*
> *[<u>Movimiento al Socialismo</u>] [(<u>MAS</u>, Bolivia)]*
> *[<u>APRA</u>] [(<u>Alianza Popular Revolucionario Americana</u>, Perú)]*

3. Political Advocacy Groups and Think Tanks:

> *[<u>ACLU</u>]*
> *[<u>Human Rights Watch</u>]*
> *[<u>Confederación Nacional del Trabajo</u>] [(<u>CNT</u>)]*

4. Professional Regulatory and Advocacy Groups:

> *[The <u>American Bar Association</u>]*
> *[The <u>American Medical Association</u>]*
> *[<u>Asociación de Médicos Municipales</u>]*

5. Charitable Organizations:

> *[The <u>Red Cross</u>]*
> *[The <u>United Way</u>]*
> *[<u>Doctors Without Borders</u>]*
> *[<u>Cruz Roja</u>}*

[*Grupo Ecológico de Puerto Vallarta*]
[*Congreso de Latinos Unidos*]

6. International Regulatory and Political Bodies:

[*NATO*]
[*The World Bank*]
[*El Banco Mundial*]
[*la ONU*]

Under the above definition, all of the following would be annotated as ORG.NonGov:

[*Hezbollah*]

[*Islamic Resistance*]

[*the "Caravan of Death*]," [*a military party* [*that*] killed 73 political prisoners]

the leading deputy of [*the Rally for Congolese Democracy*], [*one of* [*the biggest rebel movements supported by Uganda*]]

[*The GOP*]

[*the Democratic Party*]

[*Republican National Committee*]

[*Rock the Vote*]

[*Rally for Congolese Democracy*]

[*College Board*]

[*National Rifle Association*]

[*NAACP*]

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

[*American Bar Association*]

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

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

Tres empleados de [*la ONU*] *en Bagdad.*

El plan de [*NATO*]

Media (MED)

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

[*Agence France Presse*]

[*Reuters* international news agency]

[*ABC News*]

[*Associated Press*]

[*Chicago Sun-Times*]

[*Road & Track* magazine]

[*National Geographic*]

[*Univisión*]

[El *Mundo Deportivo*]

[*Telemundo*]

Religious (REL)

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

We will treat "Islamic" and "Muslim" equally. Both are considered to be ORG-REL.

> *…said Archbishop Khajag Barasamian, head of [the Diocese of [the Armenian Church in America, [whose] headquarters are in Manhattan]].*
>
> [*German Bishops Conference*]
>
> [*Islamic* militants]
>
> [El *Vaticano*]

Medical-Science (SCI)

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

> [*Overseas Chinese Physics Institute*]
>
> [*Massachusetts General Hospital*]
>
> [medical and health *groups*]
>
> [*Centro Nacional de Investigaciones Científicas*]
>
> [*Hospital Infantil de México Federico Gómez*]

Sports (SPO)

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

> *A group of survivors belonging to [a German ski club in Vilseck, Germany] …*
>
> [*Philippines Olympic Committee*]
>
> [*National Hockey League*]
>
> [El club *Celaya*], de [la segunda *división* de México]
>
> [una *liga* para mujeres en Puerto Rico]

[Club Atlético Boca Juniors]

3.2.2 Organization Subtype Trumping Rules

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

Less Specific   More Specific

⟵——————————————————————⟶

REL

GOV

NonGov   EDU SCI

COM   ENT SPO

                                     MED

Exceptions to Trumping Rules

GPE military

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

Media Conglomerates

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

Medical Schools and Research Labs

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

Soft Science Research Institutions

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

Boy Scouts

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

3.2.3 Organization Entities used in Person Contexts

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

However, there seems to be little to be gained in the usual case by thus "reaching inside the organization" to posit a mention of a Person entity. It seems better to adopt the view that organizations can be agentive, and take action on their own. Only when something in the context draws particular attention to the people within the organization should a separate mention of a Person entity be marked.
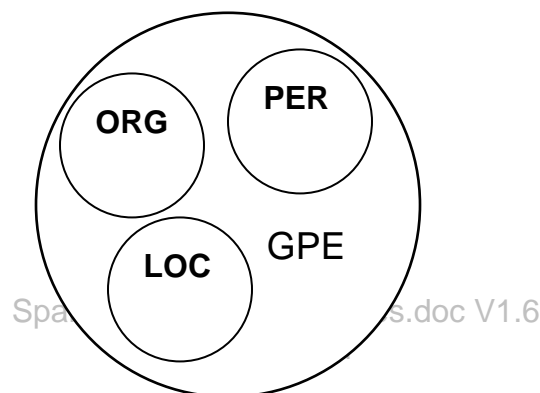
Sets of people who are not formally organized into a unit are to be treated as a Person entity rather than an Organization entity. It is often difficult to tell the difference between Organization entities and collections of individuals tagged as PER.Group entities. Examples of organization-like nouns which are *not* organizations are "employees," and "crew." Although the members of a company or crew may work together in an organized and even hierarchical fashion, the groups are not organizations by themselves.

3.2.4 First Person Pronouns Referring to Organizations

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

3.3 Geographical/Social/Political Entities (GPE)

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

In this sentence,

> *The people of France welcomed the agreement.*

there are two mentions

> *[The <u>people</u> of France]*   *GPE*
> *[France]*   *GPE*

Likewise,

> *[El <u>pueblo</u>] [francés]*

is two mentions of the same entity.

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:

> *[<u>Italia</u>] derrotó a [<u>Francia</u>] 5-3 en tiros penales tras la prórroga*

These are to be recorded as distinct entities, not as mentions of the GPE entity. Thus, in this example, both "*Francia*" and "*Italia*" would evoke Organization/Sport entities.

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

Continent
Taggable mentions of the entireties of any of the seven continents: North America, South America, Antarctica, Europe, Asia, Africa, and Australia (África, América del Norte, América del Sur, Antártica, Asia, Europa y Oceanía).

> *Many people in [<u>North America</u>] 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].*
> *…the conversion to Christianity of the [Roman] emperor Constantine…*
> *La escritora [panameña] va a hablar de su nuevo libro.*
> *[GPE.ORG El gobierno [NAMPOST.GPE chino]] anunció hoy que…*

State-or-Province

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

> *[Salzburg] governor Schausberger said…*
> *[el Yucatán]*
> *[La Rioja] produce mucho vino.*

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.*
> *[l'Alt Camp] es uno de [los comarques] de Cataluña.*

Population-Center

Taggable mentions of the entireties of any GPE below the level of County-or-District. This includes names of cities and towns.

> *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 Alps.]*
> *Mi hermana está estudiando en [Madrid] este semestre.*

Sometimes a phrase will contain two mentions of the same entity,
> *[La capital uruguaya de [Montevideo]].*

In this example both mentions are coreferenced, and refferential.

GPE-Cluster

Named groupings of GPEs that can function as political entities.

> *[Eastern Europe]*
> *…and the government tried to redefine [the West].*
> *[the European Union]*
> *[Southeast Asia]*

[el Medio Oriente]]

[Latinoamérica]

Special

A closed set of GPEs for which the conventional labels do not straightforwardly apply. This subtype is currently restricted to Palestine, the Palestinian Authority, "Mainland China", Taiwan, and Native American reservations.

3.3.2 GPE-like Locations and Organizations

Incidental, non-political clusters of GPEs should be marked Location.

[el sur de los Estados Unidos]

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 "*Dallas, Texas*" evokes one entity for each region. Thus "*Dallas, Texas*" evokes one entity for the population center
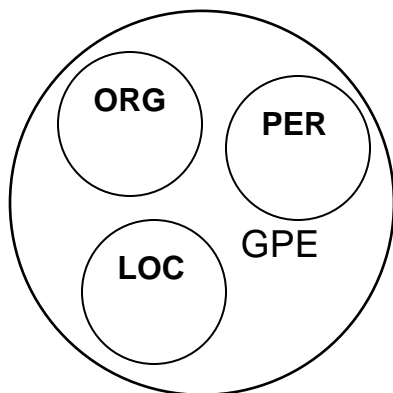
[Dallas, Texas]

and a second one for the state

[Texas]

both mentions are referential.

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. 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 vacations in August.
- **GPE.LOC** - The world leaders met in France yesterday.
- **GPE.GPE** - France produces better wine than New Jersey.

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

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

> *[France] produces better wine than New Jersey.*   GPE Role (whole nation)
>
> *[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 *Mexico* agrees to something, *Mexico* 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.*
>
> *En una histórica decisión, [**GPE.ORG** Cuba] aceptó ayuda de EEUU para paliar daños del huracán "Wilma".*

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.*
>
> *[**GPE.ORG** El gobierno colombiano] y el Ejército de Liberación Nacional (ELN) anunciaron el pasado lunes un acuerdo para reunirse y avanzar en el camino de la paz.*

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. All GPE premodifiers will be assigned the role GPE as explained below. (**nb- while premodifiers do occur in Spanish, we will be dealing almost exclusively with postmodifiers, which will be treated equivalently**).

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** Cubans] 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.*

*[**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.PER** Los <u>españoles</u>] son una gente muy hospitalaria.*

Muchos [PER <u>españoles</u>] murieron en los ataques en Madrid en 2004.

*Para la mayoría de [**GPE.PER** los <u>chinos</u>], el monje Da Mo fue también el creador del Kung-fu…*

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]*
>
> *las afueras de [**GPE.LOC** Culiacán], la capital de Sinaloa*
>
> *Uno de los mayores grupos de transporte urbano en [**GPE.LOC** Brasil]*

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

> *[**GPE.ORG** Dallas, [**GPE.LOC** Texas]] announced its decision to build a subway system.*
>
> *Horas después, desde [**GPE.LOC** Toronto, [**GPE.LOC** Canadá]], el director ejecutivo de la Asociación de Jugadores, Bob Goodenow, dejó en claro que él creía que Bettman podía detener el paro.*

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 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/Postmodifiers

Premodifiers/Postmodifiers are inherently vague and difficult to decompose. For this reason, all GPE premodifiers/postmodifiers will be assigned the role GPE.GPE.

> *[**GPE.GPE** Israeli] troops*
>
> *[**GPE.GPE** New York] policemen*
>
> *[**GPE.GPE** U.S.] Commander-in-Chief*
>
> *[**GPE.GPE** U.S.] surveillance aircraft*
>
> *[**GPE.GPE** Iraqi} flag*
>
> *The [**GPE.GPE** California] company*
>
> *El Ministerio del Exterior de [**GPE.GPE** Rusia]*
>
> *El presidente de [**GPE.GPE** Costa Rica}*
>
> *otras tradicionales decoraciones [**GPE.GPE** chinas]*
>
> *los líderes de la [**GPE.GPE** UE]*

Military Activity

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

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

> [**GPE.GPE** Cuba] *proveyó asistencia militar a* [**GPE.GPE** Angola] *durante la guerra civil que surgió luego de su independencia en 1976.*

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

Places defined on a geographical or astronomical basis which are mentioned in a document and do not constitute a political entity give rise to Location entities.

These include, for example, the solar system, Mars, the Hudson River, Mt. Everest, and Death Valley.

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

3.4.1 Subtypes for Locations

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

Address

A location denoted as a point such as in a postal system or abstract coordinates ("31° S, 22° W"). The name of a location in a postal system is also an address.

> *[3600 Market Street]*
>
> *un [cruce] cerca de la comisaría de Tikrit*

Boundary

A one-dimensional location such as a border between GPEs or other locations.

> *Armenia occupies a politically challenging swath of the southern Caucasus, with [borders shared by Turkey, Azerbaijan, and Georgia]*
>
> *la [frontera] con Panamá*

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]…*
>
> *abusos que, según Castro, comete Europa o Estados Unidos en todo [el mundo]*

Water-Body

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

> *[The Missouri River]*
>
> *[el Océano Pacífico]*

Land-Region-natural

Geologically or ecosystemically designated, non-artificial locations.

> *…a politically challenging swath of [the southern Caucasus]…*
>
> *…las [montañas] rodeadas por un manto de árboles*
>
> *…las [selvas] del oeste de Colombia*

Region-International

Taggable locations that cross national borders.

>*…the main threats to [southern <u>Africa</u>] over the coming 15 years*
>*los países [<u>andinos</u>]*

Region-General

Taggable locations that do not cross national borders.

>*…from neighboring Bavaria, in [southern <u>Germany</u>]…*
>*…en [otra <u>parte</u> de la ciudad]…*
>*…[una <u>región</u> a 65 km al oeste de Ulan Bator]*

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 <u>center</u> of the country] but at [its <u>periphery</u>]"*

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:

>*"The rabbit is hiding behind that rock."*
>*"I have an idea in my head."*
>*"Tu camisa está encima de la cama."*

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."*
>*"Él puso la lámpara en su propio lugar."*

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

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

3.5 Facilities

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

3.5.1 Subtypes for Facilities

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

Airport

A facility whose primary use is as an airport.

> *[New York's LaGuardia Airport] has been a nightmare this year.*
>
> *[El Aeropuerto de Madrid-Barajas] está situado en el noreste de Madrid, a 12 kilómetros del centro.*
>
> *Por quinto día consecutivo amanecieron hoy bloqueados [los aeropuertos].*

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]…*
>
> *...los chinos temen que la situación provoque una oleada de nuevas inversiones en [las fábricas de textiles chinas]…*
>
> *...[tres nuevas plataformas petroleras] en la cuenca de Campos*

Building-or-Grounds

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

> *…at [the national archives].*
>
> *[The Berlin Wall]*
>
> *the parades at [Disneyland]*
>
> *Se ofrece al público la oportunidad de visitar [la Casa Blanca] en grupos de 10 o más personas.*
>
> *[La cárcel de [la base naval estadounidense de Guantánamo, Cuba,]] es un lugar…*

Unas [1200 minas estatales] de tamaño mediano contribuyen al 15 por ciento de la producción total.

Subarea-Facility

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

…two men who rented [an Aden apartment]…

La explosión afecto también a [una oficina de la Fiscalía]

Path

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

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

[Telephone lines] were knocked down…

...[las principales carreteras [que] comunican la sur andina ciudad peruana de Andahuaylas]

3.6 Vehicle

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

3.6.1 Subtypes for Vehicles

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

Air

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

…[military helicopters] float overhead…

[diez aviones de transporte C-295]

Land

Vehicles designed to locomote primarily upon land.

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].

[los vehículos militares]

…la gasolina que usan [los vehículos particulares]

Water

Vehicles designed to locomote primarily on or submerged in water.

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

[cuatro corbetas] y otras cuatro pequeñas naves

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]…

La bomba fue encontrada en [el baúl del coche].

Underspecified

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

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

[Israeli tanks and gunships]

…cuatro corbetas y [otras cuatro pequeñas naves]

3.7 Weapon

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

3.7.1 Subtypes for Weapons

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

Blunt

Blunt weapons are those designed or used as bludgeoning instruments.

He was beaten repeatedly with [a baseball bat]…

Tres campesinos se enfrentaron con [piedras], machetes y armas caseras

Exploding

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

> *…a small boat packed with [explosives]…*

> *[Una bomba] colocada en la carretera fue detonada alrededor de las 7:45 am….*

Sharp

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

> *[Knife]*

> *Tres campesinos se enfrentaron con piedras, [machetes] y armas caseras.*

Chemical

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

> *[Sarin gas]*

> *[gas lacrimógeno]*

Biological

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

> *[Anthrax]*
> *[Viruela]*

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…*

> *[un revólver] desde el interior de una mochila*

Projectile

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

> *[Bullets]*
>
> *sus milicianos lanzaron [seis <u>morteros</u>]*
>
> *se atribuyo la autoría del lanzamiento de [<u>cohetes</u> caseros] contra los asentamientos en el Gush Katia*

Nuclear

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

> *[Nuclear <u>missiles</u>], [hydrogen <u>bombs</u>], [dirty <u>bombs</u>], [deliberately sabotaged nuclear <u>facilities</u> leading to catastrophic meltdowns]*
>
> *Irán haya intentado enriquecer uranio para desarrollar secretamente [<u>bombas</u> atómicas].*

Underspecified

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

> *They searched him for [a <u>weapon</u>].*
>
> *U.N. [<u>arms</u>] inspectors*
>
> *Saddam Hussein tenía [<u>armas</u> secretas]*
>
> *la lucha contra [las <u>armas</u> de destrucción masiva]*

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 _lawyer_] would take that case.
>
> [No _one_] has claimed responsibility.
>
> There are [no confirmed _suspects_] yet, but officials say several Middle East groups are expected to be investigated.
>
> El candidato de Bush al Supremo dice que [ningún _presidente_] está por encima de la ley.
>
> "En vista de que [no hay un _muelle_], los pasajeros serán bajados a tierra con la ayuda de balsas."
>
> Braga dijo que [_ninguno_ de ellos] era Ravix.

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_].
>
> 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]…

Mi hermana es [médico].

Igor Sechin, [subsecretario de gabinete] y [antiguo aliado de Putin], se desempeña como presidente de la junta de Rosneft

un soldado [ecuatoriano]

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]

[Igor Sechin], subsecretario de gabinete y antiguo aliado de Putin, se desempeña como presidente de la junta de Rosneft

[El Ministerio Público] ha solicitado para cada uno de [los siete acusados] una pena de diez años de prisión.

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.

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.

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

[Las futbolistas] tenemos que tocar puertas para que nos apoyen para tener una liga en México

[Empresarios privados] obtuvieron propiedades a precios muy bajos poco después de la caída de la Unión Soviética en 1991.

Agregó que hay preocupación en el país porque está llegando ayuda alimentaria con transgénicos para el consumo de [niños en edad preescolar] y [mujeres embarazadas].

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 people] will participate in the parade.*
>
> *I don't know [how many people] came.*
>
> *Do you know [how many people] came?*
>
> *We will elect [five new officials].*
>
> *[Algunos participantes del Mercado] obtuvieron activos estatales multimillonarios utilizando diferentes trucos.*
>
> *[Los trabajadores] terminarán el tendido de la línea ferroviaria Qinghai-Tíbet a finales de este año*
>
> *Rice tiene planeado también ponerse en contacto con [funcionarios de la Unión Europea y de la ONU].*

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

> *A few,*
>
> *some (plural),*
>
> *most,*
>
> *more than half*
>
> *unos*
>
> *algunos*
>
> *la mayoría*

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 people] are going to attend the rally.*
>
> *[Más de 500 efectivos de la Fuerza de Seguridad Iraquí] arrestaron a unos 65 terroristas en el distrito de Al- Rasheed el 11 de abril*

From the context in which they appear, the under-specificity of these quantifications may be ascertained; however, the appearance of these strings often indicates 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 **Underspecified**. By partitioning these truly ambiguous cases into the USP category, annotators will be able to make clearer distinctions between the other four categories, thus improving consistency.

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

> *Ayer conocí a algunas personas que les encanta el ajedrez.*

the extent of the mention is the entire phrase:

> *[algunas personas que les encanta el ajedrez]*

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. In English, possessive endings ('s) are treated as separate tokens, and contractions are split (so that "*we're*" becomes the two tokens "*we*" and "*'re*"). Likewise in Spanish, we will regard both "del" (de + el) and "al (a + el) in the same way such that:

> *Del gobierno* is tagged as *d[el gobierno]*
>
>   and
>
> *Vamos al estadio de fútbol* is tagged as *Vamos a[l estadio de fútbol]*

Extents must begin at the beginning of a token and end at the end of a token.

Conjoined Mentions that are Modified

In constructions of conjoined mentions that share the same premodifiers or postmodifiers, each of the conjoined heads will be tagged form the head of a single mention. The following examples will yield two tagged mentions.

[Julio y Pedro quienes les gusta jugar fútbol]	
[<u>Julio</u> y Pedro quienes les gusta jugar fútbol]	Name mention
[Julio y <u>Pedro</u> quienes les gusta jugar fútbol]	Name mention

Bill Clinton and Jimmy Carter who are both former presidents	
[<u>Bill Clinton</u> and Jimmy Carter who are both former presidents]	Name mention
[Bill Clinton and <u>Jimmy Carter</u> who are both former presidents]	Name mention

[La salud de mujeres y niños nicaragüenses]	
[La salud de <u>mujeres</u> y niños nicaragüenses]	Nominal mention
[La salud de mujeres y <u>niños</u> nicaragüenses]	Nominal mention

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>].
[<u>Pedro Banchón</u>] era [un <u>agitador</u> profesional].

5.1.3 Names (NAM)

Proper nouns and nicknames of people, places, locations, facilities, organizations, vehicles, GPEs, and weapons.

[<u>John</u>]
[Defense Secretary <u>William Cohen</u>]
[The <u>Jeluzoon Refugee Camp</u> near [<u>Ramallah</u>]]
[El presidente <u>Jean Bertrand Aristide</u>]

pan [Bimbo]

[el Programa Mundial de Alimentos]

5.1.4 Quantified Nominal Constructions (NOM)

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

[the Yemeni port of Aden]

[the southern tip of Africa]

[a crowd of angry Muslims]

[el departamento de Leon]

[una casa de apuestas]

[la población d[el país]]

5.1.5 Unquantified (Bare) Nominal Mentions (BAR)

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

[lawyers]

[US officials]

offices in [foreign countries]

[Americans] eagerly await the results of the election.

[efectivos de la fuerza norteamericana]

[niños en edad preescolar]]

5.1.6 Pronouns (PRO)

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

[he]

[they]

[her]

[l]

[ella]

[ellos

[nosotros]

[usted]

This also includes demonstrative pronouns such as:

[this]

[that]

[eso/a/os/as]

[esto/a/os/as]

Since pronouns are commonly dropped or are implied in Spanish, just tag those that are in the text.

Reflexives: The reflexive morpheme "-se" should be marked as a pronoun if the entity it refers to is taggable as an entity.

> *De su lado Dahlane [**PRO** <u>se</u>] reunira el sábado en la noche con el Ministro isreaíi de Defensa Shaul Mofaz.*

> *Abbas debe reuniur[**PRO** <u>se</u>] el domingo con miembros de la Organización de Liberación Palestina.*

5.1.7 WH-Question Words and Specifiers (WHQ)

WH-question words and the specifiers: donde, cual, quien/quienes, que, cuyo

> *[<u>who</u>]*
>
> *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*
>
> *...dijo la agencia, [<u>la cual</u>] agrego que un oficial de seguridad murió en el enfrentamiento.*
>
> *Entre los herido figuran Wilfredo Reyes y Saidy Urdaneta, [<u>quienes</u>] fueron trasladados hasta el hospital capitalino Periferico de Coche.*
>
> *...la góndola, de color anaranjado [<u>que</u>] perdió los frenos*
>
> *se reunió con Leonid Abramov, con [<u>quien</u>] discutió la cuestión chipriota.*
>
> *Cuando llegue a la última fila de la tumba, [<u>donde</u>] yace mi hijo, vi una mancha negra sobre la piedra.*
>
> *los grupos extremistas judíos [<u>cuyos</u>] funcionarios de seguridad temen que puedan intensificar sus acciones...*
>
> *...la Comisión Pastoral de la Tierra, [<u>cuyo</u>] presidente, el obispo Toman Balduino, emitió un documento...*

5.1.8 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*
>
> *[<u>dos</u>] están en investigación*

4000 soldados han regresado a casa, mientras que [el <u>resto</u>] se esta retirando hacia el valle Bekaa.

*Según RSF, además de los 53 periodistas asesinados, 15 colaboradores de los medios de comunicación también murieron, y al menos [**HLS** <u>907</u>] fueron detenidos, y [**HLS** <u>1,146</u>] agredidos o amenazados.*

5.1.9 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]

[<u>algunos</u> estudiantes de la universidad]

Bono es [<u>uno</u> de los cinco ministros españoles que viajarán a Estados Unidos]

[la <u>mayoría</u>] de los iraquíes

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

5.2 Premodifier/Postmodifier Mentions

Pre/postmodifier mentions are those mentions which occur in a modifying position before or after 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/postmodifier 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 LOCs and GPEs in their pre/postmodifier positions. Transformations of names or nominal mentions for any other types of entities are not taggable.

Please be aware that in Spanish, almost all modifiers will be postmodifiers. Because of the nature of appositive constructions (APP) in Spanish, titles and honorifics that in English would be tagged as premodifiers would not be in

Spanish, such that whereas in English, the "President" in *[[President] Bush]* is a premodifier, "presidente" in *[[El presidente] [Bush]]* is the head of NOM mention.[2]

Taggable	Not Taggable
the [**NAMPRE** Russian] foreign minister	Stalinist nations
el Ministro de Asuntos Exteriores [**NAMPOST** ruso]	
Ministro de Asuntos Exteriores de [**NAM** Rusia]*	
[**NOMPRE** mountain] regions	mountainous regions
regiones de [**NOM** montañas]*	regiones [montañosas]
[**NOMPRE** government] grants	federal grants
concesiones [**NOMPOST** gubernamentales]	concesiones federales
concesiones del [**NOM** gobierno]*	
[**NAMPRE** Tuscan] food	Machiavellian strategy
comida [**NAMPOST** toscana]	
[**NOMPRE** Alpine] villages	
aldeas [**NOMPOST** alpestres]	
the [**NAMPRE** Microsoft] [**NOMPRE** court] case	
el proceso legal de [**NAM** Microsoft]*	
the [**NOMPRE** state] budget this year	
el presupuesto de [**NOM** l estado]*	

5.2.1 Name or Nominal Classification

For each premodifier/postmodifier mention, we must decide whether that mention is a Name (NAMPRE/NAMPOST) or a Nominal (NOMPRE/NOMPOST). If the mention would be tagged as a Name if it were to appear as the head of a noun phrase, we will mark it as a NAMPRE/NAMPOST mention. If the mention would be tagged as a Nominal if it were to appear as the head of a noun phrase, we will mark it as a NOMPRE/NOMPOST mention.

[Israeli] troops	NAMPRE
[Republican] voters	NAMPRE

---

[2] See section 5.3.1 for more information on appositive constructions (APP).

[[US] Secretary of State] Powell	US – NAMPRE, Secretary – NOMPRE
[Mr]. Clinton	NOMPRE
[rooftop] decks	NOMPRE
24 turistas [suizos]	NAMPOST
sopas [Maggi]	NAMPOST
La empresa [estatal] Petróleos de Brasil	NOMPOST
La reunión [ministerial]	NOMPOST

5.2.2 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, NAMPRE, ATR*
>
> *Press – ORG.Commercial, NOMPRE, ATR*
>
> *White House press secretary – PER, NOMPRE, ATR*
>
> *White House press secretary Scott McClellan – PER, NAM, SPC*

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 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.

> *[[Mr. Black, 58], [a victim of the terrorist assault]], told the Associated Press…*
>
> *[[the show's production company], [*
>
> *[[the heavy lift ship], [Blue Marlin]]*
>
> *[[Sauache County], [home of the Watchtower]]*
>
> *[[la empresa] [Corcel]]*
>
> *[[El arquitecto francés], [Jerome de Boisseson]]*
>
> *[[El director técnico del fútbol chileno], [Juvenal Olmos]*

[[El señor] [Banchón]][3]

An appostive construction can also consisto of two embedded APP's

[[[El alcalde] [Tony Rodríguez]], [un recien llegado] [que] [se] unio a inversionistas chinos para levantar uno de las docenas de aserraderos de Novo Progresso.]]

[El alcalde] [Tony Rodríguez]] is tagged as an APP with two Referential parts, and the entire mention is also tagged as a corefferenced APP.

5.3.2 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]

[La gente [que] no sabe leer ni escribir] se llama analfabeta

[[Jim Gilchrist], [[un contador jubilado] y [ex veterano de Vietnam]] [que] está ayudando a reclutar Minutemen a través del país]

[[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 ARCs are not tagged. Because an embedded apposition is requisite to ARCs, 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

---

[3] Please note that all mentions that include both a determiner (el, la) and an honorific (señor, doctor, presidente, etc.) are tagged as APP. When the determiner is not included, these mentions are tagged as either NAM or NOM.

when the name of one entity is used to refer to another entity. The sections below outline several common examples. This is not an exhaustive list. Any example of this kind of reference should be identified. We will coreference the mention with the entity to which the mention refers in the context and indicate that this is an example of Nickname Metonymy by selecting the check box.

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

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

> *Pero e camino de[**GPE.ORG** <u>Ankara</u>] para realizar su sueño largamente deseado de unirse al bloque de 25 naciones es todavía difícil y largo, envista del tema de derechos humanos y Chipre*

In this example there are two senses of the word Ankara: the city Ankara and the government of Turkey. We will tag the government of Turkey (GPE.ORG) mention and coreference it with the Turkey entity. If there is a later mention of the city of Ankara (for example, La reunión se terminó en {la ciudad} esta mañana), it would be a GPE.LOC mention of the Ankara entity.

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

6.2 City name for Sports Team

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

> *"Nosotros esperamos solucionar este problema en los próximos dos partidos (ante [ORG <u>Bolivia</u>] y [ORG <u>Venezuela</u>]), que serán trascendentales."*

7 Cross-Type Metonymy

Cross-Type Metonymy occurs when more than one aspect of an entity is referenced in a document. For example, entities of type Organization often have a physical entity of type Facility associated with them. These two incarnations of the same entity will be tagged as type Organization when the textual reference is directly referring to the organization and as type Facility when the mention refers to the physical building.

At the entity annotation stage, we will group entities of different types together with a Metonymy relation when they refer to different aspects of the same underlying entity.

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

Examples

In this example, there are two mentions of a hospital. The first mention is referencing the physical building or hospital facility. The second references the organization that runs or administrates the hospital.

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

We will annotate examples like this as follows.

Type	Argument 1	Argument 2
Metonymy	St. John Macomb Hospital	hospital