ACE (Automatic Content Extraction) English Annotation Guidelines for Entities

Version 6.6 2008.06.13

Linguistic Data Consortium

http://projects.ldc.upenn.edu/ace/

Table of Contents

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*

Entities are limited to the following seven types:

- Person (PER) - Person entities are limited to humans. A person may be a single individual or a group.
- Organization (ORG) - Organization entities are limited to corporations, agencies, and other groups of people defined by an established organizational structure.
- Geo-political Entity (GPE) - 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 (LOC) - Location entities are limited to geographical entities such as geographical areas and landmasses, bodies of water, and geological formations.
- Facility (FAC) - Facility entities are limited to buildings and other permanent man-made structures and real estate improvements.

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.

Metadata imported from other sources is not part of the source signal from which ACE systems are presumed to extract information. Therefore no information is to be extracted from such metadata. Hence, in newswire documents, material in headlines and slug sections is not to be tagged. In broadcast data and telephone speech data, only the transcribed speech is to be tagged. Added information, i.e. <TURN> tags or speaker identification tags, is NOT to be tagged. In weblog and newsgroup files, we do not tag the poster name.

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. 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]* [1]
>
> *[The political cat of the year]*
>
> *She's known as [the brain of the family]*

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.

> *[Bill Clinton]*
>
> *[Edmund Pope]*
>
> *[The President of the U.S.]*
>
> *The police found [[his] body]*

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 as described below. This will include family names and ethnic and religious groups that do not have a formal organization unifying them.

> *[Analysts]*
>
> *[IBM s lawyers]*

---

[1] *For the remainder of the document, square brackets [ ] will indicate the extent of an Entity Mention and underlining will be used to indicate its head. Only the part of the extent that illustrates the example being discussed will be annotated. Complete annotations of the examples can be found at the end of the guidelines.*

[The family]

[The house painters]

[The linguists under the table]

[The Kennedys]

[The Arabs]

[The Christians]

[the Hmong individuals]

[the ethnic Albanians]

PER.Indeterminate

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

[You] want to go to the party.

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

3.1.2 Titles, Honorifics, and Positions

In English, titles and most honorifics precede the name. We will not consider these to be part of the name of a Person. We will annotate these as mentions in their own right. The parts of titles are taggable if they refer to entities. For example, in the string *"US Secretary of State Madeline Albright"*, there would be four mentions of three distinct entities.

[Secretary of State Madeline Albright]	PER
[Secretary of State]	
[US]	GPE
[State]	ORG

Saints and other religious figures

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

St. Christopher

[St. Christopher]	PER
[St.]	

Pope John Paul II

[Pope _John Paul II_]	PER
[Pope]	

References to "God" will be taken to be the name of this entity for tagging purposes. If it is used as a descriptor rather than a name, it will be considered a nominal mention. Note that capitalization information may not be available in speech transcripts.

> _If you believe in [god], you must_   name mention
>
> _Although he felt like he was [a god], he_   nominal mention

"Empty" Job Titles

Sometimes job titles refer to an "empty" position, as in:

> _She is running for President._

In cases like this, we will not tag President because it refers to the job or position and not to the person holding it.
Note, however, that this is different from:

> _Who will be [the next President]?_

In this case, we will tag [the next *President*] as PER.USP, because the reference is more to a unknown/hypothetical person than to a position or office.

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

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

> _[Batman] has become a popular icon_
>
> _[Adam West] s costume from Batman the TV series_

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

> _Morris the cat_
>
> _Snuggle, the fabric softener bear_

3.1.4 Body Parts

Body parts are taggable ONLY in cases in which the body part can reasonably define the entire person, as in the case of pieces of corpse recovered after attacks, accidents, etc.

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 office] is investigating the disaster to determine if criminal charges could be filed.*
>
> *Putin, a former [KGB] agent, defended [the court [that] convicted Pope and [the security services]],*
>
> *The Financial Accounting Standards Board will take no conclusive action on its current project on business combinations until [Congress] has reconvened in 2001*
>
> *[The US navy] now says the USS Cole was being refueled when an explosion ripped through it in Yemen last week, killing 17.*
>
> *[the administration]*

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 [TechSource Marine Industries in State College, PA.], and an associate were…*

Educational (EDU)

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

> *Pope, 54, a retired U.S. Navy officer from State College, PA., had worked for [the Applied Research Laboratory at [Pennsylvania State University]].*
>
> *[NDSU] and [University of Minnesota] weeds specialist Alan Dexter says 98% of the plant survived.*

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 [<u>Alton</u>], [<u>Solace</u>] frequently headlines in Celtic festivals.

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>Al Aqsa Martyr's Brigade</u>]

> [<u>Tamil Tigers</u>]

2. Political Parties:

> [<u>Republican Party</u>]

> [<u>Labour Party</u>]

3. Political Advocacy Groups and Think Tanks:

> [<u>ACLU</u>]

> [The <u>Cato Institute</u>]

4. Professional Regulatory and Advocacy Groups:

> [The <u>American Bar Association</u>]

> [The <u>American Medical Association</u>]

5. Charitable Organizations:

> [The <u>Red Cross</u>]

> [The <u>United Way</u>]

> [<u>Doctors Without Borders</u>]

6. International Regulatory and Political Bodies:

> [<u>UN</u>]

[NATO]

[The World Bank]

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

three of the [U.N]. workers stationed in East Timor

[International Monetary Fund] aid

[Hizbollah]

[Islamic Resistance]

[Rally for Congolese Democracy]

[the ``Caravan of Death],'' [a military party [that] killed 73 political prisoners]

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

[Institutional Revolutionary Party]

[The GOP]

[the Democratic Party]

[the Socialist People's Party]

[Republican National Committee]

[Rock the Vote]

[American Medical Association]

[American Council on Education]

[College Board]

[National Rifle Association]

[American Diabetic Association]

[NAACP]

[National Center for Public Policy and High Education]

[American Bar Association]

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

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

Media (MED)

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

[Agence France Presse]

[Reuters international news [agency]]

[Time magazine]

[abc news]

[Associated Press]

[Chicago Sun-Times]

[Road & Track magazine]

[National Geographic]

Religious (REL)

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

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

[German Bishops Conference]

[The Vatican]

Medical-Science (SCI)

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

[Overseas Chinese Physics Institute]

[Gulf Coast Research Laboratory]

[Massachusetts General Hospital]

[medical and health groups]

Sports (SPO)

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

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

[Saudi Soccer Federation]

[Taekwondo Association]

[Philippines Olympic Committee]

[national hockey league]

[San Francisco 49ers]

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. 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.

3.2.3.1 Military/Police

We make exception to the above rule for certain military entities. We have found that "troops", "forces", and "police" are commonly used in the same form multiple times in a document, and it is impossible to tell when they are referring to the same group of people, so co-referencing becomes very difficult and inconsistent.

Because the same mention string can be used multiple times in different contexts without distinguishing the PER entities involved, we have decided these mentions are better tagged as referring to the ORG of which they are members.

We will tag any mention of "troops", "police", "forces", etc. as ORG.GOV, where it is not explicitly referring to a partial subset of the organization.

> *[U.S. <u>troops</u>] entered Baghdad yesterday - BAR.ORG.GOV.SPC*
>
> *[the coalition <u>forces</u>] took the bridge to Basra - BAR.ORG.GOV.SPC*
>
> *[A few <u>troops</u>] were left to man the base. – NOM.PER.Group.SPC*
>
> *[Chicago <u>police</u>] arrested the suspect.. – BAR.ORG.GOV.SPC*

We will tag any structured subset of military organizations as ORG.GOV.

> *[the <u>squad</u> of Marines]*
>
> *[the 101st <u>division</u>]*

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.
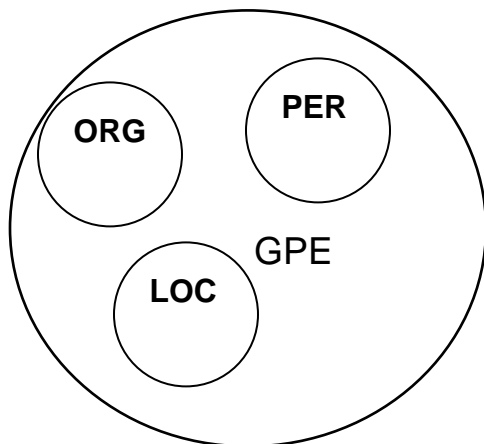
In this sentence,

> *The people of France welcomed the agreement.*

there are two mentions

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

The mention of the population of France is marked GPE, rather than PER. These mentions would be coreferenced, 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. See section 6, "Nickname Metonymy", for further discussion.

3.3.1 Subtypes for GPEs

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

Continent

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

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

[Antartica]

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…

[Australia]

[the axis of evil]

[Historical China] (co-referenced with today's China)

[the Soviet Union] (NOT co-referenced with Russia)

State-or-Province

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

[Salzburg] governor Schausberger said…

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.

Population-Center

Taggable mentions of the entireties of any GPE below the level of County-or-District.

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]].

GPE-Cluster
Named groupings of GPEs that can function as political entities.

> *[Eastern Europe]*
>
> *[the European Union]*
>
> *[the Middle East]*
>
> *[Southeast Asia]*
>
> *[Latin America]*
>
> *…and the government tried to redefine [the West].*
>
> *[the Arab World]*
>
> *[the Asia-Pacific region]*

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.

> *[the southern United States]*

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]*

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.
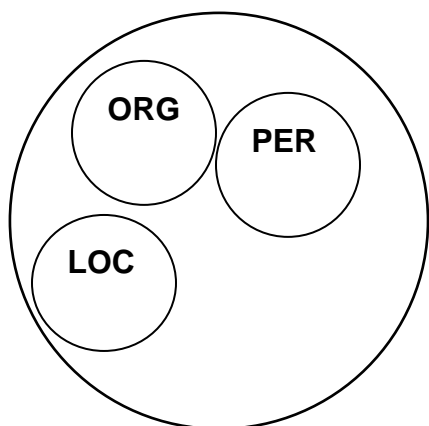
- **GPE.ORG** - <u>France</u> signed a treaty with Germany last week.
- **GPE.PER** - <u>France</u> vacations in August.
- **GPE.LOC** - The world leaders met in <u>France</u> yesterday.
- **GPE.GPE** - <u>France</u> produces better wine than New Jersey.

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 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.

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

Governments

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

> But *{[GPE.ORG] the Russian <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 modifying government organizations, like *New York's police department* and *Kentucky's fire marshall'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] <u>Switzerland</u>}'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.

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]** <u>Iraq</u>**}**, 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]** <u>United States</u>**}** _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.

> **{[**<u>New York</u>**]}** *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]** <u>Ohio</u>**}**), 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.

> *[Capitol 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-International

Taggable non-named locations that cross national borders.

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

Region-General

Taggable locations that do not cross national borders.

> *In [another part of the city],…*

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

> *…from neighboring Bavaria, in [southern Germany]…*

> *[airspace]*

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:

> *"Your coat is under the dog."*

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

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

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

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

> *"He put the lamp back in its place."*

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

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

3.5 Facilities

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

3.5.1 Subtypes for Facilities

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

Airport

A facility whose primary use is as an airport.

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

[the military <u>airbase</u>]

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 <u>refinery</u>] to [the <u>smelter</u>]…*

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 [<u>Disneyland</u>]*

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-facilitiy, but other portions of buildings, such as walls, windows, or doors, are not tagged.

> *…two men who rented [an Aden <u>apartment</u>]…*

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 <u>streets</u>].*
>
> *[Telephone <u>lines</u>] were knocked down*

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).

In part-whole constructions, genericity should be tagged the same for both part and whole.

> [one of the biggest rebel groups supported by Uganda] ORG.NON.SPC

> [the biggest rebel groups supported by Uganda] ORG.NON. SPC

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.

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.

4.3 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]

> [At least four people] were injured.

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.*

Even if the author may intend to use a GEN reading, if he/she refers to all members of a set rather than the set itself, use the SPC tag.

> *[All ACE annotators] are intelligent*   *[All ACE annotators] = SPC*
>
> *[ACE annotators] are intelligent*   *[ACE annotators] = GEN*

4.5 Under-specified Referential (USP)

An entity is USP if it is impossible to determine its referent. 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].*

 [You] know, I didn't even realize...

Imprecise quantifications *occasionally* 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.

> *[Over two hundred thousand people] participated in the riots.*

In cases where the author makes mention of an entity whose identity would be difficult to locate, and then conflates it with multiple other fuzzy mentions, all mentions are tagged as USP.

> *[Sources] said…*

> *[Officials] reported…*

When used as in the above examples, multiple mentions of *[sources]* and *[officials]* in a document should be coreferenced (see section 5 below).

Mentions of *[a source]* or *[an official]* would be SPC.

If a GEN or SPC reading is possible, the USP tag should not be used. However, the annotator may still come across cases which are impossible to classify. In these cases, the 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.

Punctuation

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.

Thus, in

> *Rep. Joe Smith (R-KY)*

The extent of the mention does not include the second parenthesis.

> *[Rep. <u>Joe Smith</u> (R-KY]*

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.

20 angry men and women

[20 angry <u>men</u> and women]	NOM.PER.GROUP
[20 angry men and <u>women</u>]	NOM.PER.GROUP

Bill Clinton and Jimmy Carter who are both former presidents

[<u>Bill Clinton</u> and Jimmy Carter who are both former presidents]	NAM.PER.IND
[Bill Clinton and <u>Jimmy Carter</u> who are both former presidents]	NAM.PER.IND

The movers and shakers in Washington

[The <u>movers</u> and shakers in Washington]	NOM.PER.GROUP
[The movers and <u>shakers</u> in Washington]	NOM.PER.GROUP

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-word item, the last word 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:

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

5.1.3 Names (NAM)

In the most obvious cases, a NAM is a name or nick-name of any entity.

> *[John]*
>
> *[Defense Secretary William Cohen]*
>
> *[The Jeluzoon Refugee Camp near [Ramallah]]*
>
> *[North Dakota State University in [Fargo]]*
>
> *[The house of representatives]*
>
> *[the US army 101ˢᵗ Airborne Division]*
>
> *[Congress]*

More borderline cases exist, however.

> *[the US Army]*
>
> *[the US Supreme Court]*

5.1.4 Quantified Nominal Constructions (NOM)

A noun quantified with a determiner, a quantifier, or a possessive. These noun phrases may contain adjectives as well.

> *[The lawyer]*
>
> *[the Yemeni port of Aden]*
>
> *[some American executives]*
>
> *[the southern tip of Africa]*
>
> *[a crowd of angry muslims]*
>
> *[these people]*
>
> *[other U.S. officials]*
>
> *[thousands of troops]*
>
> *[this year's Miss America]*
>
> *[Saddam's troops]*

NOM mentions are in general only allowed one word in the head extent. The main exception to this are LOC mentions that consist of GPE name and directional modifier.

[the southwestern United States]

This entity was originally a GPE.NAM, but the modifier causes it to become a LOC. "The United States" cannot be decomposed into a head and modifier, so it must be the entire head.

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]]

[representatives of the company's Tananger branch]

[Americans] eagerly await the results of the election.

When singular, the noun phrase often refers to a type.

[Man] is a strange beast.

5.1.6 NAM vs. NOM

Some ambiguities can arise when trying to make a NAM-NOM distinction. It may appear that a NOM is being used to name something, or that a NAM mention may be decomposable into a few NOMs.

A general property of NAMs is that they are defined to pick out one particular entity as a referent. They are unique identifiers, like "Vladimir Putin" or "The United States."

NOMs, on the other hand, define an entire category. They can pick out a referent which belongs to that category, but only after disambiguating it from all other potential members of the category. If a nominal mention is used as an individual reference in a discourse, the head noun often has to be "individualized" via quantification and/or qualification with determiners, adjectives, relative clauses, etc.

[Vladimir Putin] sat at the table.　　　 *[Vladimir Putin] NAM*

[The man] sat at the table.　　　　　 *[The man] NOM*

When NAMs do carry a determiner, the determiner is a definite article and is not separable from the NAM. The definite article cannot be replaced with other determiners, quantifiers, or possessives.

[The New York Times] ran the article.

Ungrammatical:

*A New York Times ran the article.

*That New York Times ran the article.

One of the trickiest parts of distinguishing NAM's and NOM's is NOM categories modified by NAM's such that they only have one referent, such as:

the Pakistani army
the Egyptian supreme court
the University of Chicago payroll department

With the GPE/ORG modifying the categories, they pick out a specific referent in each NOM category. It is hard to decide whether the whole string should be treated as a NAM, or as a NOM mention with GPE/ORG NAMPRE.

Looking up the real name of the entity is not a good option for any entity type, since consistency is more important than accuracy and annotation speed can't afford to drop.

To annotate this kind of entity, we will follow the following steps:

Step 1: ORG NAM or NOM?
Some ORGs are unambiguously NAM, as they automatically pick out one specific entity, not a member of a set.

[Nazareth Academy] - NAM.ORG.EDU.SPC
[the Danger Danger Gallery] - NAM.ORG.COM
[the United States Armed Forces] – NAM.ORG.GOV

Some ORGs are unambiguously NOM, as they could not be considered the name of an organization, only a type of organization.

[the [U.S]. military] – NOM.ORG.GOV.SPC

When it's possible for an ORG to be NAM or NOM, default to the word count.

Rule : ORG's with more than two words are NAM's, and ORG's with one or two words are are **most often** NOM's (see trumping rules below).

Thus the following are NOM's:
-army
-supreme court
-city council

-foreign ministry
-health ministry
-defense department

But the following are NAM's:
-center for disease control
-ministry of health
-automobile manufacturer's association
-communications development office
-developmental concepts and doctrine center
-census data center

Trumping Rules:

a)   Acronyms

Any ORG's with a corresponding acronym are labeled NAM

[European Union (EU]
[American Airlines (AA]

These examples may be considered unambiguously NAM, but there may be more difficult cases which this rule will clarify.

Step 2: NAMPRE NOM vs. NAM

Once it is known whether the modified ORG is NOM or NAM, the next step is to decide whether the entire mention is NAM + NOM or NAM. For this, we will follow these rules:

1:   ORG ORG/GPE --> ORG.NAM

[Oxfam Japan] NAM
[Abbot Laboratories Phillippines] NAM
[Diagnostic Division, Abbot Laboratories] NAM

2: ORG of/at ORG/GPE --> ORG.NAM

[Agricultural Bank of China]   NAM
[the Payroll Department at University of Pennsylvania] NAM

3: ORG/GPE ORG

The Pakistani army
Pakistan Supreme Court
Thailand's ministry of health
Abbott Laboratories Diagnostic Division

a)   ORG/GPE ORG.NOM

i)   GPE-adj ORG.NOM --> GPE ORG.NOM

[the Pakistani <u>army</u>]   NOM
[Pakistani] NAMPRE

ii)   ORG/GPE-noun ORG.NOM --> ORG.NAM

[Pakistan Supreme Court]   NAM
[Abbott Laboratories Diagnostic Division]   NAM

b)   ORG/GPE's ORG NAM/NOM --> GPE's ORG.NOM

[Thailand's health <u>ministry</u>]   NOM
[Thailand's <u>ministry</u> of health] NOM
[Abbott Laboratories' Diagnostic <u>Division</u>] NOM

c)   ORG/GPE ORG.NAM --> ORG.NAM

[Chinese Center for Disease Control]   NAM
[U.S. Centers for Disease Control] NAM

d)   ORG/GPE possessive/adjective ORG

 Any [some adjective/possessive] ORG.NAM --> ORG.NAM

[The present-day <u>CDC</u>] NAM

Trumping Rules:

ORG.USP --> ORG.NOM

[the proposed <u>China Development Bank</u>] NOM

Any hypothetical, future, etc. organization is automatically tagged NOM regardless of any modifiers or other constructions. The modifiers' inclusion in the head will depend, however, on whether the construction would be NAM PRE NOM or NAM if not USP.
[the proposed [Chinese] Development <u>Bank</u>]   NOM
[Chinese]   NAMPRE

Plural ORG's

Any ORG's that are plural are automatically considered NOM's:

The Chinese, U.S., and Japanese Centers for Disease Control and Prevention

Mixed-Type Lists

ORG's which occur in a list format as above need to match in type. This exception may cause a ORG which would usually be NAM to be considered NOM, as in:

[*The Israeli Army, the Japanese Army, and the [US] Army*] NOM

GPE/ORG in Mention Head
Note that if the modified ORG is considered a NAM, the rule is to include the GPE or ORG as part of the head of the name, even though the ORG could *be a stand-alone NAM.*

Annotators may often find the two versions of names in the same file.

[European Automobile Manufacturers' Association] NAM.ORG
[Automobile Manufacturers' Association]   NAM.ORG

[Chinese Center for Disease Control and Prevention]   NAM.ORG
[Center for Disease Control and Prevention] NAM.ORG

[China Southern Air Holding Company] NAM.ORG
[Southern Air Holding Company] NAM.ORG

[European Automobile Manufacturers' Association Beijing Office] NAM.ORG

5.1.6 Pronouns (PRO)
Pronouns with the exception of wh-question words and the specifier 'that'.

[*he*]

[*they*]

[*no one*]

[*I*]

[*That*]*'s not mine. (demonstrative "that" not specifier "that" used to introduce a relative clause)*

[*These*] *have more terminals than the other top-ranking airports*

[*his own*]

[*everyone*]

5.1.7 WH-Question Words and Specifiers (WHQ)
WH-question words and the specifier 'that'.

[*who*]

the executive, [who] spoke on the condition of anonymity

the Russian navy, [which] waited several days before attempting to launch any rescue mission.

Germany, [where] they were taken yesterday

[Who] is the president of Brazil?

[Whose] underwear is this?

As a general rule, coreference a WHQ in a question with its answer.

[Who] was [the first president]?   *WHQ E1-REF, NOM E1-REF*

[George Washington] was..   *NAM E1-REF*

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 toughest]

[more than 30]

[many] on both sides

[60%] said

[sixty percent] said

[35] were injured

They will [each] pay $5.

Bare demonstratives followed by a relative clause (or modified in some way) should be tagged HLS.

[Those who would disagree] won't be happy.

[Those present at the meeting] noticed.

We must help [those in need]

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.

[some of the lawyers]

[one of the houses]

[half of the team]

[all of them]

[sixty percent of the participants]

5.1.9.1 Non-Partitives

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.

To determine this, consider whether, if the WHOLE element of a PTV construction were omitted, the PART element would resemble a HLS mention. If the head of the resulting construction would be another mention type (e.g., BAR, NOM, NAM, etc.) then the construction is *not* PTV.

> *[a <u>few</u>] were planning to attend* HLS.PER.Group.SPC
>
> *[A <u>few</u> of them] were planning..* PTV.PER.Group.SPC

Two Entity Mentions

Examples of non-partitives (two entity mentions):

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

Redundant Embedded Mentions

Each of these is a pair of co-referenced mentions

> *[a <u>group</u> of [<u>people</u>]]*
>
> *[a variety of [non-profit <u>organizations</u>]]*
>
> *[the <u>city</u> of [<u>Basra</u>]]*

Non-quantified Nouns

The following may look like PTV's, but they are not.

> *[hundreds of <u>troops</u>]* NOM
>
> *[thousands of <u>refugees</u>]* NOM
>
> *[a lot of <u>people</u>]* NOM

If the PART element would resemble a HLS mention, but it cannot be constructed without "of", then it is a Non-Quantified Noun, and should be tagged NOM.

> *[Five people] = NOM*
>
> *[<u>Five</u> of [the <u>people</u>]] = PTV, [the <u>people</u>] = NOM*

But

thousands people* = not grammatical

[thousands of <u>people</u>] = BAR

5.2 Pre-modifier Mentions

Premodifier mentions are those mentions which occur in a modifying position before another word(s). Their taggability is not determined by taggability of the modified word.

However, their taggability can be affected by the modified word, in cases like:

criminal charges

Islamic nations

The word "criminal" can either refer to a PER entity, or to "of or relating to crime". In the phrase "criminal charges", it takes on the second meaning.

"Islamic" is meant to refer to the religion of Islam, while "Muslim" refers to the people of Islam. "Islamic" is sometimes used interchangeably with "Muslim" when modifying PER entities, and then we would tag it PER. However, if "Islamic" is modifying other entity type, it is not taggable.

If there is ambiguity, we would go with the PER meaning, such as:

[feminist] groups

Also, we do not tag the "X" in "X" department, i.e:

[the state department]

We tag "state" in "Secretary of State", because it is referring to the department, and "state" in "state property", because it is referring to the GPE entity, but we consider this "state" too abstract to tag.

5.2.1 Construction of Pre-modifier mention

To be taggable, the construction of the premodifier mention must be identical to its BAR, NOM, or NAM construction, **unless** it is a NAMPRE PER.Group, LOC, ORG, or GPE mention.

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	
[school] vouchers	
the [state] budget this year	

NAM adjectives

The exception to this rule of transformations is for NAM mentions of LOC's, GPE's, ORG's, and PER.Group's in their premodifier positions.

Taggable Names	Adjectives	Nouns
Taggable	GPE LOC ORG *PER.Group	PER GPE FAC LOC ORG
Not Taggable	FAC PER.Individual PER.Indeterminate	——

*ethnicities and religious affiliations

> [Islamic] militants.
>
> [Jewish] traditions.
>
> [Arabic] movements
>
> [democratic] senators
>
> [American] views
>
> [Himalayan] explorers

PER.Group ethnicities and religious affiliations in NAMPRE constructions will be tagged GEN and ATR. Unless it is clear that the NOM mention is the same generic group, the NAMPRE will not be coreferenced with the entity it modifies.

> [the [Hmong] individuals]   - not coreferenced
>
> [Christian [people]] celebrate Christmas - coreferenced

5.2.1 NAMPRE or NOMPRE

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

[Israeli] troops	NAMPRE
[Republican] voters	NAMPRE

[[US] *Secretary* of State] *Powell*	[US] NAMPRE [Secretary] NOMPRE
[*Mr*]. *Clinton*	NOMPRE
[*rooftop*] *decks*	NOMPRE
[*airline*] *regulators*	NOMPRE
[*Chrysler*] *factories*	NAMPRE
[*union*] *leaders*	NOMPRE
[*subway*] *schedule*	NOMPRE
[*muslim*] *extremists*	NAMPRE
[*Jewish*] *organizations*	NAMPRE
[*Arabic*] *movements*	NAMPRE

In accordance with the rule above, if an ORG falls under NOM in our NAM/NOM rules, it should be NOMPRE, but if under NAM, it should be NAMPRE.

> *[state department] spokesman*   *NOMPRE.ORG.GOV*
>
> *[US state department] spokesman*   *NAMPRE.ORG.GOV*

5.2.1.1 NAMPRE PER.Group mentions

The exception to this rule is that we tag PER.Group religious and ethnic groups NAMPRE, although we tag them NOM or BAR when they appear as the head of a noun phrase.

> *a majority of [the Kurds]*   *NOM.PER.Group.GEN**
>
> *[Shiite Muslims] believe that..* *BAR.PER.Group.GEN*
>
> *[Shiite] Muslims believe that..* *NAMPRE.PER.Group.GEN*
>
> *[Muslim] people believe that.. NAMPRE.PER.Group.GEN*

This operates on a similar principle as tagging:

> *[Americans]*   *BAR.GPE.Nation.SPC, Role=PER*
>
> *The [American] people NAMPRE.GPE.Nation.SPC, Role=GPEl*

The difference is that there is no external NAM entity associated with PER.Groups, as there is "America" for "Americans".

5.2.1.2 NOMPRE.PER mentions

Taggable NOMPRE PER.Group mentions should be tagged PER.Indet, not PER.Group.

This is because the premodified form does not distinguish singular and plural, and they do not have an external identity as a group (unlike the NAMPRE.PER ethnic and religious groups).

> *[rebel] groups] - NOMPRE.PER.Indet.GEN*

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

5.2.3 Post-modifiers
There are some rare cases of post-modifiers in English. These should be tagged as pre-modifiers.

all things Seattle

Extent	Type	REF-ATR
[Seattle]	NAMPRE.GPE.PopCenter.SPC.GPE	E1-ATR

the two villages, Arab and Jewish

Extent	Type	REF-ATR
[the two villages, Arab and Jewish]	NOM.GPE.PopCenter.SPC	E1-REF
[Arab]	NAMPRE.PER.GROUP.SPC	E2-ATR
[Jewish]	NAMPRE.PER.GROUP.SPC	E3-ATR

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.

> [[Bill], [John's lawyer]]
>
> [[Mr. Black, 58], [a victim of the terrorist assault]], told the Associated Press…

[[the show's production <u>company</u>], [<u>Celador</u>]]

[[the heavy lift <u>ship</u>], [<u>Blue Marlin</u>]]

[[<u>Sauache County</u>], [<u>home</u> of the Watchtower]]

[[We] [Richmonders]]

[[[Pittsburgh], a US airways <u>hub</u>]] recently paid to revamp the airport including shopping malls in the terminals

The second element is always ATR while the first is REF.

5.3.1.1 Multi-part Appositives

In cases where there are more than two elements, all elements after the first are ATR.

Egypt, O Great, land of grace, cradle of civilization

Extent	Type	REF-ATR
[Egypt, O Great, land of grace, cradle of civilization]	APP.GPE.NATION.SPC	E1-REF
[<u>Egypt</u>]	NAM.GPE.NATION.SPC	E1-REF
[O Great, <u>land</u> of grace]	BAR.GPE.NATION.SPC	E1-ATR
[<u>cradle</u> of civilization]	BAR.GPE.NATION.SPC	E1-ATR

5.3.1.2 Mis-matched Appositives

Iliescu and Nastase, ex-president and ex-prime minister

Extent	Type	REF-ATR
Iliescu and Nastase, ex-president and ex-prime minister	APP.PER.GROUP.SPC	E1-REF
[<u>Iliescu</u>]	NAM.PER.IND.SPC	E2-REF
[<u>ex-president</u>]	BAR.PER.IND.SPC	E2-ATR
[<u>Nastase</u>]	NAM.PER.IND.SPC	E3-REF
[ex-prime <u>minister</u>]	BAR.PER.IND.SPC	E3-ATR

Bob, the director, and Bill, the producer, who both thought...

Extent	Type	REF-ATR
[Bob, the director]	APP.PER.IND.SPC	E1-REF
[<u>Bob</u>]	NAM.PER.IND.SPC	E1-REF
[the <u>director</u>]	NOM.PER.IND.SPC	E1-ATR
[Bill, the producer]	APP.PER.IND.SPC	E2-REF
[Bill]	NAM.PER.IND.SPC	E2-REF
[the <u>producer</u>]	NOM.PER.IND.SPC	E2-ATR
[<u>who</u>]	WHQ.PER.GROUP.SPC	E3-REF

5.3.2.3 Non-Appositives

Appositives must have equivalent elements capable of standing on their own.

The following are not appositives:

> *Argentina, Brasil, Chile and Mexico, members of XYZ*

> *Sotheby's and Christies', the world's biggest auction houses*

None of the separate entity mentions in the first element are equivalent to the second element, which is the set of all of them.

These are also not appositives

> *The American Civil Liberties Union (ACLU)*

> *He, himself had known it was true.*

> *President George Bush.*

> *Israeli newspaper Ha'aretz.*

The following non-appositive is actually a complex post-modifier construction, so make sure to tag the 2nd element ATR.

> *George Bush, president of the United States*

> *[George Bush, president of the United States]*   *NAM E1-REF*

> *[president of the United States]*   *BAR E1-ATR*

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]*

> *[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.

5.4 Referential and Attributive Mentions

5.4.1.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]…

Premodifying mentions are always ATR, regardless of whether they are coreferenced with the modified entity.

> [*U.S*]. president George W. Bush
>
> [*U.S*]. helicopters surrounded the city

Regardless of the truth of the attribute, the attributive mention is co-referenced with its corresponding referent.

> Hillary Clinton is [the future *president*]. (Co-referenced)
>
> Hillary Clinton, [the future *president* of the US] (Co-referenced)
>
> Bush is [the future *president* of the US]. (Co-referenced)
>
> I think that Hillary Clinton will be [the future *president*]. (Co-referenced)
>
> Hillary Clinton may be [the future *president*]. (Co-referenced)
>
> Bush said that Hezbollah is [a terrorist *organization*]. (Co-referenced)
>
> Abbas said that Hezbollah is not [a terrorist *organization*]. (NOT co-referenced)

In the final example the attribute is explicitly NOT assigned to the referent and thus not co-referenced.

5.4.2. Referential (REF)

An entity is referential when it refers to an entity—whether real or hypothetical—rather than a property or attribute of an entity.

> [That *man*] is a teacher.
>
> [Israeli *troops*] are well-trained.
>
> [*I*] need [five *people*] to go to the conference.
>
> [The *United States*] is a violent country.

6 Nickname Metonymy

Metonymy occurs when a speaker uses a reference to one entity to refer to another

entity (or entities) related to it. For example, in the sentence below *Beijing* is a capital city name that is used as a reference to the Chinese government:

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

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

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

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

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

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

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

6.2 GPE 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.

> [*America*] *brought home the gold. – NAM.ORG.SPO*

> [*Pittsburgh*] *won 30-27 - NAM.ORG.SPO*

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*

9 Fully Annotated Examples

In cases where tagging is context dependent, the most likely tagging(s) is/are given and an asterisk (*) is used after the tag OR in cases where the tagging is extremely context dependent, the tag is left as (?).

9.1: Introduction

Joe Smith

Extent	**Type**	**REF-ATR**
[Joe Smith]	NAM.PER.Ind.SPC	

the guy wearing a blue shirt

Extent	**Type**	**REF-ATR**
[the guy wearing a blue shirt]	NOM.PER.Ind.SPC	

He

Extent	Type	REF-ATR
[He]	PRO.PER.Ind.SPC*	

9.3: Entity Types and Subtypes

9.3.1: Persons

He is a real turkey

Extent	Type	REF-ATR
[He]	PRO.PER.Ind.SPC	E1-REF
[a real turkey]	NOM.PER.Ind.SPC	E1-ATR

The political cat of the year

Extent	Type	REF-ATR
[The political cat of the year]	NOM.PER.Ind.SPC*	

She's known as the brain of the family

Extent	Type	REF-ATR
[She]	PRO.PER.Ind.SPC	E1-REF
[the brain of the family]	NOM.PER.Ind.SPC	E1-ATR
[the family]	NOM.PER.Group.SPC	E2-REF

Bill Clinton

Extent	Type	REF-ATR
[Bill Clinton]	NAM.PER.Ind.SPC	

Edmund Pope

Extent	Type	REF-ATR
[Edmund Pope]	NAM.PER.Ind.SPC	

The police found his corpse

Extent	Type	REF-ATR
[The police]	NOM.ORG.GOV.SPC	E1-REF
[his corpse]	NOM.PER.IND.SPC	E2-REF
[his]	PRO.PER.IND.SPC	E2-REF

The President of the U.S.

Extent	Type	REF-ATR
[the president of the U.S.]	NOM.PER.Ind.SPC*	
[the U.S.]	NAM.GPE.Nation.SPC	

Analysts

Extent	Type	REF-ATR
[Analysts]	BAR.PER.Group.?	

IBM's lawyers

Extent	Type	REF-ATR
[IBM's lawyers]	NOM.PER.Group.?	E1
[IBM]	NAM.ORG.COM.SPC	E2-REF

The squad of Marines

Extent	Type	REF-ATR
[The squad of Marines]	NOM.ORG.GOV.SPC*	E1-REF*
[Marines]	BAR.ORG.GOV.SPC*	E1-REF*

The family

Extent	Type	REF-ATR
[The family]	NOM.PER.Group.SPC*	

The house painters

Extent	Type	REF-ATR
[The house painters]	NOM.PER.Group.SPC	
[house]	NOMPRE.FAC.Building.GEN*	

The linguists under the table

Extent	Type	REF-ATR
[The linguists under the table]	NOM.PER.Group.SPC	

 The Kennedys'

Extent	Type	REF-ATR
[The Kennedys]	NAM.PER.Group.SPC	

The Arabs

Extent	Type	REF-ATR
[The Arabs]	NOM.PER.Group.SPC*	

The Christians

Extent	Type	REF-ATR
[The Christians]	NOM.PER.Group.SPC*	

the Hmong individuals

Extent	Type	REF-ATR
[the Hmong individuals]	NOM.PER.Group.SPC*	E1-REF
[Hmong]	NAMPRE.PER.Group.GEN	E2-ATR

the ethnic Albanians

Extent	Type	REF-ATR
[the ethnic Albanians]	NOM.PER.Group.SPC*	

You want to go to the party

Extent	Type	REF-ATR
[You]	Pro.PER.INDET*.SPC*	

If you believe in god, you must…

Extent	Type	REF-ATR
[you]	PRO.PER.IND.USP*	E1-REF
[god]	NAM.PER.IND.SPC	E2-REF
[you]	PRO.PER.IND.USP*	E1-REF

Although he felt like he was a god, he…

Extent	Type	REF-ATR
[he]	PRO.PER.IND.SPC	E1-REF
[he]	PRO.PER.IND.SPC	E1-REF
[a god]	NOM.PER.IND.SPC	E1-ATR
[he]	PRO.PER.IND.SPC	E1-REF

She is running for President

Extent	Type	REF-ATR
[She]	PRO.PER.IND.SPC	E1-REF

Who will be the next President?

Extent	Type	REF-ATR
[Who]	WHQ.PER.IND.USP	E1-REF
[the next President]	NOM.PER.IND.USP	E1-ATR

Batman has become a popular icon

Extent	Type	REF-ATR
[Batman]	NAM.PER.Ind.SPC	E1-REF
[a popular icon]	NOM.PER.Ind.SPC	E1-ATR

Adam West's costume from Batman the TV series

Extent	Type	REF-ATR
[Adam West]	NAM.PER.Ind.SPC	

9.3.2: Organizations

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

Extent	Type	REF-ATR
[The Salzburg prosecutor's office]	NOM.ORG.GOV.SPC	E1-REF
[The Salzburg prosecutor]'s	NOM.PER.Ind.SPC	E2-REF
[Salzburg]	NAMPRE.GPE.PopCenter.SPC.GPE	E3-ATR

Putin, a former KGB agent, defended the court that convicted Pope and the security services,…

Extent	Type	REF-ATR
[Putin]	NAM.PER.Ind.SPC	E1-REF
[Putin, a former KGB agent]	APP.PER.Ind.SPC	E1-REF
[a former KGB agent]	NOM.PER.Ind.SPC	E1-ATR
[KGB]	NAMPRE.ORG.GOV.SPC	E2-ATR
[the court that convicted Pope and the security services]	NOM.ORG.GOV.SPC	E3-REF
[that]	WHQ.ORG.GOV.SPC	E3-REF
[Pope]	NAM.PER.Ind.SPC	E4-REF
[the security services]	NOM.ORG.GOV.SPC*	E5-REF

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

Extent	Type	REF-ATR
[The Financial Accounting Standards Board]	NAM.ORG.GOV.SPC	
[its]	PRO.ORG.GOV.SPC	
[Congress]	NOM.ORG.GOV.SPC	

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

Extent	Type	REF-ATR
[The US navy]	NAM.ORG.GOV.SPC	E1-REF
[Yemen]	NAM.GPE.Nation.SPC.LOC	E4-REF
[17]	HLS.PER.Group.SPC	E5

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

Extent	Type	REF-ATR
[Pope, who owns TechSource Marine Industries in State College, PA]	NAM.PER.Ind.SPC	E1-REF
[who]	WHQ.PER.Ind.SPC	E1-REF

[TechSource Marine Industries in State College, PA]	NAM.ORG.COM.SPC	E2-REF
[State College, PA]	NAM.GPE.PopCenter.SPC	E3-REF
[PA]	NAM.GPE.State.SPC	E4-REF
[an associate]	NOM.PER.Ind.SPC	E5-REF

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

Extent	Type	REF-ATR
[Pope, 54]	NAM.PER.IND.SPC	E1-REF
[Pope, 54, a retired U.S. Navy officer from State College, PA]	APP.PER.IND.SPC	E1-REF
[a retired U.S. Navy officer from State College, PA]	NOM.PER.IND.SPC	E1-ATR
[U.S. Navy]	NAMPRE.ORG.GOV.SPC	E2-ATR
[State College, PA]	NAM.GPE.PopCenter.SPC.GPE	E4-REF
[PA]	NAM.GPE.State.SPC.LOC	E5-REF
[the Applied Research Laboratory at Pennsylvania State University]	NAM.ORG.EDU.SPC	E6-REF
[Pennsylvania State University]	NAM.ORG.EDU.SPC	E7-REF

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

Extent	Type	REF-ATR
[NDSU and University of Minnesota weeds specialist *Alan Dexter*]	NAM.PER.IND.SPC	E1-REF
[NDSU and University of Minnesota weeds specialist]	NOMPRE.PER.IND.SPC	E1-ATR
[NDSU]	NAMPRE.ORG.EDU.SPC	E2-ATR
[University of Minnesota]	NAMPRE.ORG.EDU.SPC	E3-ATR

the Roundabout Theater Company is calling its new facility in Times Square the American

Airlines Theater

Extent	Type	REF-ATR
[the Roundabout Theater Company]	NAM.ORG.ENT.SPC	E1-REF
[its]	PRO.ORG.ENT.SPC	E1-REF
[its new facility in Times Square]	NOM.FAC.Building.SPC	E2-REF
[Times Square]	NAM.FAC.Building-Grounds.SPC	E3-REF
[the American Airlines Theater]	NAM.FAC.Building.SPC	E2-ATR

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

Extent	Type	REF-ATR
[the famous Irish group the Chieftains]	APP.ORG.ENT.SPC	E1-REF
[the famous Irish group]	NOM.ORG.ENT.SPC	E1-REF
[the Chieftains]	NAM.ORG.ENT.SPC	E1-ATR
[Irish]	NAMPRE.GPE.Nation.SPC	E2-ATR
[Alton]	NAM.ORG.ENT.SPC	E3-REF
[Solace]	NAM.ORG.ENT.SPC	E4-REF
[Celtic]	NAMPRE.PER.GROUP.GEN	E5-ATR

three of the [U.N]. workers stationed in East Timor

Extent	Type	REF-ATR
[three of the U.N. workers stationed in East Timor]	PTV.PER.GROUP.SPC*	E1-REF
[the U.N. workers stationed in East Timor]	NOM.PER.GROUP.SPC	E2-REF
[U.N]	NAMPRE.ORG.NON.SPC	E3-ATR
[East Timor]	NAM.GPE.Nation.SPC	E4-REF

the ``Caravan of Death,'' a military party that killed 73 political prisoners

Extent	Type	REF-ATR
[the ``Caravan of Death,'' a military party that killed 73 political prisoners]	APP.ORG.NON.SPC	E1-REF
[the ``Caravan of Death]	NAM.ORG.NON.SPC	E1-REF

[a military party that killed 73 political prisoners]	NOM. ORG.NON.SPC	E1-ATR
[that]	WHQ. ORG.NON.SPC	E1-REF
[military]	NOMPRE.ORG.GOV.GEN	E2-ATR
[73 political prisoners]	NOM.PER.GROUP.SPC	E3-REF

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

Extent	Type	REF-ATR
[the Rally for Congolese Democracy, one of the biggest rebel movements supported by Uganda]	APP.ORG.NON.SPC	E1-REF
[the Rally for Congolese Democracy]	NAM.ORG.NON.SPC	E1-REF
[one of the biggest rebel movements supported by Uganda]	PTV. ORG.NON.SPC	E1-ATR
[the biggest rebel movements supported by Uganda]	NOM.ORG.NON.SPC	E2-REF
[Uganda]	NAM.GPE.Nation.SPC.ORG	E3-REF

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

Extent	Type	REF-ATR
[The Red Cross]	NAM.ORG.NON.SPC	E1-REF
[about 15 people]	NOM.PER.GROUP.SPC	E2-REF
[at least three people]	NOM.PER.GROUP.SPC	E3-REF

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

Extent	Type	REF-ATR
[the Financial Accounting Standards Board, the private-sector body based in Norwalk, Conn., that sets the nation's accounting standards]	ARC.ORG.NON.SPC	E1-REF

[the <u>Financial Accounting Standards Board</u>]	NAM.ORG.NON.SPC	E1-REF
[the private-sector <u>body</u> based in Norwalk, Conn]	NOM.ORG.NON.SPC	E1-REF
[<u>that</u>]	WHQ.ORG.NON.SPC	E1-REF
[<u>Norwalk</u>, Conn]	NAM.GPE.PopCenter.SPC.LOC	E2-REF
[<u>Conn</u>]	NAM.GPE.State.SPC.LOC	E3-REF
[the <u>nation</u>]'s	NAM.GPE.Nation.SPC.GPE	E4-REF

Reuters international news agency

Extent	Type	REF-ATR
[Reuters international news <u>agency</u>]	NOM.ORG.MED.SPC	E1-REF
[<u>Reuters</u>]	NAMPRE.ORG.MED.SPC	E2-REF

Time magazine

Extent	Type	REF-ATR
[<u>Time</u> magazine]	NAM.ORG.MED.SPC	

abc news

Extent	Type	REF-ATR
[<u>abc news</u>]	NAM.ORG.MED.SPC	

Associated Press

Extent	Type	REF-ATR
[<u>Associated Press</u>]	NAM.ORG.MED.SPC	

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

Extent	Type	REF-ATR
[Archbishop <u>Khajag Barasamian</u>, head of the Diocese of the Armenian Church in America, whose headquarters are in Manhattan]	NAM.PER.IND.SPC	E1-REF

[head of the Diocese of the Armenian Church in America, whose headquarters are in Manhattan]	BAR.PER.IND.SPC	E1-ATR
[Archbishop]	NOMPRE.PER.IND.SPC	E1-ATR
[the Diocese of the Armenian Church in America, whose headquarters are in Manhattan]	NOM.ORG.REL.SPC	E2-REF
[the Armenian Church in America, whose headquarters are in Manhattan]	NAM.ORG.REL.SPC	E3-REF
[whose]	WHQ.ORG.REL.SPC	E3-REF
[America]	NAM.GPE.Nation.SPC.LOC	E4-REF
[whose headquarters]	NOM.FAC.Building-Grounds.SPC	E5-REF (maybe coref with E3)
[Manhattan]	NAM.LOC.Reg-Gen.SPC	E6

The Vatican

Extent	Type	REF-ATR
[The Vatican]	NAM.ORG*.REL.SPC	

[Islamic militants]

Extent	Type	REF-ATR
[Islamic militants]	BAR.PER*.GROUP	E1-REF
[Islamic]	NAMPRE.PER.GROUP.GEN*	E2-ATR

Overseas Chinese Physics Institute

Extent	Type	REF-ATR
[Overseas Chinese Physics Institute]	NAM.ORG.SCI.SPC	

medical and health groups

Extent	Type	REF-ATR

[medical and health groups]	BAR.ORG.SCI	

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

Extent	Type	REF-ATR
[A group of survivors belonging to a German ski club in Vilseck, Germany]	NOM.PER.GROUP.SPC*	E1-REF
[survivors belonging to a German ski club in Vilseck, Germany]	BAR.PER.GROUP.SPC*	E1-REF
[a German ski club in Vilseck, Germany]	NOM.ORG.SPO.SPC	E2-REF
[German]	NAMPRE.GPE.NATION.SPC.GPE	E3-ATR
[Vilseck, Germany]	NAM.GPE.PopCenter.SPC.LOC	E4-REF
[Germany]	NAM.GPE.NATION.SPC.LOC	E3-REF

Philippines Olympic Committee

Extent	Type	REF-ATR
[Philippines Olympic Committee]	NAM.ORG.SPO.SPC	

San Francisco 49ers

Extent	Type	REF-ATR
[San Francisco 49ers]	NAM.ORG.SPO.SPC	

9.3.3: Geographical/Social/Political Entities (GPE)

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

Extent	Type	REF-ATR
[North America]	NAM.GPE.CONTINENT.SPC	E1-REF
[Many people in North America]	NOM.PER.GROUP.USP	E2-REF

Hospital officials said all eight survivors were German.

Extent	Type	REF-ATR
[Hospital officials]	BAR.PER.GROUP.SPC*	E1-REF
[Hospital]	NOMPRE.ORG.SCI.SPC	E2-ATR
[all eight survivors]	NOM.PER.GROUP.SPC	E3-REF
[German]	NAMPRE.GPE.NATION.SPC.GPE	E4-ATR

…the conversion to Christianity of the Roman emperor Constantine…

Extent	Type	REF-ATR
[Roman]	NAMPRE.GPE.NATION.SPC.GPE	E1-ATR
[the Roman emperor Constantine]	APP.PER.IND.SPC	E2-REF
[the Roman emperor]	NOM.PER.IND.SPC	E2-REF
[Constantine]	NAM.PER.IND.SPC	E2-ATR

Salzburg governor Schausberger said…

Extent	Type	REF-ATR
[Salzburg governor Schausberger]	NAM.PER.IND.SPC	E1-REF
[Salzburg]	NAMPRE.GPE.STATE.SPC.GPE	E2-ATR
[Salzburg governor]	NOMPRE.PER.IND.SPC	E1-ATR

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

Extent	Type	REF-ATR
[Palm Beach]	NAMPRE.GPE.County.SPC.GPE	E1-ATR
[Volusia]	NAMPRE.GPE.County.SPC.GPE	E2-ATR
[Palm Beach and Volusia counties]	BAR.GPE.County.SPC.LOC	E3-REF

John Martin, ABC news, Washington

Extent	Type	REF-ATR
[John Martin, ABC news, Washington]	NAM.PER.IND.SPC	E1-REF
[ABC news]	NAM.ORG.MED.SPC	E2-REF
[Washington]	NAM.GPE.PopCenter.SPC.LOC	E3-REF

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

Extent	Type	REF-ATR
[women in New Delhi]	BAR.PER.GROUP.USP	
[New Delhi]	NAM.GPE.PopCenter.SPC.LOC	

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

Extent	Type	REF-ATR
[Norbert Karlsboeck, mayor of Kaprun, a town some 50 miles south of Salzburg in the central Austrian Alp]	APP.PER.IND.SPC	E1-REF
[Norbert Karlsboeck]	NAM.PER.IND.SPC	E1-REF
[mayor of Kaprun, a town some 50 miles south of Salzburg in the central Austrian Alp]	BAR.PER.IND.SPC	E1-ATR
[Kaprun, a town some 50 miles south of Salzburg in the central Austrian Alp]	APP.GPE.PopCenter.SPC.GPE	E2-REF
[Kaprun]	NAM.GPE.PopCenter.SPC.GPE	E2-REF
[a town some 50 miles south of Salzburg in the central Austrian Alp]	NOM.GPE.PopCenter.SPC.GPE	E2-ATR
[Salzburg]	NAM.GPE.PopCenter.SPC.LOC	E3-REF
[the central Austrian Alp]	NAM.LOC.Land-region-natural.SPC	E4-REF

the southern United States

Extent	Type	REF-ATR
[the southern United States]	NOM.LOC.REG-Gen.SPC	

NATO peacekeepers arrived in the valley before nightfall.

Extent	Type	REF-ATR
[NATO peacekeepers]	BAR.PER.GROUP.SPC	E1-REF
[NATO]	NAMPRE.ORG.NON.SPC	E2-ATR
[the valley]	NOM.LOC.Land-Region-Natural.SPC	E3-REF

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

Extent	Type	REF-ATR
[Ritter]	NAM.PER.IND.SPC	E1-REF
[Iraq]	NAM.GPE.NATION.SPC.ORG	E2-REF
[inspectors]	BAR.PER.GROUP.USP	E3-REF
[eight of Saddam Hussein's presidential palaces]	PTV.FAC.BUILDING.SPC	E4-REF
[Saddam Hussein's presidential palaces]	NOM.FAC.BUILDING.SPC	E5-REF
[Saddam Hussein]	NAM.PER.IND.SPC	E6-REF

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

Extent	Type	REF-ATR
[the Russian government]	NOM.GPE.NATION.SPC.ORG	E1-REF
[Russian]	NAMPRE.GPE.NATION.SPC.GPE	E1-ATR
[many politicians]	NOM.PER.GROUP.USP	E2-REF
[the United States]	NAM.GPE.NATION.SPC.ORG*	E3-REF
[they]	PRO.PER.GROUP.USP	E4-REF

New York's police department

Extent	Type	REF-ATR
[New York's police department]	NOM.ORG.GOV.SPC	
[New York]	NAM.GPE.Pop-Center*.SPC.ORG	
[police]	NOMPRE.ORG.GOV.GEN	

The coast of Britain

Extent	Type	REF-ATR
[The coast of Britain]	NOM.LOC.Land-reg-nat.SPC	
[Britain]	NAM.GPE.NATION.SPC.LOC	

U.S. warplanes flew over Afghanistan

Extent	Type	REF-ATR
[U.S]	NAMPRE.GPE.NATION.SPC.GPE	E2-ATR

[Afghanistan]	NAM.GPE.NATION.SPC.LOC	E3-REF

Northern portions of Montana

Extent	Type	REF-ATR
[Northern portions of Montana]	BAR.LOC.Reg-Gen.SPC*	
[Montana]	NAM.GPE.State.SPC.LOC	

Dallas, Texas announced its decision to build a subway system.

Extent	Type	REF-ATR
[Dallas, Texas]	NAM.GPE.PopCenter.SPC.ORG	
[Texas]	NAM.GPE.STATE.SPC.LOC	
[its]	NAM.GPE.PopCenter.SPC.ORG	
[a subway system]	NOM.FAC.PATH.USP	

Prime Minister of Britain

Extent	Type	REF-ATR
[Prime Minister of Britain]	BAR.PER.IND.SPC*	
[Britain]	NAM.GPE.NATION.SPC.GPE	

U.S. Commander-in-Chief

Extent	Type	REF-ATR
[U.S. Commander-in-Chief]	BAR.PER.IND.SPC*	E1-REF
[U.S]	NAMPRE.GPE.NATION.SPC.GPE	E2-ATR

U.S. surveillance aircraft

Extent	Type	REF-ATR
[U.S]	NAMPRE.GPE.NATION.SPC.GPE	E2-ATR

The Palestinian Authority has banned rallies that are pro-Iraq, but that ban has been widely ignored

Extent	Type	REF-ATR
[The Palestinian Authority]	NAM.ORG.GOV.SPC	E1-REF
[Iraq]	NAM.GPE.NATION.SPC.GPE	E2-ATR

Six days into the Nagano Games, one Alpine event the women's super-G won on Wednesday by Picabo Street of the United States has been completed.

Extent	Type	REF-ATR
[Nagano]	NAMPRE.GPE.PopCenter.SPC.GPE	E1-ATR
[women]	BAR.PER.GROUP.GEN	E2-REF
[Picabo Street of the United States]	NAM.PER.IND.SPC	E3-REF
[the United States]	NAM.GPE.NATION.SPC.GPE	E4-REF

"This is going to be a brutal fight," said Rep. Thomas C. Sawyer (D- 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.

Extent	Type	REF-ATR
[Rep. Thomas C. Sawyer (D- 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]	NAM.PER.IND.SPC	E1-REF
[Rep]	NOMPRE.PER.IND.SPC	E1-ATR
[D]	NAM.ORG.NON.SPC	E2-REF
[Ohio]	NAM.GPE.STATE.SPC.GPE	E3-REF
[who]	WHQ.PER.IND.SPC	E1-REF
[those who believe the ongoing debate played a role in Riche's departure]	HLS.PER.GROUP.SPC	E4-REF
[who]	WHQ.PER.GROUP.SPC	E4-REF
[Riche]	NAM.PER.IND.SPC	E5-REF

9.3.4: Locations

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

Extent	Type	REF-ATR

[Armenia]	NAM.GPE.NATION.SPC.LOC	
[a politically challenging <u>swath</u> of the southern Caucasus, with borders shared by Turkey, Azerbaijan, and Georgia]	NOM.LOC.REG-GEN.SPC	
[the southern <u>Caucasus</u>]	NAM.LOC.LAND-NAT.SPC	
[<u>borders</u> shared by Turkey, Azerbaijan, and Georgia]	BAR.LOC.BOUNDARY.SPC	
[<u>Turkey</u>]	NAM.GPE.NATION.SPC.LOC	
[<u>Azerbaijan</u>]	NAM.GPE.NATION.SPC.LOC	
[<u>Georgia</u>]	NAM.GPE.NATION.SPC.LOC	

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

Extent	Type	REF-ATR
[eclipse <u>fans</u>]	BAR.PER.GROUP.GEN	
[the <u>sun</u>]	NOM.LOC.CELESTIAL.SPC	

In Armenia, the three of them will join other, similar delegations from around the world…

Extent	Type	REF-ATR
[<u>Armenia</u>]	NAM.GPE.NATION.SPC.LOC	E1-REF
[the <u>world</u>]	NOM.LOC.CELESTIAL.SPC	E2-REF
[the <u>three</u> of them]	HLS.PER.GROUP.SPC	E3-REF
[the three of <u>them</u>]	PRO.PER.GROUP.SPC	E3-REF
[other, similar delegations from around the world]	NOM.PER.GROUP.SPC	E4-REF

In another part of the city,…

Extent	Type	REF-ATR
[another <u>part</u> of the city]	NOM.LOC.REG-GEN.SPC	
[the <u>city</u>]	NOM.GPE.PopCenter.SPC	

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

Extent	Type	REF-ATR
[Hans Wallner, the director	APP.PER.IND.SPC	

of the tourist region of Kaprun]		
[Hans Wallner]	NAM.PER.IND.SPC	
[the director of the tourist region of Kaprun]	NOM.PER.IND.SPC	
[the tourist region of Kaprun]	NOM.LOC.REG-GEN.SPC	
[tourist]	NOMPRE.PER.IND.GEN	

9.3.5: Facilities

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

Extent	Type	REF-ATR
[new york's la guardia airport]	NAM.FAC.AIRPORT.SPC	E1-REF
[new york]	NAM.GPE.PopCenter.SPC	E2-REF
[a nightmare]	NOM.FAC.AIRPORT.SPC	E1-ATR

…the train ran directly from the oil refinery to the smelter…

Extent	Type	REF-ATR
[the oil refinery]	NOM.FAC.PLANT.SPC	
[the smelter]	NOM.FAC.PLANT.SPC	

…at the national archives

Extent	Type	REF-ATR
[the national archives]	NAM.FAC.BUILDING.SPC	

…two men who rented an Aden apartment…

Extent	Type	REF-ATR
[two men who rented an Aden apartment]	NOM.PER.GROUP.SPC	
[who]	WHQ.PER.GROUP.SPC	
[an Aden apartment]	NOM.FAC.Subarea.SPC	
[Aden]	NAMPRE.GPE.PopCenter.SPC.GPE	

9.4: Entity Class

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

Extent	Type	REF-ATR
[no confirmed suspects]	NOM.PER.GROUP.NEG	E1-REF
[officials]	BAR.PER.GROUP.USP	E2-REF
[several Middle East groups]	NOM.ORG.NON.SPC	E3-REF
[Middle East]	NAMPRE.GPE.CLUSTER.SPC.GPE	E4-ATR

Lee Hawk Seder is Jerusalem Bureau Chief for the Washington Post

Extent	Type	REF-ATR
[Lee Hawk Seder]	NAM.PER.IND.SPC	E1-REF
[Jerusalem Bureau Chief for the Washington Post]	BAR.PER.IND.SPC	E1-ATR
[Jerusalem Bureau]	NOMPRE.ORG.MED.SPC	E2-ATR
[Jerusalem]	NAMPRE.GPE.PopCenter.SPC.GPE	E3-ATR
[the Washington Post]	NAM.ORG.MED.SPC	E4-ATR

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

Extent	Type	REF-ATR
[a naval force]	NOM.ORG.GOV.GEN	E1-REF
[Japan's equivalent of a naval force]	NOM.ORG.GOV.SPC	E2-REF
[Japan]	NAM.GPE.NATION.SPC.GPE	E3-REF
[the Japan Maritime Self-Defense Force]	NAM.ORG.GOV.SPC	E2-ATR

9.5: Mention Types/Mention Levels

9.5.1: Simple Mentions

I met some people yesterday who love chess.

Extent	Type	REF-ATR
[I]	PRO.PER.IND.SPC	E1-REF
[*some people yesterday who love chess*	NOM.PER.GROUP.SPC	E2-REF
[who]	WHQ.PER.GROUP.SPC	E2-REF

Rep. Joe Smith (R-KY)

Extent	Type	REF-ATR
[Rep]	NOMPRE.PER.IND.SPC	E1-ATR
[Rep. Joe Smith (R-KY	NAM.PER.IND.SPC	E1-REF
[R]	NAM.ORG.NON.SPC	E2-REF
[KY]	NAM.GPE.STATE.SPC	E3-REF

Bill Clinton and Jimmy Carter who are both former presidents

Extent	Type	REF-ATR
[Bill Clinton and Jimmy Carter who are both former presidents]	NAM.PER.IND.SPC	E1-REF
[Bill Clinton and Jimmy Carter who are both former presidents]	NAM.PER.IND.SPC	E2-REF
[who]	WHQ.PER.GROUP.SPC	E3-REF
[former presidents]	NOM.PER.GROUP.SPC	E3-ATR

The house of representatives

Extent	Type	REF-ATR
[The house of representatives]	NAM.ORG.GOV.SPC	

the Yemeni port of Aden

Extent	Type	REF-ATR
[the Yemeni port of Aden]	NOM.GPE.PopCenter.SPC	E1-REF
[Aden]	NAM.GPE.PopCenter.SPC	E1-REF

[Yemeni]	NAMPRE.GPE.NATION.SPC	E2-REF

a crowd of angry muslims

Extent	Type	REF-ATR
[a crowd of angry muslims]	NOM.PER.GROUP.SPC*	E1-REF
[angry muslims]	BAR.PER.GROUP.SPC*	E1-REF

the Russian navy, which waited several days before attempting to launch any rescue mission

Extent	Type	REF-ATR
[the Russian navy, which waited several days before attempting to launch any rescue mission]	NOM.ORG.GOV.SPC	E1-REF
[which]	WHQ.ORG.GOV.SPC	E1-REF
[Russian]	NAMPRE.GPE.NATION.SPC.GPE	E2-ATR

Who is the president of Brazil?

Extent	Type	REF-ATR
[Who]	WHQ.PER.IND.USP*	E1-REF
[the president of Brazil]	NOM.PER.IND.USP*	E1-ATR
[Brazil]	NAM.GPE.NATION.SPC.GPE	E2.REF

They will each pay $5

Extent	Type	REF-ATR
[They]	PRO.PER.GROUP.SPC*	E1-REF
[each]	HLS.PER.GROUP.SPC	E1-REF

Those who would disagree won't be happy.

Extent	Type	REF-ATR
[Those who would disagree]	HLS.PER.GROUP.USP	E1-REF
[who]	WHQ.PER.GROUP.USP	E1-REF

Those present at the meeting noticed.

Extent	Type	REF-ATR

[Those present at the meeting]	HLS.PER.GROUP.SPC	E1-REF

We must help those in need

Extent	Type	REF-ATR
[We]	PRO.PER.GROUP.USP*	E1-REF
[those in need]	HLS.PER.GROUP.GEN	E2-REF

two members of the team

Extent	Type	REF-ATR
[two members of the team]	NOM.PER.GROUP.SPC*	E1-REF
[the team]	NOM.PER.GROUP.SPC	E2-REF

the bottom of the ocean

Extent	Type	REF-ATR
[the bottom of the ocean]	NOM.LOC.REG-GEN.SPC	E1-REF
[the ocean]	NOM.LOC.WATER.SPC	E2-REF

some of the lawyers

Extent	Type	REF-ATR
[some of the lawyers]	PTV.PER.GROUP.SPC*	E1-REF
[the lawyers]	NOM.PER.GROUP.SPC	E2-REF

all of them

Extent	Type	REF-ATR
[all of them]	PTV.PER.GROUP.SPC	E1-REF
[them]	PRO.PER.GROUP.SPC	E2-REF

sixty percent of the participants

Extent	Type	REF-ATR
[sixty percent of the participants]	PTV.PER.GROUP.USP	E1-REF
[the participants]	NOM.PER.GROUP.SPC	E2-REF

9.5.2 Pre-modifier Mentions

Islamic militants

Extent	Type	REF-ATR
[Islamic militants]	BAR.PER.GROUP.SPC*	E1-REF
[Islamic]	NAMPRE.PER.GROUP.SPC	E2-ATR

Arabic movements

Extent	Type	REF-ATR
[Arabic]	NAMPRE.PER.GROUP.SPC	E1-ATR

democratic senators

Extent	Type	REF-ATR
[democratic senators]	BAR.PER.GROUP.SPC*	E1-REF
[democratic]	NAMPRE.ORG.NON.SPC	E2-ATR

Himalayan explorers

Extent	Type	REF-ATR
[Himalayan explorers]	BAR.PER.GROUP.SPC*	E1-REF
[Himalayan]	NAMPRE.LOC.NATURAL.SPC	E2-ATR

Republican voters

Extent	Type	REF-ATR
[Republican voters]	BAR.PER.GROUP.SPC*	E1-REF
[Republican]	NAMPRE.ORG*.NON*.SPC*	E2-ATR

rooftop decks

Extent	Type	REF-ATR
[rooftop decks]	BAR.FAC.SUB.GEN*	E1-REF
[rooftop]	NOMPRE.FAC.SUB.GEN	E2-ATR

airline regulators

Extent	Type	REF-ATR
[airline regulators]	BAR.PER.GROUP.GEN*	E1-REF

[airline]	NOMPRE.ORG.COM.GEN	E2-ATR

union leaders

Extent	Type	REF-ATR
[union leaders]	BAR.PER.GROUP.SPC*	E1-REF
[union]	NOMPRE.ORG.NON.SPC*	E2-ATR

9.5.3 Complex Constructions

Bill, John's lawyer

Extent	Type	REF-ATR
[Bill, John's lawyer]	APP.PER.IND.SPC	E1-REF
[Bill]	NAM.PER.IND.SPC	E1-REF
[John's lawyer]	NOM.PER.IND.SPC	E1-ATR
[John]	NAM.PER.IND.SPC	E2-REF

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

Extent	Type	REF-ATR
[Mr. Black, 58, a victim of the terrorist assault]	APP.PER.IND.SPC	E1-REF
[Mr. Black, 58]	NAM.PER.IND.SPC	E1-REF
[a victim of the terrorist assault]	NOM.PER.IND.SPC	E1-ATR
[Mr]	NOMPRE.PER.IND.SPC	E1-ATR
[terrorist]	NOMPRE.PER.IND*.SPC	E2-ATR
[the Associated Press]	NAM.ORG.MED.SPC	E3-REF

the show's production company, Celador

Extent	Type	REF-ATR
[the show's production company, Celador]	APP.ORG.ENT.SPC	E1-REF
[the show's production company]	NOM.ORG.ENT.SPC	E1-REF
[Celador]	NAM.ORG.ENT.SPC	E1-ATR

We Richmonders

Extent	Type	REF-ATR
[We Richmonders]	APP.PER.GROUP.SPC*	E1-REF
[We]	PRO.PER.GROUP.SPC*	E1-REF
[Richmonders]	NAM.PER.GROUP.SPC*	E1-ATR

Pittsburgh, a US airways hub recently paid to revamp the airport including shopping malls in the terminals

Extent	Type	REF-ATR
[*Pittsburgh, a US airways hub*]	APP.GPE.PopCenter.SPC.ORG	E1-REF
[*Pittsburgh*]	NAM.GPE.PopCenter.SPC.ORG	E1-REF
[*Pittsburgh, a US airways hub*]	NOM.GPE.PopCenter.SPC.ORG	E1-ATR
[US airways]	NAMPRE.ORG.COM.SPC	E2-REF
[the airport]	NOM.FAC.Airport.SPC	E3-REF
[shopping malls in the terminals]	BAR.FAC.Building-Grounds.SPC	E4-REF
[the terminals]	FAC.FAC.Subarea.SPC	E5-REF

The American Civil Liberties Union (ACLU)

Extent	Type	REF-ATR
[The American Civil Liberties Union (ACLU]	NAM.ORG.NON.SPC	E1-REF
[ACLU]	NAM.ORG.NON.SPC	E1-REF

He, himself had known it was true.

Extent	Type	REF-ATR
[He, himself]	PRO.PER.IND.SPC	E1-REF
[himself]	PRO.PER.IND.SPC	E1-REF

Israeli newspaper *Ha'aretz.*

Extent	Type	REF-ATR
[Israeli newspaper Ha'aretz]	NAM.ORG.MED.SPC	E1-REF

[Israeli newspaper]	NOMPRE.ORG.MED.SPC	E1-ATR
[Israeli]	NAMPRE.GPE.NATION.SPC.GPE	E2-ATR

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

Extent	Type	REF-ATR
[The former company CEO, John Fisher, who fired the board of directors]	ARC.PER.IND.SPC	E1-REF
[The former company CEO]	NOM.PER.IND.SPC	E1-REF
[John Fisher]	NAM.PER.IND.SPC	E1-ATR
[who]	WHQ.PER.IND.SPC	E1-REF
[company]	NOMPRE.ORG.COM.SPC	E2-ATR
[the board of directors]	NOM.ORG.COM.SPC	E3-REF
[directors]	BAR.PER.GROUP.SPC*	E4-REF

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

Extent	Type	REF-ATR
[John Richards, the party leader who attended the DNC convention]	ARC.PER.IND.SPC	E1-REF
[John Richards]	NAM.PER.IND.SPC	E1-REF
[the party leader]	NOM.PER.IND.SPC	E1-ATR
[who]	WHQ.PER.IND.SPC	E1-REF
[party]	NOMPRE.ORG.NON.SPC	E2-ATR
[DNC]	NAMPRE.ORG.NON.SPC	E3-ATR

9.5.4 Referential and Attributive Mentions

John is a lawyer.

Extent	Type	REF-ATR
[John]	NAM.PER.IND.SPC	E1-REF
[a lawyer]	NOM.PER.IND.SPC	E1-ATR

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

Extent	Type	REF-ATR
[Richard Zolinger]	NAM.PER.IND.SPC	E1-REF
[a weed specialist in North Dakota State University]	NOM.PER.IND.SPC	E1-ATR
[part of an International Monitoring Network]	NOM.PER.IND.SPC	E1-ATR
[North Dakota State University]	NAM.ORG.EDU.SPC	E2-REF
[an International Monitoring Network]	NOM.ORG.NON.SPC	E3-REF

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

Extent	Type	REF-ATR
[David Lule, the head of light entertainment for Pearson Television, one of the largest international distributors of programming]	APP.PER.IND.SPC	E1-REF
[David Lule]	NAM.PER.IND.SPC	E1-REF
[the head of light entertainment for Pearson Television, one of the largest international distributors of programming]	NOM.PER.IND.SPC	E1-ATR
[Pearson Television, one of the largest international distributors of programming]	APP.ORG.COM.SPC	E2-REF
[Pearson Television]	NAM.ORG.COM.SPC	E2-REF
[one of the largest international distributors of programming]	PTV.ORG.COM.SPC	E2-ATR
[the largest international distributors of programming]	NOM.ORG.COM.SPC	E3-REF

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

Extent	Type	REF-ATR

[No one]	PRO.PER.Indet.NEG	E1-REF
[our]	PRO.GPE*.NATION.SPC.ORG*	E2-REF
[a force for peace and for stability]	NOM.GPE*.NATION.SPC.ORG*	E2-ATR

I think that Hillary Clinton will be the future president.

Extent	Type	REF-ATR
[I]	PRO.PER.IND.SPC	E1-REF
[Hillary Clinton]	NAM.PER.IND.SPC	E2-REF
[the future president]	NOM.PER.IND.SPC	E2-ATR

Hillary Clinton may be [the future president]. (Co-referenced)

Bush said that Hezbollah is [a terrorist organization]. (Co-referenced)

Abbas said that Hezbollah is not [a terrorist organization]. (NOT co-referenced)

[That man] is a teacher.

[Israeli troops] are well-trained.

[I] need [five people] to go to the conference.

[The United States] is a violent country.