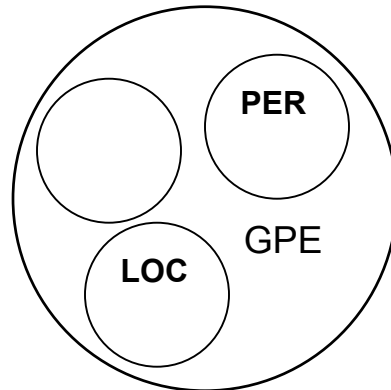organization. In these cases, annotators should mark first person plural pronouns as Organization mentions, and not as Person mentions.

3.3 Geographical/Social/Political Entities (GPE)

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

In this sentence,

 The people of France welcomed the agreement.

there are two mentions

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

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

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

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

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

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

3.3.1 Subtypes for GPEs

We will further classify GPE entities with the following subtypes. 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.*

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…*

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 <u>West</u>].

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, and Native American reservations.

3.3.2 GPE-like Locations and Organizations

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

[the southern <u>United States</u>]

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

[<u>NATO</u>] 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
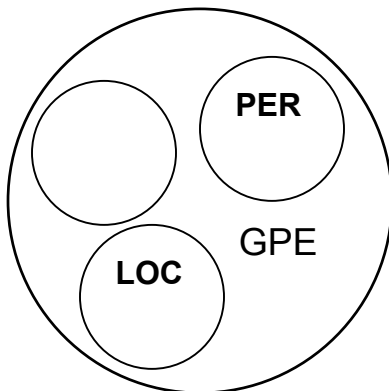
[<u>Dallas</u>, Texas]

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

GPEs and Government Organizations

GPEs modifying government organizations, like *New York police department* and *Kentucky fire 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]*** *Switzerland}'s supreme court.*

Note that the GPE mention is a possessive in this example and not a premodifier. All GPE premodifiers will be assigned the role GPE as explained below.

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] Israeli} troops*
> *{[GPE.GPE] New York} policemen*
> *Prime Minister of {[GPE.GPE] Britain}*
> *{[GPE.GPE] New York} attorney*
> *{[GPE.GPE] U.S.} Commander-in-Chief*
> *{[GPE.GPE] U.S.} surveillance aircraft*
> *{[GPE.GPE] Iraqi} flag*
> *The {[GPE.GPE] California} company*

Military Activity

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

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

Activities Associated with GPEs

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

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

Athletes, Sports Teams, and GPEs

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

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

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

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

Political associations

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

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

3. 4 Locations

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 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]…

3.4.2 Sub-parts of Locations and GPEs

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

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

3.4.3 Non-Locations

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

"Your coat is under the dog."

"The rabbit is hiding behind that rock."

"I have an idea in my head."

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

"He dropped the logs on the ground."

"He put the lamp back in its place."

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

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

3.5 Facilities

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

3.5.1 Subtypes for Facilities

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

Airport

A facility whose primary use is as an airport.

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

Plant

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

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

Building-or-Grounds

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

> *…at [the national archives].*
>
> *[The Berlin Wall]*
>
> *the parades at [Disneyland]*

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 apartment]…*

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…*

3.6 Vehicle

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

3.6.1 Subtypes for Vehicles

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

Air

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

> *…[military <u>helicopters</u>] float overhead…*

Land

Vehicles designed to locomote primarily upon land.

> *[The <u>train</u>] is reported to be about 600 yards inside the mountain.*

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

Water

Vehicles designed to locomote primarily on or submerged in water.

> *…the Oct. 12th bombing of [the <u>U.S.S. Cole</u>].*

Subarea-Vehicle

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

> *...[the <u>compartment</u> of the train]…*