ACE (Automatic Content Extraction)
Arabic Annotation Guidelines for Entities

Version 6.0 2006.06.15

Linguistic Data Consortium

http://www.ldc.upenn.edu/Projects/ACE/
1 Introduction ........................................................................................................4
  Basic Concepts ..................................................................................................4
2 Text to Annotate.................................................................................................5
3 Entity Types and Subtypes ................................................................................5
  3.1 Persons (PER) ............................................................................................5
    3.1.1 Subtypes for Person ............................................................................6
    3.1.2 Titles, Honorifics, and Positions ...........................................................7
    3.1.3 Fictional characters, names of animals, and names of fictional animals .......................................................................................................................8
  3.2 Organizations ..............................................................................................9
    3.2.1 Subtypes for Organizations ...................................................................9
    3.2.2 Organization Subtype Trumping Rules ...............................................15
    3.2.3 Organization Entities used in Person Contexts ......................................16
    3.2.4 First Person Pronouns Referring to Organizations .............................16
  3.3 Geographical/Social/Political Entities (GPE) .............................................16
    3.3.1 Subtypes for GPEs .............................................................................17
    3.3.2 GPE-like Locations and Organizations ...............................................20
    3.3.3 Formulaic GPE Constructions: Nested Region Names .......................21
    3.3.4 GPE Mention Roles ...........................................................................21
  3.4 Locations ..................................................................................................26
    3.4.1 Subtypes for Locations .......................................................................26
    3.4.2 Sub-parts of Locations and GPEs.......................................................28
    3.4.3 Non-Locations ...................................................................................28
  3.5 Facilities ....................................................................................................29
    3.5.1 Subtypes for Facilities.........................................................................29
  3.6 Vehicle ......................................................................................................31
    3.6.1 Subtypes for Vehicles .........................................................................31
  3.7 Weapon .....................................................................................................33
    3.7.1 Subtypes for Weapons........................................................................33
4 Entity Class ......................................................................................................36
  4.1 Negatively Quantified (NEG) .....................................................................36
  4.2 Non-referential/Attributive/Ascriptive (ATR) ...............................................37
  4.3 Specific Referential (SPC) ..........................................................................38
  4.4 Generic Referential (GEN) ........................................................................38
  4.5 Under-specified Referential (USP) ..............................................................39
5 Mention Types/Mention Levels ........................................................................40
  5.1 Simple Mentions ........................................................................................41
    5.1.1 Mention Extent ....................................................................................41
    5.1.2 Mention Head .....................................................................................42
    5.1.3 Names (NAM) .....................................................................................43
    5.1.4 Nominal Constructions (NOM) ............................................................43
    5.1.5 Pronouns (PRO) .................................................................................44
    5.1.6 WH-Question Words and Specifiers (WHQ) .......................................44
5.1.7 Headless Mentions (HLS) ................................................................. 45
5.1.8 Partitive Constructions (PTV) ......................................................... 45
5.1.9 Postmodifier Mentions (NAMPRE, NOMPRE) ................................. 45
5.2 Complex Constructions ................................................................... 47
  5.2.1 Appositive Constructions (APP) .................................................. 47
  5.2.2 Complex Constructions taking a Relative Clause (ARC) ............... 48
6 Nickname Metonymy ........................................................................... 48
  6.1 Capital City or Government Seat (FAC) Names standing in for Country’s Government .......................................................... 49
  6.2 City name for Sports Team ............................................................... 49
7 Cross-Type Metonymy ........................................................................ 49
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:**

جو سميث

**Joe Smith**

**Nominal Mention:**

[ The guy wearing a blue shirt ]

**Pronoun Mentions:**

هو

He

Entities are limited to the following seven types:

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

2 Text to Annotate

Only material between <TEXT> and </TEXT> tags is to be annotated. In newswire documents, material in headlines and slug sections is not to be tagged. In broadcast news, only the transcribed speech is to be tagged; added information, such as that within <TURN> tags or speaker identification tags, is not to be tagged.

3 Entity Types and Subtypes

3.1 Persons (PER)

Each distinct person or set of people mentioned in a document refers to an entity of type Person. For example, people may be specified by name ("John Smith"), occupation ("the butcher"), family relation ("dad"), pronoun ("he"), etc., or by some combination of these. Dead people and human remains are to be recorded as entities of type Person. 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.

[التعليم السياسي لهذا العام]
[The political cat of the year]

[النشرة التركية] في العراق
The Turkish forces attacked Iraq

[أعضاء هيئة التدريس]
[The staff]

[بعض الطلاب]
[Some students]
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.

- Young man aged 21
- Major General Hussein Tantawy

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

- Al Manar lawyers
- The prisoners in Israeli prisons
- Two were killed yesterday in an accident in Amman
- Two Egyptians were injured by the Israeli army shooting
- [The Arabs] NOM-PER-Group
- [The Muslims] NOM-PER-Group
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. We are not expecting to see such subtypes in Arabic.

3.1.2 Titles, Honorifics, and Positions

In English, as well as in Arabic, 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. The two person mentions are co referential, but the title Secretary of State is attributive while the proper name Madeline Albright is tagged as Specific.

[الرئيس الاسد]
[President Al-Asad]

<table>
<thead>
<tr>
<th>Secretary of State</th>
<th>[وزارة الخارجية الأمريكية مادلين اوبرایت]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Secretary of State Madeline Albright]</td>
<td>PER Atr.</td>
</tr>
<tr>
<td>ظهير</td>
<td>[وزارة الخارجية الأمريكية مادلين اوبرایت]</td>
</tr>
</tbody>
</table>

Saints and other religious figures

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

[[The prophet] Moses, peace be upon him]

<table>
<thead>
<tr>
<th>The prophet</th>
<th>[نفي موسى عليه الصلاة وسلم]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[The prophet] NOM.PER.Indiv.ATR</td>
<td></td>
</tr>
</tbody>
</table>
The prophet [Moses], peace be upon him

NAM.PER.Indiv.SPC

The catholic church is discussing considering Pope John Paul the second as one of the saints

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]… nominal mention

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

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

[Batman] has become a popular icon

A lot of kids love Bakar
Bokloz teaches the kids how to behave

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.

الفيل حيوان البف
الأسد ملك الغابة

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 Supreme Court] in Washington] agreed to discuss the appeal

اعتقلت [شرطة] السعودية [ارهابيين]
[[The Saudi [police] arrested two terrorists]

أغلقت الولايات المتحدة [سفارتها] في الدوحة]
The United States closed [[it’s embassy] in Duha]

قامت [شرطة] المصرية [بتعقب فلول الإرهابيين]
The Egyptian police is looking for the terrorists

اعتمدت [الخارجية] اللبنانية [ثلاث سفراء جديد]
The Lebanese ministry of exterior approved three new ambassadors yesterday

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.

LG Electronics declared that it is expecting to sell in Moraco’s market with more than $36 millions this year

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.

The Lebanese Universities have branches of European and American Universities

The Egyptian schools are crowded with students

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.

[Miamy group] signed an agreement with Rotana production

[Naguib Al Rihany theater group] presented the greatest comedy shows

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:

   [كتاب شهداء الأقصى]  
   [Al Aqsa Martyr's Brigade]

2. Political Parties:

   [الحزب الجمهوري]  
   [Republican Party]

   [حزب العمل]  
   [Labour Party]

3. Political Advocacy Groups and Think Tanks:

   [الاتحاد الأمريكي للحريات المدنية]  
   [ACLU]

   [معهد كاتو]  
   [The Cato Institute]

4. Professional Regulatory and Advocacy Groups:

   [الجمعية المحاماة الأمريكية]  
   [The American Bar Association]

   [الجمعية الطبية الأمريكية]  
   [The American Medical Association]

   [نقابة المهندسين]  
   [The engineering syndicate]

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

The conflict between [Taliban] and the Afghan protestor is running few kilometers from the borders of this former Sovit Republic in Asia.

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.

Millions of Arab watch [LBC and ART channels]

[ Gizaera channel ]

[Al Hayal newspaper]

[New York times]

[VOA from Washington]

Religious (REL)

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

[The German Bishop Union]

[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.

وتقدم [العيادات البرازيلية] خدماتها في الخارج وتسعى إلى جذب أكبر عدد من النساء الأجنبية إلى البرازيل للاختصاصات 해외 العمليات الجماعية.

[The Brazilizn clinics] offering services abroad and trying to attract the biggest number of female foreigners to operate plastic surgeries

The physicians face serious injuries and desasters on a daily bases and short in number of employee in the hospitals

[The Palestinian hospitals] are suffering of serious problems, as the physicians face serious injuries and desasters on a daily bases and short in number of employee in the hospitals

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.

[Al Negma] won Shabab Al Sahel yesterday in Lebanese tournament

[Milan ] reached the semifinal in soccer
3.2.2 Organization Subtype Trumping Rules

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

Less Specific    More Specific

```
<table>
<thead>
<tr>
<th>REL</th>
<th>GOV</th>
<th>NonGov</th>
<th>EDU</th>
<th>SCI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COM</td>
<td></td>
<td>ENT</td>
<td>SPO</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MED</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

Exceptions to Trumping Rules

**GPE military**

The military organizations connected to a GPE’s government will be tagged Government (GOV), for example: US Air Force.

**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.

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

In this sentence,
The people of France welcomed the agreement.

There are two mentions:

[The people of France]  GPE
[France]  GPE

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

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

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


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. Please refer to 6.2 for more discussions.

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.

The report included names of 24 countries mostly from [Asia], Latin America, and Southern America

[Europe] is trying to unite
**Nation**
Taggable mentions of the entireties of any nation.

[Saudi] agreed to wave the Yemeni dept

Two Egyptians were injured by the [Israeli] army

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

[Cairo Governorate]

Two more accidents in [Kalyoubia Governorate] north of Cairo where a man and four women died

[Salzburg] governor Schausberger said…

The liberty agreement was signed in [Pennsylvania]

**County-or-District**
Taggable mentions of the entireties of any county, district, prefecture, or analogous body of any state/province/canton. Usually a County-or-District is bigger than a city, but smaller than a State-or-Province. Note in a lot of Arabic countries, there is not such infrastructure. You can find such mentions in files which reports news in other countries though.

Miami County

Los Angeles County

[Golan Heights]
Population-Center
Taggable mentions of the entireties of any GPE below the level of County-or-District, including cities, villages in Arabic countries. Units that are smaller than villages are considered as LOC-Region-General. Please refer below for more information.

The ambassadors met at the [Pakistani capital]

At least ten persons were injured in [Ashmoun city] in road accident

The International Institute for Strategic Studies in [London] declared that international weapon tradings os slowed down

The American Embassy's spokesman declared Tuesday that American secretary of states Madelyn Olbright will arrive to [Rhyad] later

It seems that Christmas celebration in [Jerusalem] where Jesus was born, will be sad event this year

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

Eastern Europe

Western Europe

the European Union

the Middle East

Southeast Asia
The Gulf should be tagged as a Location/Region /natural when it refers to the sea, but as GPE/ Clusters when referring to countries around the Gulf.

**Special**
A closed set of GPEs for which the conventional labels do not straightforwardly apply.

The [Palestinian] police arrested Palestinian suspect that was dealing with Israel.

The following entities are also tagged as GPE /Special

Northern Ireland
Taiwan

**3.3.2 GPE-like Locations and Organizations**
Incidental, non-political clusters of GPEs should be marked Location. Please refer to 3.4 for more discussion on this.

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 "Provo, Utah" evokes one entity for each region. Thus "Provo, Utah" evokes one entity for the population center (with mention "Provo, Utah") and a second one for the state (with mention "Utah"). In Arabic, due to the absence of punctuation, there are several variants of this structure:

\[
\text{طربلس، لبنان} \quad \left[ \begin{array}{c} \text{Trabols, Libanon} \\ \text{القاهرة، مصر} \end{array} \right] \quad \text{Cairo, Egypt}
\]

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.

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).

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

\[
\begin{array}{c}
\text{ORG} \\
\text{PER} \\
\text{LOC} \\
\text{GPE}
\end{array}
\]

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

Somalese officials declared that [their government] made an agreement with Ethiopia, the former enemy, to work together

Governments

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

But [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. In Arabic however, a parallel structure of Switzerland’s supreme court “is literally "court the switzerland the supreme" where Switzerland” is a postmodifier which is tagged as GPE.GPE. Below has a detailed discussion on this issue.

The department said Sonabend can appeal to {[GPE.ORG] Switzerland}’s supreme court. (possessive)
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.

[The Japanese] have a considerable responsibility for the wars of the first half of the century

In this example, the phrase the Japanese may be interpreted as the population of Japan, or the government of Japan, or the Japanese military, or even some part of the Japanese population. If the annotator believes that the phrase in question refers to the population of the GPE, or most of the population of a GPE, then the annotation should be GPE.PER and the mention is a name mention. However, if the annotator believes the phrase refers to a group of people, then PER is the assigned annotation and the mention is nominal because it does not refer to the name of a person. Examples:

Butler says those sanctions could end soon if [the Iraqis] allow the inspectors to do their job.

The Missouri will come to rest near the memorial for the USS Arizona, which was sunk by [the Japanese] during the attack on Pearl Harbor.

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

Our reporter in our regional office in [Asia], Elishary, said that dismissing president Astradad may take six months.

Declared in [Riyadh] city in Saudi Arabia that 3 British were injured in an explosion in their car.
No one declared his responsibility on any of the two accidents that happened after one week of throwing a bomb on the building of the American Embassy in Sanaa.

France has an area of 547,090 square kilometers.

U.S. warplanes flew over Afghanistan.

In nested mentions of the form [child].[parent], the parent GPE always takes a LOC role; the child's role depends on context. Please refer to 3.3.3 for detailed discussion.

Dateline mentions of GPEs are given a location role.

Beirut – Lebanon 1 - 5 – 2005

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 Postmodifiers (Premodifiers in English)

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

Israel] troops

New York] policemen

British ] Prime Minister
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 [Soviet Union] invaded 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-[Iraq], but that ban has been widely ignored.

Athletes, Sports Teams, and GPEs
Athletes and teams are associated with GPE.GPEs as in example below.

Lance Amstrong of the United States has won the Tour of Fance 7 times.

However, when a GPE name is used as a team name (as in Boston beat Philly), the entity is marked as a Nickname Metonymy. Please refer to 6.2 for further discussion.

[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 [Ohio],
Magdy El Sayad National-[Beny Sweif] said that the government will support low income people

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.

17 Foad St.

Boundary
A one-dimensional location such as a border between GPE’s or other locations.

The Israeli army declared Friday evening that he stopped terrorist group from the Lebanese [borders]

Celestial
A location which is otherworldly or entire-world-inclusive The rule of thumb is to tag world, earth, globe in addition to all other planets as LOC celestial

Arabic-Entities-Guidelines.doc V6
2006.06.08
26
Scientists said that [Mars] may began as lakes after discovery of layers of rocks similar to those on [earth]

The space ship moves in a very high speed around the [earth]

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

Egyptian Saudi air and sea military trainings began yesterday above the [red sea]

Land-Region-natural
Geologically or ecosystemically designated, non-artificial locations.

170 were killed after fire in train carrying scaters to a winter resort in [Alp mountains]

Region-International
Taggable locations that cross national borders.

Travelling [abroad] of the officials in Taliban up to ministers is banned

Region-General
Taggable locations that do not cross national borders.

The Russian leader in the area said last week that his forces will change the strategy and will leave the bases that secure her and will put small units [all over the area] instead

Thursday afternoon an attack with weapons on their cars in northern [West Bank]

Travelling [abroad] of the officials in Taliban up to ministers is banned

The Russian leader in the area said last week that his forces will change the strategy and will leave the bases that secure her and will put small units [all over the area] instead

Thursday afternoon an attack with weapons on their cars in northern [West Bank]
At least twelve persons were killed in an attack in [the Indian area of Kashmir].

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.

In Arabic there are two structures of this kind of mentions that should be treated differently in annotation.

1) The Sudan the southern where the southern is an adjective

2) South-genetive Sudan where south is a nominal (Arabic Translation needed)

1) is rare in Arabic. When you do see this structure, it should be annotated like: [the Sudan the southern] where Sudan is the head of a NOM mention and is of LOC-Region-General; in 2)When we mean south of Sudan Republic, there are two taggable mentions: [South-genetive Sudan] where south is head (NOM-LOC-Region-General) and [Sudan] (NAM-GPE-Nation)., but 3) When the text mean the southern part of Sudan republic as In the following example we have only one mention "Sudan" for the whole extent. It will be Nom LOC

[South the Sudan]

The Arabic construction of 20 and 3) are exactly the same in Arabic, but it could be understood from the text in most of cases. In some other cases it is too hard to distinguish between them.

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:

"معطفك تحت الكلب"
"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.

Hundreds of Palestinians were [hospitalized] in Israeli hospitals for treatment

The Israeli army ordered the Palestinians to move out of the [offices of the common communication in West Bank and Ghaza strip]

Saudi business man is trying to avoid his delivery to USA charged by exploding [the two American Embassies]

56 were arrested in the early morning from [their houses]

The relatives of the victim and the injured tried to fire [the gas station in the city]

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

The first session of the trial was attended by the lawyers in [the 5th. Charge room in the military court]
An artist was able to hear the witnesses in Montreal from his jail in Seattle through the closed circuit cameras.

**Path**

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

The Israeli forces was preventing the people from moving around the main streets and the Israeli forces was preventing the people from moving around the main streets.

[Telephone lines] 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 helicopters] float overhead...

The former Pakistani Prime minister Nawaz Sherif arrived to Gedda in a private Saudi plane.

A group of injured Palestinians arrived to Iraq on a Palestinian plane for treatment in the Iraqi hospitals.
**Land**
Vehicles designed to locomote primarily upon land.

The military broadcast said that [the car] driver was not injured, but he was in shock.

The Israeli troops attacked the Lebanese harbour by [an amphibian truck] and the driver was not injured, but he was in shock.

The armed Greece guy that was holding 35 persons in bus between Athens and Corinthia came out of [the bus] alone.

A driver of [truck loaded with construction materials] was injured on Wednesday by Israeli shooting while he was passing through Fatma’s Gate.

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

We expect to finish constructing [the submarines] in Southern Korea in 2009.

Kiewit behaved seriously when it holed [5 boats] suspected to transport Iraqi goods.

**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.

[compartment of the train]...

[compartment of the train]...
Underspecified
Vehicles whose subtype is not specified in the text, or sets of vehicles of different subtypes.

We have [202,000 vehicles] in our fleet

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

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

Blunt
Blunt weapons are those designed or used as bludgeoning instruments

The Israeli forces killed three of the [rock] shooting Palestinians by shooting them in West Bank and Gaza strip

The game was stopped for 15 min after throwing a [plastic stick] towards Milan’s goal keeper

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

A bomb exploded close to Jewish settlement in Gaza

The Indian police forces in Kashmir launched the tearing gas bombs on the demonstrators
Sharp
Sharp weapons are those designed or used to cut, slash, jab, & hack

The protestor Deryoush Farouhar and his wife were killed by stabbing with a knife.

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

[Sarin gas]

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

The letter was contaminated with white powder that proved to contain anthrax.

Shooting
Shooting weapons are weapons designed or used to send projectile objects at great speed for the purpose of causing harm.

The guy came out of the bus without the gun that threatened the hostages with

The officers were able to bring in piston into the plane, and the plane was hijacked with this piston.

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

An official in the hospital said that Moayad reached the hospital injured with bullet in the head.
Tehran declared that she has self satisfaction from [the missiles land-sea]

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

[Crude missiles]
[Nuclear missiles]
[Hydrogen bombs]
[Dirty bombs]

Officials in the American army declared that a fire destroyed base supporting [the nuclear missiles] belonging to the army in North Dakota

and said, we support Russia to confirm slow down the level of [the strategic nuclear weapons]

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

They searched him for [a weapon].

A number of Palestinians were injured by the Israeli troop’s fire, through conflicts and attacks with [weapons] in different places.
4 Entity Class

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

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

A mention is attributive if the mention (a) states a property or properties about an entity referenced by another mention within the same sentence – often as an appositive to or part of a predicate on the other mention – or (b) qualifies an entity through immediate modification within the same phrase.

Referential mentions are further divided into generic and non-generic classes. A generic mention refers to a class/kind/species of objects or a typical representative of that class/kind/species and does not point to or pick out any specific individual object(s) of that class/kind/species. So if any property predicates on a generic mention, it means the entire class referred to by the mention has that property, or all/most/any members of that class have the property.

A non-generic referential mention refers to one or more individual member entities of a particular class. The entity or entities can be accounted for by pointing (specific) or cannot be precisely accounted for (underspecified).

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

4.1 Negatively Quantified (NEG)

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

No sensible lawyer would take that case.

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

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

They are not [lawyers].

The class of ‘lawyers’ in this example would be ATR.

**4.2 Non-referential/Attributive/Ascriptive (ATR)**

An entity is ATR when it is not being used to refer, but rather to attribute some property or attribute to some entity. The titles (3.1.2), the nominal mention in apposition (5.4.1) and some postmodifiers (5.1.9) all fall into this class. Note that not all postmodifiers are tagged as ATR. For example, in

[[The Israeli] army]

Israeli is an ATR; but in

[[The Arab] summit]

Arab is consideres as a SPC. The difference between these two structures is that in “the Army the Israeli”, both mentions have determiners (the), while in “summit the Arab” which is an Edafa, only the postmodifier has determiner (the). In the following examples where the title is followed by proper name, the title is annotated as ATR.

[[President Al-Asad]]

[The chemist Samir ]

If the title was followed by an adjective, it is annotated as SPC, as shown in the following examples:

[[The Egyptian president]]

[الرئيس المصري]
[The smart physician]
[الطبيب المهمل]
[The careless student]

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.

[France] produces better wine than New Jersey.

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

[Three Israeli soldiers]

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 or class 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. In fact, the subject NP in all the following examples has a generic reading:

[The lawyers] are honest

[The engineers are creative]

[The nurses take care of the patients]

What is common to these examples is the predicate: its a kind-level predicate, meaning the predicate describes a property of a kind of entities. Individual- and species-level predicates are also triggers:
Non-subject NP's are harder to decide. Again their interpretation is dependent on their interaction with the predicate that contains them:

Here are some Arabic examples of Generic mentions:

[المعرضات] يجب أن يكونوا عطفين

[The nurses] should be kind

[القادة] يتمتعون بشخصيات قوية

[The leaders] have strong personality

4.5 Under-specified Referential (USP)

We reserve the term underspecified for non-generic non-specific reference. Underspecified references include quantified NP's in modal, future, conditional, hypothetical, negated, uncertain, question contexts (in all cases the entity/entities referenced cannot be verified, regardless of the amount of “effort”).

[بعض الأمريكيان لا يحبون البيتزا]

[Some of the Americans] don’t like pizza

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

- A few,
- some (plural),
- most,
- more than half

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

[ألف من الناس] سيحضروا المؤتمر

[thousands of people] will attend the conference.

If the phrase was in the past tense as in the following example, we will tag it as specific

[ألف من الناس] حضروا المؤتمر

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

While we try to define the other four categories as precisely as possible, annotators may still encounter NPs that cannot be classified. In these cases, annotators should make these NPs Underspecified. By partitioning these truly ambiguous cases into the USP category, annotators will be able to make clearer distinctions between the other four categories, thus improving consistency.

5 Mention Types/Mention Levels

For each entity, we record and coreference all mentions of the entity. As mentioned, an entity refers to an object or set of objects in the world. A mention is a reference to an entity. A single entity can have multiple representations, such as proper name, noun phrase, pronoun or even clitic. If multiple mentions refer to the same entity, we need to coreference them. (In the tool, the way of coreference is to put all mentions of an entity in the same row).

Mentions will frequently be nested; that is, they will contain mentions of other entities. For example, the phrase

\[
\text{[President [Al Assad]} \] \text{will give talk tomorrow. He will state all his achievements and the problems that he still looking for solution to them.}\n\]

\[
\text{The president will be accompanied by most of the ministers}\n\]

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,

\[
\text{[President [Lahoud]} \] \text{[who is running for new presidency period]}\n\]

Mentions are categorized at several levels. At the top level there are two major types, simple and complex mentions. This top level distinction is motivated by the fact that some mentions have complex syntactic structures that cannot be easily annotated without breaking the syntactic analyses and/or information loss.
Subcategorizations of mentions types are syntactically motivated. The following table lists all the types we distinguish for Arabic.

<table>
<thead>
<tr>
<th>Simple Mentions</th>
<th>Named (NAM, NAMPRE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nominal (NOM, NOMPRE)</td>
</tr>
<tr>
<td></td>
<td>Pronominal (PRO)</td>
</tr>
<tr>
<td></td>
<td>Headless nominal (HLS)</td>
</tr>
<tr>
<td></td>
<td>Partive Constructions (PTV)</td>
</tr>
<tr>
<td></td>
<td>WH-Question words and specifiers (WHQ)</td>
</tr>
<tr>
<td>Complex Mentions</td>
<td>Apposition Constructions (APP)</td>
</tr>
<tr>
<td></td>
<td>Complex constructions taking a relative clause (ARC)</td>
</tr>
</tbody>
</table>

We will tag the mention as PTV only when it contains a number and followed by “off” as in:

[خمسة من الأطفال]

[Five of the kids]

But for other cases it will be tagged as NOM as in:

[بعض من الأطفال]

[Some of the kids]

**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

[[الرئيس [مبارك] [الذي شّهد نفسه لفترة رئاسية جديدة]]

[[President [Mubarak] [who] is running for new presidency period]]

the extent of the mention is the entire phrase:

الرئيس مبارك الذي رشّح نفسه لفترة رئاسية جديدة

The extent includes all the modifiers of a nominal phrase, including prepositional phrases and relative clauses.
Generally speaking, tokens are broken at white space, and each item of punctuation is treated as a separate character. As a rule, we do not include punctuation such as commas, periods, and quotation marks in the extent of a mention unless words included within the extent continue on after the punctuation mark.

Conjoined Mentions that are Modified
In constructions of conjoined mentions that share the same premodifiers or postmodifiers, each of the conjoined heads is tagged as the head of a single mention. The following examples will yield two tagged mentions.

<table>
<thead>
<tr>
<th>Arabic</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 رجل وامرأة غضبي</td>
<td>20 angry men and women</td>
</tr>
<tr>
<td>20</td>
<td>Nom mention</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Arabic</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>و 20</td>
<td>and 20</td>
</tr>
<tr>
<td>رجل وامرأة غضبي</td>
<td>angry men and women</td>
</tr>
<tr>
<td>20</td>
<td>Nom mention</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Arabic</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>and 20</td>
<td></td>
</tr>
<tr>
<td>آرﺕﺮ وﺟﻴﻤﻲ</td>
<td>Bill Clinton and Jimmy Carter who are both former presidents</td>
</tr>
<tr>
<td>آرﺕﺮ</td>
<td>Name mention</td>
</tr>
<tr>
<td>وﺟﻴﻤﻲ</td>
<td>Name mention</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Arabic</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>واشنطن</td>
<td>Washington</td>
</tr>
<tr>
<td>واشنطن</td>
<td>The movers and shakers in Washington</td>
</tr>
<tr>
<td>واشنطن</td>
<td>Nominal mention</td>
</tr>
</tbody>
</table>

5.1.2 Mention Head
In addition to the extent of the nominal phrase, the head of the phrase must be marked. In

[اعضاء] هيئة التدريس قامو برحلة
[The staff] went for trip
the full mention is
አعضاء هيئة التدريس
[The staff]
and the head is [عضاء هيئة التدريس]. If the syntactic head of the phrase is a multi-token item, the first right token is marked. If the head is a proper name, however, then the whole extent of the name is considered to be the head. In the following examples, the mention is enclosed in brackets and the head is underlined:
أصبح [فردين سمير] رئيس الوزراء الجديد
Fred Smith] became [the new Prime Minister].
تم اختيار [ابراهيم أبكرموفي الثالث] للاستاذة الوظيفية
The job fell to [Abraham Abercrombie III].

5.1.3 Names (NAM)
Proper nouns and nicknames.
[John] 
[Defense Secretary William Cohen] 
[وزير الدفاع ولیام کوہن]
[المعسكر لا حنین الجنزور بجوار امام الله]
[The Jeluzoon Refugee Camp near Ramallah]
[جامعة نورث داکوتا ستیت فی فارجو]
[North Dakota State University in Fargo]

5.1.4 Nominal Constructions (NOM)
A noun quantified with a determiner, a quantifier, or a possessive.
[الشیر العائلة] كلمته مسموعة
[The biggest of the family] has strict word
[الأشداء] هم من مارسوا الرياضة
[The toughest] are those who practices sports
[المحامي]
[The lawyer]
5.1.5 Pronouns (PRO)

Pronouns with the exception of wh-question words and the specifier ‘that’. A pronominal paradigm if Arabic consists of 12 forms: In singular and plural, the 2nd and 3rd persons differentiate gender, while the 1st person does not. In the dual, there is no 1st person, and only a single form for each 2nd and 3rd person. Traditionally, the pronouns are listed in order 3rd, 2nd, 1st.

**Personal pronouns**

<table>
<thead>
<tr>
<th>Person</th>
<th>Singular</th>
<th>Plural</th>
<th>Dual</th>
</tr>
</thead>
<tbody>
<tr>
<td>3rd (m)</td>
<td><em>huwa</em></td>
<td><em>hum</em></td>
<td><em>humā</em></td>
</tr>
<tr>
<td>3rd (f)</td>
<td><em>hiya</em></td>
<td><em>hunna</em></td>
<td></td>
</tr>
<tr>
<td>2nd (m)</td>
<td><em>anta</em></td>
<td><em>antum</em></td>
<td><em>antumā</em></td>
</tr>
<tr>
<td>2nd (f)</td>
<td><em>anti</em></td>
<td><em>antunna</em></td>
<td></td>
</tr>
<tr>
<td>1st</td>
<td><em>ana</em></td>
<td><em>naḥnu</em></td>
<td>(n/a)</td>
</tr>
</tbody>
</table>

**Notes:**

In Arabic the pronoun is normally not used. Instead, there are a lot of enclitic forms of the pronoun which are affixed to nouns (representing genitive case, i.e. possession) and to verbs (representing accusative, i.e. a direct object). For example in the "katabtu" the suffix "tu" shows that it is the first person singular and there is no need to use the pronoun "anaa". Enclitic pronouns are not taggable.

5.1.6 WH-Question Words and Specifiers (WHQ)

WH-question words and the specifier ‘that’.

*who*

المدير [الذي] تحدث عن

the executive, *who* spoke on the condition of anonymity

الجندي الروسي [الذي]نتظر عدة أيام قبل ان يطلب الطوارئ

[202x165] some American executives

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

Germany, [where] they were taken yesterday

5.1.7 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 quantifiers such as the numbers in the following. Note that in Arabic, there are similar structures like the English expression “the toughest”. We treat these as Nominals rather than Headless mentions.

[more than 30] [أكثر من 30]

[35] were injured [35] جُرح

5.1.8 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] [بعض من المحامين]

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

Examples of non-partitives (two entity mentions):

[عضويين من الفريق] عضويين من الفريق

Two members of the team

5.1.9 Postmodifier Mentions (NAMPRE, NOMPRE)
Postmodifier mentions (the premodifiers in English) are those mentions which occur in a modifying position before another word(s). It is immaterial whether or not the word being modified is a taggable entity.

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

[القوات الإسرائيلية]
[The Israeli troops]

[The Lebanese publishers]

[The German language]

[The British books]

<table>
<thead>
<tr>
<th>Taggable</th>
<th>Not Taggable</th>
</tr>
</thead>
<tbody>
<tr>
<td>[منطقة[الجبيل]] [mountain] regions</td>
<td>[مناطق الجبلية] mountainous regions</td>
</tr>
<tr>
<td>[منحة[الحكومة]] [government] grants</td>
<td>[منح الحكومة] federal grants</td>
</tr>
</tbody>
</table>

5.1.9.1 Name or Nominal Classification

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

| [القوات[الأسرائيلية]] [Israeli] troops | NAMPRE                      |
| [الناخبين[الجمهوريين]] [Republican] voters | NAMPRE                      |
| [مصانع[كرايسلر]] [Chrysler] factories   | NAMPRE                      |
5.2 Complex Constructions

The purpose of complex constructions is to identify difficult regions where the simple mention extent rules do not apply. We do not identify heads for complex constructions. Within the extent of a complex construction, simple mentions will be annotated. Each of these complex constructions has rules for extent.

5.2.1 Appositive Constructions (APP)

A construction which consists of two or more full mentions which refer to (or predicate on) the same entity. The two mentions are placed side by side, with one explains or characterize the other. Superficially this looks similar to a noun modifying another noun. However, they are different in that in an appositive construction, the first NP and the second NP roughly refers to the same entity, whereas this is not the case in a noun-noun compound. In annotation, the APP-mention itself has no head-assignment.

اﻟﺮوﺱﻲ اﻟﺮﺉﻴﺲ  
[The Russian president] [Valdmir Putin]

[رئيس روسيا ] [ بوتين]
[The Russian president] [ Putin]

[امين عام الأمم المتحدة] [ كوفي عنان]
[The secretary general to the UN] [Kofi Anan]

Another structure in Arabic similar to App constructions is the construction of proper names with titles (3.1.2 and 3.1.3), such as “the president Putin” which is not considered as an apposition. The difference is that when the nominal head has modifiers, we treat the whole phrase as an APP; when the nominal head has no modifiers, we tag the nominal as the head of a NOM mention and the proper name as a NAM mention. Below is the annotation for these two structures.

Title with proper name

<table>
<thead>
<tr>
<th>[البابا ] [ بوناوس الثاني]</th>
<th>C NOM.ATR PER.Individual</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Pope [John Paul II]</td>
<td>NOM.ATR</td>
</tr>
<tr>
<td>[John Paul II] NAM SPC</td>
<td></td>
</tr>
</tbody>
</table>

Appostion:

<table>
<thead>
<tr>
<th>[الرئيس الروسي بوتين ]</th>
<th>PER.Individual</th>
</tr>
</thead>
<tbody>
<tr>
<td>[The Russian president ]</td>
<td>APP.SPC</td>
</tr>
<tr>
<td>Putin</td>
<td></td>
</tr>
</tbody>
</table>
5.2.2 Complex Constructions taking a Relative Clause (ARC)
An ARC-construction is an appositional construction with an adjacent relative clause that is an adjunct to the entire apposition, rather than just to the Nominal mention or Name mention. 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.


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 pure
metaphoric mentions as in “he is my sun” is not treated as metonymy, but as attributive nominal. The sections below outline several common examples. This is not an exhaustive list. Any example of this kind of reference should be identified. We will coreference the mention with the entity to which the mention refers in the context and indicate that this is an example of Nickname Metonymy by selecting the check box.

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

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

ایﺮان

[Beijing] will not continue sales of anti-ship missiles to Iran.

In this example there are two senses of the word Beijing: the city Beijing and the government of China. We will tag Beijing as 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 City name for Sports Team

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

[America] brought home the gold.

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 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 administers 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.

<table>
<thead>
<tr>
<th>CE</th>
<th>Argument 1</th>
<th>Cross type metonymy</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE1(Arabic)</td>
<td>مستشفى القديس جون ماكومب</td>
<td>CE1(ORG. Med/Sci)</td>
</tr>
<tr>
<td>CE2 (Arabic)</td>
<td>المستشفى hospital</td>
<td>CE2 (FAC. Buid/Ground)</td>
</tr>
</tbody>
</table>
Appendix:

Some annotation examples:

**FAC.Building**

<table>
<thead>
<tr>
<th>Name</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>berkeley marina</td>
<td>FAC.Building</td>
</tr>
<tr>
<td>gitmo</td>
<td>FAC.Building</td>
</tr>
<tr>
<td>guantanamo bay</td>
<td>FAC.Building</td>
</tr>
</tbody>
</table>

**FAC.Subarea**

<table>
<thead>
<tr>
<th>Name</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>rose garden</td>
<td>FAC.Subarea</td>
</tr>
</tbody>
</table>

**GPE.Cluster**

<table>
<thead>
<tr>
<th>Name</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>balkans</td>
<td>GPE.Cluster</td>
</tr>
<tr>
<td>Commonwealth</td>
<td>GPE.Cluster</td>
</tr>
<tr>
<td>East</td>
<td>GPE.Cluster</td>
</tr>
<tr>
<td>EU</td>
<td>GPE.Cluster</td>
</tr>
<tr>
<td>european union</td>
<td>GPE.Cluster</td>
</tr>
<tr>
<td>group of eight industrial powers</td>
<td>GPE.Cluster</td>
</tr>
<tr>
<td>middle east</td>
<td>GPE.Cluster</td>
</tr>
<tr>
<td>mideast</td>
<td>GPE.Cluster</td>
</tr>
<tr>
<td>south asia</td>
<td>GPE.Cluster</td>
</tr>
<tr>
<td>southeast asia</td>
<td>GPE.Cluster</td>
</tr>
<tr>
<td>West</td>
<td>GPE.Cluster</td>
</tr>
</tbody>
</table>
### GPE.Continent

<table>
<thead>
<tr>
<th>Name</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>GPE.Continent</td>
</tr>
<tr>
<td>Antarctic</td>
<td>GPE.Continent</td>
</tr>
<tr>
<td>antarctica</td>
<td>GPE.Continent</td>
</tr>
<tr>
<td>Asia</td>
<td>GPE.Continent</td>
</tr>
<tr>
<td>asian</td>
<td>GPE.Continent</td>
</tr>
<tr>
<td>Europe</td>
<td>GPE.Continent</td>
</tr>
<tr>
<td>north america</td>
<td>GPE.Continent</td>
</tr>
</tbody>
</table>

### GPE.County

<table>
<thead>
<tr>
<th>Name</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>fairfax county</td>
<td>GPE.County</td>
</tr>
</tbody>
</table>

### GPE.Nation

<table>
<thead>
<tr>
<th>Name</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>America</td>
<td>GPE.Nation</td>
</tr>
<tr>
<td>Australia</td>
<td>GPE.Nation</td>
</tr>
<tr>
<td>axis of evil</td>
<td>GPE.Nation</td>
</tr>
<tr>
<td>cayman islands</td>
<td>GPE.Nation</td>
</tr>
<tr>
<td>china</td>
<td>GPE.Nation</td>
</tr>
<tr>
<td>congo</td>
<td>GPE.Nation</td>
</tr>
<tr>
<td>england</td>
<td>GPE.Nation</td>
</tr>
<tr>
<td>iran</td>
<td>GPE.Nation</td>
</tr>
<tr>
<td>iraq</td>
<td>GPE.Nation</td>
</tr>
<tr>
<td>israel</td>
<td>GPE.Nation</td>
</tr>
<tr>
<td>Korean</td>
<td>GPE.Nation</td>
</tr>
<tr>
<td>spain</td>
<td>GPE.Nation</td>
</tr>
</tbody>
</table>

### GPE.Pop

<table>
<thead>
<tr>
<th>Name</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>atlanta</td>
<td>GPE.Pop</td>
</tr>
<tr>
<td>Name</td>
<td>Decision</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>baghdad</td>
<td>GPE.Pop</td>
</tr>
<tr>
<td>hong kong</td>
<td>GPE.Pop</td>
</tr>
<tr>
<td>key largo</td>
<td>GPE.Pop</td>
</tr>
<tr>
<td>macao</td>
<td>GPE.Pop</td>
</tr>
</tbody>
</table>

**Gaza**

<table>
<thead>
<tr>
<th>Name</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaza</td>
<td>GPE.Pop</td>
</tr>
</tbody>
</table>

**GPE.Special**

<table>
<thead>
<tr>
<th>Name</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>mainland (China)</td>
<td>GPE.Special</td>
</tr>
<tr>
<td>palestine</td>
<td>GPE.Special</td>
</tr>
<tr>
<td>Taiwan</td>
<td>GPE.Special</td>
</tr>
</tbody>
</table>

**GPE.State**

<table>
<thead>
<tr>
<th>Name</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>gaza strip</td>
<td>GPE.State</td>
</tr>
<tr>
<td>northern ireland</td>
<td>GPE.State</td>
</tr>
<tr>
<td>west bank</td>
<td>GPE.State</td>
</tr>
</tbody>
</table>

**LOC.Land-Reg**

<table>
<thead>
<tr>
<th>Name</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>azores islands</td>
<td>LOC.Land-Reg</td>
</tr>
<tr>
<td>canary islands</td>
<td>LOC.Land-Reg</td>
</tr>
<tr>
<td>continental slope</td>
<td>LOC.Land-Reg</td>
</tr>
<tr>
<td>diego garcia</td>
<td>LOC.Land-Reg</td>
</tr>
<tr>
<td>ellis island</td>
<td>LOC.Land-Reg</td>
</tr>
<tr>
<td>everest</td>
<td>LOC.Land-Reg</td>
</tr>
<tr>
<td>him la yas (Himalayas)</td>
<td>LOC.Land-Reg</td>
</tr>
<tr>
<td>korean peninsula</td>
<td>LOC.Land-Reg</td>
</tr>
<tr>
<td>mount everest</td>
<td>LOC.Land-Reg</td>
</tr>
<tr>
<td>neichangshan islands</td>
<td>LOC.Land-Reg</td>
</tr>
<tr>
<td>Name</td>
<td>Decision</td>
</tr>
<tr>
<td>--------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>sekamian plateau</td>
<td>LOC.Land-Reg</td>
</tr>
<tr>
<td>urbuinano island</td>
<td>LOC.Land-Reg</td>
</tr>
</tbody>
</table>

**LOC.Reg-Gen**

<table>
<thead>
<tr>
<th>Name</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>central jakarta</td>
<td>LOC.Reg-Gen</td>
</tr>
<tr>
<td>continental ~usa</td>
<td>LOC.Reg-Gen</td>
</tr>
<tr>
<td>east coast</td>
<td>LOC.Reg-Gen</td>
</tr>
<tr>
<td>east texas</td>
<td>LOC.Reg-Gen</td>
</tr>
<tr>
<td>green zone</td>
<td>LOC.Reg-Gen</td>
</tr>
<tr>
<td>midwest</td>
<td>LOC.Reg-Gen</td>
</tr>
<tr>
<td>new england</td>
<td>LOC.Reg-Gen</td>
</tr>
<tr>
<td>North</td>
<td>LOC.Reg-Gen</td>
</tr>
<tr>
<td>Northeast</td>
<td>LOC.Reg-Gen</td>
</tr>
<tr>
<td>siberia</td>
<td>LOC.Reg-Gen</td>
</tr>
<tr>
<td>South</td>
<td>LOC.Reg-Gen</td>
</tr>
<tr>
<td>south florida</td>
<td>LOC.Reg-Gen</td>
</tr>
<tr>
<td>Southeast</td>
<td>LOC.Reg-Gen</td>
</tr>
<tr>
<td>southside</td>
<td>LOC.Reg-Gen</td>
</tr>
<tr>
<td>southwest</td>
<td>LOC.Reg-Gen</td>
</tr>
<tr>
<td>upper east side</td>
<td>LOC.Reg-Gen</td>
</tr>
</tbody>
</table>

**LOC.Reg-Int**

<table>
<thead>
<tr>
<th>Name</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>arctic</td>
<td>LOC.Reg-Int</td>
</tr>
<tr>
<td>eurasia</td>
<td>LOC.Reg-Int</td>
</tr>
<tr>
<td>far east</td>
<td>LOC.Reg-Int</td>
</tr>
<tr>
<td>greater asia</td>
<td>LOC.Reg-Int</td>
</tr>
<tr>
<td>indochina</td>
<td>LOC.Reg-Int</td>
</tr>
<tr>
<td>Keshmere (Kashmir)</td>
<td>LOC.Reg-Int</td>
</tr>
<tr>
<td>Manchiku</td>
<td>LOC.Reg-Int</td>
</tr>
<tr>
<td>manchuria</td>
<td>LOC.Reg-Int</td>
</tr>
</tbody>
</table>
### LOC.Water

<table>
<thead>
<tr>
<th>Name</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>arabian sea</td>
<td>LOC.Water</td>
</tr>
<tr>
<td>atlantic</td>
<td>LOC.Water</td>
</tr>
<tr>
<td>chesapeake bay</td>
<td>LOC.Water</td>
</tr>
<tr>
<td>euphrates river</td>
<td>LOC.Water</td>
</tr>
<tr>
<td>indian ocean</td>
<td>LOC.Water</td>
</tr>
<tr>
<td>khor abdallah river</td>
<td>LOC.Water</td>
</tr>
<tr>
<td>khoser river</td>
<td>LOC.Water</td>
</tr>
<tr>
<td>lake geneva</td>
<td>LOC.Water</td>
</tr>
<tr>
<td>lake superior</td>
<td>LOC.Water</td>
</tr>
<tr>
<td>pacific ocean</td>
<td>LOC.Water</td>
</tr>
<tr>
<td>red sea</td>
<td>LOC.Water</td>
</tr>
<tr>
<td>san francisco bay</td>
<td>LOC.Water</td>
</tr>
<tr>
<td>strait of gibraltar</td>
<td>LOC.Water</td>
</tr>
<tr>
<td>tigris river</td>
<td>LOC.Water</td>
</tr>
<tr>
<td>yellow sea</td>
<td>LOC.Water</td>
</tr>
</tbody>
</table>

### Land Areas Named by Water References

<table>
<thead>
<tr>
<th>Name</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caribbean</td>
<td>LOC.Water, LOC.Reg-Int</td>
</tr>
<tr>
<td>gulf</td>
<td>LOC.Water, GPE.Cluster</td>
</tr>
<tr>
<td>mediterranean</td>
<td>LOC.Water, LOC.Reg-Int</td>
</tr>
<tr>
<td>persian gulf</td>
<td>LOC.Water, GPE.Cluster</td>
</tr>
</tbody>
</table>

### ORG.Non

<table>
<thead>
<tr>
<th>Name</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>kizilay</td>
<td>ORG.Non</td>
</tr>
</tbody>
</table>

### Palestine

<table>
<thead>
<tr>
<th>Name</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>palestinians</td>
<td>PER.Group, GPE.Special</td>
</tr>
<tr>
<td>Name</td>
<td>Decision</td>
</tr>
<tr>
<td>------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>palestinian authority</td>
<td>ORG.Gov</td>
</tr>
<tr>
<td>PLC</td>
<td>ORG.Gov</td>
</tr>
</tbody>
</table>

**NOT TAGGABLE**