

# **ACE (Automatic Content Extraction) Arabic Annotation Guidelines for Entities**

Version 5.3.3 2005.05.31

Linguistic Data Consortium

<http://www ldc upenn edu/Projects/ACE/>

1 Introduction .....	4
Basic Concepts .....	4
2 Text to Annotate .....	4
3 Entity Types and Subtypes .....	5
3.1 Persons .....	5
3.1.1. Subtypes for Person .....	5
3.1.2 Titles, Honorifics, and Positions .....	6
3.1.3 Fictional characters, names of animals, and names of fictional animals.....	6
3.1.4 Groups of people .....	7
3. 2 Organizations .....	7
3.2.1 Subtypes for Organizations .....	7
3.2.2 Organization Subtyping Trumping Rules.....	13
3.2.3 Organization Entities used in Person Contexts .....	13
3.3 Geographical/Social/Political Entities (GPE) .....	14
3.3.1 Subtypes for GPEs .....	15
3.3.2 GPE-like Locations and Organizations.....	16
3.3.4 GPE Mention Roles .....	17
3. 4 Locations .....	21
3.4.1 Subtypes for Locations .....	21
3.4.2 Sub-parts of Locations and GPEs .....	22
3.4.3 Non-Locations .....	22
3.5 Facilities .....	23
3.5.1 Subtypes for Facilities .....	23
3.6 Vehicle .....	24
3.6.1 Subtypes for Vehicles.....	24
3.7 Weapon .....	25
3.7.1 Subtypes for Weapons .....	25
4 Entity Class .....	26
4.1 Negatively (NEG) .....	26
4.2 Specific (SPC) .....	27
4.3 Generic (GEN) .....	27
4.4 Under-specified (USP) .....	27
5 Mention Types.....	28
5.1 Mention Extent .....	28
5.2 Mention Head.....	29
5.3 Mention Types.....	29
5.3.1 Names (NAM) .....	29
5.3.2 Nominal Constructions (NOM) .....	29
Nominal Mentions with Multiple-word Heads (MWH) .....	30
5.3.3 Pronouns (PRO).....	31
Headless Mentions (HLS) .....	31
Partitive Constructions (PTV) .....	31
5.3.4 Premodifier Mentions (Post-modifiers in Arabic) .....	31
5.4 Complex Constructions .....	32
5.4.1 Conjoined Mention Constructions (MNH).....	32
5.4.2 Appositive Constructions (APP) .....	33
5.4.3 Mr. and Mrs. Smith Constructions .....	33
6 Nickname Metonymy.....	34

6.1 Capital City or Government Seat (FAC) Names standing in for Country's  
Government ..... 34  
6.2 GPE name for Sports Team..... 34

# 1 Introduction

*The Entity Detection task requires that selected types of entities mentioned in the source data be detected, their sense disambiguated, and that selected attributes of these entities be extracted and merged into a unified representation for each entity.*

## Basic Concepts

*An entity is an object or set of objects in the world. A mention is a reference to an entity. Entities may be referenced in a text by their name, indicated by a common noun or noun phrase, or represented by a pronoun. For example, the following are several mentions of a single entity:*

**Name Mention:** Joe Smith جو سميث

**Nominal Mention:** the guy wearing a blue shirt الفتى الذي يرتدي قميص أزرق

**Pronoun Mentions:** he, him هو هم

*For Phase 3 of ACE, entities are limited to the following seven types:*

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

*For each entity, the annotation records the type of the entity (PER, ORG, GPE, LOC, FAC, VEH, and WEA), subtype, class, and all the textual mentions of that entity.*

## 2 Text to Annotate

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

## 3 Entity Types and Subtypes

### 3.1 Persons

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

There are a number of words that are ambiguous as to their referent. For example, nouns, which normally refer to animals or non-humans, can be used to describe people. If it is clear to the annotator that the noun refers to a person in a given context, it should be marked as a Person entity.

He is [a real turkey]

إنه [سلحفاة] لا يتحمل المواجهة

[The political cat of the year]

الثعلب السياسي لهذا العام

She’s known as [the brain of the family]

من المعروف عنها أنها [العقل المدبر للعائلة]

#### 3.1.1. Subtypes for Person

##### **PER.Individual:**

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

President Clinton

الرئيس كلينتون

Edmund Pope

ادمون بوب

The President of the U.S.

رئيس الولايات المتحدة الامريكية

##### **PER.Group:**

If the Person entity refers to more than one person, tag it as PER.Group.

Analysts

المحللون

Christians

المسيحيين

IBM’s lawyers

محاموا شركة اي بي ام

##### **PER.Indefinite**

If from the context you can’t judge whether the Person entity refers to one or more than one person, tag it as PER.Indefinite. We are not expecting to see such examples 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” (وزيرة الخارجية الأمريكية مادلين أولبرايت), there would be four mentions of three distinct entities.

[Secretary of State <u>Madeline Albright</u> ] [وزيرة الخارجية مادلين أولبرايت]	PER
[Secretary of State] [وزيرة الخارجية]	
[US] [الأمريكية]	GPE
[State] [الخارجية]	ORG

#### 3.1.2.1 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]

[النبي]	[موسى] عليه الصلاة والسلام]
Name	Nom
Per	Per
Indiv	Indiv
<b>SPC</b>	<b>ATR</b>

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.

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

..... إذا كنت تؤمن [بالله] فلا بد

Although he felt like he was [a god]... nominal mention

..... على الرغم من أنه كان يشعر بأنه [له]

### 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

[باتمان] أصبح صورة شعبية

[Adam West]'s costume from Batman the TV series

زي [آدم ويست] من مسلسل باتمان التلفزيوني

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.

توم وجيري Tom and Jerry  
كعبول الأكل (Bear) Ka'bool the Gluttonous

### 3.1.4 Groups of people

*Groups of people are to be considered an entity of type Person.Group unless the group meets the requirements of an organization or a GPE described below.*

[The family]  
[العائلة]  
[The house painters]  
[دهانو المنزل]  
[The linguists under the table]  
[محللي اللغة تحت المنضدة]

*In the following example the underlined head will be tagged as Headless*

[الهجوم اسفر عن عشرون قتيلا]

#### 3.1.4.1 Family Names

*Family names are to be tagged as Person.*

[The Kennedy]'s  
[عائلة كينيدي]

#### 3.1.4.2 Ethnic and Religious Groups

*Ethnic groups of people and religious groups that do not have a formal organization unifying them will be considered entities of type Person.*

[The Arabs]  
[العرب]  
[The Christians]  
[المسيحيون]

## 3. 2 Organizations

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

### 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. Governing bodies of GPE's are included in this subtype.*

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

[مكتب المدعي العام لسالزبورج] يتحرى حاليا عن الكارثة لتحديد امكانية توجيه تهمة اجرامية

Putin, a former {KGB} agent, defended [the court] [that] convicted Pope and [the security services],...

بوتين العميل السابق [الكي.جي.بي.] دافع عن [المحكمة] [التي] اتهمت البابا و [الخدمات الأمنية]

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

لن يتخذ مجلس المحاسبين القانونيين أي اجراء في مشروع دمج الأعمال الحرة حتي اجتماع [الكونجرس] في 2001

## **Commercial (COM)**

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

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

أعلن بنك الكويت التجاري أرباحه السنوية  
أعلنت والت ديزني إنتاجها لمشروع فيلم رسوم متحركة جديد

## **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.*

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

بوب البالغ من العمر 54 عام وهو ضابط بحرية متقاعد من ستيت كوليج، بنسلفانيا، قد عمل في [مختبر الأبحاث التطبيقية في جامعة بنسلفانيا ستيت]

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

أخصائي الحشيش لدى جامعة نورث داكوتا ستيت وجامعة ميناسوتا آلن دكستير يقول إن 98% من النبتة عاشت

## **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 exceptions are all best annotated as Commercial Organizations.*

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

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

ملايين العرب يشاهدون قنوات ال أل.بي.سي. وال أي.آر.تي  
عقدت فرقة ميامي الكويتية اتفاقا مع شركة روتانا للإنتاج الفني

## **Non-Governmental Organizations (NonGov)**

*NonGov 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:**

KKK  
Al Aqsa Martyr's Brigade  
Tamil Tigers

عشيرة كو كلوكس  
كتائب شهداء الأقصى  
نمور التاميل

### **2. Political Parties:**

Republican Party  
Labour Party

الحزب الجمهوري  
حزب العمل

### **3. Political Advocacy Groups and Think Tanks:**

ACLU  
The Cato Institute  
الاتحاد الأمريكي للحريات المدنية  
معهد كاتو

### **4. Professional Regulatory and Advocacy Groups:**

The American Bar Association  
The American Medical Association

جمعية المحامين الأمريكيين  
الجمعية الطبية الأمريكية

## 5. Charitable Organizations:

The Red Cross  
UNICEF  
Doctors Without Borders

الصليب الأحمر  
صندوق الأمم المتحدة للأطفال  
أطباء بلا حدود

## 6. International Regulatory and Political Bodies:

UN  
NATO  
The World Bank

الأمم المتحدة  
منظمة حلف الشمال الأطلسي  
البنك الدولي

*Under the above definition, all of the following would be annotated as Organizations of Subtype NonGov:*

three of the [U.N]. workers stationed in East Timor  
ثلاثة من عمال [الأمم المتحدة] الموضوعين في تيمور الشرقية  
[International Monetary Fund] aid  
معونات [صندوق النقد الدولي]  
[Hizbollah]  
[حزب الله]  
[the ``Caravan of Death], [a military party [that] killed 73 political prisoners]  
[قافلة الموت]، [حزب عسكري قتل ثلاثا وسبعين سجيناً سياسياً]  
[Institutional Revolutionary Party]  
[الحزب التأسيسي الثوري]  
[The GOP]  
[الحزب الجمهوري]  
[the Democratic Party]  
[الحزب الديمقراطي]  
[the Socialist People's Party]  
[الحزب الاشتراكي الشعبي]  
[Republican National Committee]  
[اللجنة الجمهورية القومية]

[American Medical Association]

[الجمعية الطبية الأمريكية]

[American Council on Education]

[المجلس الأمريكي للتربية]

[College Board]

[المجلس الجامعي]

[National Rifle Association]

[الجمعية القومية للبندقية]

[American Diabetic Association]

[الجمعية الأمريكية للبول السكري]

[National Center for Public Policy and High Education]

[المركز القومي للسياسة العامة والتعليم العالي]

[American Bar Association]

[جمعية المحامين الأمريكيين]

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

ذكر [الصليب الأحمر] أن نحو خمسة عشر شخصا تمكنوا من الفرار ولكن ثلاثة على الأقل ماتوا

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

...

## **Media (MED)**

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

Agence France Presse

وكالة الأنباء الفرنسية (اف ب)

Reuters international news agency

وكالة رويترز للأنباء

Time magazine

جريدة التايمز

abc news

اذاعة ايه بي سي

Associated Press

وكالة الاسوشيتد بريس

قناة الجزيرة

جريدة الحياة

نيو يورك تايمز

ال بي بي سي  
مجلة سيدتي

### **Religious (REL)**

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

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

[German Bishops Conference]

[The Vatican]

...قال رئيس الأساقفة خجق باراساميان رئيس [أسقفية] الكنيسة الأرمنية الأمريكية التي مقرها في مدينة مانهاتن]]...  
اتحاد الأساقفة الألمان

الفاتيكان

### **Medical-Science (SCI)**

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

Overseas Chinese Physics Association

Gulf Coast Research Laboratory

Massachusetts General Hospital

جمعية الفيزياء الصينية عبر البحار

معهد الدراسات النووية

مستشفى مبارك

### **Sports (SPO)**

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

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

Saudi Soccer Federation

Taekwondo Association

Philippines Olympic Committee

national hockey league

San Francisco 49ers

فريق القادسية الكويتي

جمعية التايكواندو

### 3.2.2 Organization Subtyping 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.



#### Exceptions to Trumping Rules

##### GPE military

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

##### Media Conglomerates

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

##### Medical Schools and Research Labs

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

##### Soft Science Research Institutions

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

##### Boy Scouts

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

### 3.2.3 Organization Entities used in Person Contexts

Whenever an organization takes an action, there are people within or in charge of the organization that one presumes actually made the decision and then carried it out. Thus

many organization mentions could be thought of as metonymically referring to people within the organization.

However, there seems to be little to be gained in the usual case by thus “reaching inside the organization” to posit a mention of a Person entity. It seems better to adopt the view that organizations can be agentive, and take action on their own. Only when something in the context draws particular attention to the people within the organization should a separate mention of a Person entity be marked.

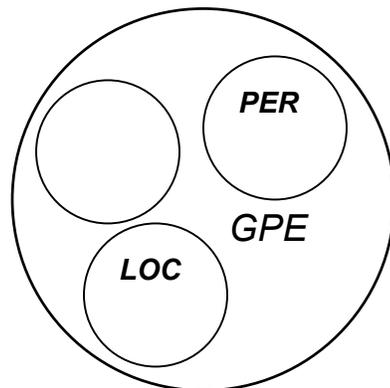
Sets of people who are not formally organized into a unit are to be treated as a person entity rather than an organization entity. It is often difficult to tell the difference between organization entities and collections of individuals tagged as person entities. Examples of organization-like nouns which are not organizations are “employees,” and “crew.” In the latter two cases, although the members of a company or crew may work together in an organized and even hierarchical fashion, the groups are not organizations by themselves.

### 3.2.4 First Person Pronouns Referring to Organizations

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

## 3.3 Geographical/Social/Political Entities (GPE)

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



In this sentence,

The people of France welcomed the agreement.

رحب شعب فرنسا بالاتفاقية

there are two mentions

[The people of France]

GPE [شعب فرنسا]

[France]

GPE [فرنسا]

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.

هزمت [نيو يورك] [بوسطن] 97-99 في الوقت الاضافي

*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.

سيواجه كثير من الناس في [أمريكا الشمالية] كسوف جزئي للشمس غدا

**Nation** – Taggable mentions of the entireties of any nation.

But the [U.S.] presence in Aden is unmistakable.

الوجود [الأمريكي] في عدن لا مجال للشك فيه

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

[Salzburg] governor Schausberger said...

قطاع غزة

**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.

مرتفعات الجولان

### **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.]

فيصل القاسم، الدوحة، قطر

استقبل وفد من المسافرين شرشيل عند وصوله القاهرة للاعتراض على اجتياحاته لحقوق المسافرين

تقع مدينة دوما شمال دمشق

### **GPE-Cluster**

*Named groupings of GPEs that can function as political entities.*

Eastern Europe أوروبا الشرقية

Western Europe أوروبا الغربية

the European Union الاتحاد الأوروبي

the Middle East الشرق الأوسط

Southeast Asia جنوب شرق آسيا

West Africa غرب إفريقيا

Latin America أمريكا اللاتينية

...and the government tried to redefine [the West].

والحكومة حاولت إعادة تعريف "الغرب" ...

### **Special**

*A closed set of GPEs for which the conventional labels do not straightforwardly apply. Initially restricted to Palestine, the Palestinian Authority, and Native American reservations.*

### **3.3.2 GPE-like Locations and Organizations**

*Other incidental, non-political clusters of GPEs are still considered Locations.*

the southern United States

جنوب الولايات المتحدة

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

[NATO] peacekeepers arrived in the valley before nightfall.

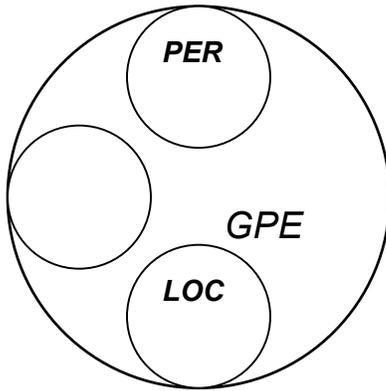
محافظو السلام من منظمة حلف الشمال الأطلسي وصلوا إلى الوادي قبل حلول الليل

### 3.3.3 Formulaic GPE Constructions: Nested Region Names

A series of nested region names, such as "Provo, Utah" evokes one entity for each region. Thus "Provo, Utah" evokes one entity for the population center (with mention "Provo, Utah") and a second one for the state (with mention "Utah").

### 3.3.4 GPE Mention Roles

Annotators need to decide for each entity mention in the text which role (Person, Organization, Location, GPE) the context of that mention invokes. The four roles are Person, Organization, Location and GPE. This judgment typically depends on the relations that the entity enters into.



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

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

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

France produces better wine than New Jersey.	GPE Role (whole nation)
تنتج فرنسا نبيذ أفضل مما تنتجه نيو جيرسي	
France's greatest national treasure	GPE Role (indistinct referent)
أعظم كنوز فرنسا الوطنية	

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.

تعتبر عودة المحقق ريتير نوع من الاختبار للمعاهدة التي وافق العراق من خلالها منح المفتشين حق كامل لدخول ثماني من قصور الرئاسة لصدام حسين

## 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.

ان [حكومة روسيا] وكثير من السياسيين سيقدمون نقدا لاذعا للولايات المتحدة في حالة اكتشافهم تجاهلها لهم

## 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.

أعلنت الإدارة أن سونابند يمكنه اللجوء الي محكمة سويسرا العليا

*Note that the GPE mention is a possessive in this example and not a premodifier (postmodifier in Arabic).*

## 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 Japanese}** have a considerable responsibility for the wars of the first half of the century

اليابانيون يحملون مسؤولية ضخمة لحروب النصف الأول من القرن

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

**{[GPE.PER] Cubans}** have been waiting for this day for a long time.

انتظر الكوبيون طويلا هذا اليوم

You and **{[GPE.PER] the [(GPE.GPE) American] people}** have a right to- to get answers.

من حقك و من حق الشعب الأمريكي الحصول على اجابات

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

وأیضا تطور بارز آخر للفرنسيين في معالجاتهم المستمرة للعدد الهائل من حيواناتهم المنزلية

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

أفاد بتلر بأنه يمكن انهاء الحظر الاقتصادي اذا سمح العراقيون للمفتشين بأداء واجبهم

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

الميزوري سوف يستقر جنب المذكرة للسفينة الحربية يو أس أس أريزونا، التي أغرقها اليابانيون خلال هجومهم على بيرل هاربور

**{[PER] idealistic Europeans}**

أوروبيون مثاليون

**{[PER] Americans}** who want to come and, and learn, uh, from the communities how to live in a community, how to take decisions among the community}

الأمريكان الذين يرغبون في الحضور و التعلم من الجماعات كيفية الحياة مع الجماعة و كيف يتم اتخاذ القرار فيما بين الجماعة

I do think there is a danger that **{[PER] some Chinese}** may underestimate American will on the Taiwan issue.

أعتقد أن هناك احتمال خطير أن [بعض الصينيين] سوف يستخفون بإرادة الأمريكية فيما يتعلق بمسألة تايوان

## **GPE.LOC**

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

The coast of **{[GPE.LOC]Britain}**

ساحل بريطانيا

**{[GPE.LOC]France}** has an area of 547,090 square kilometers

إن مساحة فرنسا 547090 كيلو متر مربع

U.S. warplanes flew over **{[GPE.LOC]Afghanistan}**

المقاتلات الجوية الأمريكية حلقت فوق أفغانستان

Northern portions of **{[GPE.LOC]Montana}**

المناطق الشمالية من مونتانا

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

## **GPE.GPE**

*GPE.GPE is used when more than one of the other GPE roles are being referenced at once or when no one role stands out in the context. Below are a few particular contexts in which GPE.GPE should always be used.*

## **GPE Premodifiers (Postmodifiers in Arabic)**

*Pre-modifiers (post-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 القوات الإسرائيلية  
{[GPE.GPE] New York} policemen رجال بوليس نيو يورك  
{[GPE.GPE] British} Prime Minister رئيس الوزراء البريطاني  
{[GPE.GPE] California} attorney محامي كاليفورنيا  
{[GPE.GPE] U.S.} Commander-in-Chief رئيس أركان أمريكا  
{[GPE.GPE] U.S.} surveillance aircraft طائرة المتابعة الأمريكية  
{[GPE.GPE] Iraqi} flag العلم العراقي  
The {[GPE.GPE] California} company شركة كاليفورنيا

## 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}.

كان من الممكن أن تستخدم المدينة بعض الإجراءات الوقائية في عام 1979 حينما قام الاتحاد السوفيتي بغزو أفغانستان

## 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.

حظرت السلطة الفلسطينية التجمعات المدافعة عن العراق و لكن هذا الحظر لم يؤخذ في الاعتبار

## Athletes, Sports Teams, and GPEs

*Athletes and teams are assoc*

*iated 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.

كان من الممكن أن تنتصر نيو يورك باصابة الهدف و لكن كريس تشايلدز ضيع ثلاثة أهداف

## 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 of a point in a postal system is also an address.

[Capital Hill] كابيتال هيل

**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].

تحتل أرمينيا موقع استراتيجي سياسيا في جنوب القوقاز و تشترك في [الحدود مع تركيا و آذربيجان و جورجيا]

**Celestial** – A location which is other worldly 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.

ومن المعروف عدم النظر مباشرة الى الشمس وقت الكسوف لأن ذلك يمكن أن يؤدي الي ضرر في العين لا يمكن اصلاحه

In Armenia, the three of them will join other, similar delegations from around [the earth]...

يدور رواد الفضاء حول الارض بسرعة مهولة

**Water\_Body** – Bodies of water, natural or artificial (man-made).

[The Missouri River] نهر ميزوري

**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

التهديدات الأساسية لجنوب إفريقيا خلال الخمس عشرة سنة المقبلة...

### **Region-General**

*Taggable locations that do not cross national borders.*

In [another part of the city],...

في (جزء آخر من المدينة)...

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

...قال هانس والنر، مدير (المنطقة السياحية في كابرون)

...from neighboring Bavaria, in [southern Germany]...

من بافاريا المجاورة في (جنوب ألمانيا)

### **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]"\*

انهم غالبا لا يسكنون في [وسط [الدولة]] و انما علي [حدودها الخارجية]

### **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."

"معطفك تحت الكلب"

"The rabbit is hiding behind that rock."

"الأرنب مختبئ خلف هذه الصخرة"

"I have an idea in my head."

"الذي فكرة في مخيلتي"

*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."

"لقد أسقط المفاتيح علي الأرض"

"He put the lamp back in its place."

"لقد أعاد المصباح الي مكانه مرة أخرى"

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  
(مطار لاجوارديا في نيو يورك) كان كارثة هذا العام

##### **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]...  
...القطار كلن ينتقل مباشرة من (مصفاة النفط) إلى المصهر...

##### **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.

[the Netzarim settlement] مستوطنة نتزاريم  
[the Berlin Wall] حائط برلين  
the parades at [Disneyland] الاستعراضات في ديزني لاند

##### **Subarea-Facility**

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

...two men who rented [an Aden apartment]...

رجلان أجرا (شقة في عدن)

### **Path**

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

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

...وعملاء سريون يجوبون (شوارع عدن)

[Telephone lines] were knocked down...

(الخطوط الهاتفية) انهارت

## **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] floating overhead...

(طائرات مروحية عسكرية) تحوم من فوق...

#### **Land**

*Vehicles designed to locomote primarily upon land.*

In one of the worst train accidents in Indian history, [a passenger train bound for Amritsar] was derailed, then struck moments later by [another train bound for Calcutta].

في واحدة من أسوأ حوادث القطارات في تاريخ الهند خرج [قطار ركاب ذاهب الي أمريتسار] عن مساره و بعدها بقليل صدمه [قطار آخر ذاهب الي كالكوتا]

#### **Water**

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

...the Oct. 12<sup>th</sup> bombing of [the U.S.S. Cole].

تفجير الثاني عشر من أكتوبر لسفينة ال يو أس أس كول

#### **Subarea-Vehicle**

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

...[the compartment of the train]...

(حجيرة القطار)

### **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] لدينا (202,000 مركبة في أسطولنا)

## **3.7 Weapon**

*A Gas, liquid, rays or device that is primarily used as device for physically harming or destroying entities (taggable or not).*

### **3.7.1 Subtypes for Weapons**

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

#### **Blunt**

*Blunt weapons are those designed or used as bludgeoning instruments*

He was beaten repeatedly with [a baseball bat]...

ضرب مرارا (بعصا البيسبول)

#### **Exploding**

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

...a small boat packed with [explosives]...

قارب صغير محمل بالمتفجرات

An international investigation led to charges against two alleged Libyan intelligence agents of putting [a bomb] on the plane.

أدى تحري دولي إلى توجيه اتهامات ضد عملي مخابرات ليبين بخصوص وضع قنبلة على الطائرة

#### **Sharp**

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

Knife سكين

#### **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 a ACE-oriented modification of the definition of chemical weapons as set out in the National Security Anti-Terror Law.*

Sarin gas غاز سارين

## Biological

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

Anthrax أنثراكس

## Shooting

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

He was caught attempting to carry [a semi-automatic pistol] through customs...

أمسك وهو يحاول حمل (مسدس نصف آلي) عبر الجمارك

## Projectile

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

Bullets رصاصات

## Nuclear

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

[Nuclear missiles], [hydrogen bombs], [dirty bombs], [deliberately sabotaged nuclear facilities leading to catastrophic meltdowns]

(صواريخ نووية)، (قنابل هايدروجينية)، (قنابل قذرة)، (مباني نووية مخربة عمدا مؤدية إلى كارثة انصهار قلوب المفاعلات النووية)

## 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]. فنتشوه من أجل سلاح.

U.N. [arms] inspectors مفتشوا أسلحة من الأمم المتحدة

[Israeli tanks and gunships] [دبابات و] [مقاتلات جوية] [إسرائيلية]

## 4 Entity Class

Each taggable entity must be assigned one of the following four classes: Negative, Specific, Generic, Underspecified.

### 4.1 Negatively (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.

(لا محامي عاقل) يمكن أن يقبل بهذه القضية

[No one] has claimed responsibility.

[ما من أحد] أعلن مسؤوليته

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

لا يوجد [متهمون محددون] بعد و لكن أعلن المسؤولون أنه سيتم استجواب مجموعات كثيرة من الشرق الأوسط

## 4.2 Specific (SPC)

*An entity is SPC when the entity being referred to is a particular, unique object (or set of objects), whether or not the author or reader is aware of the name of the entity or its anchor in the (local) real world.*

[John's lawyer] won the case.

كسب [محامي جون] القضية

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

بعد ظهر اليوم قامت [مجموعة من المسلمين الغاضبين] بحرق [فندق]

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

إن [لي هوك سيدر] هو رئيس مكتب القدس [لجريدة الواشنطن بوست]

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

[معهد دراسات الحرب والسلام] ب[جامعة كولومبيا]

## 4.3 Generic (GEN)

*An entity is GEN when the entity being referred to is not a particular, unique object (or set of objects). Notice that the mention in question is still understood as being referential in that it points to objects rather than saying of another object that it 'has that property' or some similar notion.*

[Lawyers] don't work for free.

لا يعمل [المحامون] مجاناً

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

يعتبر ما بين 231 و 264 قدم من الماء ضحلاً [للغواصات]

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

يتزامن الشعور بضرورة هذا الاجتماع مع اغتياض الجانبين (الإسرائيلي) و (الفلسطيني) من عنف يوم أمس

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

تحصل [الجماعات المتطرفة] على كثير من الدعم و القوة في هذه الأيام

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

ما تملكه اليابان من مقابل (لقوة بحرية) ملقب رسمياً ب"القوة البحرية اليابانية للدفاع عن النفس"

## 4.4 Under-specified (USP)

*We reserve the term underspecified for non-generic non-specific reference.*

*Underspecified references include quantified NP's in modal, future, conditional, hypothetical, negated, uncertain, question contexts (in all cases the entity/entities referenced cannot be verified, regardless of the amount of "effort").*

[Many people] will participate in the parade.

سيشارك [كثير من الناس] في المسيرة

I don't know [how many people] came.

لا أعلم [كم شخصاً] حضر

Do you know [how many people] came?

هل تعلم (كم شخصاً) حضر؟

We will elect [five new officials].

سوف ننتخب [خمسة أعضاء جدد]

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

A few, some (plural), most, more than half, etc.

*It is important to note that these words don't always indicate that an entity is USP. 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

*For each entity, we record and coreference all mentions of the entity. Mentions will frequently be nested; that is, they will contain mentions of other entities. For example, the phrase*

The president of Ford رئيس فورد

*is a mention of an entity of type Person, and contains the name "Ford", a mention of an entity of type Organization. It is even possible for a noun phrase to contain an embedded mention of the same entity. For instance, the phrase*

The historian who taught herself COBOL

*evokes a Person entity with three mentions: the entire phrase, "who", and "herself".*

### 5.1 Mention Extent

*The extent of a mention consists of the entire noun phrase. In case of structures where there is some irresolvable ambiguity as to the attachment of modifiers, the extent annotated should be the maximal extent. In the case of a discontinuous constituent, the extent goes to the end of the constituent, even if that means including tokens that are not part of the constituent. Thus, in*

I met some people yesterday who love chess.

قابلت بعض الأشخاص الذين يحبون الشطرنج أمس

*the extent of the mention is the entire phrase:*

[Some people yesterday who love chess]

(بعض الأشخاص الذين يحبون الشطرنج) أمس

The extent includes all the modifiers of a nominal phrase, including prepositional phrases and relative clauses.

Each item of punctuation is treated as a separate character. As a rule, we do not include punctuation such as commas, periods, and quotation marks in the extent of a mention unless words included within the extent continue on after the punctuation mark. Possessive endings ('s) are treated as separate tokens, and contractions are split (so that "we're" becomes the two tokens "we" and "re"). Extents must begin at the beginning of a token and end at the end of a token.

## 5.2 Mention Head

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

the hurricane destroyed the new glass-clad skyscraper

الإعصار دمر ناطحة السحاب الجديدة المكسوة بالزجاج

the full mention is

the new glass-clad skyscraper

ناطحة السحاب الجديدة المكسوة بالزجاج

and the head is skyscraper (ناطحة). If the syntactic head of the phrase is a multi-token item, the last token (first token in Arabic) is marked. If the head is a proper name, however, then the whole extent of the name is considered to be the head. In the following examples, the mention is enclosed in brackets and the head is underlined:

[Fred Smith] became [the new Prime Minister].

أصبح [فريد سميث] [رئيس الوزراء الجديد]

The job fell to [Abraham Abercrombie III].

تم اختيار [إبراهام أوبركرومبي الثالث] لتلك الوظيفة

## 5.3 Mention Types

### 5.3.1 Names (NAM)

Proper nouns and nicknames.

[John] (جون)

[Defense Secretary William Cohen] [وزير الدفاع وليام كوهين]

[The Jeluzoon Refugee Camp near [Ramallah]]

[معسكر لاجئين الجلزون بجوار [رام الله]]

[North Dakota State University in [Fargo]]

[جامعة نورث داكوتا ستيت في [فارغو]]

[The house of representatives] [مجلس الشيوخ]

### 5.3.2 Nominal Constructions (NOM)

A nominal is a noun phrase headed by a common noun [The lawyer]

[المحامي]  
[the Yemeni port of Aden]  
[ميناء عدن اليمني]  
[some American executives]  
[بعض الأمريكان التنفيذيين]  
[the southern tip of Africa]  
[الجزء الجنوبي لقارة أفريقيا]  
[a crowd of angry muslims]  
[مجموعة من المسلمين الغاضبين]

### **Nominal Mentions with Multiple-word Heads (MWH)**

*A construction in which two or more heads are employed.*

[20 angry men and women]  
[عشرون رجلا وامرأة غاضبين]  
[20 angry men and women]  
[عشرون رجلا وامرأة غاضبين]  
[Palestinian towns and villages]  
[المدن و القرى الفلسطينية]  
[Palestinian towns and villages]  
[المدن و القرى الفلسطينية]  
[Hollywood actors and writers]  
[ممثلين و مؤلفين مدينة هوليوود]  
[Hollywood actors and writers]  
[ممثلين و مؤلفين مدينة هوليوود]

### **Decision Rules: Name versus Nominal**

1. *When there is a taggable pre-modifier, then we will assume the smallest possible head for the containing entity and tag the pre-modifier entity separately. In cases where there is a GPE pre-modifier, we will always try to capture the GPE separately.*

بوتين العميل السابق [لكي.جي.بي.] دافع عن [المحكمة] [التي] اتهمت البابا و [الخدمات الأمنية]

2. *We will assume that a mention is a nominal unless there is compelling evidence to indicate a single non-compositional mention.*

*In the following, there will be two entities. (The heads are indicated in **bold**).*

The **Arizona** Supreme Court

محكمة (2) اريزونا (1) العليا

*The first entity (1) is a NAME mention of the GPE Arizona, the second is a NOMINAL mention of the ORG Arizona Supreme Court (2).*

### 5.3.3 Pronouns (PRO)

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

[he] [هو]

[they] [هم]

### Headless Mentions (HLS)

Headless mentions are constructions in which the nominal head is not overtly expressed. Although these mentions are technically headless, we will assign as head the right most premodifier (postmodifier) that falls directly before the spot where the head would be.

[the toughest] [الأشد]

[more than 30] [أكثر من 30]

[many] on both sides [كثيرون] من الجانبين

[sixty percent] said [ستون بالمئة] قال

[35] were injured [35] جرح

### 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 and will classify the mention as a pronoun.

[some of the lawyers] [بعض من المحامين]

[one of the houses] [أحد المنازل]

[half of the team] [نصف الفريق]

[each of them] [كل منهم]

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

Examples of non-partitives (two entity mentions):

two members of the team

عضوين من الفريق

### 5.3.4 Premodifier Mentions (Post-modifiers in Arabic)

Premodifier mentions are those mentions, name or nominal, which occur in a modifying position before another word(s).

#### Taggability

We will tag all premodifiers that refer to taggable entities. It is immaterial whether or not the word being modified is a taggable entity.

We will only tag premodifiers when the construction of the premodifier mention is 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.

Taggable	Not Taggable
the [Russian] foreign minister	Stalinist nations
وزير الخارجية [الروسي]	الامة الستالينية
[mountain] regions	mountainous regions
مناطق [الجبل]	المناطق الجبلية
[government] grants	federal grants
منح [الحكومة]	المنح الفدرالية
[Tuscan] food	Machiavellian strategy
غذاء [توسكاني]	الفكر الميكافلي
[Alpine] villages	
قرى [البينية]	
the [Microsoft] [court] case	
قضية [مايكروسوفت] المحكمة *	
[bomb]-making materials	
مواد صناعة [القنابل]	
the [state] budget this year	
ميزانية [الولاية] لهذا العام	

### Name or Nominal Classification

For each Premodifier mention, we must decide whether that mention is a Name or a Nominal. 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 Name 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 Nominal mention.

[Israeli] troops	Name	حشود اسرائيلية
[Republican] voters	Name	ناخبين جمهوريين
[US] Secretary of State] Powell	US – Name, Secretary – Nominal	وزير الخارجية الامريكية باول
[Mr]. Clinton	Nominal	السيد كلينتون
[rooftop] decks	Nominal	
[airline] regulators	Nominal	
[Chrysler] factories	Name	مصانع كريسلر
[union] leaders	Nominal	قادة النقابة
[subway] schedule	Nominal	جدول القطار

## 5.4 Complex Constructions

### 5.4.1 Conjoined Mention Constructions (MNH)

A construction which consists of two (or more) full entity mentions conjoined. The fashion in which this construction appears is: quantification, relative clauses, prepositional phrases, and/or modifiers distributing over two or more entity mentions.

For the purposes of this annotation task, we will annotate these constructions as two separate mentions and will include any modifiers in the extent of the mention to which they are closest.

[[15 men] & [20 women] [who] stormed the barracks] were captured...  
 تم القبض على [[الخمسة عشر رجلا] و [العشرين سيدة] [الذين] فجروا الناقلة]  
 [[Kennedy], [Leiberman], [Gephardt], & [Kerry] all] attended the fundraiser.  
 إن [[كنيدي] و [ليبرمن] و [جفاردت] و [كري] جميعهم] حضروا حفل جمع التبرعات  
 [[Elderly women] and [young children] in the refugee camp]...  
 [ [النساء المسنات] و [الأطفال الصغار] في معسكر اللاجئين ]  
 [Approximately 80 [immigrant teenagers] and 40 [Swedish natives]] died in the blaze.  
 مات في موجة البرد [حوالي] [80 مهاجر في سن المراهقة] [و 40 سويدي الأصل]

### 5.4.2 Appositive Constructions (APP)

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

[[Bill], [John's lawyer]]  
 [[بيل] [محمي جون]]  
 [[Mr. Black], 58, [a victim of the terrorist assault]], told the Associated Press...  
 [[السيد بلاك]، 58 سنة، [وضحية الهجوم الإرهابي]، قال لوكالة الأنباء أسوشيتد برس...  
 [[the show's production company], [Celador]]  
 [[شركة إنتاج العرض] [سيلادور]]  
 [[the heavy lift ship], [Blue Marlin]]  
 [[باخرة الحمولات الثقيلة] [بلو مارتن]]  
 [[Sauache County], [home of the Watchtower]]  
 [[بلدة سواتش] [مقر برج المراقبة]]

### 5.4.3 Mr. and Mrs. Smith Constructions

In these constructions, three distinct entities are referenced.

Mr. Smith  
 Mrs. Smith  
 both Mr. And Mrs. Smith

To capture this information, we will tag these constructions as follows.

Mr. and Mrs. Smith		
[Mr]	السيد	Premod mention of Mr. Smith
[Mrs]	السيدة	Premod mention of Mrs. Smith
[Mr. and Mrs. <u>Smith</u> ]	لا يستعمل هذا التركيب في العربي	Name mention of the couple

## 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.

بكين لن تستمر في بيع القاذفات المضادة للسفن الي ايران

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

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

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

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

أعرب وزير الدفاع وليام كوهين اليوم عن سعادته بأن **بيجينج** لن تستمر في بيع الصواريخ المضادة للمراكب لأيران و ذلك في نهاية جولته التي استمرت 4 أيام كما أوضح أنه حقق تقدما في العلاقات الصينية الأمريكية من الناحية الحربية

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

*Common examples of government seats used to refer to the nation’s government are “The White House” and “The Kremlin”. We will tag these*

### 6.2 GPE name for Sports Team

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

**{[ORG]America}** brought home the gold.

عادت أمريكا بالذهب