

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, NOMPRES).....	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:

[الرجل ذو القميص الازرق]

[*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.

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

تدخلت [القوات التركية] في العراق
The Turkish forces attacked Iraq

[أعضاء هيئة التدريس]
[The staff]

[بعض الطلاب]
[Some students]

[3 جنود اسرَائِيلِيين]
Three Israeli soldiers

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.

[شَاب في الحادية والعشرين من العمر]
Young man aged 21

[القائد العام للقوات المسلحة حسين طنطاوي]
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.

[محاموا شركة المنار]
Al Manar lawyers

[المعتقلين بالسجون الإسرائيلي]
The prisoners in Israeli prisons

قتل [اثنين] في حادث الطريق امس في عمان
Two were killed yesterday in an accident in Amman

اصيب [مصريان] برصاص الجيش الاسرائيلي
Two Egyptians were injured by the Israeli army shooting

[العرب]
[The Arabs] NOM-PER-Group

[المسيحيون]

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]

[Secretary of State] [وزيرة الخارجية الأمريكية مادلين اولبرايت]	
[Secretary of State Madeline Albright] [وزيرة الخارجية الأمريكية مادلين اولبرايت]	PER Atr.

Saints and other religious figures

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

[النبي موسى عليه الصلاة والسلام]

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

[النبي موسى عليه الصلاة والسلام] [The prophet] NOM.PER.Indiv.ATR	PER
---	-----

<p>[[النبي [موسى] عليه الصلاة والسلام] [[The prophet] [Moses], peace be upon him] NAM.PER.Indiv.SPC</p>	
---	--

الكنيسة الكاثوليكية تبحث في ادراج [البابا يوحنا بولس الثاني] في عداد القديسين
The catholic church is discussing considering Pope John Paul the second as one of the saints

<p>[[البابا [يوحنا بولس الثاني] [Pope] NAM.PER.Indiv.ATR</p>	PER
<p>[[البابا [يوحنا بولس الثاني] [Pope John Paul II] NAM.PER.Indiv.SPC</p>	

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.

..... إذا كنت تؤمن [بالله] فلا بد
If you believe in [god], you must... name mention

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

[باتمان] أصبح صورة شعبية

[Batman] has become a popular icon

الكثير من الاطفال يعشقون [بكار]

A lot of kids love Bakar

[بِقَاطِ] يعلم الاطفال حسن التصرف

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 اليف

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.

وافقت [المحكمة العليا] في واشنطن [على النظر في احدى دعاوى الاستئناف

[[The Supreme Court] in Washington] agreed to discuss the appeal

اعتقلت [الشرطة] السعودية [اراهابيين

[[The Saudi [police] arrested two terrorists

أغلفت الولايات المتحدة [سفارتها] في الدوحة [

The United States closed [[it's embassy] in Duha]

قامت [الشرطة] المصرية [بتعقب فلول الارهابيين

The Egyptian police is looking for the terrorists

اعتمدت [الخارجية] اللبنانية [ثلاث سفراء جدد امس

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.

اعلنت [[شركة «ال جي الكترونيكس»] الكورية الجنوبية] انها تتوقع تحقيق مبيعات في سوق المغرب تتجاوز 36 مليون دولار السنة الجاري

[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.

[[الجامعات] اللبنانية] بها فروع لجامعات اوروبية وامريكية

The Lebanese Universities have branches of European and American Universities

[[المدارس] المصرية] مزدحمة بالتلاميذ

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.

عقدت [[فرقة ميامي] الكوبينية] اتفاقا مع شركة روتانا للإنتاج الفني

[Miami group] signed an agreement with Rotana production

[[فرقة نجيب الريحاني] المسرحية] قدمت اروع الاعمال الكوميديا

[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:

[كتائب شهداء الأقصى]

[*Al Aqsa Martyr's Brigade*]

[نمور التاميل]

[*Tamil Tigers*]

2. Political Parties:

[الحزب الجمهوري]

[*Republican Party*]

[حزب العمل]

[*Labour Party*]

[الحزب الوطني الديمقراطي]

3. Political Advocacy Groups and Think Tanks:

[الاتحاد الأمريكي للحريات المدنية]

[*ACLU*]

[معهد كاتو]

[*The Cato Institute*]

4. Professional Regulatory and Advocacy Groups:

[جمعية المحامين الأمريكيين]

[*The American Bar Association*]

[الجمعية الطبية الأمريكية]

[*The American Medical Association*]

[نقابة المهندسين]

[*The engineering syndicate*]

5. Charitable Organizations:

[الصليب الأحمر]
[The Red Cross]

[صندوق الأمم المتحدة للأطفال]

[اليونيسيف]

[UNICEF]

[أطباء بلا حدود]
[Doctors Without Borders]

6. International Regulatory and Political Bodies:

[الأمم المتحدة]
[UN]

[منظمة حلف الشمال الأطلسي]
[NATO]

[البنك الدولي]
[The World Bank]

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

ان معارك بين [الطالبان] والمعارضة الافغانية تدور حاليا على بعد بضعة كيلومترات من حدود هذه الجمهورية السوفياتية السابقة في آسيا الوسطى.

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

[معاونات صندوق النقد الدولي]
[International Monetary Fund] aid

[حزب الله]
[Hizbollah]

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.

ملايين العرب يشاهدون [قنوات ال آل بي بي سي. وال أي آر تي]
Millions of Arab watch [LBC and ART channels]

[قناة الجزيرة]
[Gizera channel]

[جريدة الحياة]
[Al Hayat newspaper]

[نيو يورك تايمز]
[New York times]

[ال بي بي سي]
[BBC]

[مجلة سيدتي]
[Sayedaty magazine]

قال [التلفزيون السوداني الرسمي] أن هجوما وقع في مسجد السنة المحمدية في قرية الجرافة بالقرب من أم درمان شمال الخرطوم
[The official Sudanese TV] declared an attack on Al Sonna Al Mohamadia mosque in Garafa village close to Om Dorman nort to Khartoum

نوالى تقديم الأنباء من [صوت أمريكا في واشنطن]
We will continue our news from [VOA from Washington]

Religious (REL)

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

[اتحاد الأساقفة الألمان]
[The German Bishop Union]

[الفاتيكان]
[The Vatican]

[ادارة الازهر]

[The Azhar administration]

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 الى جذب اكبر عدد من النساء الاجنبيات الى البرازيل لاختصاصهن لعمليات التجميل

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

يواجه الأطباء إصابات معقدة وأحداثا مأساوية بشكل يومي ونقصا في عدد العاملين في [المستشفيات]

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

إن [المستشفيات الفلسطينية] تعاني من مصاعب جمة حيث يواجه الأطباء إصابات معقدة وأحداثا مأساوية بشكل يومي ونقصا في عدد العاملين في المستشفيات

[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

[جمعية الفيزياء الصينية عبر البحار]

[Chinese Overseas Physics Society]

[معهد الدراسات النووية]

[The Neoclear Research Institute]

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.

[فاز] النجمة [علي شباب الساحل امس في الدوري اللبناني]

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

وصل [فريق ميلان] الي نصف النهائي في كرة القدم

[Milan] reached the semifinal in soccer

[لجنة الفلبين الأولمبية]

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

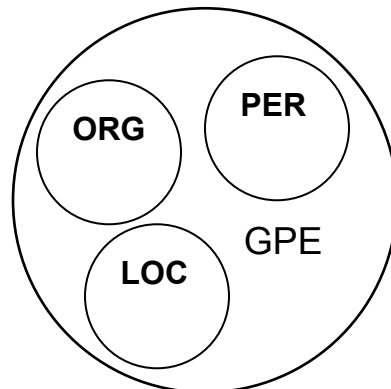
Whenever an organization takes an action, there are people within or in charge of the organization that one presumes actually made the decision and then carried it out. Thus many organization mentions could be thought of as metonymically referring to people within the organization.

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:

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

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

ويشمل التقرير أسماء أربع وعشرين دولة معظمها من [آسيا] وأمريكا الوسطى والجنوبية

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

تحاول [أوروبا] ان تتحد

[Europe] is trying to unite

Nation

Taggable mentions of the entirety of any nation.

وافقت **[السعودية]** على تخفيف أعباء الديون المتركمة على اليمن
[Saudi] agreed to wave the Yemeni dept

اصيب مصريان برصاص الجيش **[الاسرائيلي]**
Two Egyptians were injured by the [Israeli] army

State-or-Province

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

[محافظة القاهرة]

[Cairo Governorate]

كما وقع حادثان اخران في **[محافظة القليوبية]** شمال القاهرة حيث لقي رجل واربعة نساء مصرعهم
Two more accidents in [Kalyoubia Governorate] north of Cairo where a man and four women died

محافظ **[سا ليسيبرج]** شواسبيرجر

[Salzburg] governor Schausberger said...

تم التوقيع علي وثيقة الاستقلال في **[بنسلفانيا]**

The liberty agreement was signed in [Pennsylvania]

County-or-District

Taggable mentions of the entirety 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

[مرتفعات الجولان]

[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.

عقد سفيراً أوزبكستان وأفغانستان في إسلام آباد اجتماعاً ثانياً في [العاصمة الباكستانية]

The ambassadors met at the [Pakistani capital]

اصيب عشرة اشخاص على الاقل بجروح في [مدينة اشمون] في حادث طريق بالامس

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

ذكر المعهد الدولي للدراسات الاستراتيجية في [لندن] أن تجارة الأسلحة الدولية انخفضت

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

صرح مصدر في السفارة الاميركية في [الرياض] اليوم الثلاثاء ان وزيرة الخارجية الاميركية مادلين اولبرايت ستصل الى العاصمة السعودية في وقت لاحق

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

يبدو أن احتفالات عيد الميلاد في [مدينة بيت لحم] بالذات والتي ولد فيها السيد المسيح ستكون مناسبة كئيبة هذا العام

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

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

Latin America أمريكا اللاتينية

The West الغرب

The East الشرق

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.

إن الشرطة [الفلسطينية] اوقفت فلسطينيا يشتبه في انه يتعامل مع اسرائيل

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

The following entities are also tagged as GPE /Special

ايرلندا الشمالية

Northern Ireland

تايوان

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.

جنوب الولايات المتحدة

the southern United States

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

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

[NATO] peacekeepers arrived in the valley before nightfall.

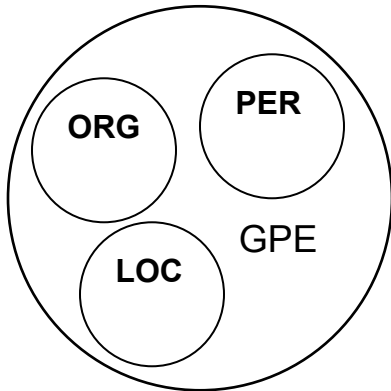
3.3.3 Formulaic GPE Constructions: Nested Region Names

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:

[طرابلس، لبنان]
Trabols, Libanon
 [القاهرة، مصر]
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.



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

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

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

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

The following sections give particular guidelines for frequently encountered cases, with examples.

GPE.ORG

GPE.ORG is used for GPE mentions that refer to the entire governing body of a GPE. It is important to differentiate between a part of the government (the executive branch, the courts) and the entire governing body. Below are some examples of contexts in which GPE.ORG should be used.

Political Communication and Decision-making

ORGs are responsible for decisions to take military actions. ORGs are also responsible for political communication events such as announcements, agreements, statements, denials, expressions of approval and disapproval, etc. So, if *China* agrees to something, *China* is a GPE.ORG.

أصدرت [الولايات المتحدة] تقريرها السنوي عن الدول الرئيسية التي تهرب المخدرات
[USA] published her annual report of the main countries that export drugs

أعلن مسؤولون صوماليون ان [حكومتهم] توصلت الى اتفاق مع اثيوبيا الخصم السابق على أن يعمل
البلدان معاً
*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.

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

أعلنت الادارة أن سونابند يمكنه اللجوء الي [المحكمة السويسرية العليا]
(postmodifier in Arabic)
*The department said Sonabend can appeal to {[GPE.ORG] Switzerland}'s
supreme court. (possesive)*

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.

[اليابانيون] يحملون مسؤولية ضخمة لحروب النصف الأول من القرن

[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:

أفاد بتلر بأنه يمكن إنهاء الحظر الاقتصادي إذا سمح [العراقيون] للمفتشين بأداء واجبهم

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

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

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.

أن مراسلتنا أليشاراي في مكتبنا الإقليمي في [آسيا] أفادت بأن عملية عزل الرئيس استرادا قد تستغرق ستة أشهر

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

أعلنت في [مدينة الرياض] في السعودية عن إصابة ثلاثة بريطانيين بجراح يوم أمس أثر وقوع انفجار في سيارة كانت تقلهم

Declared in [Riyadh] city in Saudi Arabia that 3 British were injured in an explosion in their car

لم يعلن أحد مسؤوليته عن أي من الحادثين الذين وقعا بعد أسبوع من إلقاء قنبلة على مبنى السفارة الأمريكية في [العاصمة اليمنية صنعاء]

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

إن مساحة **[فرنسا]** 547090 كيلو متر مربع
[France] has an area of 547,090 square kilometers

المقاتلات الجوية الأمريكية حلقت فوق **[أفغانستان]**
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.

[طرابلس لبنان]
[Trabols, Libanon]
[القاهرة مصر]
[Cairo, Egypt]

Dateline mentions of GPEs are given a location role.

[بيروت – لبنان 1-5-2005]

[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.

القوات **[الاسرائيلية]**
[Israel]} troops

رجال بوليس **[نيو يورك]**
[New York] policemen

رئيس الوزراء **[البريطاني]**
[British] Prime Minister

طائرة المتابعة [الأمريكية]
[**U.S.**] surveillance aircraft

العلم [العراقي]
[**Iraqi**] flag

مواجهات اندلعت بعد ظهر اليوم الاربعاء بين فلسطينيين وبين الجيش [الإسرائيلي] في مدينة بيت لحم وضواحيها في الضفة الغربية
Conflicts on Wednesday in Jorshlim and suburbs between Palestinians and the [**Israeli**] army

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.

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

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.

حظرت السلطة الفلسطينية التجمعات المدافعة عن [العراق] ولكن هذا الحظر لم يؤخذ في الاعتبار

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 Armstrong 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.

[نيويورك فريق قوي]

[**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.

"ستكون المناقشات عنيفة" قال توماس ساير مندوب الحزب الديمقراطي - [أوهايو]

"This is going to be a brutal fight," said Rep. Thomas C. Sawyer [**Ohio**],

صرح مجدي الصياد - الحزب الوطني - [بنى سوييف] - بان الحكومة ستعزز محدودتي الدخل

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 شارع فؤاد

17 Foad St.

Boundary

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

اعلن الجيش الاسرائيلي مساء اليوم الجمعة احباط محاولة تسلل "المجموعة ارهابية" على [الحدود] مع لبنان

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

قال علماء إن [الكوكب المريخ] ربما كان في بداياته أرضاً للبحيرات بعد اكتشاف طبقات

لصخور رسوبية على غرار تلك الموجودة على [الأرض]

Scientists said that **[Mars]** may began as lakes after discovery of layers of rocks similar to those on **[earth]**

يدور الصاروخ حول **[الأرض]** بسرعة مهولة

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

Water-Body

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

بدأت امس تدريبات عسكرية مصرية سعودية مشتركة جوية وبحرية في **[البحر الاحمر]**

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

Land-Region-natural

Geologically or ecosystemically designated, non-artificial locations.

قتل مائة وسبعون شخصا بعد اندلاع حريق في قطار يحمل متزلجين إلى منتجع شتوي في **[جبال الألب]**

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

Region-International

Taggable locations that cross national borders.

ويحظر ايضا سفر المسؤولين في حركة طالبان حتى حدود رتبة وزير، الى **[الخارج]**

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

Region-General

Taggable locations that do not cross national borders.

وكان القائد الروسي في المنطقة قد قال الأسبوع الماضي إن قواته ستغير استراتيجيتها بمغادرة القواعد التي تكفل سلامتها ونشر وحدات صغيرة في **[جميع أنحاء المنطقة]**

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

بعد ظهر اليوم الخميس حدث هجوم مسلح على سياراتهم في **[شمال الضفة الغربية]**

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

قتل اثني عشر شخصا على الأقل في هجوم في **[الجزء]** الذي تسيطر عليه الهند من كشمير

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

[السودان الجنوبي]

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

[جنوب السودان]

[شمال الرياض]

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-genitive 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

[جنوب السودان]

[*South the Sudan*]

The Arabic construction of 2) 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."

"الأرنب مختبئ خلف هذه الصخرة"

"The rabbit is hiding behind that rock."

"لدي فكرة في مخيلتي"

"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:

"لقد أسقط المفاتيح علي الأرض"

"He dropped the logs on the ground."

"لقد أعاد المصباح الي مكانه مرة أخرى"

"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.

[مطار لاجوارديا] في نيو يورك كان كارثة هذا العام

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.

...القطار كلن ينتقل مباشرة من [مصفاة النفط] إلى المصهر...

...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.

مئات الفلسطينيين أدخلوا إلى [المستشفيات الإسرائيلية] لتلقى العلاج

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

أمر الجيش الإسرائيلي الفلسطينيين بمغادرة [مكاتب الاتصال المشتركة في الضفة الغربية وقطاع غزة]

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

رجل أعمال سعودي يسعى إلى تجنب تسليمه إلى الولايات بتهم تتعلق بتفجير [السفارتين الأمريكيتين]

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

56 اعتقلوا فجرًا من [منزلهم]

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

وحاول اقرباء القتيل والجرحى اشعال النار في [محطة للبنزين في المدينة]

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.

الجلسة الاولى من المحاكمة جرت في حضور المحامين امام [غرفة الاتهام الخامسة في المحكمة العسكرية]

The first session of the trial was attended by the lawyers in [the 5th. Charge room in the military court]

وقد تمكن رسام من [**زئزئته** في سياتل] من الاستماع الى اقوال الشهود في مونتريال بفضل نظام كاميرات بحلقة مغلقة

*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.

كانت القوات الاسرائيلية تمنع المواطنين من التنقل عبر [**الشوارع الرئيسية**]

*The Israeli forces was preventing the people from moving around [**the main streets**]*

[**الخطوط الهاتفية**] انهارت

*[**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.

[**طائرات مروحية عسكرية**] تحوم من فوق

*... [**military helicopters**] float overhead...*

وصل الى جدة امس، رئيس الوزراء الباكستاني السابق نواز على متن [**طائرة سعودية خاصة**]

*The former Pakistani Prime minister Nawaz Sherif arrived to Gedda in a [**private Saudi plane**]*

وصول مجموعة من الفلسطينيين الجرحى إلى العراق على متن [**طائرة فلسطينية**] للعلاج في المستشفيات العراقية

*A group of injured Palestinians arrived to Iraq on [**Palestinian plane**] for treatment in the Iraqi hospitals*

Land

Vehicles designed to locomote primarily upon land.

وافادت الاذاعة العسكرية ان سائق [السيارة] لم يصب الا انه كان في حال صدمة

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

هاجمت القوات الإسرائيلية الميناء اللبناني بواسطة [حافلة برمائية]

The Israeli troops attacked the Lebanese harbour by [an amphibian truck]

خرج اليوناني المسلح الذي كان يحجز 35 شخصا من [الحافلة] بمفرده

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

وكان سائق [شاحنة] محملة بمواد البناء [قد اصيب الاربعاء بجروح برصاص اسرائيلي اطلق عليه اثناء مروره قرب "بوابة فاطمة"

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.

ويتوقع ان ينتهي بناء [الغواصات] الذي سيجري في كوريا الجنوبية في 2009

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

ان الكويت تصرفت على نحو خطير عندما احتجزت [خمسة قوارب] يشتبه أنها كانت تنقل سلعاً عراقية

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.

[حجيرة القطار]

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

Underspecified

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

لدينا [202,000 مركبة] في أسطولنا
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

قتلت القوات الإسرائيلية ثلاث فلسطينيين من [رماة **الحجارة**] بإطلاق النار عليهم في الضفة الغربية وقطاع غزة

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

وقد توقفت المباراة لمدة ربع ساعة بعد ان رميت [عصا من البلاستيك] باتجاه حارس ميلان

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

قارب صغير محمّل [بالمفجرات]
...a small boat packed with [**explosives**]...

[**قنبلة**] انفجرت بالقرب من مستوطنة يهودية في غزة

[A **bomb**] exploded close to Jewish settlement in Gaza

أطلقت قوات الشرطة الهندية في كشمير امس [قنابل الغاز المسيل للدموع] علي المتظاهرين

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

واغتيل المعارض دريوش فاروهار وزوجته بروانة طعنا [بالسكين]

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.

[غاز سارين]

[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.

وجد بالخطاب مسحوق ابيض ثبت انه يحتوي علي [انثراكس]

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.

ونزل الرجل من الحافلة من دون [البندقية التي هدد بها الرهائن]

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

تمكن الضابطين من ادخال [مسدس] الى طائرة الركاب وتم خطف الطائرة بهذا المسدس

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

قال مسؤول في المستشفى ان مؤيد وصل المستشفى وقد اصيب [بعيار ناري] في الرأس

An official in the hospital said that Moayad reached the hospital injured with [bullet] in the head

أعلنت طهران انها وصلت الى عتبة الاكتفاء الذاتي في صنع [صواريخ أرض - بحر]
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.

[صواريخ نووية]
[Nuclear missiles]

[قنابل هيدروجينية]
[Hydrogen bombs]

[قنابل قذرة]
[Dirty **bombs**]

أعلن مسؤولون في الجيش الاميركي ان حريقاً دمر منشأة مساندة [للسواريخ النووية] تابعة للجيش في ولاية نورث داكوتا

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

وقال: اننا نشاطر روسيا بالتأكيد الحرص على خفض مستوى [الاسلحة النووية الاستراتيجية] 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.

فتشوه من أجل [سلاح]
They searched him for [a **weapon**].

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

أصيب عدد من الفلسطينيين بنيران جنود إسرائيليين خلال اشتباكات ومعارك [بالأسلحة] في عدة أماكن
Number of Palestinians were injured by the Israeli troop's fire, through conflicts and attacks with [**weapons**] in different places

وكان رئيس الدولة اليمنية دعا لارسال [السلاح] والمقاتلين الى الشعب الفلسطيني
*The Yemeni president asked to send weapons and troops to the
Palestinian people*

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.

لا يوجد [محامي عاقل] يتراجع في مثل هذه القضية
[No sensible lawyer] would take that case.
لم يدعي [احد] بمسئوليته عن الحادث
No [one] has claimed responsibility.

لا يوجد [متهم محدد] حتي الان لكن المسؤولين قالوا ان عديد من مجموعات الشرق الاوسط من المتوقع ان تخضع للتحريات

*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.

إنهم ليسوا [بمحامين]

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

[الجيش الاسرائيلي]

[[The Israeli] army]

Israeli is an ATR; but in

قمة [العرب]

[The Arab] summit

Arab is considered 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 Edefa, only the postmodifier has determiner (the). In the following examples where the title is followed by proper name, the title is annotated as ATR.

[الرئيس الاسد]

[President Al-Asad]

[الدكتور علي]

[Dr. Ali]

[الكيميائي سمير]

[The chemist Samir]

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

[الرئيس المصري]

[The Egyptian president]

[الطبيب البارع]

[*The smart physician*]

[الطالب 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.

تنتج [فرنسا] نبيذ أفضل مما تنتجه نيو جيرسي

[*France*] produces better wine than New Jersey.

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

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

[3 جنود اسرانييليين]

[*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:

[المحامون مخلصون في عملهم]

[*The lawyers*] are honest

[المهندسون مبتكرون]

[*The engineers* are creative]

[الممرضات يعتنون بالمرضى]

[*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:

[الأطباء يحبون قراءة الأبحاث الحديثة]

[*The physicians* like to read new researches]

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:

الممرضات [الممرضات] يجب ان يكونوا عطفين

[The nurses] should be kind

القادة [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 الأمريكان لا يحبون البيتزا

[Some of the Americans] don't like pizza

سبعة أشخاص أصيبوا في الحادث، ثلاثة منهم إصابتهم خطيرة

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.

الآف 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

الآف 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).

سيأتي [[الرئيس] الاسد] خطاب باكر و سيستعرض فيه انجازاته و المشاكل التي يبحث لها عن حل و سيصحب الرئيس اغلب الوزراء

***[President [Al Assad]]** will give talk tomorrow. He will state all his achievements and the problems that he still looking for solution to them. The president will be accompanied by most of the ministers*

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

[[رئيس] فورد]
[The president of [Ford]]

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

[[الرئيس] لحدود]
[President [Lahoud]]
[[الرئيس] مبارك] [الذي] رشح نفسه لفترة رئاسية جديدة
[[President [Mubarak]] [who] is running for new presidency period]

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.

Simple Mentions	Named (NAM, NAMPRE)
	Nominal (NOM, NOMPRES)
	Pronominal (PRO)
	Headless nominal (HLS)
	Partive Constructions (PTV)
	WH-Question words and specifiers (WHQ)
Complex Mentions	Apposition Constructions (APP)
	Complex constructions taking a relative clause (ARC)

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 رشح 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 رشح 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.

20 رجل وامرأة غضبي	
[20 رجل وامرأة غضبي] [20 angry <u>men</u> and women]	Nom mention
[20 رجل وامرأة غضبي] [20 angry men and <u>women</u>]	Nom mention

بيل كلينتون وجيمي كارتر من الرؤساء السابقين	
[بيل كلينتون وجيمي كارتر من الرؤساء السابقين] [Bill Clinton and Jimmy Carter who are both former presidents]	Name mention
[بيل كلينتون وجيمي كارتر من الرؤساء السابقين] [Bill Clinton and Jimmy Carter who are both former presidents]	Name mention

الحمالون والتباعون في واشنطن	
[الحمالون والتباعون في واشنطن] [The movers and shakers in Washington]	Nominal mention
[الحمالون والتباعون في واشنطن] [The movers and shakers in Washington]	Nominal mention

5.1.2 Mention Head

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

[اعضاء] هيئة التدريس] قامو برحلة

[The staff] went for trip

the full mention is

[اعضاء هيئة التدريس]

[The staff]

and the head is **اعضاء**. 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 **[فريد سميث]** رئيس الوزراء الجديد

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

تم اختيار **[ابراهيم أبركرومبي الثالث]** لتلك الوظيفة

The job fell to **[Abraham Abercrombie III]**.

5.1.3 Names (NAM)

Proper nouns and nicknames.

[John] [جون]

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

[معسكر لاجئين الجلزون بجوار **[رام الله]**]

[The Jeluzoon Refugee Camp near [Ramallah]]

[جامعة نورث داكوتا ستيت في **[فارغو]**]

[North Dakota State University in [Fargo]]

5.1.4 Nominal Constructions (NOM)

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

[كبير العائلة] كلمته مسموعة

[The biggest of the family] has strict word

[الاشداء] هم من مارسوا الرياضة

[The toughest] are those who practices sports

[المحامي]

[The lawyer]

[بعض التنفيذيين الأمريكيين]
[some American executives]

[الجزء الجنوبي لقارة أفريقيا]
[the southern tip of Africa]

5.1.5 Pronouns (PRO)

Pronouns with the exception of wh-question words and the specifier 'that'. A pronominal paradigm in 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

Person	Singular	Plural	Dual
3rd (m)	huwa	hum	humā
3rd (f)	hiya	hunna	
2nd (m)	anta	antum	antumā
2nd (f)	anti	antunna	
1st	ana	naḥnu	(n/a)

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] انتظر عدة ايام قبل ان يطلب الطوارئ

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

المانيا [حيث] اعتقلوا بالامس

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 [جرح 35]

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):

[عضوين من الفريق]

Two members of the team

5.1.9 Postmodifier Mentions (NAMPRE, NOMPRES)

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 الاسرائيلية]

[The Israeli troops]

[الناشرون اللبنانيون]

[The Lebanese] publishers

[اللغة الألمانية]

[The German] language

[الكتب الإنجليزية]

[The British] books

Taggable	Not Taggable
وزير الخارجية [الروسي] the [Russian] foreign minister	الامم الستالينية Stalinist nations
مناطق [الجبل] [mountain] regions	المناطق الجبلية mountainous regions
منح [الحكومة] [government] grants	المنح الحكومية federal grants
قري [الالب] [Alpine] villages	Machiavellian strategy

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 (NOMPRES). 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 NOMPRES mention.

[القوات الاسرائيلية] [Israeli] troops	NAMPRES
[الناخبين الجمهوريين] [Republican] voters	NAMPRES
[مصانع كرايكلر] [Chrysler] factories	NAMPRES

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 characterizes 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.

[الرئيس الروسي][بوتين]
[The Russian president][Valdmir Putin]]
 [رئيس روسيا] [بوتين]
[The Russian president] [Putin]]

[امين عام الامم 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

[[البابا] يوحنا بولس الثاني]] C [Pope [John Paul II] NOM.ATR	PER.Individual
[يوحنا بولس الثاني] [John Paul II] NAM. SPC	

Apposition:

[الرئيس الروسي بوتين] [The Russian president Putin] APP.SPC	PER.Individual
--	----------------

[الرئيس الروسي] [The Russian president]Nom. Atr.	
[بوتين] [Putin] NAM. SPC	

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.

[[[رئيس روسيا][بوتين][الذي] اصبح في السلطة عام 2000]

[[[The Russian **president**] [**Putin**]]APP [who] took power since 2000]ARC

[رئيس روسيا بوتين الذي اصبح في السلطة عام 2000] [The Russian president Putin who took power in 2000] ARC	PER.Individual
[[رئيس [روسيا][بوتين]] [[The Russian]president]Putin] APP	
[رئيس روسيا] NOM.PER [The Russian president]	
[بوتين] [Putin]	
[الذي] [who]	

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.

[بكين] لن تستمر في بيع القاذفات المضادة للسفن الي ايران

[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.

[امريكا] عادت بالذهب

[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 administrates the hospital.

قال المتحدث الرسمي [للمستشفى] ان سعيد 42 عام لقي حتفه بعد ساعة في [مستشفى القديس جون
ماكومب] و المتهم توفي بعد ذلك في نفس الليلة

*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.

CE	Argument 1	Cross type metonymy
CE1(Arabic)	مستشفى القديس جون ماكومب St. John Macomb Hospital	CE1 (ORG. Med/Sci)
CE2 (Arabic)	للمستشفى hospital	CE2 (FAC. Buid/Ground)

Appendix:

Some annotation examples:

FAC.Building

Name	Decision
berkeley marina	FAC.Building
gitmo	FAC.Building
guantanamo bay	FAC.Building

FAC.Subarea

Name	Decision
rose garden	FAC.Subarea

GPE.Cluster

Name	Decision
balkans	GPE.Cluster
Commenwealth	GPE.Cluster
East	GPE.Cluster
EU	GPE.Cluster
european union	GPE.Cluster
group of eight industrial powers	GPE.Cluster
middle east	GPE.Cluster
mideast	GPE.Cluster
south asia	GPE.Cluster
southeast asia	GPE.Cluster
West	GPE.Cluster

GPE.Continent

Name	Decision
Africa	GPE.Continent
Antarctic	GPE.Continent
antarctica	GPE.Continent
Asia	GPE.Continent
asian	GPE.Continent
Europe	GPE.Continent
north america	GPE.Continent

GPE.County

Name	Decision
fairfax county	GPE.County

GPE.Nation

Name	Decision
America	GPE.Nation
Australia	GPE.Nation
axis of evil	GPE.Nation
cayman islands	GPE.Nation
china	GPE.Nation
congo	GPE.Nation
england	GPE.Nation
iran	GPE.Nation
iraq	GPE.Nation
israel	GPE.Nation
Korean	GPE.Nation
spain	GPE.Nation

GPE.Pop

Name	Decision
atlanta	GPE.Pop

baghdad	GPE.Pop
hong kong	GPE.Pop
key largo	GPE.Pop
macao	GPE.Pop

Gaza

Name	Decision
Gaza	GPE.Pop

GPE.Special

Name	Decision
mainland (China)	GPE.Special
palestine	GPE.Special
Taiwan	GPE.Special

GPE.State

Name	Decision
gaza strip	GPE.State
northern ireland	GPE.State
west bank	GPE.State

LOC.Land-Reg

Name	Decision
azores islands	LOC.Land-Reg
canary islands	LOC.Land-Reg
continental slope	LOC.Land-Reg
diego garcia	LOC.Land-Reg
ellis island	LOC.Land-Reg
everest	LOC.Land-Reg
him la yas (Himalayas)	LOC.Land-Reg
korean peninsula	LOC.Land-Reg
mount everest	LOC.Land-Reg
neichangshan islands	LOC.Land-Reg

sekamian plateau	LOC.Land-Reg
urbuinano island	LOC.Land-Reg

LOC.Reg-Gen

Name	Decision
central jakarta	LOC.Reg-Gen
continental ~usa	LOC.Reg-Gen
east coast	LOC.Reg-Gen
east texas	LOC.Reg-Gen
green zone	LOC.Reg-Gen
midwest	LOC.Reg-Gen
new england	LOC.Reg-Gen
North	LOC.Reg-Gen
Northeast	LOC.Reg-Gen
siberia	LOC.Reg-Gen
South	LOC.Reg-Gen
south florida	LOC.Reg-Gen
Southeast	LOC.Reg-Gen
southside	LOC.Reg-Gen
southwest	LOC.Reg-Gen
upper east side	LOC.Reg-Gen

LOC.Reg-Int

Name	Decision
arctic	LOC.Reg-Int
eurasia	LOC.Reg-Int
far east	LOC.Reg-Int
greater asia	LOC.Reg-Int
indochina	LOC.Reg-Int
Keshmere (Kashmir)	LOC.Reg-Int
Manchiu	LOC.Reg-Int
manchuria	LOC.Reg-Int

LOC.Water

Name	Decision
arabian sea	LOC.Water
atlantic	LOC.Water
chesapeake bay	LOC.Water
euphrates river	LOC.Water
indian ocean	LOC.Water
khor abdallah river	LOC.Water
khoser river	LOC.Water
lake geneva	LOC.Water
lake superior	LOC.Water
pacific ocean	LOC.Water
red sea	LOC.Water
san francisco bay	LOC.Water
strait of gibraltar	LOC.Water
tigris river	LOC.Water
yellow sea	LOC.Water

Land Areas Named by Water References

Name	Decision
Caribbean	LOC.Water, LOC.Reg-Int
gulf	LOC.Water, GPE.Cluster
mediterranean	LOC.Water, LOC.Reg-Int
persian gulf	LOC.Water, GPE.Cluster

ORG.Non

Name	Decision
kizilay	ORG.Non

Palestine

Name	Decision
palestinians	PER.Group, GPE.Special

palestinian authority	ORG.Gov
PLC	ORG.Gov

NOT TAGGABLE

Name	Decision
Latin	NOT TAGGABLE