

EDT Metonymy Annotation Guidelines

Version 2.2 – 2001-06-05

1 Introduction

This document summarizes the changes and additions to the ACE EDT annotation guidelines resulting from the exploration of metonymy initiated at the ACE kickoff meeting at the end of November, 2000, and pursued through the “Camp 0/A/B” series of guideline development and exploratory annotation efforts. Portions of this document are edited versions of guidelines produced by Camp A and Camp B during that experiment. As soon as possible, the material in this document should be integrated with the EDT task definition, the addenda to that document edited by Adam Meyers, and the relevant sections of Lisa Ferro’s nominal entity specification to produce an updated task definition and a combined guidelines document.

The remainder of this introduction summarizes the proposed two-fold handling of metonymy, and Section 2 then describes the resulting changes that will need to be made in the Task Definition and Annotation Guidelines. Sections 3 and 4 give the more detailed proposed guidelines for the classic metonymies and for the roles to be marked on GPEs.

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.

ACE EDT annotation distinguishes two types of metonymic usage: classic metonymies and composites. Classic metonymies make reference to two entities, one explicit and one indirect reference. Common examples are cases of capital city names standing in for national governments, as shown above. Other common examples involve facilities and organizations, which are closely related in that organizations typically have facilities, and facilities are typically owned and administered by organizations. Thus when a facility is mentioned, the organization is sometimes also referenced. So, in *the museum announced its new exhibit*, the entity *museum* is a facility that houses artwork, but in this context it is the organization running the museum that is doing the announcing. In cases like this, where both entities are expressed by the same phrase, two entity mentions should be marked, one for each of the corresponding references. If only one entity is expressed, then only one entity mention is marked. In the above example, the annotator would mark mentions of a FAC and an ORG entity for *the museum*.

The second metonymy type applies only to GPEs. GPEs are composite in nature, typically having a government, a populace, and a geographic location, as well as some more abstract notion of statehood. France, for example, is a composite of its people, its physical territory and its government. In some contexts one of those attributes of France might be emphasized more than the others. In these cases, annotators should assign to

each mention a role indicating which attribute is salient, rather than marking another mention of a separate entity. So, in the example *he visited France*, the GPE mention *France* is assigned a role LOC. This captures the sense of France that is most dominant given the context. GPE mentions may be assigned GPE, ORG, PER, or LOC roles. The notation “X.Y” is used to denote a mention with role Y referring to an entity of type X.

2 Changes to the Task Definition and Annotation Guidelines Document

At the highest level, the proposed approach is to annotate metonymy (other than that involving the aspects of GPEs) directly with overlapping mentions, while the aspects of mentions of GPE entities will be distinguished by assigning mention roles. This section outlines the resulting changes that will need to be made in the task definition and annotation guidelines document.

Section 2 of the Task Definition document lists four subtasks within the EDT complex. A fifth subtask will need to be added to this list, “the recognition of mention roles”, defined as an optional task that measures a system’s ability to correctly determine the predominant role for each mention of a composite entity. In the current version of the guidelines, only GPE entities will be assigned roles. Role recognition performance will be measured only for those GPE mentions that are correctly detected.

A number of changes will have to be made in the Annotation Guidelines Section 6.2, where the types of entities are defined. The last paragraph of Section 6.2.2 currently states that a mention of an Organization entity that refers to the buildings housing the organization or the people who belong to it should still be marked as simply an Organization entity mention. In the new version, that paragraph will be replaced with text stating that mentions of Organization entities will always be marked as Organizations, but that in cases where they refer to other entities like facilities or persons, they should also be marked as mentions of those other entities.

Similarly, in Section 6.2.5 on Facilities, the last paragraph will be replaced with text stating that mentions of Facility entities will always be marked as such, but when they are used to refer to organizations or people, they will also be marked as mentions of those other entities.

In Section 6.2.3 on GPEs, a sentence will be added stating that GPE mentions will be annotated with roles that specify the predominant aspect of the composite GPE that is being referenced in each case, and pointing to a new section 6.7 that will briefly describe how GPE roles are determined.

Other more minor changes may also need to be made in the Task Definition and Annotation Guidelines document, for example, noting that clusters of GPEs (Europe, Africa, ...) will now be annotated as GPEs rather than as LOCs.

3 Classic Metonymies (Annotated with Separate Mentions)

Classic metonymies are to be annotated with two separate mentions, one for each of the entities referred to. This naturally means that each of those mentions will need to be linked appropriately to any other mentions of that entity in the document.

For example, there is a building (a FAC) called the “Holocaust Memorial Museum” but the name of this building is also often used to refer to the organization that runs its business in that building. Thus, in a sentence like the following, “the museum” would be marked as two mentions, one associated with the FAC entity and the other associated with the ORG entity.

But Lerman also added that **{[FAC][ORG] the museum}** would not extend Arafat the formal courtesies that are routine for other world leaders.

If, elsewhere in the document, a mention of “the museum” occurred in the context “New windows were ordered for the museum”, that mention would be marked as an additional mention of the same FAC entity referred to above, but not as an additional mention of the ORG entity.

In cases like the above, where two mentions are marked on the same text, annotators are to specify which of the two mentions is the “literal” one and which the “intended” metonymic one. The Alembic Workbench will support this by allowing the properties “literal” or “intended” to be added to mentions. In examples in these guidelines, the literal mention will always be listed first. Both the literal and the intended mentions, with the entities underlying them, will be counted in the scoring.

The remainder of this section outlines specific annotation guidelines for metonymy in different contexts.

3.1 *Capital City for Governmental GPE*

Cases in which the capital city is used to refer to the nation’s government are marked as true metonyms. (Because two separate GPEs are involved, this is not an exception to the general rule that GPEs are marked as one entity with a role rather than as two entities.)

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

In this example there are two mentions covering the word Beijing. The GPE.GPE is a mention of the city Beijing and the GPE.ORG is a mention of China. The GPE.ORG mention is a mention of the same China entity that would be referred to by other GPE mentions of “China” that might be found elsewhere in the document. Also 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 same Beijing entity referred to by the GPE.GPE mention in the above example.

3.2 *Metonymies Involving ORG Base Entities*

There is a table (see the Pilot Study task definition, Section 6.2.5) that specifies a “base” type for various kinds of entities. Mentions of entities with ORG base types like schools, restaurants, or churches are sometimes used to refer to the organization itself, and sometimes used to refer to the facility that houses that organization. Every mention of such an entity is to be marked (at least) as a mention of an entity of its base type. A second mention of a different type should also be marked if the context invokes a metonymic entity. Thus a mention whose base type is ORG but that is used in a FAC context will have mentions of both of those two entities associated with it.

Below are some examples of ORGs that refer either to a single base type entity, or else to both a base type and metonymic type entity.

Example 1

Universities have an ORG base type so both mentions of the university in 1A and 1B invoke an ORG entity. But 1B also invokes a FAC entity because it refers to the site.

- A. Lee Jung Hoon, a political science professor at **[[ORG-1] Yonsei University]**... (From 9801.162)
- B. Thousands of parochial school and college students are joining this year's demonstration, including 1,500 high school students from across the country who spent last night at **[[ORG-2] [FAC-3] Catholic University]**. (From 9801.175)

Example 2

Embassies have an ORG base type so both 2A and 2B invoke an ORG entity. But 2A also invokes a FAC entity because FACs, not ORGs have gates.

- A. ...a few hundred ethnic Albanians laid a black wreath at the gate of **[[ORG-4] [FAC-5] Yugoslavian embassy]**. (From APW19980308.0201)
- B. “Our Ministry of Defense is working very hard with **[[ORG-6] the U.S. Embassy in Bogota]** to get the information together,” Cano said. (From 9801.382)

3.3 *Special Guideline for ORG Entities used in PER Contexts*

Whenever an organization takes an action, there are people within or in charge of the organization whom 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 PER metonymy. 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 PER entity be marked.

3.4 *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 ORG mentions, and not as PER mentions.

Example: “Obviously, it is deeply discouraging for the family after 22 months, but there is no proof of life. But there is no evidence in either direction _ that there is proof of life or death,” said a Foreign Office spokesman, speaking with customary anonymity.” **{[ORG] We}** will continue to do everything **{[ORG] we}** can to establish what has happened.”

3.5 *Metonymies Involving FAC Base Entities*

The same approach used for ORG entity mentions that refer to an associated FAC should also be used when a FAC entity mention refers to an associated ORG.

Here are two examples from the same document (9801.266):

- A. Competing self-images of victimhood have long prevented Israelis and Arabs from acknowledging the full weight of each other's historical tragedies, and many Arab leaders have resisted efforts to lure them to **{[FAC-7] the museum}** and the similar Yad Vashem memorial in Jerusalem.
- B. Lerman, reached at his New Jersey home, said the subject of Arafat and Israel's talks with the Palestinian Authority still profoundly divided U.S. and world Jewry and "we believe **{[FAC-8] [ORG-9] the museum}** should not get involved in a political dispute where half of the people are for something and half are against it."

Since museums have a FAC base type, both examples A and B invoke a FAC entity. But example B also invokes an ORG entity because it is the organization that should not get involved in the dispute.

Note in the above examples that FAC mention 7 and FAC mention 8 refer to the same FAC entity, as shown in the following table of entities and mentions:

Entity 1: **{[FAC-7] the museum}**, **{[FAC-8] the museum}**

Entity 2: **{[ORG-9] the museum}**

Another common class of FAC metonymies is found when named buildings are used to refer to the organizations based there:

Example: It is unlikely **{[FAC] [ORG] the White House}** would nominate a successor who did not support sampling, and equally unlikely Republican leaders would look favorably on such a candidate.

3.6 *Special Rule for Offices and Branches*

Because the term “office” is often used to refer to an organization, as in “the Office of the Attorney General,” the base type for offices will be ORG. When the context suggests a

reference to the physical entity, the entity should be marked both ORG and FAC. Examples that are ambiguous as to whether a facility or an organization is intended should be marked metonymically, with both an ORG and a FAC mention. Thus in the following example the office is marked both ORG and FAC because it is unclear whether the context suggests that the investigators are from the physical office or from the organization.

Example: Investigators from **{[ORG-9] [FAC-10] the Kentucky state fire marshal's office}**.

(In that particular example, *Kentucky* would also be marked, so that the full annotation for that phrase would be *{[ORG-9] [FAC-10] the {[GPE.ORG] Kentucky} state fire marshal's office}*.)

The same general guidelines apply to other facility terms like “branches” (as in the local branch of a bank).

3.7 Metonymies Involving LOC Base Entities

Entities whose base type is LOC can also be used in metonymic senses. In the following example, “the world” has literal type LOC but intended type PER, and thus is annotated with two separate mentions:

Example: **{[LOC] [PER] The whole world}** was watching.

4 Guidelines for GPEs and GPE Roles

First, note that the definition for GPE entities here is a modified version of the ACE Pilot Study definition. GPEs will now also include all four aspects of the composite entity: population, government, physical location, nation (or province, state, county, city, etc.). So, any of those four aspects of GPEs will be marked GPE with whatever role is appropriate in context. For example, below, the mention of the population of France is marked GPE, rather than PER:

{[GPE.PER] The people of {[GPE.GPE] France}} welcomed the agreement.

Thus any mention of an aspect of a GPE is to be marked as a mention of a GPE entity with an appropriate role. Annotators need to decide for each entity mention in the text which role the context of that mention invokes. This judgment typically depends on the relations that the entity enters into. For example, in “The world leaders met in France yesterday”, the annotation is GPE.LOC because the relation that is conveyed is that France is the venue where the leaders met.

Even if more than one aspect of the entity is invoked by the context, only one role should be assigned. This usually occurs in the case of conjoined predicates. For example,

Washington is preparing for potentially massive demonstrations against the World Bank and the International Monetary Fund as ministers from those organizations arrive for Sunday's opening session.

In the above example, it is the government of Washington (ORG) that is preparing for the demonstrations, but ministers will arrive at the location Washington. In these cases, the annotator should assign a role based on the closest local predicate. In this example, only

the ORG role should be assigned to Washington because “preparing...” is the local predicate and invokes an ORG reading.

Note that 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).

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

4.1 GPEs Modifying People and Artifacts

Pre-modifiers are inherently vague and difficult to decompose. For this reason, all GPE pre-modifiers of people and artifacts will be assigned the role GPE.GPE. For the sake of consistency, the corresponding post-modifiers should also be marked GPE.GPE. For example, *[[GPE.GPE] French] president* should be marked in the same way as *president of [[GPE.GPE] France]*. More examples of GPEs modifying people include:

[[GPE.GPE] Israeli] troops
[[GPE.GPE] New York] policemen
Prime Minister of [[GPE.GPE] Britain]
Picabo Street of [[GPE.GPE] the United States]
[[GPE.GPE] New York] attorney
[[GPE.GPE] U.S.] Commander-in-Chief

GPEs modifying artifacts should also be marked GPE.GPE. Common artifacts modified by GPEs include but are not limited to vehicles, weapons, and flags. Some examples follow:

[[GPE.GPE] U.S.] surveillance aircraft
[[GPE.GPE] Iraqi] flag

4.2 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, rather than exclusively a location or government.

Example: The Palestinian Authority has banned *pro-[[GPE.GPE] Iraq] rallies*, but that ban has been widely ignored.

4.3 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.

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

4.4 Political Communication and Decision-making

On the other hand, 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.

Example: 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.

4.5 Political associations

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

Example: "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 believes the ongoing debate played a role in Riche's departure.

4.6 Embedding

GPE names embedded in government entities have a GPE role. For example, in *the British government*, *British* is a GPE.GPE. This annotation conveys the relationship between nation and government. Similarly, GPEs are often embedded in GPEs so that in examples like *the Israeli settlement*, *Israeli* is marked GPE.GPE. However, if the text strongly supports a LOC reading, the embedded GPE should be marked GPE.LOC. For example, in *the heartland of America*, *America* is a GPE.LOC because a locative relation is conveyed. In the following example, all three GPE mentions have such a LOC role:

Example: Meanwhile, secretary of state Madeleine Albright, Berger and defense secretary William Cohen announced plans to travel to *an unnamed* **[[GPE.LOC] city]** *in the* **[[GPE.LOC] US]** **[[GPE.LOC] heartland]** next week, to explain to the American people just why military force will be necessary if diplomacy fails.

4.7 Athletes, Sports Teams, and GPEs

Athletes **and teams** are associated with GPE.GPEs as in *Picabo Street of the United States* below.

Example: 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 location name is used as a team name (as in *Boston beat Philly*), the entity is marked as a metonymy, with the Literal mention being the city and the Intended mention being the team.

Example: {[GPE.GPE-Lit] [ORG-Int] New York} had a shot to win but Chris Childs missed a three.

In addition, because all GPEs are assigned a role, the Literal GPE mention is assigned a GPE role

4.8 GPEs modifying organizations

In cases where GPEs modify organizations, the organizations are considered to be located in that GPE. Those GPEs should be marked GPE.LOC. So, in *New York corporation*, *New York* gets a GPE.LOC markup.

Example: *The {[GPE.LOC] California} company* also asked that CAI be ordered to pay restitution to CSC "in an amount to be determined at trial."

4.9 Governments

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

Example: But *{[GPE.ORG] the Russian government}* and many politicians will be stridently critical of the United States if they believe they are being ignored.

(In that particular example, *Russian* would also be marked, so that the full annotation for that phrase would be *{[GPE.ORG] the {[GPE.GPE] Russian} government}*, and the two GPE mentions would be coreferential.)

4.10 GPEs and Government Organizations

GPEs modifying government organizations, like *New York police department* and *Kentucky state 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.

Example: The department said Sonabend can appeal to *{[GPE.ORG] Switzerland}'s supreme court*.

5 Pronouns

Pronouns that refer to GPEs are marked as mentions of the same entity as their antecedent, but are assigned the role invoked by the context of the pronoun, which may not be the same as the role of the antecedent GPE.

Composite Example: The president flew to {[GPE.LOC] Israel} to meet with {[GPE.GPE] its} Prime Minister.

Similarly, in the case of classic metonymies (where two entities are created), pronoun annotation is determined in part by the link to the antecedent and in part by the context in which the pronoun appears. If the antecedent is a classic metonymy, the pronoun will be a mention of the same entity as either the literal mention or the intended mention of the antecedent.

Metonymy Example: Thousands of parochial school and college students are joining this year's demonstration, including 1,500 high school students from across the country who spent last night at {[ORG-Literal][FAC-Intended] Catholic University}. {[FAC] It}'s in Georgetown.

In some cases, the antecedent is not a metonymy but the context of the pronoun invokes an entity with a type that is different from that of the antecedent. In such cases, in addition to the mention of the new entity, the annotator should also mark the pronoun as a literal mention of the antecedent entity. (This allows us to maintain the connection between the pronoun and the antecedent.)

Metonymy Example: {[FAC] The museum} is located on 45th Street. {[FAC-Literal] [ORG-Intended] They} just hired a new guard.

6 GPE Clusters to be Treated as GPEs

The Pilot Study Guidelines distinguish between GPEs and LOCs based on whether or not the entity has a central government. According to this definition, Europe, Asia, the Middle East and all continents are LOCs. However, like GPEs these entities consist of a populace, a well defined physical territory, and in some cases (like Europe), have an organizing body (the European Union) associated with it. Because of their similarities to GPEs, these entities appear in contexts similar to those of GPEs. For example:

President-elect Kim Dae Jung today blamed much of **Asia's** devastating financial crisis on governments that "lie" to their people and "authoritarian" leaders who place economic growth ahead of democratic freedoms. [9801.404]

Many of the leaders of **Asian** society have been saying that military dictatorship was the way and democracy was not good for their nations," Kim said. [9801.404]

They concentrated only on economic development," he said, without singling out any nations but referring to "**Asian**-style democracy," in which governments are built around a strong leader who controls economic policy. [9801.404]

For this annotation task, geographical entities comprised of a reasonably well defined set of GPEs will be considered GPEs rather than LOCs. Following is a non-exhaustive list of entities that were LOCs in the Pilot Study, but should be GPEs for this task.

Asia, Europe, the Middle East, Palestine, all continents.

Note that while the base type of these entities is now GPE, they may still take a .LOC role in certain contexts, just like any other GPE.

7 EDT Mention Extent Rules

In certain cases where a name mention is involved in a classic metonymy, the annotator may be tempted to mark two mentions over a span of text with one head. For example:

{[FAC] The {[ORG] McDonald's} where I used to buy breakfast} has closed.

In the example above, there is only one head, *McDonald's*, but the context makes it tempting to distinguish the organization from the facility by creating two mentions with different extents.

While there may be some advantages to this approach, it would raise the following issues: 1) If the above span represented two mentions with different extents, it would not be a case of metonymy. There would be one FAC mention and one ORG mention rather than one mention used to represent another. 2) It would force reevaluation of our mention spans for other entity types as well. For example, *southern France* is now one mention because there is one head. If we started to differentiate between the mention represented by the head and the mention represented by the entire noun phrase, then *southern France* would have to represent two mentions.

For these reasons, annotation should be consistent with the Pilot Study Task Definition, which states that one mention is created for each unique head.

{[ORG-Literal][FAC-Intended] The McDonald's where I used to buy breakfast} has closed.

8 Generics

Generics will be annotated separately at NYU by Adam Meyers' annotation team. Generics will be generated in the APF form of the annotation but current plans don't call for them to be counted in the scoring.