Simple Named Entity Guidelines

IL9 V1.0

Linguistic Data Consortium

(Based largely on the MUC-7 NE, LCTL English V6.4, BOLT-LRL Turkish SNE V1.8, DEFT Rich ERE Entity V2.4 Guidelines)

Change Log from V1.0

- (initial version)

Table of Contents

1 Introduction

This annotation task, Simple Named Entity tagging, is an important, fundamental annotation for the project. The goal of Simple Named Entity (SNE) annotation is to read carefully through textual documents for the **names** of certain types of things (entities), to find and **mark** where all such entities' names appear in the text, and then to tell what **type** of entity the names refer to. An **entity** is some real object in the world – for instance, a place or a person. A **name** is a word or group of words that refers to a specific and unique object (or group of objects) by its proper name, nickname, alias, acronym or abbreviation, or other alternate name. A "Named Entity" is simply an entity which has been mentioned by name in a document. The basic features of Simple Named Entity annotation are:

1. We will only annotate names. For this task, a name is understood as an expression that identifies and directly refers to a specific, unique entity. Names can be proper names, abbreviations, nicknames, aliases, acronyms, or other alternate names.

2. We will only tag names that refer to four (4) specific types of entities: persons (PER), organizations (ORG), geo-political entities (GPE), and physical locations (LOC).

3. A name may be tagged (annotated) regardless of which parts of speech it is, as long as the name is valid for annotation according to these guidelines.

4. The annotated text extents will always coincide with token boundaries –selection of sub-token strings of text (e.g., partial words inside a token) will not be possible.

5. We will not make overlapping or nested annotations – annotations will not be made within or over other annotations. That is, names will not be broken down into other, shorter names for annotation.

6. Annotation of any expression will depend on its intended meaning in the specific context of the document.

7. We will annotate exhaustively – we will tag each and every name in each document, as long as it is valid according to these guidelines.

2 General Approach to Finding Named Entities

NOTE: In these guidelines, examples of valid annotations – names which we would tag (annotate) – are indicated in **[square brackets]**. Invalid or "non-taggable" examples – words which we would not annotate in this task – are shown with ~~**strikethrough**~~.

The general order of decision-making for annotation is summarized here:

1. Read the document carefully, looking for possible names.

2. If you see a name mentioned in the text, and it is not within the extent of another name, highlight the extent of the name to select it for annotation.

3. Decide whether the entity being named is actually one of the four valid types of entity (PER, ORG, GPE, LOC), determine which one it is, and tag it with the correct type.

4. Continue through the document until all taggable Named Entity mentions are annotated. This means even if the same entity is mentioned once (say, "United States") and then again later (e.g., as "U.S." and then later as "America"), we must tag all of these mentions of the entity.

You should tag everything that you recognize as a name (of our four types), even if it is not in your language. If it is a name in a foreign language, but you know that it is a name and you know which type it is, you should tag it.

2.1 Names, Entities, and Mentions

In Simple Named Entity annotation, we only tag **names** – and we only tag the names of certain types of **entities** (a unique thing or object or set of objects in the world). **Names** are expressions (a word or words) that identify and directly refer to specific entities. A name can be a proper name, abbreviation, nickname, alias, acronym, or another alternate name. For example, names of entities mentioned in a document might include:

```
Kalisa
Kigali
Umugi wa Kigali "City of Kigali"
Umugezi wa Nyabarongo "Nyabarongo River"
Minisiteri y'uburezi "Ministry of Education"
Bob Austin
Eiffel Tower
```

```
Big Apple
IBM
Yankees
Coca-Cola Bottling Co.
Médecins Sans Frontières
Uganda
Bowdon, Georgia      (two separate entities)
Mt. Fuji
Kremlin
Kennedys
```

Words that are not actually being used as names are not valid for annotation in SNE. This means that we cannot tag words which do not in themselves act as names – words which do not specifically identify or refer to a unique or particular entity or group of entities should not be tagged. General-sounding words like common nouns, such as in "the ~~company~~", "the ~~president~~", "the ~~victims~~", "the ~~plaza~~" could not be tagged, even if we knew exactly to which Named Entity these words were actually referring.

Acronyms and other abbreviations of proper names, nicknames, aliases, and other alternative names will be tagged as names:

`IBM`	acronym for International Business Machines
`Big Blue`	alias for International Business Machines
`Big Board`	alias for New York Stock Exchange
`The Boss`	nickname for Bruce Springsteen
`Big Apple`	nickname for New York City
`Red Sox`	short for the Boston Red Sox
`Sears`	short for Sears Roebuck and Co.
`Cali.`	abbreviation for California

A **mention** is simply a single, particular spot where an entity is mentioned in a document. Names may be considered for annotation regardless of which syntactic function the name fulfils – no matter which part of speech they are mentioned in. For example, in English, we can often tag the adjective forms of names, such as in "[Korean] cars" or "[American] travelers". Names of a valid entity may be mentioned more than once in a document, and we would need to tag each and every valid mention of its name or names.

Whether any particular mention of a name can be tagged (is valid or not for annotation) is a decision which generally depends on:

- to which **type of entity** the name refers (see Section 2.2 – 2.3, and

Section 3 below)

- whether or not the mention of the name appears within the mention **extent** of another name (see Section 5.1)

➔ **Guiding Principle**: If you know that any word or phrase is a name (as defined in these guidelines), then annotate it – and if you know it's not a name, then don't annotate it.

When tagging a name, we tag the whole name and nothing but the name. The extent we annotate must include all the parts of the name that appear in that mention. We also try to avoid words, characters, letters that are not part of the name.

The taggable **extent** of a name includes all letters or characters of the word or words that make up the name as it is mentioned in the text. However, the taggable extent will normally exclude any article (e.g., "the"), title (e.g., Mrs.), or other modifiers before or after the name itself. These are excluded because they are not part of the entity's actual name (e.g. Bill Clinton's name is "Bill Clinton" not "the former president Bill Clinton"). More examples of mention extents follow.

Incorrect extents:

- ~~[the Eiffel Tower]~~$_{LOC}$
  ➔ Do not include articles.
- ~~[Coca-Cola]~~ Bottling Co.$_{ORG}$
- ~~[Coca-Cola Bottling]~~ Co.$_{ORG}$
  ➔ Incomplete mentions: part of the name has been left out.
- ~~[Bowdon, Georgia]~~$_{GPE}$
  ➔ Two distinct entities must be tagged separately.
- Mt. ~~[Fuji]~~$_{GPE}$
  ➔ Incomplete: "Mt. (Mount)" is an accepted part of the name.
- the ~~[famous Lincoln Memorial]~~$_{LOC}$
  ➔ Cannot include a modifier ("famous") which is not part of the name itself.
- ~~[Bill and Hillary Clinton]~~$_{PER}$
  ➔ Two entites' names must be tagged as separate mentions.
- ~~[Sean]~~$_{PER}$ ~~["Puffy"]~~$_{PER}$ ~~[Combs]~~$_{PER}$
  ➔ When alternate names (such as nicknames or abbreviations) appear between parts of a full proper name, they should not be separated – tag them as one mention.
- ~~[Sean "Puffy" Combs, a.k.a. Puff Daddy or P. Diddy]~~$_{PER}$
  ➔ Different names of the same entity must each be tagged as separate mentions if they can be separated.

Correct extents:

- [Kalisa]$_{PER}$
- Musenyeri [Rucyahana]$_{PER}$ "Bishop Rucyahana"
- [Kigali]$_{GPE}$
- [Umugi wa Kigali]$_{GPE}$ "City of Kigali"
- [Umugezi wa Nyabarongo]$_{LOC}$ "Nyabarongo River"
- [Minisiteri y'uburezi]$_{ORG}$ "Ministry of Education"
- [Bob Austin]$_{PER}$
- former president [George W. Bush]$_{PER}$
- the [Eiffel Tower]$_{LOC}$
- [IBM]$_{ORG}$
- the [Yankees]$_{ORG}$ (sports team)
- [Coca-Cola Bottling Co.]$_{ORG}$
- The [Big Apple] $_{GPE}$
- [Uganda]$_{GPE}$
- [Bowdon]$_{GPE}$, [Georgia]$_{GPE}$
- [Mt. Fuji]$_{GPE}$
- the [States]$_{GPE}$ (as a nickname for the US)
- the [Kremlin]$_{ORG}$ announced that …
- the [North]$_{GPE}$ (for 'North Korea')
- the so-called [Northern Cyprus Chess Federation]$_{ORG}$
- the famous [Lincoln Memorial]$_{LOC}$
- the incomparable [Steven Spielberg]$_{PER}$
- the [US State Department]$_{ORG}$
- the [Kennedys]$_{PER}$
- [Bill]$_{PER}$ and [Hillary Clinton]$_{PER}$
  ➔ When a phrase refers to multiple named entities, mark each entity separately.
- [Sean "Puffy" Combs]$_{PER}$, a.k.a. [Puff Daddy]$_{PER}$ or [P. Diddy]$_{PER}$

Names for some types of entities may include words which are otherwise used as general, common nouns. Be sure to include such words in the full extent of a name if they are typically considered or understood as part of the **name** in your language. For example, such words can often be included in the English names of many LOCs and some ORGs, such as in:

[**Umugezi** wa Nyabarongo] "Nyabarongo River"
[**Minisiteri** y'uburezi] "Ministry of Education"
[**Ishuri** rya Nyanza] "Nyanza School"
[Amazon **River]**
[Lake Michigan]

[Eiffel **Tower]**
[Statue of Liberty]
[Apple **Inc.**]
[Myers Elementary **School]**

In general, use the Reasonable Reader Rule to help decide whether any word or character should be included in the extent of the name you annotate:

Reasonable Reader Rule: If a reasonable person reading the document would understand that any text must be, or is highly likely to be, part of the name you are annotating, then you should include that text in the name you tag. Conversely, if a reasonable person reading the document could not determine whether or not that text must be, or is highly likely to be, part of the name, then do not include it in the mention. Whether or not any text should be annotated as part of the mention of a name should be decided according to a reasonable, intuitive interpretation of the document.

2.2 Valid Entity Types

We will only tag names that refer to four (4) specific types of entities – persons (PER), organizations (ORG), geo-political entities (GPE), and locations (LOC). Only entities belonging to these four types are valid for SNE annotation:

PERSON (PER): Person entities are individual humans (or sets of individual humans) identified by name, nickname, or alias.

ORGANIZATION (ORG): Organization entities are formal groups defined by an established organizational structure such as corporations, professional associations, government agencies, and other institutions.

GEOPOLITICAL (GPE): GPE entities are composite entities, meaning there are multiple criteria that must be present to make something a GPE. GPEs consist of (1) a physical location, (2) a unifying government, and (3) a population. All three of these elements must be present for an entity to be tagged as a GPE. GPEs may include things like countries, provinces, counties, cities, and districts, but not things like voting wards, parishes, or geographical (non-political) regions.

LOCATION[1] (LOC): Location entities are non-politically-defined places. This includes natural features or geographically-defined places (like

[1] In SNE, LOC encompasses the facility (FAC) entity type that may be differentiated in other projects.

islands, bodies of water, mountains, international or local regions, etc.) as well as artificial (human-made) structures (like airports, streets, canals, factories and monuments).

These four valid entity types are discussed in greater detail and with examples in Sec. 3 below.

We will not annotate any other types of named entities outside of these four. Thus, we cannot tag names for things such as:

- animals
- products
- books or other publications
- inanimate objects
- artifacts (artificial objects which are not structures)
- events
- ideas
- religions
- nationalities and ethnic groups (unless they correspond to a GPE)
- languages
- monetary units

Some examples:

- my ~~Taurus~~ is one of the latest car models
  ➔ Product is not a valid entity type
- I saw it in the ~~New York~~ ~~Times~~.
  ➔ Publication is not a valid entity type
- ~~Arabic~~ is a major international language.
  ➔ Language is not a valid entity type
- The ~~Olympics~~ were on last week.
  ➔ Event is not a valid entity type
- The ~~Gypsies~~ have been persecuted through history.
  ➔ Ethnicity is not a valid entity type

As you can see, sometimes you will need to consider the context carefully in order to decide what type of entity is being talked about – whether a name refers to one of these four taggable types of entity or not, and if it does, which of the four types of entity it may be.

2.3 Tag for Meaning

Sometimes a name does not actually refer to the entity we think it might at first glance. For instance, “Chicago” could mean either a GPE, a city – as in “arrived in [Chicago]” – or an organization (ORG) – as in “[Chicago] beat [L.A.] last night”. Likewise, “Cambridge University” could refer to either an

ORG – as in “she got her degree from [Cambridge University]” – or a LOC – as in “walked around [Cambridge University]”. Think about what entities are meant by the names in these examples:

1. [South Korea] beat [Japan] in the World Cup semifinals.
2. We had to go to the [U.S. Embassy] in Beirut to get our visas.
3. The [U.S. Embassy] in Beirut issued a statement…

In example 1, “South Korea” and “Japan” actually refer to the soccer teams – which are organizations (ORG) – and not to the countries themselves, so both should be tagged as ORG mentions. In examples 2 and 3, the “U.S. Embassy” refers to different types of entities: it means a building or physical location (LOC) in example 2, while it means the embassy’s institutional organization (ORG) in example 3.

In order to decide whether and how to tag anything we must first understand what the words mean within that specific context. In other words, annotation decisions depend on how the words are being used in the immediate sentence and within the context of that particular document. This means reading carefully to understand how words are actually being used in the context. We need to be clear about what the author intends an expression to mean, what the author is actually referring to. We call this rule **Tag for Meaning**.

We cannot tag certain names because they are **generic** – meaning they used to refer to a **general** category of person or thing. A generic mention cannot be tagged, whether it refers to the entire general category of entities or to a single entity from a general category. For example:

- ~~Americans~~ love fast food.
- The ~~French~~ enjoy wine.
- Of course a ~~Finn~~ would say that…
- Most ~~South Americans~~ speak ~~Spanish~~.
  ➔ note that “Spanish” here does not refer to the GPE Spain but to the language, so it is not taggable either.
- ~~Christians~~ help each other.
- ~~Arabs~~ around the world speak Arabic.
- Luckily, the ~~Australians~~ made it to the barbeque on time.
- The ~~Italians~~ have joined us on the bus tour.

However, if a name modifies another word (even if the modified word itself

is not a taggable entity type), we will tag it. This includes when an entity name occurs in the form of an adjective. Examples:

- The [Italian]$_{GPE}$ soldiers joined us on the bus tour.
- [Bridgestone]$_{ORG}$ profits
- the [Clinton]$_{PER}$ administration
- [Treasury]$_{ORG}$ bonds and securities
- [U.S.]$_{GPE}$ exporters
- [Apple]$_{ORG}$ computers
- [Texas]$_{GPE}$ intermediate crude oil
- a [China]$_{GPE}$ film festival
- the [American]$_{GPE}$ companies
- [Cuban]$_{GPE}$ citizens
- [Greek]$_{GPE}$ food

Sometimes a single sentence might include both usages, the taggable modifier and the untaggable generic noun. For example, what can be tagged in the following sentence?

> Thirty people have died in a week of unseasonally heavy rains, including five Indians and four Pakistanis and a United Arab Emirates family that was swept away in their car, police said Saturday.

In this example, we cannot tag Indians or Pakistanis at all, but it is correct to tag United Arab Emirates as GPE. We cannot tag “Indians” or “Pakistanis” as PER because they are generic terms, and we cannot tag them as GPE because the words “Indians” and “Pakistanis” refer to people (not the countries). But we can tag “United Arab Emirates” as GPE in this example because it is a country that is modifying the word “family”. So, the correct annotation for this sentence has only one named entity tagged:

- [United Arab Emirates]$_{GPE}$

> Mu bishwe n'imyuzure harimo abahindi batanu, abanyapakisitani bane, abapashitu batatu n'umuryango ukomoka muri Leta zunze ubumwe z'Abarabu watwawe n'umuvu uri mu modoka yawo.
> “The dead from the storms included five Indians, four Pakistanis, three Pashtos and a United Arab Emirates family who were washed away in their car”

- [Leta zunze ubumwe z'Abarabu]$_{GPE}$

Here are some additional examples:

- ~~Umuhinde~~ "A native of India"
  - Not taggable
- ~~Umunyagisaka~~ "A native of Gisaka"
  - Not taggable
- ~~Umunyamerika~~ "A native of America"
  - Not taggable
- [India]$_{GPE}$
- [Indian]$_{GPE}$ citizens
- the [Indian]$_{GPE}$ citizens
- [Indian]$_{GPE}$ embassy
- 4 ~~Indians~~
  - Not taggable
- 4 [Indian]$_{GPE}$ citizens
- the ~~Italians~~
  - Not taggable
- the [Italian]$_{GPE}$ soldiers

The use of a plural form of the adjectival country name to refer to generic citizens is a type of construction that is very common in English, but it might not be so common in your language. Be aware of plural generic adjectives like this, if they do appear in the text.

Keep in mind that meaning depends on context and usage, so take care in using your judgment, and think carefully about what exactly is being referred to in each case. Consider the following text extent, for example:

> The [greens]$_{ORG}$ are always pushing for a more stringent climate policy.

"greens" in this case could be understood as meaning a general category of Green Party-leaning voters or supporters, or even some group of Green Party politicians – neither of which would be taggable. However, in this sentence it is most likely being used as shorthand – a shortened version of "the Green Party" (ORG).

2.4 Locative Markers

Locative markers are generally written with a space as a separate token, and so they are not included in the extent of named entities. If it happens that a locative marker does appear written attached to the name, it should then be included in the extent of the named entity.

Locations are marked by one of three locative noun class markers (mu, ku,

and i).

- Aratembera ava i [Kigali]$_{GPE}$ ajya i [Butare]$_{GPE}$ "He is travelling from Kigali to Butare"

2.5 Possessive Compound Noun Construction or Associative Construction

A multi-token associative or possessive construction can also be a name. If the entire structure is a name of a taggable entity, the extent of the name should include the entire structure, for example [city of New York].

[Ishyamba rya Nyungwe]$_{LOC}$
forest of Nyungwe
'Nyungwe Forest'

[Umugi wa Kigali]$_{GPE}$
city of Kigali
'City of Kigali'

[Minisiteri y'imari]$_{ORG}$
ministry of finance
'Ministry of Finance'

[Umugezi wa Nyabarongo]$_{LOC}$
river of Nyabarongo
'Nyabarongo River'

[Mukayuhi]$_{PER}$
Wife-Yuhi
'Yuhi's wife'

In the last example, only "Yuhi" is a taggable PER name (because the wife's name is not present). However, because "Makayuhi" is a single token, we will tag the whole token in order to capture Yuhi's name.

3 Entity Types

NOTE: For simplicity, in the examples under the following sections we will generally only show annotation marks for the entity type being discussed. For example, under the section on locations (LOC), GPE names that happen to occur in the examples may not appear in [square brackets], even though we would need to tag those GPE names in the actual task.

3.1 Person (PER) Names

Person (PER) entities are limited to individual humans (or sets of individual humans) identified by name, nickname, or alias.

A person may be specified by all or part of a personal name, a nickname, or an alias. Family or clan names should also be tagged as person (PER) entities. Names of deceased people should be tagged as person (PER). We will **not** tag religious deities, supernatural beings, mythical persons, nor fictional human characters (appearing in movies, television, books, video games, etc.) as persons. Some examples of taggable and untaggable person (PER) mentions:

- [Kalisa]$_{PER}$
- Abana ba [Kalisa]$_{PER}$ "Kalisa's kids"
- Musenyeri [Rucyahana]$_{PER}$ "Bishop Rucyahana"
- [Mukayuhi]$_{PER}$
  ➔ Note that this example is tagged for Yuhi's name
- umuryango wa [Kennedy]$_{PER}$ "Kennedy family"
- Bwana [John Kerry]$_{PER}$ "Mister John Kerry"
- [Hillary Rodham Clinton]$_{PER}$
- Perezida [Barack Obama]$_{PER}$ "President Barack Obama"
- [Ban Ki-moon]$_{PER}$
- former prime minister [Gordon Brown]$_{PER}$
- the incomparable [Steven Spielberg]$_{PER}$
- the [Bushes]$_{PER}$
  ➔ Note that this is a family name, and refers to a definite and specific set of people.
- [Bill]$_{PER}$ and [Hillary Clinton]$_{PER}$
  ➔ When a phrase refers to multiple named entities, mark each entity separately.
- [Sean "Puffy" Combs]$_{PER}$, a.k.a. [Puff Daddy]$_{PER}$ or [P. Diddy]$_{PER}$
- the famous [~~Lincoln~~ Memorial]$_{LOC}$
  ➔ Although "Lincoln" is the last name of a person entity, because it is inside the name of another entity, the "Lincoln Memorial", it cannot be tagged. Only the longer, "containing" named entity can be tagged.
- "…played the ~~[Yankees]$_{per}$~~ today…"
  ➔ N.Y. Yankees baseball team = organization (ORG) type (this is tagged as an ORG and is not tagged as a PER group)
- the so-called ~~Northern Cyprus Chess Federation~~
  ➔ organization (ORG) type (this is tagged as an ORG and is not tagged as a PER group)

Titles, roles, and honorifics such as "Mrs." and "President" are not part of a person's name:

- Correct: Chairman [Mao]$_{PER}$
- Wrong: ~~[Chairman Mao]$_{PER}$~~

- Correct: Dr. [Oz]$_{PER}$
- Wrong: ~~[Dr. Oz]$_{PER}$~~

- Correct: [GlobalCorp]$_{ORG}$ Vice President [John Smith]$_{PER}$
- Wrong: [GlobalCorp]$_{ORG}$ ~~[Vice President John Smith]$_{PER}$~~

However, generation suffixes – such as "Junior (Jr.)", "Senior (Sr.)", "the Third (III)", etc. – must be included in the extent of a person's name:

- Correct: Dr. [Martin Luther King, Jr.]$_{PER}$
- Wrong: Dr. ~~[Martin Luther King]$_{PER}$~~, Jr.

- Correct: [Thomas P. O'Neill III]$_{PER}$, Esq.
- Wrong: ~~[Thomas P. O'Neill]$_{PER}$~~ III, Esq.

Some groups of people – such as movements (e.g. 'Occupy Wall Street' or 'Black Lives Matter') – may not clearly have a formally organized structure (so cannot be tagged as ORG). They may however, still refer to a definite, specific set of people, and be tagged as a person group (PER). Use your best judgment as to whether to tag them as ORG or PER. We will generally try to make fewer assumptions about the organizational aspects – if it does not definitely have an organizational structure, we will tag it as a group of people, with the PER type:

- [Occupy Wall Street]$_{PER}$
- the [Tea Party]$_{PER}$

For more on organizations (ORG) and how they differ from less formally-organized groups of persons (PER), see Section 3.2 below.

Remember that names of kinds or categories of people – nationalities, ethnicities, religious groups, etc. – are not taggable. The person (PER) type only applies to proper names or nicknames by which any specific person (or group of people) are distinctively known or addressed. E.g., the following are not PER mentions:

- ~~Americans~~ love fast food.
- ~~Christians~~ help each other.

- ~~Arabs~~ around the world speak Arabic.

Note that sometimes a word that appears to refer to a group of people may in fact be meant to refer to an organization (ORG), depending on context.

- The [Socialists]$_{ORG}$ have a lot to lose this year.
- [Republicans]$_{ORG}$ are mostly conservative.

3.2 Organization (ORG) Names

We will tag all proper name mentions of organizations (ORG) – groups with an officially- or formally-defined organizational structure. Organizations include the following subtypes:

- Governmental
- Commercial, Educational, Scientific, Medical
- Media
- Religious, Social, Advocacy
- Sports

Though we will not be labeling subtypes, it is useful to consider examples of each of them. For simplicity, we will generally only show annotations for ORG types in this section.

Governmental (includes Political, Quasi-Governmental, Military, and Para-Military Groups)

- [Minisiteri y'uburezi]$_{ORG}$ "Ministry of Education"
- [Minisiteri y'imari]$_{ORG}$ "Ministry of Finance"
- Akora muri [ministeri y'ubutabera]$_{ORG}$ "He works for the Ministry of Justice"
- [Labour Party]
- the [Socialist People's Party]
- [Democratic National Committee]
- the [ACLU]
- The [Cato Institute]
- [NATO]
- [NYPD]
- The [World Bank]
- three of the [U.N.] workers stationed in East Timor
- [International Monetary Fund] aid
- [Hizbollah]
- [Islamic Resistance]
- [Rally for Congolese Democracy]

- [Institutional Revolutionary Party]
- [Al Aqsa Martyr's Brigade]
- [Tamil Tigers]
- the [Caravan of Death],a military party that killed 73 political prisoners
- the leading deputy of the [Rally for Congolese Democracy], one of the biggest rebel movements supported by Uganda
- Putin, a former [KGB] agent, defended the court that convicted Pope and the security services.
- The [Financial Accounting Standards Board] will take no conclusive action on its current project on business combinations until [Congress] has reconvened in 2001.
- ~~Government~~ officials at the site said… .
  - ➔ Some mentions may not clearly refer to a specific named entity but to a general kind of organization.
- The [Socialists] have a lot to lose this year.
- [Greens] are mostly progressive.
  - ➔ For the last two examples here, note that while some mentions appear to be for persons, in some contexts they may be understood as shorthand referring to the ORGs. You will need to decide which type to use based on the usage of the name in the document context – use the Tag for Meaning rule.

Commercial

- [Ikibuga cy'indege cya Kigali]$_{ORG}$ "Kigali Airport"
  - o Note this could be a LOC mention, depending on the context
- the [Roundabout Theater Company] is calling its new facility in Times
- [Pixar], the award-winning animation company
- the [American Airlines Theater]
- Pope, who owns [TechSource Marine Industries] in State College
- Like the famous Irish group the [Chieftains], [Solas] frequently headlines in Celtic festivals.
- …the ~~Campbell Soups~~ of the world…
  - ➔ Name mentions referring to a generalized kind of entity are not taggable.

Educational, Scientific, Medical

- [George Washington University]

- [Overseas Chinese Physics Institute]
- [Gulf Coast Research Laboratory]
- A coalition of medical and health groups from [Massachusetts General Hospital]
- Pope had worked for the [Applied Research Laboratory] at [Pennsylvania State University].
- [NDSU] and [University of Minnesota] weeds specialist Alan Dexter says 98% of the plants survived.
- [Médecins Sans Frontières]
- [MSF]
- [Doctors Without Borders]

Media and Social Media

Organizations that give their name to a publication, media program or outlet (whether printed or digital) may only be tagged as ORGs when it's clear that the organization itself is being referred to, and not the publication. Possible examples:

- [Associated Press]
- [Agence France Presse]
- [Deutsche Welle]
- [ABC] news
- the offices of the [Chicago Sun-Times]
- an contributor to [National Geographic]

Mentions of publications, websites, media outlets, and TV or radio programs are not themselves considered direct references to organizations. You will need to use the Tag for Meaning rule to decide whether a mention can be tagged or not:

- The [New York Times]$_{ORG}$ announced that it has named a new CEO.
- Bob enjoys reading the ~~[New York Times]~~$_{ORG}$ on Sunday.
- [Facebook]$_{ORG}$ is headquartered in Menlo Park, CA.
- Isn't the [BBC]$_{ORG}$ is partly funded by the British government?
- i saw on ~~[facebook]~~$_{ORG}$ there was something on the ~~[bbc]~~$_{ORG}$ saying the earth had exploded

Religious, Social, Advocacy

- [German Bishops Conference]
- [Rock the Vote]
- [American Medical Association]
- [American Council on Education]
- [National Rifle Association]
- [American Diabetic Association]
- [NAACP]
- [American Bar Association]
- [National Center for Public Policy and High Education]
- The [Red Cross] said about 15 people managed to escape…

Sports

- [Taekwondo Association]
- [Philippines Olympic Committee]
- [national hockey league]
- [San Francisco 49ers]

Do not tag event names, even if they refer to events that occur on a regular basis. Mentions of events such as conferences, summits, and sports competitions cannot be tagged, even if the events are associated with institutional structures. Only names which refer to the institutional structures themselves — organizing committees, etc. — may be tagged:

- "…won the ~~World Cup~~ last year,…"
- [FIFA]$_{ORG}$

Sets of people who are not formally organized into a unit with organizational structure shouldn't be treated as an ORG entity. This distinction can sometimes be difficult. <u>If in doubt, do not label it as ORG</u>, but PER. See Section 3.1 above.

Most organizations have not only an organizational structure, but a physical location. For any cases where a name can be used to refer to the location or the organization, you should tag the name based on the meaning in the sentence. For instance, museums are primarily organizations but are also housed in a specific building or facility. So while we often tag museums as ORG (organization) entities, there are cases when a particular example might function more like a LOC (location). In cases like this, annotators should tag the named entity based on the way it is used in the sentence. Examples of annotation:

- The [Guggenheim Museum]$_{ORG}$ announced a new acquisition.
- The [Guggenheim Museum]$_{LOC}$ was designed by Wright.
- Thirty people were wounded in the bomb blast in front of the city's [Gulshan Hotel]$_{LOC}$.
- A [Gulshan Hotel]$_{ORG}$ spokesman called the incident a tragedy.

3.3 Geopolitical Entity (GPE) Names

Geo-Political Entities are politically-defined populated territories (such as countries, provinces, states, cities, districts, etc.). For something to be taggable as a GPE, it must include all three of the core elements: a political organization, a population, and physical territory.

Compound expressions, in which GPE names may or may not be separated by a comma, such as in English, should be tagged as separate instances of GPE.

[Kaohsiung]$_{GPE}$, [Taiwan]$_{GPE}$
[Johannesburg]$_{GPE}$, [South Africa]$_{GPE}$
[München]$_{GPE}$, [Deutschland]$_{GPE}$

Countries of countries, such as 'the [European Union]$_{GPE}$' and 'the [United Kingdom]$_{GPE}$' will be annotated as GPEs, since they have all three GPE components (i.e., a single, unified or overall government as well as population and territory). The same applies to contested areas – such as Taiwan – as long as they have all three elements, but not to organizations like the G8 or ASEAN, as they do not have a unified, controlling government.

Sometimes the context makes it appear that the mention of a geo-political unit (such as the capital) or a government location is referring specifically to the government of the GPE. In these cases we still tag the mention as a GPE. For instance:

- [Iraq]$_{GPE}$ signed a treaty with [Kuwait]$_{GPE}$.
- [Washington]$_{GPE}$ discussed economic policies with [Moscow]$_{GPE}$ at the summit.
- The government of [France]$_{GPE}$ welcomed the agreement.
- [India]$_{GPE}$ is interested in strengthening economic ties with the [US]$_{GPE}$.
- The Premier said [China]$_{GPE}$ would continue on a path of economic liberalization.
- [Turkey]$_{GPE}$ regards [Northern Cyprus]$_{GPE}$ as a

sovereign country.
- [Washington]$_{GPE}$ announced a new tax policy today.

Sometimes a GPE mention may appear to refer more strictly to the physical location, but in such cases we still tag it as a GPE, for example:

- We went to [France]$_{GPE}$ for our vacation.
- They delivered the supplies to [Pakistan]$_{GPE}$

Notice that GPEs in adjective form can be tagged if modifying a noun:

- [American]$_{GPE}$ soldiers
- [Korean]$_{GPE}$ cars

Here are some additional examples of mentions of GPE names:

- Aratembera ava i [Kigali]$_{GPE}$ ajya i [Butare]$_{GPE}$ "He is travelling from Kigali to Butare"
- [Umugi wa Kigali]$_{GPE}$ "City of Kigali"
- [Ubuhinde]$_{GPE}$ bwavuze ko buzongera gutanga ibiribwa n'izindi mfashanyo bukoresheje indege "India said it will resume air drops of food and other aid"
- Imyuzure ikomeye muri [Bangladeshi]$_{GPE}$ yagabanyije cyane umusaruro w'ibihingwa byakomokaga muri ako karere "Severe floods in Bangladesh lowered the region's agricultural output"
- Bwana John Kerry, umunyamabanga wa Leta muri Leta zunze ubumwe [z'Amerika]$_{GPE}$, yavugiye muri [Omani]$_{GPE}$ "Mr John Kerry, the Secretary of State of the US, said that thirty people have died in Oman"
- Isoko rinini kuruta ayandi rya [Muscat]$_{GPE}$ ryongeye gufungura kuwa gatandatu "Muscat's largest market reopened Saturday"
- Inkubi y'umuyaga yo mu rwego rwa gatanu izayogoza agace k'amajyepfo [y'ubushinwa]$_{GPE}$ vuba aha "A category 5 typhoon will hit southern China soon"
- Leta y'u [Rwanda]$_{GPE}$ yakwirakwije abasirikare bazobereye mu guhangana n'ibiza mu duce twose twayogojwe n'umwuzure "The Rwandan government dispatched national guards to flood-affected areas"
- Abantu benshi bakoze imyigaragambyo mu mihanda ya [Cairo]$_{GPE}$, umurwa mukuru wa [Misiri]$_{GPE}$ "More people marched in the streets of Cairo, the

capital city of Egypt”

- Yatembereye cyane muri manda ye: kuva muri Aziya y'amajyepfo kugera muri [Afghanistani]$_{GPE}$, mu burasirazuba bwo hagati ukagera [Haiti]$_{GPE}$ “She traveled heavily during her tenure: from South Asia to Afghanistan, from the Middle East to Haiti”
- the conversion to Christianity of the [Roman]$_{GPE}$ emperor Constantine
- [Salzburg]$_{GPE}$ governor Schausberger said…
- Recounts are only just beginning in [Palm Beach]$_{GPE}$ and [Volusia]$_{GPE}$ counties.
- The economic boom is providing new opportunities for women in [New Delhi]$_{GPE}$.
- …said Norbert Karlsboeck, mayor of [Kaprun]$_{GPE}$, a town some 50 miles south of [Salzburg]$_{GPE}$ in the central [Austrian]$_{GPE}$ Alps
- [France]$_{GPE}$ produces better wine than [New Jersey]$_{GPE}$.
- [Israeli]$_{GPE}$ troops

Keep in mind that the extent of GPEs will match token boundaries. Depending on how the source texts are tokenized, the annotated extents of the names in the examples below would likely have to be:

- The [Palestinian]$_{GPE}$ leaders have banned rallies that are [pro-Iraq]$_{GPE}$
  ➔ (here, “pro-” is inseparable from “Iraq”, as they are part of the same token)
- [France’s] $_{GPE}$ greatest national treasure
  ➔ (here, the “‘s” is inseparable from “France”, as they form the same full token)

Sometimes the names of GPE entities may be used to refer something other than the GPE itself (or its core elements, the government, people, or the territory). The most common examples of this are for sports teams:

- [New York]$_{ORG}$ defeated [Boston]$_{ORG}$ 99-97 in overtime.

In these cases, we adopt the rule of *Tag for Meaning*: annotate an expression according to its meaning within the specific context of the document. So, in the example above, both ‘New York’ and ‘Boston’ are to be understood as mentions of ORG entities.

Note that GPE names nested within sports team names (ORG) are not tagged separately (as GPEs), because names cannot be annotated within other names. Example:

- the [Philadelphia Eagles]$_{ORG}$

Remember, names of languages are not taggable as GPE mentions:

- ~~French~~ is spoken in much of Africa.
- All nations of the former [Yugoslavia]$_{GPE}$ have ~~Serbian~~-speaking regions.
- ~~Arabic~~ is a major international language.
- The most widespread [Indian]$_{GPE}$ languages are ~~Hindi, Marathi, Tamil, Urdu, Bengali~~, and ~~Telugu~~.

3.4 Location (LOC) Names

Locations (LOC) are places that are not GPEs, and include natural geographical features and artificial structures. Examples of locations include continents, islands, heavenly bodies, airports, highways, street names and addresses, factories, seas, rivers, reservoirs, canals, national parks, mountains, plains, deserts, monumental structures, and regions or places that are otherwise non-politically-defined. For instance:

- Yatembereye cyane muri manda ye: kuva muri [Aziya y'amajyepfo]$_{LOC}$ kugera muri Afghanistani, mu [burasirazuba bwo hagati]$_{LOC}$ ukagera Haiti "She traveled heavily during her tenure: from South Asia to Afghanistan, from the Middle East to Haiti"
- [Ishyamba rya Nyungwe]$_{LOC}$ "Nyungwe Forest"
- [Umugezi wa Nyabarongo]$_{LOC}$ "Nyabarongo River"
- Nambutse [Nyabarongom]$_{LOC}$ "I crossed the Nyabarongo (River)"
- Yaguye mu [Kivu]$_{LOC}$ "He fell in Kivu (Lake)"
- [Ikibuga cy'indege cya Kigali]$_{LOC}$ "Kigali Airport"
    - Note this could be an ORG mention, depending on the context
- Inkomere zoherejwe mu [bitaro bya Kigali]$_{LOC}$ "The injured victims were sent to General Hospital of Kigali"
    - Note this is a LOC context for the hospital, but in a different context it could be an ORG mention
- Amakimbirane yerekeranye [n'inyanja yo mu majyepfo y'Ubushinwa]$_{LOC}$ ashingiye ku gace k'ubutaka ndetse

n'amazi ahuriyeho ibihugu byinshi byigenga byo muri ako karere "The South China Sea disputes involve both island and maritime claims among several sovereign states within the region"

- Inkubi y'umuyaga yo mu rwego rwa gatanu uzayogoza [Ubushinwa bw'amajyepfo]$_{LOC}$ vuba aha "A category 5 typhoon will hit South China soon"
- Indwara ya Zika yakwiriye mu bihugu [by'Amerika y'amajyepfo]$_{LOC}$ "Zika spread in South American countries"
- Many people in [North America]$_{LOC}$ saw a partial solar eclipse yesterday.
- NASA's journey to [Mars]$_{LOC}$
- the collapse of the newly-constructed [Teton Dam]$_{LOC}$
- The [Walt Whitman Bridge]$_{LOC}$ remained closed
- repairs began on a 10-mile stretch of the [Alaskan Pipeline]$_{LOC}$
- The [Berlin Wall]$_{LOC}$
- We went to the [Musée d'Orsay]$_{LOC}$ in Paris

Remember that place names may also be organization names, and vice versa, so be sure to tag the name based on how it's used in the sentence:

- watching the parades at [Disneyland]$_{LOC}$
- A spokesperson for [Disneyland]$_{ORG}$ theme park announced today…
- navigating new york's [la guardia airport]$_{LOC}$ was a nightmare for travelers this year
- Officials of [La Guardia Airport]$_{ORG}$ reported that…
- The [Pentagon]$_{LOC}$ is a government building in Arlington, Va.
- The [Pentagon]$_{ORG}$ issued a statement about the incident.
  - ➔ Here, a structure or building name (normally understood as a location) is actually being used to refer to an organization housed in that facility, so the name should be tagged as an ORG

Some locations do not have any corresponding political organization, such as Asia, the Middle East, or Polynesia. These should be tagged as LOC. For example,

- Inkubi y'umuyaga yo mu rwego rwa gatanu uzayogoza [Ubushinwa bw'amajyepfo]$_{LOC}$ vuba aha "A category 5 typhoon will hit South China soon"
- Inkubi y'umuyaga yo mu rwego rwa gatanu izayogoza

agace k'amajyepfo [y'ubushinwa]$_{GPE}$ vuba aha "A category 5 typhoon will hit southern China soon"
    - Note that there is a difference between the named region "South China" and the mention that is a description rather than a name for the region ("southern China")
- Yatembereye cyane muri manda ye: kuva muri [Aziya y'amajyepfo]$_{LOC}$ kugera muri Afghanistani, mu [burasirazuba bwo hagati]$_{LOC}$ ukagera Haiti "She traveled heavily during her tenure: from South Asia to Afghanistan, from the Middle East to Haiti"
- Further headache for Obama after collapse of [Middle East]$_{LOC}$ peace talks
- [Southeast Asia's]$_{LOC}$ muted reaction to the 2016 Defense White Paper

Oftentimes, place names are modified by words like "Southern", "lower", "West", and so on. When these modifiers are part of a location's official or accepted name they should be tagged as part of the location's name. For instance:

- Indwara ya Zika yakwiriye mu bihugu [by'Amerika y'amajyepfo]$_{LOC}$ "Zika spread in South American countries"
- [South America]$_{LOC}$
- [Lesser Antilles Islands]$_{LOC}$
- [East China Sea]$_{LOC}$

Even if the place name does not have "official" status but has an agreed-upon or widely-accepted definition and is in very frequent use, the string should be tagged as a location (LOC), as in:

- Yatembereye cyane muri manda ye: kuva muri [Aziya y'amajyepfo]$_{LOC}$ kugera muri Afghanistani, mu [burasirazuba bwo hagati]$_{LOC}$ ukagera Haiti "She traveled heavily during her tenure: from South Asia to Afghanistan, from the Middle East to Haiti"
- the [Middle East]$_{LOC}$
- the [West Bank]$_{LOC}$
- [Sub-Saharan Africa]$_{LOC}$
- [Eastern Europe]$_{LOC}$

When a term is not a well-known or generally-accepted name and the modifier is simply being used temporarily to describe some area of interest

within a named entity, do not tag them as part of the name. The names they modify should be tagged instead, with the correct entity type. E.g.:

- Inkubi y'umuyaga yo mu rwego rwa gatanu izayogoza agace k'amajyepfo [y'ubushinwa]$_{GPE}$ vuba aha "A category 5 typhoon will hit southern China soon"
- the southeastern [U.S.]$_{GPE}$
- upper [Yellow River]$_{LOC}$
- northern [Iowa]$_{GPE}$

Such place names can sometimes be tricky. If you are not sure whether or not a modifier is part of an official or well-known name, you should include the modifier as part of the place name.

→ **Guiding Principle**: If you know any word or phrase is a name, then tag it, and if you know it's not a name, then don't tag it.

4 Annotating Informal Data

We are annotating both formal text data (such as news articles which are usually written by professional journalists with careful editing) and informal text data (such as discussion forum threads, weblogs or tweets from Twitter, which are usually written by regular people like you and me without any careful editing). We will apply several guiding principles in annotating informal text.

4.1 Misspellings in Informal Text

It is very common for DF data to contain plenty of misspellings and shorthand. We still want to capture all instances of named Entities, even if they are spelled incorrectly. Consider the following sentence: the Lningguistic Data Consotrium in located in Phladelphia, pennsylvnia. Although the three named Entities are misspelled, we would still annotate "Lningguistic Data Consotrium" as an ORG, and "Phladelphia" and "Pennsylvnia" as GPEs.

4.2 Quoted Text in Discussion Forum Data

For discussion forum data, each document is a thread, with posts from multiple people. You won't be able to see the post boundary, since it is not needed for Simple Named Entity annotation. You may see repeated content because people quote original posts from others in the same thread. Please do annotate all names even if they are repeated in the same documents.

4.3 Twitter Data

4.3.1 Annotating Twitter Handles/Usernames (@) in Tweets

In Twitter documents, we'll annotate Twitter handles that are found in the text. Twitter users tag other users by typing the '@' symbol followed by their username. For this annotation task, we'll tag all usernames as named entities, and will include the '@' symbol (due to the full token rule). For example, from the text @WhiteHouse, we would annotate it as an ORG or LOC, depending on its usage in the context. For example,

I went to [@WhiteHouse]$_{LOC}$ for a meeting
[@WhiteHouse]$_{ORG}$ selected a new candidate.

Similarly, we would annotate [@Gr8Annot8r] as a PER and [@Colorado] as a GPE. However, names nested in usernames should not be tagged. For example, @IloveNY will be tagged [@IloveNY]$_{PER}$, but not [@IloveNY]$_{GPE}$.

4.3.2 Annotating Entities in Hashtags (#) in Tweets

It is very common for Tweets to contain hashtags, as this categorizes a user's posts. Similar to how we handle Twitter handles, we will not decompose content in hashtags. If a hashtag contains taggable named entities, we will tag the whole hashtag as the extent of such entity. In the hashtag, #iloveNY, you should annotate #iloveNY as a GPE. # is included in the extent.

What is different from how we treat Twitter handles is that for hashtags, if they contain more than one named entity, we would tag the hashtag tokens more than once, with the same extent. For example, in the hashtag #LDCinPhilly,

- [#LDCinPhilly]$_{ORG}$ referencing the entity "LDC"
- [#LDCinPhilly]$_{GPE}$ referencing the entity "Philadelphia"

If a hashtag contains two named entities of the same type, we tag the hashtag twice. For example, in the hashtag #PhillyinPA,

- [#PhillyinPA]$_{GPE}$ referencing the entity "Philadelphia"
- [#PhillyinPA]$_{GPE}$ referencing the entity "Pennsylvania"

5 Special Cases

5.1 No Nested Mentions

We generally cannot make overlapping or nested annotations – annotations within other annotations. Names cannot be broken down into smaller parts for annotation: we cannot tag a word within a longer name – even if that word is also a name itself. When the name of one entity contains within it another entity name, do not pull out the name of the other entity and mark it separately. Only tag the larger entity. For example, we cannot tag the word "European" inside "[European Union]" – even though "European" would be a valid location (LOC) entity by itself. Likewise, the GPE names in the following examples could not be tagged:

- [University of ~~[Minnesota]~~$_{GPE}$]$_{ORG}$
- [~~[Northern Cyprus]~~$_{GPE}$ Chess Federation]$_{ORG}$

When a GPE is nested in a name, sometimes it is difficult to tell whether the GPE is part of the longer name or whether there are two separate names. This occurs for many ORG and some LOC or GPE names. For example,

- US State Department
- Chicago Police Department
- French Alps
- British parliament
- Chinese congress
- French Guiana

The decision varies from case to case and may require checking external resources such as Wikipedia or searching the names on the internet to decide the extent of the names. You should limit the time you check online sources to 2-3 minutes. For the examples above, we might have:

- [US State Department]$_{ORG}$
- [Chicago Police Department]$_{ORG}$
- [French]$_{GPE}$ [Alps]$_{LOC}$
  ➔ The French Alps are the parts of the Alps mountain range within France, a range which is shared with other countries.
- [British parliament]$_{ORG}$
- [Chinese]$_{GPE}$ congress

➔ “Chinese congress” is not a name, as the name of China’s congress is the “National People's Congress”. However, “Chinese” is tagged as referring to the named entity “China”.

- [French Guiana]$_{GPE}$

5.2 Questions of Extent

When a phrase refers to more than one named entity, mark each entity separately. For instance, each of these examples contains two entities:

- [North]$_{GPE}$ and [South Korea]$_{GPE}$ signed the agreement.
- [Jimmy]$_{PER}$ and [Rosalyn Carter]$_{PER}$
- [North]$_{LOC}$ and [South America]$_{LOC}$

However, be careful not to split apart proper names that contain a conjunction. For instance,

the [Fish and Wildlife Service]$_{ORG}$

is the name of one organization and should be tagged as a single named entity (it's not referring to separate organizations with the names “Fish Service” and “Wildlife Service” but rather to the single “Fish and Wildlife Service” organization).

5.3 Token Extents in the Tool

The documents that you will annotate are processed so that words and punctuation are segmented into **tokens**. Each token will generally be a word, but may be a part of words in certain languages (such as an East Asian character). [2]

It is possible that you will find tokens in the text where a whitespace is missing and which include extra text attached to a word you want to select. This cannot be completely avoided. Please select any and all tokens that contain part of the name you want to tag, even if one of those tokens contains something extra. For instance, depending on how tokens are made for your language, a possessive marker may be included in the token

[2] The annotation tool that we use for entity annotation enforces tokenization on mention selection in order to prevent any selection of partial tokens. Regardless of what extent of characters is selected in the tool, the extent will automatically be adjusted to align with the ends of the selected token(s). For languages that are delimited by whitespace, mentions will start from the first letter after a whitespace and end at the last letter before a whitespace. For languages that do not rely on whitespace for delimitation, mention extent will coincide with the token boundaries provided by automatic tokenization output.

for a name. If the possessive marker is part of the token, the annotation tool will select the whole token even if you don't select the possessive marker. E.g., if you select [Canada] in

> … in Canada's national interest.

and if the "'s" is part of the token for "Canada", then the tool will highlight and select the whole token [Canada's] for you.

Extra text attached to a word inside a token may include punctuation (e.g., commas, periods, etc.) or even more than one word. Any extra "attached" punctuation and/or words within a token you select will simply stay included in the annotated mention extent. For example, we would want to select [Hawai'ian Islands]$_{LOC}$ in

> after the tsunami hit the Hawai'ian Islands,rescue operations…

However, if ",rescue" is part of the token that includes "Islands", then the tool would highlight and select the whole token [Islands,rescue]. The complete mention extent for this annotation must then become [Hawai'ian Islands,rescue]$_{LOC}$.

As a special case, if you find that the names (or parts of the names) of more than one named entity are joined in the same token, we will need to tag that token more than once. Please make exactly one annotation for each valid entity in the token. Here's an example of annotating two valid entities' names within one token:

> …before the Iran-Iraq war…
>
> - Entity1, Iran, tagged as: `[Iraq-Iran]GPE`
> - Entity2, Iraq, tagged as: `[Iraq-Iran]GPE`

5.4 Annotation Uncertainty

In some cases, you may encounter examples that you don't know how to handle. If so, you should proceed as follows:

- If it's an example not covered in the guidelines, note it in your copy of the guidelines and let your supervisor know about it.
- If it's a case where there's something wrong with the file you're working on, please mark the file as broken and leave a note in the pop-up text box to comment on the reason.

Your supervisor might not be available when you have a question or issue, so it's important for you to write down the problem so it can be

resolved later.

When you encounter a problem or have a question about a particular word or phrase and you can't get an immediate answer, you can check the entity as “Difficult” and make a note to address the group. That will let us easily find it later when we try to resolve the problem.

If the whole file is problematic (i.e., corrupted, font problems), stop working on it and mark the file as broken and leave a note in the pop-up text box to comment on the reason.