Guía de anotación de entidades nombradas medievales

Introducción
- ¿Qué es el reconocimiento de entidades nombradas (NER)?
- ¿Por qué un reconocedor de entidades nombradas para español medieval?
- La importancia de la anotación manual

Etiquetas
- Personas
  - Nombres de persona mondos y lirondos <persName>
  - Apodos y sobrenombres
    - Cuando el apodo aparece acompañando al nombre original de la persona <addName>
    - El apodo aparece solo (sin el nombre original de la persona) type="nickname"
  - Nombres de deidades, santos y figuras religiosas <persName type="_deity">
  - Santo padre, Criador, Cristo: todos los nombres de Dios <persName type="_nickname_deity">
- Roles
  - Cargos, títulos y profesiones <roleName>
  - Por la gracia de Dios <addName type="_legitimacy">
  - Relaciones de parentesco <roleName type="_family">
  - Don, señor, excelentísimo <roleName type="honorific">
  - Combinaciones de persName+roleName+placeName.
    - El caso Diego Sánchez de Bustamante o Alfonso de Molina
    - El caso El rey don Alfonso de Castilla
  - El caso de señor: todos sus etiquetados posibles
- Lugares
  - Países, ciudades, pueblos, regiones etc <placeName>
  - Edificios, instalaciones y monumentos <placeName type="_facility">
  - Accidentes geográficos (ríos, montañas, etc) <geogName>
  - Lugares mitológicos <placeName type="_mythological">
- Organizaciones
  - Pueblos, colectivos y agrupaciones <orgName type="_collective">
  - Organizaciones vs lugares: el problema de la metonimia
- Miscelanea <name>

Introducción

El objetivo de esta guía es establecer los criterios para anotar manualmente las entidades nombradas de un corpus de castellano medieval.

¿Qué es el reconocimiento de entidades nombradas (NER)?

El reconocimiento de entidades nombradas es el proceso por el que un programa informático detecta automáticamente las menciones a entidades (personas, lugares, instituciones, cargos, etc) que se mencionan en un texto dado. El reconocimiento de entidades nombradas nos informa de qué y de quiénes se habla en un texto y es una de las tareas fundamentales en la extracción de información textual.

¿Por qué un reconocedor de entidades nombradas para español medieval?

La rama del PLN (Procesamiento de Lenguaje Natural) ha tratado extensamente las dificultades y mecanismos para poder extraer automáticamente las apariciones de entidades nombradas de un texto, fundamentalmente aplicado al reconocimiento de entidades nombradas en textos escritos en lenguas modernas, como el español actual.

Sin embargo, la extracción de entidades nombradas es una tarea especialmente interesante para aplicar también a textos históricos. El reconocimiento de entidades nombradas aplicado a textos antiguos (como documentos notariales, crónicas, obras literarias y otros textos de época) nos abre la posibilidad de ubicar localizaciones en obras literarias clásicas, rastrear las apariciones y relaciones entre personajes históricos o descubrir las diferentes denominaciones por las que ha sido conocido un mismo lugar a lo largo del tiempo.

Los textos medievales, no obstante, presentan algunas particularidades que complican su procesamiento automático: no es ya que sigan unas normas diferente a la actuales, sino que en la época que fueron escritos muchos de estos textos no existía un estándar ortográfico fijado. Esta ausencia de estandarización se traduce en textos con constantes vacilaciones ortográficas y poca homogeneidad en el uso de mayúsculas; rasgos todos ellos utilizados como criterios fundamentales por los reconocedores de entidades nombradas convencionales para detectar entidades. Puesto que en general podemos decir que la relación entre las variedades históricas y variedad lingüística estándar actual no siempre resulta transparente, nos encontramos con que la detección de entidades nombradas aplicada al procesamiento de textos antiguos presenta una casuística propia que es necesario abordar a la hora de hacer un reconocedor de entidades nombradas que analice adecuadamente los textos en castellano medieval.

Esta guía de anotación forma parte del desarrollo de HisMeTag, un reconocedor de entidades nombradas diseñado específicamente para la detección de entidades en textos medievales. El desarrollo de HisMeTag se inscribe dentro del proyecto POSTDATA (Poetry Standardization and Linked Open Data), en concreto en el módulo de creación de recursos de PLN aplicados al dominio de la poesía.

La importancia de la anotación manual

Para poder evaluar la calidad de nuestro reconocedor de entidades nombradas medievales (HisMeTag) necesitamos partir de un corpus de textos medievales en el que las entidades nombradas que aparezcan hayan sido previamente anotadas por un humano. De este modo, podremos medir la capacidad de HisMeTag para etiquetar satisfactoriamente las entidades nombradas en comparación con el etiquetado hecho manualmente por un humano. Este corpus de textos en español medieval será, por lo tanto, nuestro corpus de evaluación.

La calidad de un corpus de evaluación anotado a mano depende directamente de la consistencia y la coherencia de la anotación humana. Las personas que anotan deben seguir siempre los mismos criterios y esos criterios deben ser seguidos por todos los anotadores ya que la coherencia en la anotación del corpus es la que nos permitirá obtener una evaluación fiable al comparar el corpus manualmente etiquetado con el etiquetado hecho por el programa. Aunque a priori pueda parecer evidente qué es una entidad nombrada y qué entidades deben ser etiquetadas, es habitual encontrarse con casos complicados al ponerse manos a la obra, tener múltiples posibilidades de anotación o que surjan divergencias de criterio entre los anotadores. El objetivo de esta guía de anotación es asegurar la coherencia en el etiquetado haciendo explícitos cuáles son los criterios que debemos seguir a la hora de anotar, cuál es el repertorio de etiquetas que vamos a utilizar y por qué tipo de anotado vamos a optar en los casos dudosos.

Etiquetas

Vamos a considerar cinco categorías de entidades nombradas: referidas a **personas** (Rodrigo Díaz), a **cargos y oficios** (notario del rey, adelantado de la frontera, juglar) a **lugares** (Burgos), a **organizaciones** (orden de Santiago) y **miscelánea** (para etiquetar aquellas entidades nombradas que sean relevantes pero que no encajen en ninguna de las categorías anteriores).

En los siguientes apartados se exponen los criterios que debemos seguir para saber cuándo un elemento del texto es una entidad nombrada que debe ser etiquetada, así como las etiquetas, atributos y en general la notación adecuada para etiquetar correctamente y, sobre todo, coherentemente cada caso. El etiquetado propuesto sigue en la medida de lo posible las etiquetas del estándar de TEI así como la jerarquía de NER de Sekine. Algunos valores de los atributos type que detallamos no cuentan con un uso previo documentado y asentado, sino que son propuestas originales de este trabajo. Hemos optado por utilizar la barra baja antepuesta para identificar aquellos valores que no tienen un uso previo y que hemos definido para este trabajo (type="_valor").

Una matización importante: las apariciones que nos interesan son aquellas entidades que aparezcan en el texto propiamente: las entidades que aparezcan en los metadatos (títulos, etiquetas referidas a la edición, parlamentos de personajes en las obras de teatro, etc) no nos interesan y no han de ser etiquetadas.

Personas

Nombres de persona mondos y lirondos <persName>

Etiquetamos bajo <persName> los nombres de persona que aparezcan en el texto. No vamos a distinguir entre nombres de pila y apellidos, puesto que esta distinción aplicada a nombres medievales puede generar más ambigüedades o errores que beneficios.

```
<persName>Celestina</persName>
<persName>Rodrigo Díaz</persName>
```

Apodos y sobrenombres

En ocasiones, las personas pueden aparecer mencionadas con apodos (el Campeador, Trotaconventos). Los apodos o sobrenombres se han de etiquetar de dos formas diferentes según el nombre original de la persona aparezca explícito o no.

Cuando el apodo aparece acompañando al nombre original de la persona <addName>

Cuando nos encontremos un nombre de persona asociado a un apodo etiquetaremos todo como <persName> y, anidado dentro del <persName>, etiquetaremos el apodo usando la etiqueta <addName>:

```
Yo soy <persName>Ruy Diaz <addName>el Cid
Campeador</addName></persName>
```

```
A <persName><addName>mynaya</addName> albarfanez</persName> mataron
le el cauallo.
```

La etiqueta <addName> puede servir además para etiquetar otros subelementos del nombre de la persona, como epítetos (*Pedro el Cruel*) o elementos especificativos como los numerales de los nombres de papas y reyes (*Inocencio IV*).

```
<persName>Pedro <addName>el Cruel</addName></persName>
<persName>Inocencio <addName>IV</addName>
```

El apodo aparece solo (sin el nombre original de la persona) type="nickname"

Cuando la mención a la persona contiene exclusivamente el apodo, anotaremos la mención como un <persName> con atributo type="nickname", que indica que la entidad no es el nombre real de la persona sino un alias o sobrenombre.

```
<persName type="nickname">Çid</persName>, en el nuestro mal vos non
ganades nada;
```

Los artículos o determinantes que aparezcan precediendo los apodos y formen parte del sobrenombre deben ir dentro de la etiqueta:

```
Así habló <persName type="nickname">Çid</persName>
Así habló <persName type="nickname">Myo Çid</persName>
Así habló <persName type="nickname">el Çid Campeador</persName>
```

Solo vamos a considerar sobrenombres aquellas denominaciones plenamente asentadas y que funcionan en la práctica como nombre de la persona. En textos medievales es habitual encontrar locuciones que sirven para denominar indirecta y poéticamente a los personajes ("Mío Cid, el que en buena hora nasco"; "Paris, el de Troya") e incluso solas ("Así habló el que en buena hora ciñó espada"). Sin embargo, vamos a ceñirnos a etiquetar exclusivamente aquellas denominaciones que respondan a la pregunta "¿cómo se llama este personaje?". Puesto que nunca responderíamos "el de Troya" o "el que en buena hora ciñó espada", vamos a dejar fuera de la etiqueta estas denominaciones (y en cambio sí etiquetaremos "Minaya" o "Cid").

Nombres de deidades, santos y figuras religiosas <persName type="_deity">

En los textos medievales es habitual encontrarse con menciones a figuras divinas (Dios, Jesus, nombres de santos). Las menciones a figuras divinas también debemos considerarlas <persName>, pero para poder distinguirlos del resto de entidades de persona vamos a añadirles el atributo type="_deity" dentro de la etiqueta. Esta marca nos permitirá después recuperar o filtrar este tipo de entidades nombradas y distinguirlas del resto de entidades de persona.

```
<persName type="_deity">Dios</persName>
<persName type="_deity">Cupido</persName>
```

Los santos se consideran <persName type="_deity"> pero con el añadido de que si van precedidos de la partícula "santo/a", "san", "arcángel", etc, debemos considerar estos elementos un <roleName>:

```
<persName type="_deity"><roleName>san</roleName> Miguel</persName>
```

```
El <persName type="_deity"><roleName>arcángel</roleName>
Gabriel</persName>
El <persName type="_deity"><roleName>arcángel</roleName>
<roleName>san</roleName> Gabriel</persName>
```

Santo padre, Criador, Cristo: todos los nombres de Dios <persName type="_nickname_deity">

Es habitual que Dios y otras divinidades aparezcan referidas con sobrenombres y denominaciones indirectas pero muy asentadas y que deben ser recogidas como entidades: el Criador, Cristo, Rey de Reyes… En estos casos nos vemos abocados a escoger si preferimos etiquetarlos como deidades (es decir, con el type="_deity") o como apodos (type="_nickname"). Para no perder ninguna de las dos informaciones, vamos a optar por usar el atributo "_nickname_deity" que recoge que la entidad etiquetada se trata tanto de una divinidad como de un sobrenombre.

```
Así lo quiso el <persName type="_nickname_deity">Criador</persName>
```

Roles

En este trabajo vamos a etiquetar bajo la etiqueta <roleName> los roles. Consideraremos roles los cargos y títulos nobiliarios (*rey de Castilla, arzobispo de Toledo, abad, papa, alcalde*) pero también los oficios y profesiones de forma más general (*caballero, vasallo, notario del rey, adelantado de la frontera, juglar*). Los roles nos permiten asociar títulos y puestos a las personas que se mencionan en el texto, y nos dan además una información muy valiosa sobre los cargos relevantes del documento y sus relaciones.

Cargos, títulos y profesiones <roleName>

Los cargos y títulos sirven para denominar personas (*la reina de Castilla, el presidente del gobierno*, *el obispo de Oviedo*). En esos casos, estas menciones las anotaremos con la etiqueta <roleName>

```
Por mandato del <roleName>rey</roleName>
```

La etiqueta <roleName> puede llevar anidadas otras etiquetas de localización o de organización:

```
El <roleName>rey de <placeName>Castilla</placeName></roleName>.
El <roleName>presidente del <orgName>gobierno</orgName></roleName>.
```

Los <roleName> pueden aparecer anidados formando parte de un <persName>:

```
<persName>Alfonso <roleName>rey de
<placeName>Castilla</placeName></roleName></persName>.
```

Estos casos están explicados con más detalle en el [apartado *Combinaciones de persName+roleName+placeName*].

Por la gracia de Dios <addName type="_legitimacy">

En textos medievales, es habitual encontrar el formalismo “por la gracia de Dios” intercalado entre el nombre y el título cuando se habla de los cargos de una persona (particularmente en el caso de los reyes):

```
Alfonso X por la gracia de Dios rey de Castilla
```

Podemos considerar que esta muletilla es un giro hecho que forma parte del <roleName>: al fin y al cabo, la muletilla “por la gracia de Dios” hace referencia al origen del que emana la legitimidad para que la persona mencionada ostente un cargo (en este caso, Dios). De hecho, nunca nos encontraríamos esta muletilla acompañando exclusivamente al nombre si no llevase un cargo asociado (**Don Alfonso por la gracia de Dios?*). Vamos, por tanto, a considerar que “por la gracia de Dios” es un giro hecho que forma parte del <roleName> (es decir, debe aparecer anidado dentro de él). Lo etiquetaremos usando la etiqueta <addName> (la misma que usamos para los epítetos y apodos), aunque añadiéndole el atributo type="_legitimacy” para distinguirlos.

```
<persName>Alfonso
<roleName>
<addName type="_legitimacy">por la gracia de <persName
type="_deity">Dios</persName>
</addName>
rey de <placeName>Castiella</placeName>
</roleName>
</persName>
```

Relaciones de parentesco <roleName type="_family">

De forma parecida a lo que ocurre con cargos y títulos, las relaciones de parentesco sirven para identificar y denominar a personas y para establecer relaciones entre ellas. En este caso también se han de etiquetar como <roleName> aunque con un atributo type="_family” para distinguirlos como un subtipo especial de <roleName>.

```
Vuestras <roleName type="_family">hijas</roleName> infantes aún son.

Mi biuda <roleName type="_family">madre</roleName> como sin
<roleName type="_family">marido</roleName> y sin abrigo se viesse...

El <roleName type="_family">fiio de
<persName>Johan</perName></roleName>
```

Igual que en el caso anterior, los roleName familiares puede aparecer solos o anidados dentro de una etiqueta <persName> cuando aparece explícito el nombre de persona:

```
Lo firma <persName>Alfonsso <roleName type="_family">fiio del
<persName><roleName>rey</roleName>
Johan</perName></roleName></persName>

<persName>
<roleName type="honorific">don~a</roleName> costanc'a
<roleName type="_family">hermana del
<persName>
<roleName>Rey</roleName>
<roleName type="honorific">don</roleName> sancho
Ordon~ez
</persName>
</roleName>
</persName>

<persName><roleName type="honorific">don~a</roleName> al-
dara sanchez. <roleName type="_family">fija del
<persName><roleName>rey</roleName> <roleName
type="honorific">don</roleName>
sancho Ramjrez de
<placeName>navarra</placeName></persName></roleName></persName>

<persName><roleName>Jnfante</roleName><roleName
type="honorific">don</roleName> Alfonso
<roleName type="_family">fiio del
<persName><roleName>rey</roleName><roleName
type="honorific">don</roleName> Ferrando
</persName>
</roleName>
<roleName>heredero de
<placeName>Castiella</placeName> e de
<placeName>Toledo</placeName> de
<placeName>Leon</placeName> e de
<placeName>Gallicia</placeName> de
<placeName>Cordoua</placeName> e de
<placeName>Murcia</placeName>
</roleName>
</persName>
```

Don, señor, excelentísimo <roleName type="honorific">

Es habitual que los nombres de persona aparezcan precedidos de formas de tratamiento (*don, señor, excelentísimo, muy noble, ilustrísima*, etc). Estas partículas deben ir dentro de la etiqueta <persName> que englobe al nombre correspondiente y a su vez identificadas como un <roleName> anidado con el atributo type="honorific".

```
<persName><roleName type=”honorific”>don</roleName>
Alfonsso</persName>
<persName><roleName type=”honorific”>señor</roleName>
Pleberio</persName>
```

A veces las formas de tratamiento no aparecen acompañando a <persName> sino a un <roleName> en los que no hay un nombre de persona explícito:

```
<roleName><roleName type=”honorific”>señor</roleName>
Alcalde</roleName>
```

No es raro encontrar este tipo de formas de tratamiento de forma aislada (sin nombre de persona o cargo al que acompañen), habitualmente como vocativos. En estos casos también los etiquetaremos:

```
Suplico a. <roleName type="honorific">U<uestra> M<erced></roleName>
reciba el pobre seruicio...
```

Combinaciones de persName+roleName+placeName.

En las menciones a personas (sobre todo en textos notariales), es habitual encontrar estructuras muy complejas en las que conviven menciones a nombres de pila, apellidos, cargos, formas de tratamiento, lugares de procedencia y relaciones de parentesco:

```
Don Odoart fiio primero e heredero del rey Henrric de Anglatierra
```

Como norma general, podemos decir que, siempre que esté explícito un nombre de pila de persona, consideraremos que la etiqueta fundamental (la que más abarca y queda más externa) es <persName>, que tendrá anidadas las demás <roleName> o <placeName>. Es decir, **siempre que aparezca una entidad nombrada formada por la combinación de nombre(s) de persona + cargos + relaciones de parentesco, la etiqueta principal (la más exterior) ha de ser siempre la de <persName>.**

```
Por mandato del <persName><roleName>rey</roleName>
Alfonso</persName>.
```

Algunas entidades pueden ser estructuras más largas y anidadas de lo esperable:

```
<persName>
<roleName type=”honorific”>Don</roleName> Odoart
<roleName type=”_family”>fiio primero</roleName> e
<roleName>heredero del
<persName><roleName>rey</roleName>
Henrric de <placeName>Anglatierra</placeName>
</persName>

Soy <persName>Aragorn,
<roleName type=”_family”>hijo de
<persName>Arathorn</persName></roleName>,
<roleName>heredero de <persName>Isildur</persName></roleName>,
```

```
<roleName>señor de los dúnedain</roleName>,
<roleName>heredero del
<orgName>trono de <placeName>Gondor</placeName></orgName>
</roleName>,
apodado <addName>Trancos</addName>,
<roleName>Capitán de los Montaraces del Norte</roleName>
</persName>
```

El caso *Diego Sánchez de Bustamante* o *Alfonso de Molina*

En textos notariales es posible encontrarse con personas denominadas por un nombre de pila seguido de “de + nombre propio”, en el que el nombre propio podría tratarse tanto de un apellido como de una localidad. En estos casos tendríamos dos etiquetados posibles, según consideremos esas estructuras apellidos sin más o localizaciones:

i.

```
<persName>Diego Sánchez de Bustamante</persName>
<persName>Alfonso de Molina</persName>
```

ii.

```
<persName>Diego Sánchez de
<placeName>Bustamante</placeName></persName>
<persName>Alfonso de <placeName>Molina</placeName></persName>
```

En estos casos ambiguos, debemos consultar la existencia de la localidad en cuestión en el tiempo del personaje en cuestión y ponderar la posibilidad de que se trate, efectivamente, de un patronímico (en ocasiones, es habitual encontrar información referente al linaje o a títulos nobiliarios asociados a la familia). En caso de no encontrar información evidente que corroboren con una cierta solidez que se trata de una localidad asociada a la procedencia del personaje, lo etiquetaremos como parte del nombre sin especificar si es un <placeName> (i).

El caso *El rey don Alfonso de Castilla*

En ocasiones, las entidades que aparecen nombradas con una mezcla de cargos, nombres de persona y procedencia pueden generarnos dudas a la hora de decidir qué elementos van con qué. Pensemos en el caso *rey don Alfonso de Castilla*. Si quisiéramos considerar que el <roleName> es *rey de Castilla* tendríamos un problema, ya que tendríamos dos posibilidades, ninguna de ellas satisfactoria:

1. Hacer una etiqueta discontinua, interrumpida por “Alfonso”.
2. Considerar todo como un <roleName> (*Alfonso* incluido), lo que incumpliría la directriz que establece que siempre que haya nombres explícitos de persona debemos anotar con <persName> como etiqueta primordial.

En estos casos vamos a considerar que el <roleName> abarca solamente *rey*, y que *Alfonso de Castilla* es todo parte del <persName>. Es decir, *de Castilla* dependería de Alfonso, no de rey, y por lo tanto no formaría parte del <roleName>.

```
<persName>
<roleName>rey</roleName>
```

```
<roleName type="honorific">don</roleName>
Alfonso de <placeName>Castilla</placeName>
</persName>
```

Esta solución (quizá a primera vista no muy ortodoxa) resuelve varios aspectos:

1. Evita tener <roleNames> discontinuos.
2. Obedece la directriz de que, cuando haya nombres de persona, <persName> sea la etiqueta principal.
3. Es coherente con otros sistemas de etiquetado de NER previos (Sekine).
4. Más allá del problema de la notación en sí, es además cuestionable que en este tipo de giros el roleName sea efectivamente "rey de Castilla". Este tipo de construcciones solo es habitual para denominar reyes y títulos nobiliarios asociados a regiones, en los que el nombre de la región actúa prácticamente como apellido o patronímico de la persona. No es infrecuente ver a estos personajes nombrados como *Isabel de Castilla* (sin que haya un *reina* antepuesto), en el que el territorio es una parte del nombre de la persona. Es significativo que este tipo de construcciones CARGO+NOMBRE+DE+TERRITORIO solo funcione con títulos nobiliarios (*la alcaldesa Manuela Carmena de Madrid* es muy improbable), por lo que consideramos que la propuesta de incluir el nombre del territorio como parte del nombre de la persona (y no del <roleName>) es semánticamente sólida, además de más práctica.

Lugares

Países, ciudades, pueblos, regiones etc <placeName>

Etiquetamos bajo <placeName> los nombres de pueblos, ciudades, países, regiones o localidades y unidades geopolíticas en general que aparezcan en el texto.
Ejemplo:

```
<placeName>Salamanca</placeName>
<placeName>Castiella</placeName>
```

Las estructuras tipo "reino de…" o "ciudad de…" deben etiquetarse como parte de la misma estructura:

```
<placeName>reyno de Murcia</placeName>
<placeName>ciudad de Toledo</placeName>
<placeName>tierras de Egipto</placeName>
```

Los <placeName> pueden aparecer anidados formando parte de nombres de persona (como <persName> o <roleName>)

```
<persName>Alfonso de <placeName>Castilla</placeName></persName>
<persName>Rodrigo Díaz de <placeName>Vivar</placeName></persName>
<roleName>rey de <placeName>Aragón</placeName></roleName>
```

Edificios, instalaciones y monumentos <placeName type="_facility">

Las menciones a edificios o monumentos que correspondan a una ubicación concreta deben aparecer también como <placeName> con el atributo type="_facility".

```
<placeName type="_facility">mezquita de Córdoba</placeName>
<placeName type="_facility">puerto de Alicante</placeName>
<placeName type="_facility">castillo de Cieza</placeName>
```

A la hora de etiquetar edificios puede asaltarnos la tentación de etiquetar absolutamente cualquier edificio que aparezca mentado en el texto. No obstante, solo debemos etiquetar como <placeName type="_facility"> aquellas denominaciones que se correspondan con el nombre de un lugar concreto: aunque Alcántara tenga varios puentes, solo el puente romano clásico es conocido como "el puente de Alcántara". Aunque existan otros puentes en Alcántara, no podemos decir que el nombre de esos puentes sea "puente de Alcántara", por lo que no deben ser etiquetados como tal.

Ante la mención de un edificio que no sepamos si debe ser etiquetado o no podemos preguntarnos: "Si nos preguntasen cuál es el nombre de este edificio, ¿responderíamos esta denominación? ¿Esperaríamos encontrar un artículo de Wikipedia bajo este título?". Si la respuesta es afirmativa, es altamente probable que nos encontremos ante una entidad que debe ser etiquetada. "Mezquita de Córdoba", "catedral de Burgos" o "puente de Alcántara" son los nombres de esos monumentos y por lo tanto son entidades nombradas que se deben etiquetar (los artículos "la", "el", etc no se consideran parte de la entidad nombrada y deben quedar fuera de la etiqueta).

Este mismo razonamiento es el que nos permite saber cuándo debemos incorporar una localización como parte de la entidad nombrada y cuando es un locativo aparte. El nombre de la BNE es "Biblioteca Nacional de España", por lo que todo ello es el nombre de la entidad. Si apareciera "Biblioteca Nacional de España de Madrid" para hablar, por ejemplo, de la sucursal de la Madrid (frente a la de Málaga, por ejemplo", deberíamos sacar fuera Madrid como un <placeName> independiente:

```
<placeName type="_facility">Biblioteca Nacional de
España</placeName>
<placeName type="_facility">Biblioteca Nacional de
España</placeName> de <placeName>Madrid</placeName>
```

Accidentes geográficos (ríos, montañas, etc) <geogName>

Los accidentes geográficos y localizaciones en general referidas a la geografía física deben etiquetarse como <geogName>.
Ejemplo:

```
<geogName>río Tormes</geogNames>
```

Al igual que los <placeName>, los <geogName> pueden aparecer anidados formando parte de nombres de persona.

```
<persName>Lázaro de <geogName>Tormes</geogName></persName>
```

Lugares mitológicos <placeName type="_mythological">

Las localizaciones de lugares religiosos o mitológicos (el Paraíso, la laguna Estigia) van etiquetados como <placeName> o <roleName> según corresponda y con el atributo type="_mythological".

```
<placeName type="_mythological">Paraíso</placeName>
<geogName type="_mythological">laguna Estigia</geogName>
```

Organizaciones

Se etiquetan como <orgName> los nombres de organizaciones, asociaciones o instituciones. En el caso de los textos medievales, los más habituales son órdenes religiosas.

```
La <orgName>orden de Calatrava</orgName>
Los <orgName>frayles de guadalupe</orgName>
```

De forma parecida a lo que ocurría respecto a los nombres de lugar integrados dentro de otros nombres de lugar, no se pondrán etiquetas anidadas en aquellos casos en los que el nombre lugar forme parte del nombre de la organización:

```
La <orgName>corona de Castilla</orgName>
La <orgName>Universidad Complutense de Madrid</orgName>
```

Pueblos, colectivos y agrupaciones <orgName type="_collective">

Siguiendo la guía TEI y los criterios de anotación de NER de Sekine, vamos a considerar organización en un sentido amplio: cualquier conjunto de personas que forme una entidad común puede ser considerada un <orgName>. *Los moros, los judíos* o *los cristianos* pueden ser ejemplos de <orgNames>. Este tipo de asociaciones las identificaremos usando el atributo type="_collective".

Otros grupo de personas cerrados, coherentes y que funcionan como una unidad también serán considerados <orgName> de este tipo. Así, *los Reyes Magos*, *los doce apóstoles* o *los santos del cielo* los etiquetaremos como <orgName type="_collective">, puesto que son entidades no muy distintas a como lo podrían ser *los Beatles*.

¡Ojo! Este tipo de menciones serán consideradas entidades cuando tengan naturaleza de entidad única en sí misma y actúen como colectivo monolítico, no cuando hagan referencia a la suma de varios individuos. Por lo tanto, cuando en *La Celestina* dice:

Oye a Salomón do dice que las mujeres y el vino hacen a los hombres renegar. Conséjate con Séneca y verás en qué las tiene. Escucha al Aristóteles, mira a Bernardo. ***Gentiles, judíos, cristianos y moros, todos en esta concordia están.***

Deberemos considerar que *gentiles*, *judíos*, *cristianos* y *moros* son <orgName type="_collective"> puesto que hace referencia a los pueblos o colectivos de forma genérica que engloba a todos los individuos:

```
<orgName type="_collective">Gentiles</orgName>, <orgName
type="_collective">judíos</orgName>, <orgName
type="_collective">cristianos</orgName> y <orgName
type="_collective">moros</orgName>, todos en esta concordia están.
```

Sin embargo, cuando en El cantar del Mío Cid dice:

```
Tres mill moros caualgan & pienssan de andar
```

No se está hablando de *los moros* como colectivo o como pueblo sino que está haciendo referencia a tres mil individuos que caracteriza como "moros". Por lo tanto, este caso no es un <orgName>.

En ocasiones puede ser difícil distinguir cuando una mención es simplemente un <roleName> en forma plural y cuando constituye un <orgName type="_collective">. Una mención a *los poetas* o *los escribanos* por ejemplo será en principio un simple <roleName> porque denomina a la suma de muchos escribanos o poetas de forma genérica, sin que el colectivo en sí tenga estructura o entidad clara, pero tendremos que estudiar cada caso en su contexto. Como siempre, los ejemplos que aporta Sekine y pensar si esperaríamos encontrar la entidad recogida como tal en la Wikipedia (y de qué manera: en singular como descripción de una profesión (como *Escriba*) o en plural como colectivo (*Los judíos*, *Masones*) puede ayudarnos a dirimir qué casos son entidad de organización y qué casos son roleName.

Para poder ver con claridad la diferencia entre un caso y otro podemos pensar en la diferencia que hay entre los casos:

`Vinieron dos beatles a la fiesta (Ringo y George).` [*dos beatles* son individuos]

`Estaban dos caballeros de la Mesa Redonda, sir Galahad y sir Lancelot` [*dos caballeros de la Mesa Redonda* son individuos]

Frente a:

`Vinieron los Beatles a la fiesta.` [*los Beatles* son una entidad]

`Estaban los caballeros de la Mesa Redonda.` [*los caballeros de la Mesa Redonda* son una entidad]

(Ojo: en estos casos los numerales nos dan la pista de que la mención no hace referencia a la entidad completa en sí, pero no debe ser el único criterio. Algunas entidades contienen numerales como parte de la entidad, como *los Cien Mil Hijos de San Luis*, *los tres Reyes Magos* o *los doce apóstoles*).

Organizaciones vs lugares: el problema de la metonimia

Una duda recurrente es la de no saber cómo etiquetar aquellos casos en los que el nombre de un lugar representa al mismo tiempo una localización física y un ente administrativo.

Ejemplos:
Moncloa decide cambiar de estrategia.
La UCM vota al nuevo rector.
Alemania invade Polonia.

En estos casos, si la localización física se corresponde con un distrito/región/localidad o cualquier otra unidad geopolítica (ciudades, países, etc) debemos considerarlo <placeName>. En caso de que la localización sea de cualquier otro tipo (universidades, instituciones, etc), optaremos por <orgName>.

```
<orgName>Moncloa</orgName> decide cambiar de estrategia.
La <orgName>UCM</orgName> vota al nuevo rector.
<placeName>Alemania</placeName> invade
<placeName>Polonia</placeName>.
```

Incluso siguiendo este criterio, algunos casos pueden resultar problemáticos: ¿el imperio romano es un lugar o una organización? Podemos seguir como pista general qué rasgos caracterizan a la entidad. Si los rasgos que definen a la entidad son coordenadas geográficas o extensión en superficie, por ejemplo, probablemente nos encontremos ante un lugar. Pero si lo que caracteriza o define a una entidad es quiénes son sus gobernantes o el tipo de cargo que ostenta su principal responsable, entonces probablemente sea una organización. *Roma* puede ser una localización, pero en el caso de *imperio romano* optaremos en organización antes que en lugar, puesto que denomina a una forma de organización política más que a una superficie (imperio romano vs república; mismo territorio, distinta organización política). Cuando decimos que se desintegra el imperio romano, hablamos de la desintegración de una estructura política, no de la aniquilación de un territorio.

Miscelanea <name>

Otras apariciones que creamos que deben ser etiquetadas como entidades nombradas pero que no quepan bien en las anteriores categorías o nos generen dudas las podemos marcar como <name> a secas.
Ejemplos:

```
<name>Colada</name> y <name>Tizona</name> (nombres de las espadas del Cid)
El <name>Evangelio</name>
Cinco <name>Pater nostres</name>, y cinco <name>Aue marias</name>
Hacer <name>Sant Juan</name>
```

(“Hacer San Juan” es una expresión hecha que hace referencia al hecho de irse de un sitio sin haber cumplido un contrato. Hace referencia a la fecha del día de San Juan, no a una persona o al santo).