

GUÍA DE ANOTACIÓN

PROYECTO GEMECO

Visibilización de brechas de género en
medios de comunicación

SINAI “SISTEMAS INTELIGENTES DE
ACCESO A LA INFORMACIÓN”

ÍNDICE

1. INTRODUCCIÓN.....	1
2. GUÍA DE ANOTACIÓN.....	1
2. 1. CUESTIONES PRELIMINARES.....	1
2.1.1. <i>Formato de los datos</i>	2
2.2. REFERENCE.....	2
2.2.1. <i>Reference unknown</i>	3
2.3. QUOTATION.....	3
2.3.1. <i>Citas directas</i>	4
2.3.2. <i>Citas indirectas</i>	4
2.3.2.1. Citas indirectas con forma de oración compleja o subordinada.....	5
2.3.2.2. Citas indirectas con forma de oración compuesta o coordinada.....	5
2.3.2.3. Citas dentro de citas.....	6
2.4. SPEAKER.....	6
2.4.1. <i>Quotespeaker</i>	7
2.5. VERB.....	7
2.6. TRIGGER.....	7
2.6.1. <i>Listado de TRIGGER</i>	8

1. INTRODUCCIÓN

Este documento describe la guía de anotación utilizada para generar el corpus anotado en español que se usará para desarrollar el sistema GeMeCo: Visibilización de brechas de Género en Medios de Comunicación, basado en el *GenderGapTracker*¹.

Se trata de un sistema automatizado en tiempo real que mide la proporción de hombres y mujeres citados en los principales medios de comunicación canadienses. La versión en español utilizará artículos publicados en seis medios de comunicación españoles: El Mundo, ABC, El País, La Vanguardia, Público y El Periódico. Esta guía toma como base la guía de anotación de *GenderGapTracker* utilizada para el inglés, adaptándola al español.

2. GUÍA DE ANOTACIÓN

2. 1. CUESTIONES PRELIMINARES

El núcleo de la anotación siempre es la cita (**QUOTATION**) que obligatoriamente debe tener una entidad (**REFERENCE**) a la que se le pueda asociar la cita. Por tanto, siempre que anotemos debemos tener obligatoriamente dos elementos: **REFERENCE** y **QUOTATION**. Es decir, una cita y alguien a quien atribuirle.

Los elementos que se anotarán son: **QUOTATION** (la cita en sí, ya sea directa (textual) o indirecta (parafraseada)) y su correspondiente **REFERENCE** (la entidad a la que se le atribuye dicha cita). De este modo, se anotan solo y exclusivamente citas que se pueda atribuir a alguna entidad presente en el texto. Para ello, nos podemos apoyar en los verbos *dicendi*, declarativos o de habla (decir, argumentar, alegar, narrar, predicar,...)².

Otros elementos circundantes a la cita son: **SPEAKER**, **VERB** y **TRIGGER**. Todos ellos se relacionan con **QUOTATION** (todos los vínculos o relaciones establecidas entre cita y demás elementos parten de la cita). Todos estos elementos mencionados se deben anotar si los hubiese.

También, es importante saber diferenciar si son palabras que ha dicho la entidad a la que pertenece la cita o si son palabras del propio periodista.

Además, se anotará el máximo de longitud posible en una cita. Ejemplo: *Llegó a confesar que [la ruptura le había hecho sufrir más dolor que el fallecimiento de Carlos Falcó por la incertidumbre de la situación].*

¹ <https://gendergaptracker.informedopinions.org/>

² Consultar la base de datos ADESSE.

Por otro lado, se ha de tener en cuenta que el punto final no se anota en la cita. En el siguiente ejemplo, el punto final se deja fuera de la cita: *la viuda de Carlos Falcó defendía que [no era conocedora de la ruptura ya que solo recibió un WhatsApp de dos líneas al que no dio mayor importancia].*

2.1.1. Formato de los datos

Cada cita debe tener un registro como el que se ilustra a continuación. Utilizaremos esta oración como ejemplo: *Mientras que el juez de la Audiencia Nacional aseguraba que habían roto el 12 de agosto.*

Annotation:

```
{
  "speaker": "el juez de la Audiencia Nacional"
  "verb": "aseguraba"
  "disparador": ""
  "quote": "habían roto el 12 de agosto"
  "speaker_index": "(x,x+32)"
  "verb_index": "(y, y+9)"
  "quote_index": "(z,z+27)"
  "reference": "Santiago Pedraz"
  "gender": "m"
}
```

El campo "gender" tiene tres posibles valores {m,f,u}

2.2. REFERENCE

Se anota la primera vez que aparece en el texto la entidad completa, es decir, con nombres y apellidos, a la cual se le puede atribuir al menos una cita. Se ha de identificar su género: **MALE**, **FEMALE** o **UNKNOWN**. Los dos primeros se asignarán a los nombres de hombre o de mujer que aparezcan como entidades con citas atribuibles. La etiqueta de **UNKNOWN** se asignará a aquellas entidades que no tengan género biológico tales como: periódicos, nombres de instituciones, partidos políticos, etc.

Además, si una entidad (**REFERENCE**) aparece varias veces en el texto solo se anota la primera vez que aparece, por tanto, las citas que le sean atribuidas se vincularán solo con esta primera aparición anotada del **REFERENCE**.

Pueden darse casos en los que la primera aparición del **REFERENCE** se dé como complemento del nombre con forma aparente de **SPEAKER**. En este caso, como el **REFERENCE** aparece después del supuesto **SPEAKER**, este último no se anotaría:

- *“Por lo tanto, siempre trabajamos en la excelencia”*(**QUOTATION**), destaca(**VERB**) la directora general (**SPEAKER**), la francesa Anne-Laure Suvage Cayrol(**REFERENCE**). - **REFERENCE** es un complemento del nombre, como en este ejemplo (el nombre Anne-Laure aparece por primera vez en el texto).

2.2.1. Reference unknown

Las citas atribuidas a empresas, grupos de personas, etc. se anotan solo en el caso de ser producidas por un colectivo concreto e identificable:

- *Fue entonces, cuando Société Générale (**REFERENCE-UNKNOWN**), banco propietario de la firma, anunció (**VERB**) su intención de adquirir LeasePlan por 4.900 millones de euros (**QUOTATION**).*
- *Desde la Asociación Española de Proveedores de Automoción (Sernauto) aseguran que el 60% de empresas de componentes continuaban afectadas por la falta de semiconductores.* Aquí la cita se atribuye a unas personas desconocidas pertenecientes a la empresa, por eso no la anotamos.
- *“Durante los primeros meses del año ha habido un retroceso de matriculaciones[...]”, constata la directiva.* Se cita un documento y no a una persona, lo dejamos sin anotar.
- *PSOE (**REF.- UNK.**) y Podemos (**REF.- UNK.**) han pedido esperar al Tribunal Supremo y a la Fiscalía para que saquen conclusiones para dar respuesta uniforme a la revisión de las sentencias.*

2.3. QUOTATION

Es la cita tanto directa como indirecta. De ella salen todas las relaciones hacia los demás elementos. La cita comienza a partir de la conjunción “que”, la cual nunca se anotará como parte de **QUOTATION**.

- *Mientras que el juez de la Audiencia Nacional aseguraba que habían roto el 12 de agosto (QUOTATION).*

2.3.1. Citas directas

Las citas directas suelen venir señaladas con comillas (tanto inglesas como españolas, dependiendo del periódico). Son las más fáciles de identificar y se incluyen en la anotación, como ocurre en este ejemplo:

- *Llegó a confesar que la ruptura le había hecho sufrir más dolor que el fallecimiento de Carlos Falcó por la incertidumbre de la situación y que “Aún no sé el motivo por el que tomó la decisión. No me ha dado explicaciones, solo la del WhatsApp. Ha sido muy traumático no poder hablar ni que se me expliquen las cosas” (QUOTATION).*

Sin embargo, existen casos donde el verbo (**VERB**) se encuentra inserto entre la cita: “---” vb “---”. No significa que existan dos citas distintas, pues es la misma, solo que el periodista, en su redacción, ha decidido cortarla poniendo el verbo en medio de la misma. En este caso, se tomará todo como una misma cita (**QUOTATION**) y, además, se anotará dentro de esta el verbo (**VERB**) con el cual se seguirá estableciendo una relación entre la cita y este.

- *“Sin duda alguna”, dice, “la generalización de la telemática aplicada a la gestión de flotas supone un paso importante”.*

En este ejemplo anotamos “Sin duda alguna”, dice, “la generalización de la telemática aplicada a la gestión de flotas supone un paso importante” como **QUOTATION**, así como “dice” como **VERB** solapándose con la anotación anterior.

2.3.2. Citas indirectas

Hablamos de cita indirecta cuando no se reproducen las palabras textuales, es por ello que formalmente no aparecen entrecomilladas. Ejemplo: *Raquel Sánchez destacó este lunes que [la incidencia de la huelga convocada por la Plataforma Nacional en Defensa del Transporte por Carretera es “mínima” y reveló que se han desplegado 50.000 agentes para evitar disturbios].*

En caso de que haya un verbo intercalado en la cita, esta se anota incluyendo el verbo, el cual también se anotará con la correspondiente etiqueta **VERB**, dentro de la cita. Además, se siguen estableciendo relaciones entre **QUOTATION** y **VERB**:

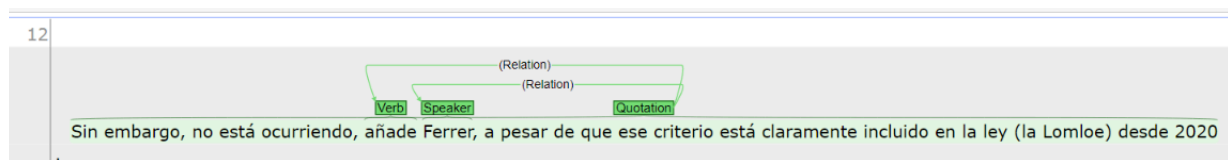


Fig. 1: ejemplo extraído de la anotación manual con la plataforma Inception.

2.3.2.1. Citas indirectas con forma de oración compleja o subordinada

Cuando tenemos una cita indirecta realizada como oración subordinada dependiente del verbo principal (**VERB**) se sacará de la cita el nexos conjuntivo subordinante (*que*). Por ejemplo, *es más, confiesa que le pidió expresamente que también fuera al suyo, algo que en su momento ocultó.*

En el caso de contar con una oración subordinada con cláusula de infinitivo, se señalará la cita a partir del verbo en infinitivo. Por ejemplo: *Mariano Rajoy* [**REFERENCE**, **MASCULINE**] *prometió* [**VERB**] *revertir esa situación* [**QUOTATION**].

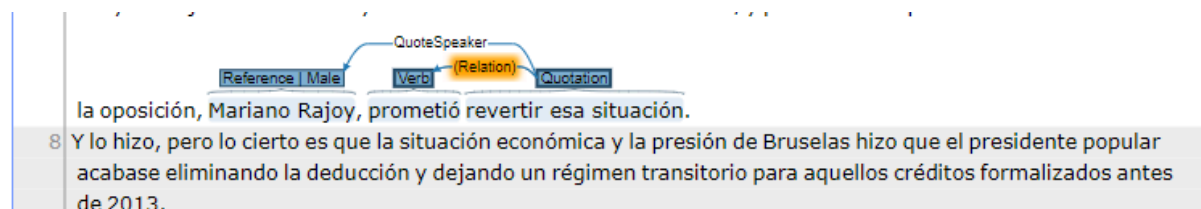


Fig. 2: ejemplo extraído de la anotación manual con la plataforma Inception.

2.3.2.2. Citas indirectas con forma de oración compuesta o coordinada

A la hora de anotar el texto, si contamos con una cita realizada sintácticamente como una oración compuesta o coordinada con la conjunción “y” hemos de señalar dos **QUOTATION** distintas que van a depender de verbos distintos, aunque pueden compartir el **SPEAKER**.

Por ejemplo, *el líder de la oposición comentó (VERB¹) que él no habría seguido esa vía y aseguró (VERB²) que no estaba de acuerdo con las decisiones tomadas en el congreso.*

Así mismo, hemos de prestar especial atención pues es muy común encontrar en las citas realizadas como una oración subordinada, dentro de esta, que se inserte una oración coordinada. Por ejemplo: *Pancho comentó a esta periodista que no sentía que Sabina le hubiese hecho “un feo”, y que entendía que no pudiese haber ido a los dos conciertos.*

En este caso, el verbo principal es *comentó*, por lo tanto, toda la oración subordinada a partir de la conjunción *que*, se anotaría como cita.

2.3.2.3. Citas dentro de citas

Cuando una cita indirecta contiene una cita exacta dentro, se anota la totalidad de la cita indirecta:

- *La viuda de Falcó (SPEAKER) ha confesado (VERB) que no borraría la etapa que ha vivido junto al juez ya que «es una persona que ha estado en mi vida. Hemos vivido cosas muy bonitas. Ha sido un año fantástico. Por el momento, no lo voy a hacer (QUOTATION)»*

2.4. SPEAKER

Este elemento se anota cuando estamos ante la aparición de la entidad en la misma oración donde se encuentra la cita una vez que ya ha sido mencionado anteriormente, es decir, el **REFERENCE** ya ha sido anotado en líneas previas a esta. Puede ser desde partes del apellido (ej: Doña, Pedraz, Castro, Roca, Lasso Vega,...), hasta lo que llamamos epítetos (la viuda, el gerente de Northgate, etc.). Estos nos remiten directamente al **REFERENCE** ya anotado previamente. Además, los pronombres también pueden funcionar como **SPEAKER**.

El sujeto que representa el **SPEAKER** se anota completo (todo el sintagma nominal incluido el artículo), si se trata de un nombre común. En el caso de que sea un nombre propio, los complementos del nombre que pueda tener no se tienen en cuenta:

- *Mientras que el juez de la Audiencia Nacional aseguraba que habían roto el 12 de agosto.*
- *Alexandru Constantin Chituta, del Museo Nacional Brukenthal (Rumanía). Aquí dejamos fuera la aposición “del Museo Nacional Brukenthal (Rumanía)”.*

En caso de tener un **SPEAKER** que haga referencia a dos o más de los **REFERENCES** anotados utilizando formas tales como: ambos, la pareja, los protagonistas, etc., se anotarán como **SPEAKER**. Por tanto, se han de establecer tantas líneas de relaciones entre la cita y las entidades como a **REFERENCES** anotados haga referencia el **SPEAKER**.

Por ejemplo: *Esther Doña (REFERENCE) y Santiago Pedraz (REFERENCE) [...] los protagonistas (SPEAKER) comunicaron ...*

2.4.1. Quotespeaker

Debido a que la herramienta utilizada no permite etiquetar dos veces la misma cadena de texto, en el caso de que la entidad (**REFERENCE**) sea el mismo que la persona (**SPEAKER**), se tendrá que añadir la etiqueta “QuoteSpeaker” a la relación entre “**QUOTATION**” y “**REFERENCE**”.

Se trata de una etiqueta que se utiliza cuando **REFERENCE** y **QUOTATION** se encuentran en la misma oración. Estos se ponen indistintamente si son citas directas o parafraseadas.

Por ejemplo: “*Para la Historia de Transilvania y Rumanía en particular, pero también para la Historia de Europa en general, si estos resultados son aceptados por la comunidad científica significará la adición de otra figura histórica importante en nuestra Historia*” (**QUOTATION**), afirma (**VERB**) Alexandru Constantin Chituta (**REFERENCE, MALE**), del Museo Nacional del Brukenthal (Rumanía).

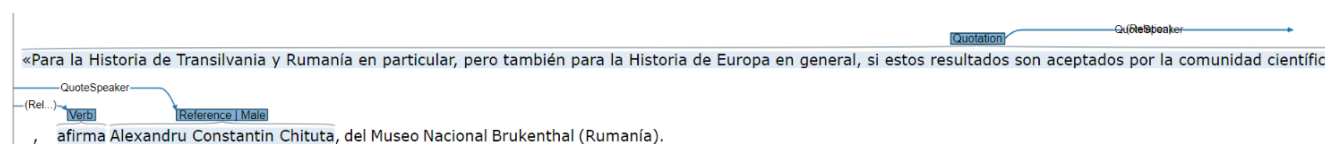


Fig. 3: ejemplo extraído de la anotación manual con la plataforma Inception.

2.5. VERB

Elemento que no tiene por qué aparecer siempre. Solo se anota el verbo principal. Ejemplos: *Llegó a **confesar**..., ...había sido **confesado**...*

2.6. TRIGGER

Funcionan como “anunciadores” de que va a darse una cita en el texto. Por ejemplo: ***Según** [TRIGGER] García Roji [SPEAKER], actualmente “un 35% del parque móvil no tiene el etiquetado que se exigirá en estas zonas” [QUOTATION],...*

Este *según* funciona como un “anunciador”, pues sin necesidad de que exista un verbo declarativo sabemos que va a venir una cita. De este modo, en el caso de encontrar lo siguiente: “*nuestra relación es imposible, hablamos algún día, cuídate y besos*” *según* **contó** poco después la viuda del Marqués de Griñón. No anotaremos *según* como **TRIGGER**, puesto que existe un verbo (*contó*) que introduce la cita y, por tanto, no está actuando como tal.

2.6.1. Listado de TRIGGER

Algunos **TRIGGER** son los siguientes que se incluyen en esta lista:

- De acuerdo con
- Según
- Teniendo en cuenta a
- Con base en
- Citando a
- Tal como
- Como lo hace notar
- Empleando las palabras de
- A juicio de
- Desde la posición de
- Dicho con palabras de
- En la opinión de
- Son de opinión
- Desde el punto de vista de
- En su opinión
- Acorde a
- En palabras de