

RUN-AS RESEARCH: BRAT INTERFACE GUIDELINE

The present guideline summarizes the steps to follow for the correct use of the annotation system interface we have built for our research.

The first interface that the created annotation system will show is a brief description of the annotation guide with some examples of the annotation (*Figure 1*).

Figure 1

Esta es la guía de anotación para el modelo FNDeepML del Proyecto LivingLang.

Este modelo de anotación representa los elementos semánticos principales de una noticia a partir de la estructura clásica WSH1 (What, When, Where, Who, Why, How), permite segmentar una noticia en sus tres partes fundamentales (headline, body, conclusion), y asociar entre sí los elementos semánticos WSH1 que están relacionados con el mismo evento.

El primer paso en la anotación consiste en segmentar la noticia. Para ello, entre cada párrafo se ubica un token especial que puede ser anotado con la etiqueta *Headline*, *Body* o *Conclusion*. Estas etiquetas marcan que a partir de ese token comienza dicha sección. Por ejemplo:

Pablo Iglesias pierde la batalla judicial por su honor contra Eduardo Inda y tendrá que pagar las costas ... La jueza cree que la información sobre sus cobros de Venezuela en un paraíso fiscal estaba contrastada ... El juzgado de Primera Instancia número 84 de Madrid ha dado un duro varapalo a Pablo Iglesias ... La magistrada Gladys López Manzanares cree ... La defensa de Pablo Iglesias exigía que ...

Es posible que una noticia no tenga ni *Headline* ni *Conclusion*, solamente *Body*.

El siguiente paso consiste en determinar todos los elementos semánticos, que son:

What: representa el núcleo evento o hecho que se narra.
When: representa el momento en que ocurre el evento.
Where: representa el lugar o locación general donde ocurre el evento.
Who: representa quiénes participan en el evento.
Why: representa los motivos o causas del evento.
How: representa el modo o forma en la que ocurre el evento.
Veamos algunos ejemplos:

- 1) Esta mañana se ha detectado el primer caso de COVID-19 en España ...
- 2) Debido a las complicaciones producidas por la enfermedad, el paciente falleció horas más tarde ...
- 3) La vacuna fue desarrollada a partir de la combinación de fármacos producidos anteriormente ...

Como se ha visto, los elementos semánticos que corresponden al mismo hecho se enlazan por una relación para diferenciar múltiples eventos o hechos que ocurren en el mismo pasaje. Es posible que la misma oración contenga más de un evento, por ejemplo:

La derrota de Donald Trump en las elecciones se suma al fracaso en contener la pandemia que ha demostrado la Casa Blanca ...

Luego de marcar todos los elementos semánticos, se procede a la verificación. Cada elemento puede ser atribuido independientemente según su criterio de veracidad, en *True*, *Fake* o *Unknown*.

Cada parte de la noticia también puede atribuirse, ya sea *Headline*, *Body*, o *Conclusion*, según el criterio de veracidad primario de esa sección.

Por último, es posible anotar también "las citas textuales". En este caso, la cita en sí no lleva un criterio de veracidad, sino que se anota la postura del artículo con respecto a la frase citada, según si está de acuerdo, desacuerdo, o se deja en blanco si no define una postura.

Veamos algunos ejemplos:

- 1) (sin indicar postura) Según el experto "la vacuna estará lista para finales de año" ...
- 2) (de acuerdo) Según ha indicado un portavoz de la OMS, "las mascarillas reducen considerablemente el riesgo de contagio", lo cuál hemos podido verificar en ...
- 3) (en desacuerdo) Trump ha mentido deliberadamente al decir "hemos ganado, y por mucho" ...

Cada *WHAT* se puede anotar con un atributo adicional (*Main-Event*) que indica si es el evento principal de la noticia.

Cada *WHO* se puede anotar con un atributo adicional (*ROLE*) que indica el rol de la entidad:

Si la entidad es responsable (causa) el evento, el rol es *SUBJECT*. Si la entidad recibe los efectos del evento, el rol es *TARGET*. También tener el rol *BOTH* si cumple ambos criterios.

Si la misma entidad se menciona más de una vez, aunque sea con diferentes formas textuales, se puede enlazar por una relación (coreference) para indicar que son el mismo concepto:

Donald Trump en un comunicado de prensa El presidente se refirió
Al título de la noticia se adiciona un atributo *TITLE-STANCE* que indica la concordancia entre este y el cuerpo. Puede ser *AGREE*, *DISAGREE*, o *UNRELATED*.

De forma opcional, se puede anotar el sujeto y predicado de una oración para asistir en el proceso de anotación.

Start annotating

To start the annotation, click the "Start annotating" button (*Figure 1*). Then, when clicking on the "Start a new pack" button (*Figure 2*), a new window opens with the raw

text to be annotated. If a pack has already started to be annotated and it has been interrupted, *Figure 3* will be directly displayed.

Figure 2

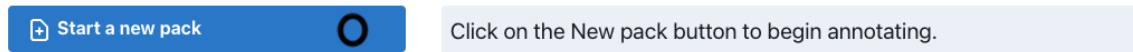
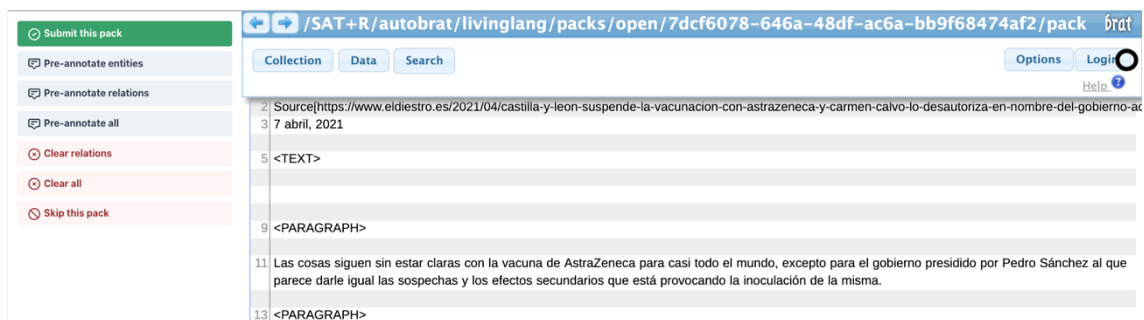


Figure 3



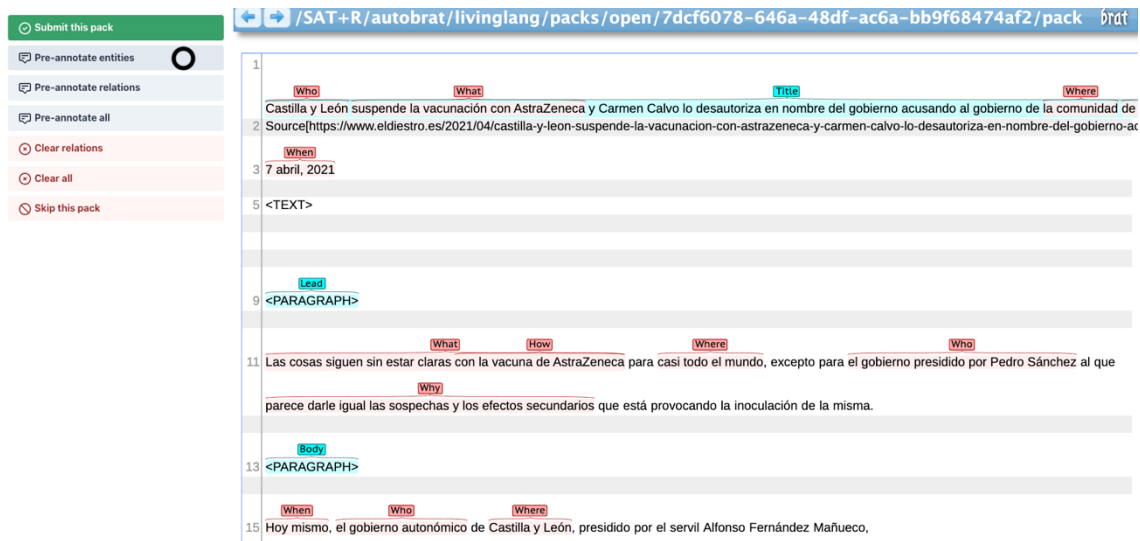
Before proceeding with the annotation, you must log in the annotation interface by clicking on the “Login” button shown in *Figure 3*. A pop-up window will be displayed as shown in *Figure 4*.

Figure 4



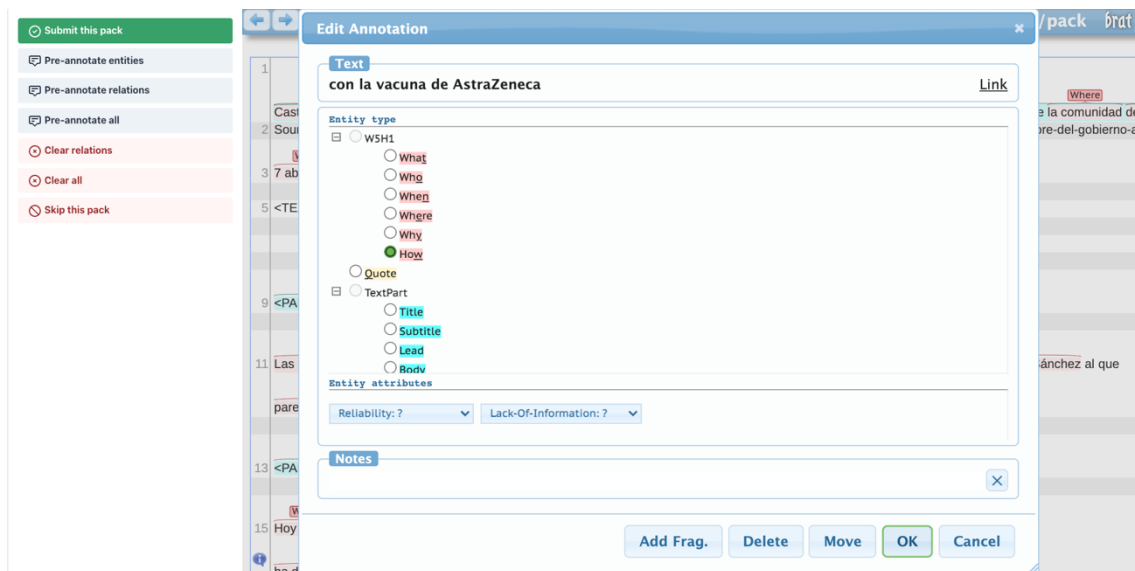
The next step requires entering the username and password assigned to the user. Press the “OK” button to introduce these personal details, as shown in *Figure 4*. Then you can pre-annotate the news by clicking on the “Pre-annotate entities” button (*Figure 5*).

Figure 5



With pre-annotation, the annotator has to start to review the annotations made by the model and annotate the reliability of the 5W1H labels. *Figure 6* shows an example of a HOW label. To access an annotation, you must double-click on the pre-annotated element or, on the contrary, if it is not annotated, select a text span and the annotation pop-up window will open automatically.

Figure 6



In *Figures 7* and *8* the annotation pop-up window shows the attributes available for the HOW label. Once the annotation of each label attribute is done, press the "OK" button.

Figure 7

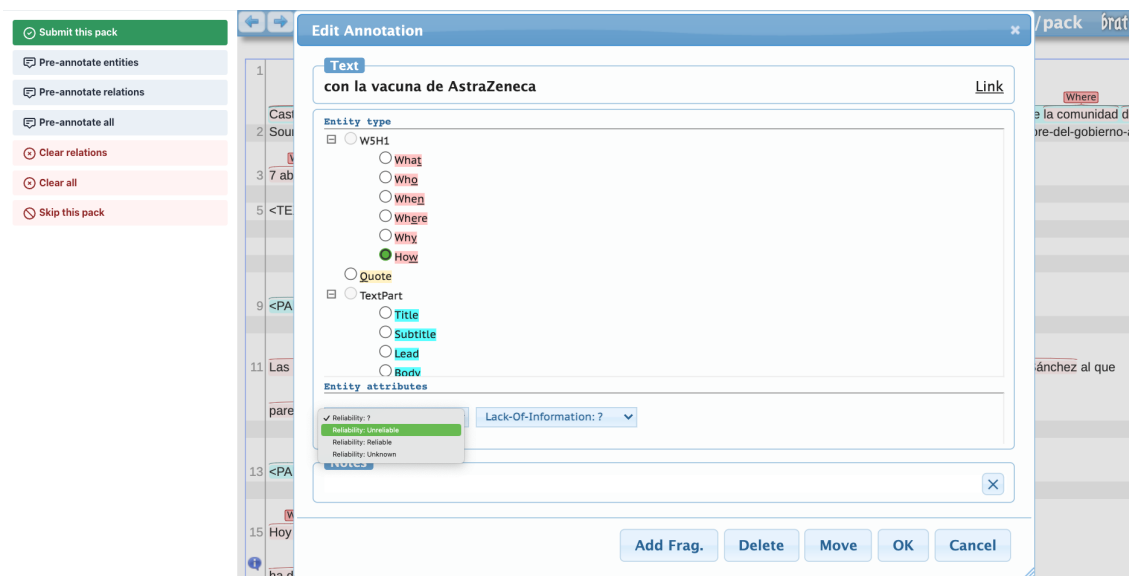


Figure 8

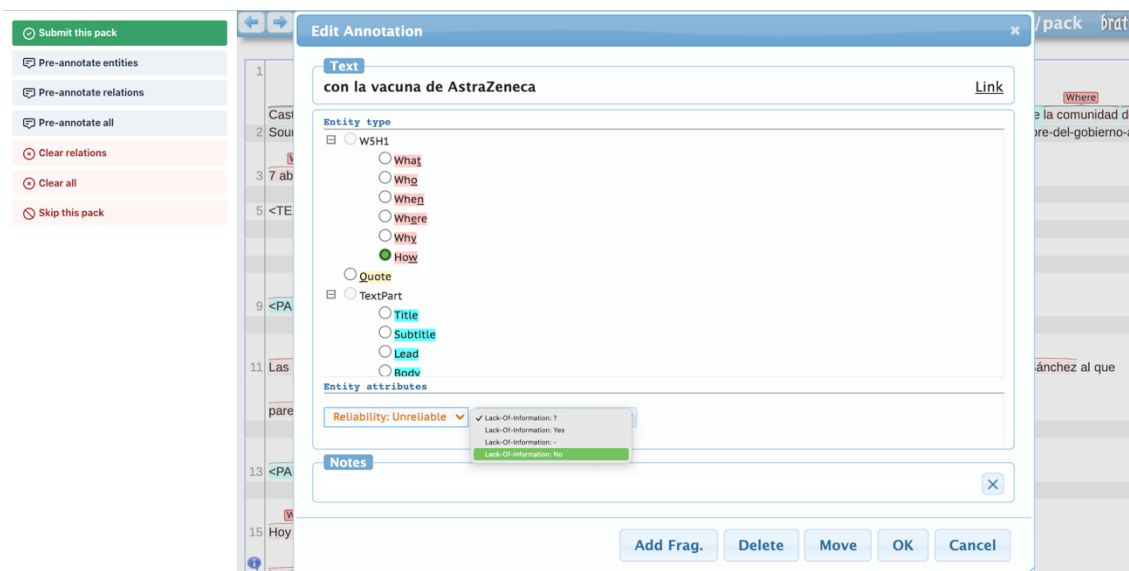
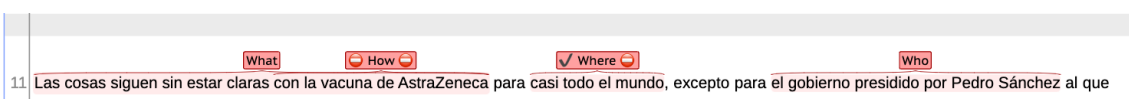


Figure 9 shows how the labels would look after annotating the corresponding attributes.

Figure 9



Once the entire news item annotation is finished, it should be stored in the annotated news folder. To perform this action, press the “Submit this pack” button. After saving the annotation, the next news item to be annotated will be displayed (if the news pack has not been completed).

Figure 10

Finally, if the news is not interesting for the annotation task or the corpus built, it can be ignored by clicking on the option “Skip this pack”. Then the active learning system suggests a new piece of news (see Figure 11).

Figure 11