

CONTRADICTION ANNOTATION GUIDELINE

This guideline has been created for the annotation of contradiction in news items. The following table of contents presents an overview of this annotation scheme: an introduction of our proposal, the different stages followed for annotating the dataset, the annotation labels and some examples.

TABLE OF CONTENTS

1. Introduction
2. Annotation stages
 - 2.1. Headlines modification
 - 2.1.1. Negation
 - 2.1.1.1. Example
 - 2.1.1.2. Example
 - 2.1.2. Antonym
 - 2.1.2.1. Example
 - 2.1.3. Numeric
 - 2.1.3.1. Example
 - 2.1.3.2. Example
 - 2.1.4. Structure
 - 2.1.4.1. Example
 - 2.2. Headline-Body relationship classification values
 - 2.3. Random mix
 - 2.3.1. Unrelated example

1. Introduction

Our proposal lies in the fact that detecting contradictions between headlines and body texts could be an important tool when fighting against disinformation in news. The automatic detection of contradictory information may allow to detect weaknesses in unreliable news and thus to make the difference between a reliable news item and a fake news.

This annotation has been applied in a corpus consisting of 7403 news items extracted from a reliable source (the news agency EFE). The dataset is in json format and it presents the following labels:

- **id**: identification number of the news item
- **headline**: main title of the news piece
- **body**: part including developed information of the news item
- **date**: date of creation of the news item
- **label_contradiction**: label indicating the contradiction relationship (or the lack of contradiction) between a headline and the body text of a news piece. This label contains several values that will be defined in detail in this guideline.

- **label:** label indicating the relationship between a headline and the body text of a news piece.

2. Annotation steps

After compiling the corpus composed of news obtained from a digital reliable source (EFE) and focused on both the economic and politics domains, the following stages has been followed for the annotation process:

2.1. Headline modification

The objective pursued at this stage of the annotation is to include specific modifications in the headline in order to contradict the body text. These modifications change the semantics of the sentence and make the headline totally contradictory to what is said in the body. The techniques used for modelling the headlines are as follows:

2.1.1. NEGATION

The negation adverb (not) is added in the correct position of the sentence, specifically before the verb in a way that affects semantically the meaning of the headline. It is important to analyse the whole sentence as there may be incompatible sentences with this type of annotation (see example 2.1.1.2.).

2.1.1.1 Example

ENGLISH

Original headline: The union representative **will discuss** Alcoa's "final proposal" tomorrow

Modified headline: The union representative **will not discuss** Alcoa's "final proposal" tomorrow

SPANISH

Original headline: El comité de empresa **debatirá** mañana la "propuesta final" de Alcoa

Modified headline: El comité de empresa **no debatirá** mañana la "propuesta final" de Alcoa

2.1.1.2. Example

ENGLISH

Original headline: Next target, budget

Modified headline: it cannot be modified (lack of verb)

SPANISH

Original headline: Próximo objetivo, presupuesto

Modified headline: no se puede modificar (falta el verbo)

2.1.2. ANTONYM

This annotation consists in finding the main verb of the sentence and changing it to an antonym. With this modification a field called “antonym” is used.

2.1.2.1 Example

ENGLISH

Original headline: The government pledges to **raise** public employees’ salaries after the elections

Modified headline: The government pledges to **cut** public employees’ salaries after the elections

SPANISH

Original headline: El Gobierno se compromete a **subir** los salarios a los empleados públicos tras los comicios

Modified headline: El Gobierno se compromete a **bajar** los salarios a los empleados públicos tras los comicios

2.1.3. NUMERIC

This annotation is only possible in headlines including dates or numbers and focuses on modifying the numeric element of a sentence: a year, a month, a day of the week, a percentage, etc. This annotation will not be possible when the sentence does not contain any figure or date (see example 2.1.3.2.).

2.1.3.1 Example

ENGLISH

Original headline: UK economy has grown by **3%** less due to Brexit, says SP

Modified headline: UK economy has grown by **5%** less due to Brexit, says SP

SPANISH

Original headline: La economía británica ha crecido un **3 %** menos por el brexit, según SP

Modified headline: La economía británica ha crecido un **5 %** menos por el brexit, según SP

2.1.3.2 Example

ENGLISH

Original headline: The nationalist and PP presidents criticise the lack of coordination and the return to work.

Modified headline: it cannot be modified (lack of figure or date)

SPANISH

Original headline: Los presidentes nacionalistas y del PP critican la descoordinación y la vuelta al trabajo

Modified headline: no se puede modificar (falta cifra o fecha)

2.1.4. STRUCTURE

The structure annotation seeks to invert the elements of a sentence by changing the order of the elements or replacing them (for example, by changing the subject and the object).

2.1.4.1 Example

ENGLISH

Original headline: **Arvind Krishna** replaces **Ginni Rometty** as IBM's CEO

Modified headline: **Ginni Rometty** replaces **Arvind Krishna** as IBM's CEO

SPANISH

Original headline: **Arvind Krishna** sustituirá a **Ginni Rometty** como consejero delegado de IBM

Modified headline: **Ginni Rometty** sustituirá a **Arvind Krishna** como consejero delegado de IBM

2.2. Headline-Body relationship classification labels

Once the headlines have been modified, the next step consists of classifying the relationship between the headline and the body text. For that purpose, four annotators trained to detect semantic relationships and contradictions have annotated several sentences by classifying them into **Compatible** (when the information of the headline and that of the body coincide) or **Contradiction** (when there is contradictory information). When the examples were classified with the label "Contradiction", the annotators had to indicate the type of contradictory relationship with the values described above: **neg** (negation), **ant** (antonym), **num** (numeric) or **str** (structure). If the annotators were of the opinion that there was no change, they must use the value "**none**", what means that any modification was used, there was no contradiction in the news item.

2.3. Random mix

For obtaining unrelated examples, an aleatory mix was carried out with all the news items. In that case, headlines were separated from their body texts and mixed in order to obtain news with no semantic relation between the headline and the body. The mixing process was verified,

so headlines and body texts could not coincide. This stage is done automatically without the use of annotators.

2.3.1. UNRELATED EXAMPLES

ENGLISH

Iberdrola buys US company PNM for 3,663 million euros

The spokesperson for Unidas Podemos in Congress, Ione Belarra, considers that the PSOE's strategy of blackmail and non-negotiation is a mistake, and has regretted that there is no other option on the table, because there are no contacts to negotiate between the two parties.

SPANISH

Iberdrola compra la estadounidense PNM con el pago de 3.663 millones de euros

La portavoz de Unidas Podemos en el Congreso, Ione Belarra, considera que la estrategia de chantaje y no negociación del PSOE es un error, y ha lamentado que no haya otra opción encima de la mesa, porque no existen contactos para negociar entre ambos partidos.