

Tagging negation guide for in Twitter negation Corpus

The project has the objective of detecting syntactic negation on Twitter.

FIRST PART

Tweet tagging with negations

In this first part of tagging, a tweet has to be highlighted only when it has a syntactic negation or when it does not. For this, a tagging tool will be used which has boxes for selection on the left side. They will select the TRUE boxes when a syntactic negation is present and the FALSE box when it is not. For this work, we considered syntactic negation as that which properly negates something in a sentence through a textual negative term. They usually are adverbs like *no, jamás, nunca*, but we also took into account prepositions such as *sin*, the determiners of *ningún*, conjunctions such as *ni*, etc. The possible variants can be written as *nel, nunka, ño*, etc.

Negative expressions that do not negate

There are discourse connectors, fillers or interjections that usually appear on tweets but that do not have a negation value in the sentence. Some examples are *no mames, sin duda, no cabe duda, no obstante, no más, nada más, de nada, nada menos, más_ que nunca, antes que nada, estoy a nada de...,ntc, npi*, etc., and any possible variations (that can be legit mistakes, or made on purpose); for example, *no mms, sn duda, dnd*. IT IS VERY IMPORTANT THAT when these elements appear in a tweet: 1) if it is a tweet that also has a syntactic negation, the tweet is marked as TRUE, 2) if it appears in a tweet without syntactic negation, the tweet is marked as FALSE.

Tweets that must be discarded

It is also possible that some tweets are not in Spanish, that are only tweet tags (@PeñanietoLord) or that only contain a link. These tweets must be discarded to eliminate them from the corpus.

SECOND PART

Syntactic negation tags

In this second part of the tagging, only the tweets that were previously identified with available syntactic negation will be analyzed. Three tags will be used to tag the syntactic negations. We do not mark the spaces before or after the negation term, only the negation

itself. The tags are inspired by the article: *La negación en español: Análisis de tipología y patrones de negación* de Martí et Al. (2016). Next, we describe the tags used in this word. The same colors are used throughout the manual to highlight the tags.

- **NegCue**: It is the negative term at a syntactic level, that is, the term that affects syntagmata and or sentences. All negative terms in the tweet must be highlighted with this tag. It is necessary to pay close attention at the moment of tagging, as there are expressions that, despite containing negative terms, they do not necessarily express a negation, or they express contrast or comparison between two or more options. In these cases, the `no_neg` tag is used (which is described later). There can be more than one negation per tweet or the negations can be related. The negative term is part of the scope.
- **Event**: It corresponds to the word or words that are explicitly negated by the negative term. They must be different to the negative term. Basically, if there is no event, there is no negation (with the exception of tweets that are responses). In most cases, it is a verb or a periphrasis (in periphrases, the event is a main verb and the auxiliar verb) since the negative term is usually an adverb. Verboids, in the same manner as gerunds, infinitives and participles are also the event of the negative term such as *sin*. In the event, clitics, adverbs or words that are closely related to the verb are included.
- **Scope**: It corresponds to the entirety of words affected by the negative term, that is to say, how much it negates the term within reach. The negative term, the related negations and the rest of affected elements are included. In the subordinated sentences with negation, whatever precedes is not taken into account, no matter if it is relative, completive or conditional.
 - No quiero que vengas.
 - Quiero que no vengas.

Basic examples of tagging including **NegCue**, **Event**, **Scope**:

- **Nada** **sucedió** ayer.
- Las trocas viejitas **nunca** **pasaran** de moda 😍❤️
- **No** **sabía** yo que Shawn se vendió a Orion por unas skins del Fortnite 😐.

Negation expression tags

Besides the tags previously described, we used the following ones to identify the type of negation expression of the negation in tweets:

- **Neg_exp**: Simple negation refers to a term of negation that is not linked to another negation term. Therefore, the scope and the event are only related to this negation term.
 - **No tengo** ganas de ir al baño ahora.
- **Neg_rel**: Related negation is a tag for negative terms that are linked to other negative term in the same sentence. The related negation is dependent on the main negative term. By being dependent on the main negative term, this tag does not have an event or scope, and is part of the scope of the first negation, so they are related. Regularly, it is used for the terms that are quantifiers.
 - **No me ha gustado nada** la película.
 - **No hay nada** de comer.
- **No_neg**: The no negation tag is used for those tweets with phrases that include negative terms, but these are not negating anything, such as the case of counterpositions, discourse markers, set phrases in Spanish, or interjections such as NO MAMES. This tag is also used in certain types of abbreviations like in the case of *número* (no.) The list of elements that are considered as no_neg can be found further in the document. Examples of the use of the tag are the following:
 - **De nada**, que tengas bonita noche.
 - **No mames**, te pasaste de Bershka.

In this example, *no mames* is used as an interjection, which is why its tagging is avoided.

- joder David Broncano esta **más** fuerte **que nunca**.

In this case, a comparative negation, rather than negating, it contrasts a before and after.

Use of multiple tags in a single tweet

There can be more than one syntactic negation in a tweet since they are not limited by the number of sentences or by the number of subordinated sentences but by the number of characters. Therefore, more than one negative term can coexist, which raises the need to highlight more than one event and one scope per tweet. In the face of this, the tagging tool is set up so the tagger has the necessary tags for the number of syntactic negations present per tweet. The following table briefly illustrates how we number the tags. The same colors are used throughout the manual to highlight the tags.

Table 1.1: Tag numbers

Neg_Exp1	Event1	Scope1	Neg_rel1	No_neg1
Neg_Exp2	Event2	Scope2	Neg_rel2	No_neg2
Neg_Exp3	Event3	<i>Scope3</i>	Neg_rel2	No_neg3

Specific rules of syntactic negation tagging in tweets

The following is a set of specific rules to tag the corpus.

- Clitics will be taken into account for the **event** since they are taken as part of the verb and include relevant information. Also, certain prepositions and pronouns such as *se* will be considered.
 - **No le dije** que lo quería
- The prepositions that introduce the verb are also taken into account for the **event**.
- Verbal periphrases (gerund, infinitive and participle) are highlighted together as **event**:
 - **No vamos a dormir**, será una noche larga.
- Care must be taken with the verbs that never form a periphrasis because they are not highlighted as in the previous instance. An example of this are the verbs *querer* and *parecer*:
 - Luis **no quería** comerse su guayaba.
- In the case of verboids (gerund, infinitive and participle), they are taken as the **event**:
 - **No actuando**, bailando.
 - **Nunca comer** así de cerdo.
 - **Nadie cansado**, todos incansados.
- In the case of eluded verbs in a negative sentence, the **event** and the scope that precede the negation were chosen to be highlighted.
 - Pensando en si **me quiere** o **no**
 - **No sabía** si **quería estar conmigo** o **no**.

In the previous example, *sabía* is the **event** and *No sabía si quería estar conmigo* is the scope of the first negation, while *quería* is the event and *quería estar conmigo* is the scope of the second negation.

- Avoid as much as possible the tagging of laughter such as *jajaja*, *hahaha*, *ggg* and/or emojis (❤️) or emoticons (*XD*).

- When a negative term appears independent and without any recoverable preceding or following verb, as if it were the case of an elision, it was chosen to tag it in the same manner as *sin*, that is, negating the closest lemma:
 - Contigo y **no** **con él**.
 - Noviembre **sin ti**, qué triste!
- The subject is always excluded from the scope
 - Ayer **no comió** Belinda.
- It is possible to divide the scope to tag it; it does not have to be a continuous phrase. A fragment can even be found before the negation and another one after it. For instance, in the sentences where a predication could exist after the negation, it is chosen to divide the scope into two.
 - **Quería** estar contigo y **no** con él.
 - **Quiera** o **no** (quiera)
- In the cases of elision of the **event** where it cannot be recovered within the negated sentence, it is chosen to highlight the event in the previous sentence in case it is present.
 - La carrera **se llama** lic en admin. **No** de empresas.
- In the cases where a coordinated negation is present (no + Ø/ni _____ + ni _____), each negative term is tagged with its respective scope and event.
 - **No** **quiero** **ni** pan **ni** vino.
 - **No** **voy** a bailar **ni** a cantar.

In both examples, *no* and *ni* share the same event but not the same scope given that they are two different syntactic negations in a single tweet. A possible paraphrasis for both sentences is the following: No quiero pan y no quiero vino y No voy a bailar y no voy a cantar.

- In cases of complex negations such as *ni nada* or *ni nadie*, *ni* is tagged as a negation, highlighting *nada* as its event.
 - **No** **quiero** pan **ni** **nada**.

In this example, *ni* is neg_exp2 and *nada* is event2.

- When the adverb *tampoco* appears followed by a verb, the adverb can be considered as an independent negation.
 - **No me gusta** el arroz, **tampoco como** manzana
- If a verb does not appear, then *tampoco* is tagged as related negation since they share the same event. The difference between *tampoco* and *ni* lies in the fact that the former is an adverb while the latter is a conjunction.
 - **No me gusta** el arroz, **tampoco** la manzana
- Bilingual tweets are always tagged provided that the negation is in Spanish.
 - When te dicen que necesitan algo de ti y **no te queda** de otra que hacer caso.
 - Fuck me up when september end... ¿así **no era**?
 - When Manuel Turizo said "es mi corazón que **no te dice** adiós, pero en mi mente te lo digo" really I felt that... Fuck! Te odio!
- *Para nada, de ningún modo, de ninguna manera, nadie nunca, nunca jamás, ni madres, para nada* are considered units for facilitating tagging:
 - #ConLosGatosNo **DE NINGÚN MODO ACEPTO** ESTO , LOS GATOS SON IMPRESCINDIBLES!!!
 - **Nadie nunca** **dijo** eso
 - Me considero una persona **para nada** **atractiva** o **bonita**
 - -De ninguna manera: —Esto empezaba a alterar sus nervios, pero **de ninguna manera iba a dejar** que se le notara— ¿Quieres algo de comer o de beber? Quizá te vendría bien descansar un poco, igual después puedes pensar con más claridad. —Tenía que conseguir que se quedase como fuera— Quiero ayudarte.
 - No es mi culpa que en sus internados hayan tenido pedos con todo mundo y que haya sido tóxico blah blah. El mío **fue** lo mejor de mi vida y **para nada tóxico**, bye con todos, putos amargados.

Exceptions and particularities

There are many unities that we don't mark as negation, therefore, some tweets will not be tagged, and they will be omitted. Those elements are the following:

- Tweets completely in a different language.
 - Damn it bro, no more for you
- Hashtags with negative terms which comprise a unit or a tendency. The remaining of the tweet can be tagged normally:
 - México-Holanda #NoEraPenal aun **no lo superó.**
- If there is a loose negative term in a single hashtag, it is tagged, ignoring the pound sign (#).
 - La mera neta I would never see myself as a barca fan . **#no lo seré.**
- Sometimes the word *no* or *no.* is used to abbreviate the word *número*. In such cases, the tweet is skipped.
 - Lo que si es que Senadores y Diputados deben intentar ampliar el **no** de Boletas foráneas
- There are also cases where the sequence *n o* is used as an abbreviation, and not as a negation.
 - JUNGKOOK EN LA GRABACION DE N.O MIRENLOO MI VIDA. SI YO LLORO POR ESTO USTEDES TAMBIEN #Jungkook #JK #BTS
- At the moment *sino* and *si no* are not tagged. This is proposed in order to avoid complications since users constantly mistake the negation after the conditional and the conjunction:
 - Valoren a una buena persona cuando la encuentran que no hay nada perfecto **si_no** almas rotas con ganas de amar.
 - Por **si no** teníais suficiente con una, aquí tenéis a cuatro. Perdón por las horas.

In the face of this, these forms were eliminated from the extraction list so they do not appear on the tweets. In case they appear, they are not tagged.

- The list of negative terms or negative units that are not taken into account are the following:

sino, si no, sno, ntc y npi.

Possible no_neg with their variations on Twitter:

- sin embargo, sinembargo, snembargo, sn embargo, sin embrgo, nomas, nd más, nadamás, ndmás, nada mas, nadamas, namas, namás, nada más, no obstante, noobstante, nobstante, no obstnte, no mames, no ma, nememes, nms, no mms, nomms, namms, na mms, na mams, ne mms, ne mams, sino, si no, sno, nada menos, nadamenos, más _ que nunca, sin duda, etc.