

## Normas de etiquetado de la negación en Corpus de Twitter

El proyecto tiene por objetivo detectar la negación sintáctica en Twitter.

### **PRIMERA PARTE**

#### **Etiquetado de Tweets con negación**

En esta primera parte del etiquetado solamente tienen que marcar cuando un tuit tiene una negación sintáctica o cuando no. Para esto utilizarán una herramienta de etiquetado que en el costado izquierdo tiene las casillas a seleccionar. Seleccionarán las casillas TRUE cuando una negación sintáctica esté presente y la casilla FALSE cuando no. Para este trabajo consideramos la negación sintáctica como aquella que propiamente niega algo en una oración mediante un término negativo textual. Suelen ser adverbios como *no*, *jamás*, *nunca*, pero también tomamos en cuenta preposiciones como *sin*, los determinantes de *ningun*, conjunciones como *ni*, etc. Las posibles variantes pueden ser escrituras como *nel nunka*, *ño*, etc.

#### **Expresiones negativas que no niegan**

Hay conectores discursivos, muletillas o interjecciones que suelen aparecer en los tuits pero que no tienen valor de negación dentro de la oración. Ejemplos de esto son *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 y cualquiera de sus posibles variaciones (que pueden ser erratas o voluntarias); por ejemplo *nomms*, *sn duda*, *dnd*. ES MUY IMPORTANTE QUE cuando estos elementos aparezcan dentro del tuit: 1) si es en un tweet que además tiene una negación sintáctica el tweet quede marcado como TRUE, 2) si aparece dentro de un tweet sin negación sintáctica, el tweet queda marcado como FALSE.

#### **Tuits que hay que descartar**

También es posible que les aparezcan tuits que no están en español, que son puras etiquetas de tuits (@PeñanietoLord) o que son solamente un link. Estos los deberán descartar para eliminarlos del corpus.

### **SEGUNDA PARTE**

#### **Etiquetas de negación sintáctica**

En esta segunda parte del etiquetado solamente se analizarán los tweets que previamente fueron identificados con negación sintáctica disponible. Se utilizarán tres etiquetas para

etiquetar la negación sintáctica en el presente trabajo. Únicamente se deben seleccionar las palabras y no los espacios a la derecha o izquierda de éstas. Las etiquetas están basadas en el artículo *La negación en español: Análisis de tipología y patrones de negación* de Martí et Al. (2016). A continuación describimos las etiquetas utilizadas en este trabajo. Se usan los mismos colores a lo largo del manual para marcar las etiquetas.

**-NegCue:** Es el término negativo a nivel sintáctico, es decir, el término que afecta a sintagmas y/o oraciones. Deberán marcarse todos los términos negativos que se encuentren en el tweet con esta etiqueta. Es necesario prestar mucha atención al momento de etiquetar, pues existen expresiones que aunque contengan términos negativos no expresan necesariamente una negación o bien expresan contraste o contraposición entre dos o más opciones. En estos casos se utiliza la etiqueta de no\_neg (la cual se describe más adelante). Puede haber más de una negación por tweet o las negaciones pueden estar relacionadas. El término negativo es parte del scope.

**-Event:** Corresponde a la palabra o palabras que se ven explícitamente negadas por el término negativo. Deben ser diferentes al término negativo. Prácticamente, si no hay evento, no hay negación (con excepción de los tuits que son respuestas). En la mayoría de los casos es un verbo o una perífrasis (en las perífrasis el event es el verbo principal y el auxiliar) ya que el término negativo suele ser un adverbio. Los verboides como los gerundios, los infinitivos y los participios son también el event del término negativo, pero igualmente pueden ser un sustantivo u otra clase de palabras para términos como *sin*. Dentro del event se incluyen los clíticos, adverbios o palabras que están estrechamente relacionadas al verbo.

**-Scope:** Corresponde a la totalidad de palabras afectadas por el término negativo, es decir, qué tanto llega a negar el término, su alcance. Se incluye al término negativo, las negaciones relacionadas, el evento y los demás elementos afectados. En las oraciones subordinadas con negación no se toma en cuenta lo que antecede a estas. No importa si es relativa, completiva o condicional.

- No quiero que vengas.
- Quiero que no vengas.

Ejemplos sencillos de etiquetado incluyendo **NegCue**, **Event**, **Scope**:

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

### Etiquetas de expresiones de negación

Además de las etiquetas ya descritas, utilizamos las siguientes para identificar el tipo de expresión de negación de la negación en tuits:

- **Neg\_exp**: Negación simple, se refiere a un término de negación que no está ligado a otro término de negación. Por lo tanto, el scope y el evento solamente está relacionado a este término de negación.

- **No** **tengo** ganas de ir al baño ahora.

- **Neg\_rel**: Negación relacionada es una etiqueta para términos negativos que están vinculados con otro término negativo dentro de una misma oración. La negación relacionada es dependiente del término negativo principal. Al ser dependiente del término negativo principal, esta etiqueta no tiene ni evento ni scope y forma parte del scope de la primera negación, por lo tanto están relacionadas. Por lo regular, se usa para los términos negativos que son cuantificadores.

- **No** **me ha gustado** **nada** la película.

- **No** **hay** **nada** de comer.

- **No\_neg**: La etiqueta no negación se utiliza para aquellos tuits con frases que incluyen términos negativos, pero éstos propiamente no están negando nada, como puede ser el caso de las contraposiciones, marcadores discursivos, frases hechas del español o interjecciones como NO MAMES. Esta etiqueta también se usa en cierto tipo de abreviaciones, como la de número (no.). La lista de elementos que se consideran como no\_neg están más adelante en el documento. Aquí unos ejemplos del uso la etiqueta:

- **De nada**, que tengas bonita noche.

- **No mames**, te pasaste de Bershka.

En este ejemplo, *no mames* es usada como interjección, motivo por el cual se evita su etiquetado.

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

En este caso es una negación comparativa, más que negar, contrapone un antes y un después.

### Uso de múltiples etiquetas en un solo tweet

En un tuit puede haber más de una negación sintáctica ya que los tuits no están limitados por el número de oraciones o por el número de oraciones subordinados, sino por la cantidad de caracteres. Por ende, pueden coexistir más de un término negativo, lo que lleva a la necesidad de marcar más de un evento y un scope por tuit. Ante esto, la herramienta de etiquetado está configurada para que el etiquetador tenga las etiquetas necesaria para la cantidad de negaciones sintácticas que se presente por tuit. En la siguiente tabla se ilustra de forma breve como numeramos las etiquetas. Se usan los mismos colores a lo largo del manual para marcar las etiquetas.

Tabla 1.1: Número de etiquetas

Neg_Exp1	Event1	Scope1	Neg_rel1	No_neg1
Neg_Exp2	Event2	Scope2	Neg_rel2	No_neg2
Neg_Exp3	Event3	Scope3	Neg_rel2	No_neg3

### Reglas específicas del etiquetado de negaciones sintácticas en tweets

A continuación, una serie de reglas específicas para etiquetar el corpus.

- Los clíticos se tomarán en cuenta para el **event**, ya que se están tomando como parte del verbo e incluyen información relevante. También ciertas preposiciones y pronombres como el *se*.
  - **No le dije** que lo quería.
- Las preposiciones que introducen al verbo también se toman en cuenta para el **event**.
- Las perífrasis verbales (gerundio, infinitivo y participio) se marcan juntas como **event**:
  - **No vamos a dormir**, será una noche larga.
- Hay que tener cuidado con los verbos que nunca hacen perífrasis porque no se marcan como el ejemplo anterior. Un ejemplo de esto son los verbos querer y parecer :
  - Luis **no quería** comerse su guayaba.
- En el caso de los verboides (gerundio, infinitivo y participio), se toman como el **event**:
  - **No actuando**, bailando.
  - **Nunca comer** así de cerdo.
  - **Nadie cansado**, todos incansados.
- En el caso de los verbos elididos dentro de una oración negativa se optó por marcar el **event** y scope que antecede a la negación:
  - Pensando en si **me quiere** o **no**
  - **No sabía** si **quería estar conmigo** o **no**.

En el ejemplo anterior, *sabía* es el **event** y *No sabía si quería estar conmigo* es el scope de la primera negación, mientras que *quería* es el evento y *quería estar conmigo* es el scope de la segunda negación.

- Evitar, en la medida de lo posible, el etiquetado de risas como *jajaja*, *hahaha*, *ggg* y/o emojis(  )o emoticones (*XD*).
- Cuando un término negativo aparece independiente y sin ningún verbo anterior o posterior recuperable, como si fuera el caso de una elisión, se optó por etiquetarlo como el *sin*, es decir, negando el lema más cercano:
  - Contigo y **no con él**.
  - Noviembre **sin ti**, qué triste!

- El sujeto siempre queda fuera del scope.
  - Ayer no comió Belinda.
- Es posible dividir el scope para etiquetarlo, no tiene que ser una frase continua. Incluso, puede encontrarse un fragmento antes de la negación y otra después de la negación. Por ejemplo, en las oraciones donde hubiera predicación después de la negación se opta por dividir en dos el scope.
  - Quería estar contigo y no con él.
  - Quiera o no (quiera)
- En los casos de elisión del event, donde este no se puede recuperar dentro de la oración negada, se opta por marcar el event en la oración anterior si es que este está presente.
  - La carrera se llama lic en admin. No de empresas.
- En los casos donde se encuentre una negación coordinada (no + Ø/ni \_\_\_\_\_ + ni \_\_\_\_\_), cada término negativo se etiqueta con su respectivo scope y evento.
  - No quiero ni pan ni vino.
  - No voy a bailar ni a cantar.

En ambos ejemplos el *no* y el *ni* comparten un mismo event, pero no un mismo scope ya que son dos negaciones sintácticas distintas en un solo tuit. Una posible reescritura para ambas oraciones es la siguiente: No quiero pan y no quiero vino y No voy a bailar y no voy a cantar.

- En casos de negaciones complejas, como *ni nada* o *ni nadie*, *ni* se etiqueta como negación, marcando a *nada* como su evento.

- No quiero pan ni nada.

En este ejemplo, *ni* es neg\_exp2 y *nada* es event2.

- Cuando aparezca el adverbio *tampoco* seguido de un verbo, dicho adverbio puede considerarse como una negación independiente.
  - No me gusta el arroz, tampoco como manzana



- Últimamente me han estado dando ganas de fumar, y **no no no no no soy yo**, te lo juro
- **Nunca nunca nunca** **había tenido** tantas ganas d pirarme a tomar x culo

### Excepciones y particularidades:

Hay varias unidades que no tomaremos en cuenta como negación, por tanto, algunos tuits quedarán sin etiquetar y se omitirán. Dichos elementos son los siguientes:

- Tuits en su totalidad en otro idioma.
  - Damn it bro, no more for you
- Los hashtag con términos negativos que conforman una unidad o una tendencia. El resto del tweet puede ser etiquetado con normalidad:
  - México-Holanda #NoEraPenal aun **no lo superó.**
- Si viene un término negativo suelto en un solo hashtag, sí se etiqueta, ignorando el signo de almohadilla (#).
  - La mera neta I would never see myself as a barca fan . **#no lo seré.**
- En ocasiones se usa la abreviatura *no* o *no.* para la palabra número, en estos casos se salta el tuit.
  - Lo que si es que Senadores y Diputados deben intentar ampliar el **no** de Boletas foráneas
- También hay casos donde la secuencia *n o* se utiliza como abreviatura y no como negación.
  - JUNGKOOK EN LA GRABACION DE **N.O** MIRENLOO MI VIDA. SI YO LLORO POR ESTO USTEDES TAMBIEN #Jungkook #JK #BTS
- De momento, no se etiquetan los *sino* y *si\_no*. Esto se plantea para evitar complicaciones, ya que los usuarios confunden constantemente la negación después del condicional y la conjunción:
  - 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.

Ante esto, se eliminaron dichas formas de la lista de extracción previendo que no salgan en los tuits. En caso de que aparezcan, no se etiquetan.

- La lista de términos negativos o unidades negativas que no se toman en cuenta son las siguientes:
  - *sino, si no, sno, ntc y npi.*

**Posibles no\_neg con sus variaciones en 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.

