

La factualidad y las oraciones condicionales en español: Un estudio de corpus

Leyre BarriosUniversitat de Lleida (España) ✉ **Glòria Vázquez**Universitat de Lleida (España) ✉ <https://dx.doi.org/10.5209/clac.79952>

Recibido: 19 de enero de 2022 • Aceptado: 12 de octubre de 2022

ES Resumen: Al describir una situación (un evento o un estado) expresamos el grado de certeza (factualidad) que poseemos sobre dicha situación, presentándola como cierta, incierta o no cierta. En este artículo analizamos el valor factual de las dos cláusulas de cerca de 500 oraciones condicionales en español con *si* procedentes de distintos periódicos españoles. Nuestra hipótesis es que es posible determinar la factualidad en estas oraciones teniendo en cuenta la presencia de *si* y los tiempos verbales de las prótasis y las apódosis. El objetivo es determinar los contextos propicios para la (no) factualidad y crear reglas formales para anotar de forma automática los valores factuales de las cláusulas de estas oraciones. Con ello, se pretende contribuir a la representación y al análisis de la factualidad de los eventos mencionados en un texto dentro de los ámbitos de la lingüística de corpus y el procesamiento del lenguaje natural. Además, se han identificado 20 combinaciones temporales que no habían sido descritas anteriormente.

Palabras clave: factualidad, certeza, oraciones condicionales, tiempos verbales.

ENG Factuality and conditional sentences in Spanish: A corpus study

Abstract: When describing a situation (an event or a state) we express the degree of certainty (factuality) we have about that situation, presenting it as certain, uncertain or not certain. In this paper we analyse the factual value of the two clauses of conditional sentences in Spanish with *si* from different Spanish newspapers. Our hypothesis is that it is possible to determine factuality in these sentences taking into account the presence of *si* and the verb tenses of protasis and apodosis. The aim is to identify the contexts conducive to (non) factuality and to create formal rules to automatically annotate the factual values of the clauses of these sentences, which can be implemented in an automatic Spanish factual annotation tool. By doing so, this study contributes to the representation and analysis of the factuality of events mentioned in a text in the field of corpus linguistics and natural language processing. In addition, 20 new temporal combinations have been identified.

Keywords: factuality, certainty, conditional clauses, verb tenses.

Sumario: 1. Introducción. 2. Las oraciones condicionales. 2.1. Condicionales de contenido. 2.2. Condicionales epistémicas. 3. Metodología. 4. Resultados. 4.1. Tiempos verbales en las prótasis condicionales. 4.2. Tiempos verbales en las apódosis condicionales. 4.3. Correlación de tiempos verbales. 5. Propuesta de formalización. 5.1. Tiempos verbales en las prótasis condicionales. 5.2. Tiempos verbales en las apódosis condicionales. 6. Conclusiones. Agradecimientos. Contribución de autoría CREdIT. Referencias bibliográficas. Anexo.

Cómo citar: Barrios, L., Vázquez, G. (2024). La factualidad y las oraciones condicionales en español: Un estudio de corpus. *Círculo de Lingüística Aplicada a la Comunicación* 98 (2024). 245-262. <https://dx.doi.org/10.5209/clac.79952>

1. Introducción

Cuando hablamos y describimos una situación (un evento o un estado), inevitablemente expresamos nuestro grado de compromiso sobre la certeza de dicha situación. Así, la podemos presentar como cierta, incierta o

como no cierta (falsa) y, para ello, empleamos diferentes marcadores lingüísticos. A dicho grado de certeza también se le ha denominado factualidad.

La representación de la factualidad de los eventos es un tema de gran interés actualmente en el ámbito de la lingüística de corpus y del procesamiento del lenguaje natural debido a la proliferación de las *fake news* y de las aplicaciones de *fact-checking*. Sin embargo, el número de corpus anotados con esta información es escaso y la mayoría de ellos se centran en el inglés (Diab *et al.*, 2009, Lee *et al.*, 2014; Saurí y Pustejovsky, 2009; Tonelli *et al.*, 2014; Van Son *et al.*, 2014). Por otro lado, hay muy pocos sistemas de anotación automática de la factualidad para las distintas lenguas. Para el español, podemos destacar el trabajo de Wonsever *et al.* (2016).

Este trabajo se enmarca dentro del proyecto TAGFACT (Alonso *et al.*, 2018; Fernández M., *et al.*, 2020; Vázquez y Fernández, 2020), en el cual se ha elaborado una herramienta automática para la anotación de la factualidad en español partiendo del estudio de textos periodísticos y basada en reglas de conocimiento lingüístico. En este artículo analizamos la factualidad de un conjunto relevante de oraciones condicionales procedentes de dichos textos y que incluyen la conjunción *si*. Para ello hemos estudiado los distintos grados de factualidad que se pueden asociar a las dos cláusulas de estas oraciones y hemos identificado los elementos textuales que aportan dicha información con el objetivo de determinar los contextos propicios para la (no) factualidad y proponer reglas formales que permitan mejorar la precisión de la etiquetación de dichas construcciones en el sistema mencionado.

Partimos de la base de que los tiempos verbales son un elemento lingüístico clave a la hora de determinar la factualidad de una oración y, en el caso de oraciones interordinadas que incluyen las condicionales, también lo es el conector. Ahora bien, nuestra hipótesis de partida es que los valores factuales asociados a los verbos de las oraciones condicionales según los tiempos verbales usados en ellas, especialmente en las prótasis, difieren en un grado no trivial respecto a los valores más frecuentemente asociados a dichos tiempos en otro tipo de oraciones. Esta es la motivación del estudio de estas oraciones en particular, que parece que presentan mayor relevancia a este respecto que otro tipo de oraciones interordinadas, como las concesivas (Barrios y Vázquez, 2020).

En este trabajo, respecto al estudio de las prótasis, hemos partido, sobre todo, de la descripción llevada a cabo por Montolío (1999). Esta autora realiza una propuesta de los valores factuales que adopta cada tiempo verbal en estas cláusulas. Así, hemos comprobado con datos extraídos de corpus los valores identificados por Montolío y hemos matizado algunas de sus afirmaciones y ampliado la casuística. Además, aportamos una propuesta de valores factuales para las apódosis, hasta ahora no planteada.

Este trabajo se estructura en seis apartados. En el apartado 2 se describirán las características de las oraciones condicionales. En el apartado 3 se presentará la metodología que se ha seguido para la extracción de los datos y el análisis. En los apartados 4 y 5 se incluirán los resultados obtenidos y la propuesta de reglas, respectivamente, y, por último, se presentarán las conclusiones.

2. Las oraciones condicionales

Las oraciones condicionales se definen prototípicamente como aquellas en las que se expresa una situación hipotética en la prótasis y sus posibles consecuencias en la apódosis (1). En español, dicha prótasis es introducida por la conjunción *si*, pero también por *como* (1b), *siempre que* (1c), etc.

(1)

- a. *Si gana hoy al Getafe se pondrá con 40 puntos por los 35 de los colchoneros (...)* (mundodeportivo.com, 06/01/2019).
- b. *Como me toque la lotería, te compraré un descapotable* (Montolío, 1999: 3700).
- c. Los perros de Viena *pueden* hacer lo que quieran *siempre que sigan* siendo perros aunque muchos perros ya no parecen serlo (Carrión, I., Danubio; ejemplo citado en R.A.E., 2009: 913).

La situación hipotética reflejada en estas prótasis se interpreta como condición, pues su realización es necesaria y suficiente para que la situación reflejada en la apódosis tenga lugar. Esta suficiencia es la que hace que esta condición se conciba como causa (Dancygier, 1998; Montolío, 1999; RAE, 2009).

En base a esta relación de causalidad entre los contenidos de las cláusulas, autores como Sweetser (1990) han propuesto una clasificación de diferentes tipos de condicionales: de contenido (1a), epistémicas (2a) e ilocutivas (2b). Las primeras son aquellas en las que la relación se produce entre situaciones (estado de cosas) y entre los miembros hay una relación de causa-efecto (1a). En las epistémicas (2a) la relación se produce como resultado de un razonamiento lógico, ya que los miembros presentan una relación premisa-conclusión. Por último, en las ilocutivas (2b) la relación se establece entre el acto de habla de la apódosis y la conclusión a la que se llega a partir del contenido expresado en la prótasis.

(2)

- a. *Si ha ganado dos elecciones en Catalunya ha sido* porque es transversal (...)
- b. *Si tienes hambre, hay* pollo en la nevera (Ducrot, 1972: 155; citado por Montolío, 1999: 3687)

Tal y como veremos, la hipoteticidad prototípica de las oraciones condicionales no siempre está presente y, por tanto, en ocasiones se predica sobre situaciones reales. En estos casos no se establecen condiciones, aunque las oraciones se siguen llamando condicionales.

A continuación, se ahondará solo sobre las características de los dos primeros tipos de condicionales, ya que en nuestro corpus no han sido halladas las condicionales ilocutivas.

2.1. Condicionales de contenido

En este tipo de condicionales la conjunción *si* adopta normalmente su función prototípica actuando como constructora de mundos posibles (Dancygier, 1998; Sweetser, 1990). Al emplearla, el hablante crea un nuevo espacio mental de carácter hipotético (o evoca uno ya introducido en el discurso previamente) en el que inserta tanto la prótasis como la apódosis, relacionadas por la relación causa-efecto. Así pues, en este tipo de oraciones típicamente no se afirma el contenido ni de la prótasis ni de la apódosis.

El orden sintáctico prototípico de estas condicionales es prótasis-apódosis (2a), si bien es posible que la apódosis aparezca antepuesta a la prótasis (3a) o incluso que la apódosis se encuentra escindida por la prótasis (3b). Cuando la prótasis se encuentra antepuesta, su carácter es temático, mientras que, si aparece pospuesta, su carácter es remático (Montolío, 1999; RAE, 2009). Si la apódosis se encuentre escindida por la prótasis, el hablante enfatiza el contenido de esta.

(3)

- a. El *ayuntamiento dispone* de parcelas para este tipo de viviendas *si hubiera* demanda (levantem.com, 05/03/2018).
- b. El rectorado de la Complutense aseguró que dichas denuncias, *si es que existieron* como tales, no *trascendieron* más allá de la propia facultad de Educación (El País, 10-I-1997; ejemplo citado en Montolío, 1999: 3655).

En relación a su factualidad, la tradición gramatical española (Alarcos, 1994; Gili Gaya, 1961; Porcar Miralles, 1993; Veiga y Mosteiro, 2006) distingue entre condicionales reales (o factuales) e irreales (o contrafactuales) teniendo en cuenta el tiempo verbal empleado en la prótasis y su carácter factual. Rojo y Montero (1983:15-16), por su parte, añaden las potenciales (o hipotéticas).

Las primeras (1a ahora recogido como 4) son aquellas en las que el hablante hipotetiza en la apódosis las posibles consecuencias del evento de la prótasis y, además, no se compromete con su realización. No obstante, son denominadas reales o factuales, ya que en ellas se hace referencia a hechos «de los que se espera un cumplimiento probable en el futuro» (Montolío, 1999: 3662), dado que la situación denotada en la prótasis, en base a experiencias del pasado (o sea, considerando hechos reales), habitualmente comporta la realización de la situación denotada en la apódosis.

- (4) *Si gana* hoy a el Getafe se *pondrá* con 40 puntos por los 35 de los colchoneros (...) (mundodeportivo.com, 06/01/2019).

Montolío (1999) recoge una serie de combinaciones temporales (o patrones) que son comunes en este tipo de oraciones. En unos se usa el presente de indicativo en la prótasis y en otros el imperfecto. El patrón prototípico de este tipo de condicionales es el presentado en la oración anterior (4): <Si + presente de indicativo + futuro>.

Otro de los patrones que señala y que pertenece al primer grupo es <Si + presente de indicativo + presente de indicativo>. Mediante este patrón el hablante hace referencia a menudo a situaciones que la sociedad tiene como verdaderas (es decir, situaciones factuales), o que suceden habitualmente, en las que se sabe que, de ocurrir la situación de la prótasis, se cumple también la de la apódosis (5a). Es decir, estamos ante condicionales genéricas o habituales. Además, en este tipo de condicionales, la conjunción *si* puede ser sustituida por *cada vez que* o *siempre que* sin alterar el significado de la oración.

Si bien no se tratan de condicionales que hacen referencia a situaciones habituales o a verdades absolutas como en el caso de (5a), también es posible encontrar este patrón en aquellas condicionales cuya apódosis está enfocada al futuro (5b) y que, por lo tanto, hacen referencia al porvenir. Ambos tipos (5a y 5b) son, en todo caso, oraciones condicionales de contenido. Sin embargo, dicho patrón también se usa en el caso de las condicionales epistémicas explicativas (2b).

(5)

- a. Si *llueve*, se mojan las calles (RAE, 2009: 906).
- b. *Si me piden* que dimita, lo *hago* y listo; (Esgueva y Cantarero 1981: 19; ejemplo citado en Montolío, 1999: 3664).

Esta autora también identifica el presente en prótasis junto con apódosis en pasado (<Si + presente de indicativo + tiempos de pasado en indicativo>) (6). No obstante, desde nuestro punto de vista, estas condicionales son ilocutivas.

- (6) *Si es así*, se *equivocaron* (Montolío, 1999: 3666).

Montolío (1999) recoge asimismo un tercer patrón, (<Si + presente de indicativo + condicional> (7)), para las condicionales reales. Este patrón hace referencia a hechos simultáneos en el presente o enfocados al

futuro, si bien la diferencia con respecto al patrón <Si + presente de indicativo + futuro> radica en que, al usar el condicional, el hablante sitúa la situación de la apódosis como menos probable.

- (7) «Si es cierto, *estaríamos* ante una de las mayores traiciones a la democracia de la historia» (lavozden galicia.com, 20/03/2017).

Respecto a los patrones con imperfecto de indicativo en la prótasis, Montolío (1999) señala <Si + imperfecto de indicativo + imperfecto de indicativo> para las situaciones habituales en el pasado (8a). Además, también se emplea en el discurso indirecto con verbo de comunicación en pasado para introducir una situación posterior a ese momento temporal del verbo de comunicación (8b).

- (8)
 a. *Seguíamos* hablando en el cuarto de estar o bien en su habitación, si, por ejemplo, papá *estaba* acostado, para no molestarle (Esgueva y Cantarero 1981: 367; ejemplo citado por Montolío 1999: 3666).
 b. Te prometí que si la semana siguiente *acababa* el dichoso capítulo, te *invitaba* a cenar (Montolío, 1999: 3667).

El imperfecto de indicativo de la prótasis puede combinarse, asimismo, con condicional en la apódosis (<Si + imperfecto de indicativo + condicional>) o con pretérito imperfecto de subjuntivo (<Si + imperfecto de indicativo + imperfecto de subjuntivo>). En el primero de ellos, el hablante hace referencia a eventos del pasado, del presente o del futuro, posteriores a un acto de habla situado en el pasado (9a), mientras que, en el segundo, relata un enunciado que en el momento de la enunciación constituyó un imperativo (9b).

- (9)
 a. El anuncio de Zapatero de que si *llegaba* a la Moncloa le *nombraría* ministro de Trabajo, sorprendió al ex secretario general de CCOO (elmundo.es, 15/03/2004).
 b. Me dijeron que (...) si *quería* que *hiciera* la reserva (Esgueva y Cantarero 1981: 236; ejemplo citado en Montolío, 1999: 3667).

Las condicionales potenciales (10), por su parte, son aquellas en las que el hablante expresa que la situación denotada en la prótasis, siempre enfocada al futuro, es probable, pero en menor medida que en el caso de las condicionales reales.

- (10) «Si *hubiera* tanto paro nos *mataríamos* los unos a los otros» (abc.es, 27/01/2014).

El pretérito imperfecto de subjuntivo puede expresar esta probabilidad debido al proceso de dislocación temporal (Rojo, 1974). Mediante este proceso, este tiempo verbal adquiere valores modales de probabilidad negativa respecto a un evento futuro, es decir, que el hablante lo utiliza cuando no está seguro sobre la realización de la situación reflejada en la prótasis.

Para expresar esta probabilidad, el patrón prototípico es <Si + imperfecto de subjuntivo + condicional> (10). Con este patrón también es posible que el hablante mitigue una petición (11).

- (11) Si usted *podiera* entregarme un ejemplar, yo le *quedaría* muy reconocido (Montolío, 1999: 3668).

Con este tiempo verbal en prótasis, Montolío (1999) recoge, asimismo, el patrón <Si + imperfecto de subjuntivo + imperfecto de indicativo> (12a). Como variante dialectal a este, esta autora menciona el patrón <Si + condicional + condicional> (12b).

- (12)
 a. «Si *gobernara* Vox, *vivíamos* sin autonomías (...)» (eldiario.es, 24/04/2019).
 b. Si *aprobaría* el carné de conducir, este verano *podríamos* viajar en coche (Montolío, 1999: 3670).

Por último, las condicionales irreales o contrafactuals se refieren a situaciones del pasado presentadas justamente al revés de cómo sucedieron (13), puesto que se enmarcan en un mundo imaginario. El hablante juega con este escenario negando un hecho que sí sucedió en el pasado. De ahí que se consideren contrafactuals, porque expresan un contenido proposicional imposible.

- (13) (...) si *hubiera* entrado *habría* sido un gol para la historia del Real Madrid (...) (marca.com, 29/05/2014).

El principal patrón señalado por Montolío (1999) para este tipo de condicionales es: <Si + pluscuamperfecto de subjuntivo + condicional compuesto / pluscuamperfecto de subjuntivo> (13, 14a). Con este tiempo verbal en prótasis también contamos con los siguientes patrones: <Si + pluscuamperfecto de subjuntivo + condicional> (14b); y <Si + pluscuamperfecto de subjuntivo + pluscuamperfecto de indicativo> (14c).

- (14)
 a. Si *hubiéramos* tenido esta reforma laboral hace dos o tres años *nos* *hubiéramos* *ahorrado* un millón de parados (elpais.com, 01/01/2013).

- b. Si el Ayuntamiento *hubiera pagado* la resolución de 2005 en los seis primeros meses, el tema *estaría* resuelto (elpais.com, 08/03/2012).
- c. El pájaro, que no estaba acostumbrado, si *hubiera podido*, *se había marchado* (Esgueva y Cantarero 1981: 333; ejemplo citado en Montolío, 1999: 3671).

Otros patrones para expresar la contrafactualidad señalados por Montolío (1999) son los siguientes: <Si + imperfecto de subjuntivo + condicional>, el cual resulta ambiguo, pues también puede ser potencial (10 recogida ahora con 15a); <Si + presente de indicativo + presente de indicativo> (15b), y <Si + pluscuamperfecto de indicativo + pluscuamperfecto de indicativo> (15c).

(15)

- a. «Si *hubiera* tanto paro nos *mataríamos* los unos a los otros» (abc.es, 27/01/2014).
- b. Se produjo una inquietante escena que, si *la ven* en el Parlamento europeo, nos *aspan* (Montolío, 1999: 3672).
- c. Si *había ganado* esa oposición, *habíamos hecho* una gran fiesta (Montolío, 1999: 3672).

2.2. Condicionales epistémicas

En las condicionales epistémicas la prótasis expresa una premisa entendida como condición suficiente para alcanzar la conclusión, de carácter lógico, expresada en la apódosis.

Montolío (1999) y Rodríguez Rosique (2008), entre otros, identifican tres subtipos de condicionales epistémicas: explicativas (2b), identificativas (16a) e irónicas (16b).

(16)

- a. Si *alguien puede* hacerlo, *esa es Ana* (Rodríguez Rosique, 2008: 145).
- b. Mira, si *eso es música*, *yo soy Beethoven* (Montolío, 1999: 3672).

En cuanto a su factualidad, ambas cláusulas son factuales, o contrafactuales en el caso de las irónicas. En las oraciones condicionales epistémicas, la conjunción *si* no se usa prototípicamente, ya que no introduce una oración que denote una situación relativa a un mundo hipotético, sino que en la prótasis se predica sobre una situación que se presenta como cierta o falsa.

En el caso de las identificativas, la estructura sintáctica es particular, ya que en la apódosis se usa un elemento anafórico que hace referencia a la prótasis y un verbo copulativo (*esa + es* en (16a)). De este modo, realmente el contenido de la prótasis es vacuo. Así pues, en estas oraciones, el contenido focalizado es el de la apódosis, que aporta la información nueva.

A diferencia de las condicionales de contenido, para las epistémicas no se han descrito los patrones temporales. No obstante, si partimos de los ejemplos aportados por los autores, podemos concluir que en ambas cláusulas se emplea indicativo. Además, la prótasis parece aparecer siempre antepuesta.

3. Metodología

Para la realización de este trabajo se han desarrollado dos fases: en la primera, se ha creado y anotado con valores factuales el corpus de estudio y, en la segunda, se han extraído las combinaciones de tiempos y valores factuales y se han formalizado las reglas para la anotación automática.

Hay que señalar que, en este estudio, solo se han analizado oraciones condicionales introducidas por *si*. Asimismo, a la hora de extraer y anotar los datos, no se ha discriminado por tipología (v. ap. 2) por varios motivos. El primero es que las clasificaciones propuestas para las oraciones condicionales no utilizan la factualidad como punto de partida, por lo que no son especialmente relevantes en nuestro análisis. En segundo lugar, de los tres tipos de condicionales que típicamente se han considerado en la bibliografía, el más frecuente en nuestro corpus es el de las condicionales de contenido (un 84,10%), seguidas de las epistémicas con una diferencia importante (un 15,90%), mientras que de las condicionales ilocutivas no hemos identificado ningún caso, como ya se ha avanzado. Por último, las combinaciones temporales identificadas en este estudio para las condicionales epistémicas son restringidas y todas ellas están presentes en las condicionales de contenido, que sí que presentan una variedad considerable.

Respecto a la primera fase, se han obtenido datos de dos corpus que recogen el empleo del español actual en el registro periodístico: el corpus Now, una sección del Corpus del Español que recoge solo textos de diarios publicados en la web (Davies, 2019), y el Corpus del Español del Siglo XXI (Corpes XXI) de la RAE (2020) (restringido a noticias). Concretamente, se ha creado un corpus compuesto por 440 oraciones condicionales cuya temática abarca principalmente la política, el deporte, la justicia y la economía.

Los marcadores lingüísticos que se han usado para determinar la factualidad de las condicionales han sido, además de la conjunción *si*, los tiempos verbales tanto de la prótasis como de la apódosis.

Las búsquedas se han diseñado a partir de los elementos de la prótasis, ya que es la cláusula que contiene la conjunción. Así pues, en primer lugar, se ha utilizado el Corpes XXI para localizar ejemplos que incluyan la conjunción mencionada combinada con cada uno de los tiempos de la conjugación verbal (con una ventana de máximo cinco palabras).

En segundo lugar, y tras una revisión manual de la muestra obtenida, se ha procedido a excluir algunas oraciones, como aquellas que presentan una complejidad sintáctica, las condicionales cuyas apódosis son subordinadas de otras oraciones o las escritas en un registro coloquial o procedentes de blogs. Las razones

de esta exclusión son que, en este primer estudio, se ha optado por ceñirnos a oraciones, por un lado, más simplificadas y, por otro, más propias del registro periodístico. También se descartaron manualmente las oraciones con la conjunción interrogativa *si*.

En tercer lugar, se ha intentado que de cada tiempo verbal en prótasis se obtuvieran un mínimo de 25 oraciones, cantidad que consideramos suficiente para poder extraer conclusiones, aunque en algunos casos ha sido mayor.

En este punto, se ha iniciado la anotación de las prótasis respecto a la factualidad. Cuando en estas 25 prótasis la variedad de valores factuales ha sido elevada, se ha repetido la búsqueda en el corpus mencionado para esos tiempos verbales con el objetivo de aumentar el número de ejemplos y poder determinar el valor más frecuente.

En cuarto lugar, se ha analizado el tiempo verbal de la apódosis de los ejemplos recopilados en la fase anterior para extraer las combinaciones temporales de ambas cláusulas. Si en algún caso no han quedado representados todos los tiempos de la conjugación en las apódosis, se ha iniciado la búsqueda de nuevos ejemplos en el corpus.

Después de seguir todos estos pasos, aquellos tiempos que no estaban representados en las apódosis se han buscado específicamente en el corpus mencionado con un patrón que incluía la conjunción y los dos tiempos verbales. Como último recurso, para aquellas combinaciones temporales sin representación o con muy baja representación se han buscado más ejemplos en el otro corpus, el Corpus del Español-Now, y se han etiquetado las nuevas frases.

Desde el principio esperábamos que el mayor número de desajustes de valores factuales vendrían por los verbos de las prótasis y por ello se priorizó que la variedad temporal en estas estuviera bien representada y de ahí que estableciéramos un mínimo de 25 casos. A partir de este corpus inicial, el siguiente paso fue asegurarnos que en las apódosis correspondientes a estas prótasis estuvieran representados todos los tiempos posibles, ampliando el corpus inicial si era necesario para dar cuenta de dicha representación. Ahora bien, no establecimos un mínimo de apódosis para cada tiempo porque eso implicaba agrandar considerablemente el corpus de prótasis y no era necesario, ya que era el foco principal de estudio y estaba muy bien representado. El número de oraciones final considerado en nuestro estudio para cada tiempo en las prótasis y en las apódosis se puede consultar en la Tabla 12 (v. Anexo).

Las etiquetas utilizadas son las del proyecto TAGFACT con las adaptaciones oportunas (v. ap. 5) y son las siguientes:

- «Compromiso positivo» [COM+]: el hablante se compromete con la realización de la situación reflejada en la cláusula y la describe como un hecho real.
- «Compromiso negativo» [COM-]: el hablante se compromete con la no realización de la situación reflejada, indicando que no tuvo o no está teniendo lugar (contrahecho).
- «No compromiso» [NO COM]: el hablante expresa incerteza respecto a la realización de la situación.
- «No aplica» [NA]: hace referencia a un mundo imaginado.

Al realizar la anotación solo se ha tenido en cuenta la información factual que aporta la morfología verbal (tiempo y modo), además de la conjunción, y no otros elementos como de probabilidad o de negación, ya que la información que aportan estos es generalizable a cualquier tipo de oración y no se espera un comportamiento especial en las condicionales, por lo que los valores de factualidad que aportan dichos elementos ya se han cubierto con reglas generales del proyecto citado. No obstante, se ha tenido en cuenta la polaridad en los casos de «compromiso» para los tiempos que incorporan el valor de contrahecho, ya que son etiquetados como [COM-]. Asimismo, cabe mencionar que para asignar los valores factuales a una de las cláusulas se ha tenido en cuenta la interpretación de la oración condicional en su totalidad.

Una vez finalizada la fase de recopilación y anotación de corpus, se ha iniciado la dedicada a la formalización. Para ello, se han comparado los valores asociados a los distintos tiempos verbales de estas frases con los valores generales asociados a estos tiempos con el objetivo de determinar cuándo es necesario formular reglas específicas para las oraciones condicionales. Estos valores generales han sido propuestos a partir de las lecturas bibliográficas (Ahern, 2008; Comajaon-Colomé y Pérez Saldanya, 2018; Escandell, 2010; RAE, 2009) y se presentan en la Tabla 1, aunque traducidos a las etiquetas usadas en este estudio. De estos tiempos, hay algunos que presentan valores fijos y otros que pueden presentar más de un valor. Estos últimos aparecen marcados en negrita y se muestra solo el más frecuente, según lo establecido en la bibliografía mencionada.

Tabla 1. Valores más frecuentes de los tiempos verbales según la bibliografía en el uso general.

TIEMPO VERBAL / FORMA NO FINITA	VALOR FACTUAL TIEMPO
/Presente de indicativo/ (VMIP)	[COM+]
/Pretérito imperfecto de indicativo/ (VMII)	
/Pretérito perfecto simple/ (VMIS)	
/Pretérito perfecto compuesto/ (VAIS VMP)	
/Pretérito pluscuamperfecto de indicativo/ (VAII VMP)	
/Futuro simple/ (VMIF)	
/Futuro compuesto/ (VMIF VMP)	

TIEMPO VERBAL / FORMA NO FINITA	VALOR FACTUAL TIEMPO
/Condicional simple/ (VMIC)	[NA]
/Condicional compuesto/ (VAIC VMP)	
/Presente de subjuntivo/ (VMSP)	
/Pretérito perfecto de subjuntivo/ (VASIS VMP)	[COM+]
/Pretérito imperfecto de subjuntivo/ (VMSI)	[NA]
/Pretérito pluscuamperfecto de subjuntivo/ (VASI VMP)	[COM-]
/Imperativo/ (VMM)	[NA]

Si al comparar los valores de nuestro corpus para cada tiempo con los de la tabla se ha observado coincidencia, dicho valor quedará asignado a partir de las reglas generales asociadas a los tiempos. En caso de que no haya coincidencia, será necesario formalizar reglas específicas que contengan el conector *si*. Se actuará así tanto para los casos de tiempos con valor único como para el valor más frecuente de los tiempos que presentan más de una posible etiqueta. Para los otros valores de estos tiempos que no son los más frecuentes tanto si son coincidentes como no con el valor general, se intentará identificar si el tiempo verbal de la otra cláusula permite crear otras reglas más específicas para asignar estos valores. Hemos usado un umbral de 10 ejemplos para proponer reglas, ya que se ha considerado que con una cifra menor de ejemplos no sería adecuado generalizar.

Además, cabe decir que en el sistema automático de anotación TAGFACT se ha incorporado un anotador morfológico y sintáctico para etiquetar los textos que actúa previamente a la etiquetación factual (Freeling 4.1; Padró y Stanilovsky, 2012). El desambiguador morfológico anota el tiempo verbal mediante las etiquetas EAGLES (1994), presentadas junto a cada tiempo verbal en la Tabla 1. En el caso de los verbos, las etiquetas pueden presentar una longitud de hasta 7 caracteres, si bien en este estudio nos hemos limitado a considerar únicamente las 4 primeras para simplificar:

- El primero de ellos se corresponde con la categoría de la palabra, es decir, al ser un verbo la primera letra es una V.
- El segundo nos indica el tipo de verbo que es: si es principal, la letra será M; si es auxiliar, la letra será A.
- El tercero señala el modo o la forma no finita: la letra I nos indica que el verbo está en indicativo, la letra S que es subjuntivo, la letra M que es imperativo, la letra C que es condicional, la letra N que es infinitivo, la letra G que es gerundio, y la letra P que es participio.
- El cuarto hace referencia al tiempo verbal, de forma que, si es presente, la letra será P; si es imperfecto la letra será I; si es futuro, la letra será F; y si es pasado (diferente al imperfecto), la letra será S.

4. Resultados

Lo primero que queremos destacar es que, en la mayor parte de las oraciones condicionales de nuestro corpus (85,91%), las prótasis están antepuestas a las apódosis, lo cual confirma lo que ya se había avanzado en la bibliografía (v. ap. 2). Por otro lado, hemos identificado 20 nuevas combinaciones temporales hasta ahora no documentadas.

Además, hemos determinado qué tiempos verbales de los patrones presentados en el apartado 3 para las prótasis y las apódosis de las oraciones condicionales son concluyentes y cuáles pueden adoptar más de un valor según la frase (v. ap. 4.1). También hemos decidido para qué tiempos es necesaria una regla específica para las oraciones condicionales, teniendo como criterio si el valor factual (ya sea único o el más frecuente) no es coincidente con el genérico asociado al tiempo verbal (v. ap. 4.2). Además, para los casos en que un mismo tiempo puede presentar más de un valor factual en cada cláusula de este tipo de oraciones, se ha estudiado hasta qué punto el tiempo verbal de la otra cláusula permite determinar cuándo aplicar cada etiqueta (v. ap. 4.3).

4.1. Tiempos verbales en las prótasis condicionales

En la Tabla 2 se muestra la distribución de los tiempos verbales de indicativo y de subjuntivo en las prótasis, así como los valores factuales extraídos de la anotación llevada a cabo. El tiempo verbal más frecuente de los seis utilizados en estas cláusulas es el pretérito pluscuamperfecto de subjuntivo (114 ejemplos), mientras que el que cuenta con menos casos es el pretérito perfecto compuesto (31 ejemplos), seguido del pretérito perfecto simple (32 ejemplos).

Tabla 2. Distribución de los valores factuales de los tiempos en prótasis condicionales.

TIEMPO VERBAL	VALOR FACTUAL VERBO (PRÓTASIS)	Nº DE EJEMPLOS / TOTAL TIEMPO PRÓTASIS
Presente de indicativo	[NA]**	97/105 (92,38%)
	[COM+]	7/105 (6,67%)
	[COM-]	1/105 (0,95%)

TIEMPO VERBAL	VALOR FACTUAL VERBO (PRÓTASIS)	Nº DE EJEMPLOS / TOTAL TIEMPO PRÓTASIS
Pretérito imperfecto de indicativo	[NA]**	37/53 (69,81%)
	[COM+]	16/53 (30,19%)
Pretérito perfecto simple	[NA]	9/32 (28,13%)
	[COM+]*	23/32 (71,87%)
Pretérito perfecto compuesto	[NA]	5/31 (16,13%)
	[COM+]*	26/31 (83,87%)
Pretérito imperfecto de subjuntivo	[NA]	47/105 (44,76%)
	[COM-]**	58/105 (55,24%)
<i>Pretérito pluscuamperfecto de subjuntivo</i>	[COM-]*	114/114 (100%)
Total		440/440

Casi todas las formas en prótasis condicionales expresan más de un valor. Si el valor más frecuente coincide con el valor reflejado en la Tabla 1, se indica con un asterisco. Sin embargo, cuando el valor más frecuente es distinto al señalado en la Tabla 1, se señala con dos asteriscos.

El único tiempo no variable es el pretérito pluscuamperfecto de subjuntivo (marcado en cursiva), ya que siempre se interpreta como «compromiso negativo» (17). Además, este valor coincide con el prototípico del tiempo verbal (v. Tabla 1).

- (17) «Al final —dicen— sacamos más que *si hubieran sido* más laxos de entrada» (lasprovincias.es, 19/06/2019). [COM-]

En cuanto a los otros cinco tiempos usados en las prótasis, son todos ellos no concluyentes. De estos, tres muestran valores más frecuentes que no coinciden con los generales para esos tiempos (v. Tabla 1), marcados con dos asteriscos: presente (7 ahora recogido como 18a, 18b y 18c) y pretérito imperfecto de indicativo (19a, y 9a ahora recogido como 19b), y pretérito imperfecto de subjuntivo (22a, y 3a ahora recogido como 22b). Para estos tiempos se formularán reglas específicas. Los otros dos tiempos usados en las prótasis condicionales, pretérito perfecto simple (20) y compuesto (2b ahora recogido como 21a, y 21b), muestran coincidencia entre su valor por defecto en estas construcciones y su valor por defecto en otras oraciones simples, por lo que no requieren reglas propias.

(18)

- «*Si* es cierto, estaríamos ante una de las mayores traiciones a la democracia de la historia» (lavozdegalicia.com, 20/03/2017). [NA]*
- Si* en algo es bueno Trump, es en definir a sus adversarios (...) (eldiario.es, 19/06/2019). [COM+]
- «La segunda tarjeta amarilla no lo era, pero ganar fue un alivio porque *si perdemos* me hubiera caído todo el peso de la derrota» (elpais.com, 01/10/2016). [COM-]

(19)

- Si* Podemos era un partido personalista, después de esta campaña es una persona con un partido (elespañol.com, 28/04/2019). [COM+]
- El anuncio de Zapatero de que *si llegaba* a la Moncloa le nombraría ministro de Trabajo, sorprendió al ex secretario general de CCOO (elmundo.es, 15/03/2004). [NA]**

(20)

- Si* Zapatero consiguió repetir la victoria fue gracias a los apoyos de los votantes nacionalistas y a la movilización de la izquierda (libertaddigital.com, 03/06/2018) [COM+]**
- «*Si cedió* sus marcas a sus hijos para eludir responsabilidad, no me parece bien» (abc.es, 03/03/2014). [NA]

(21)

- Si ha ganado* dos elecciones en Catalunya ha sido porque es transversal (...) (eldiario.es, 11/10/2018). [COM+]*
- «*Si no han entendido* que el feminismo está muy lejos de excluir a los hombres, no han entendido nada» (eldiario.es, 04/03/2019). [NA]

(22)

- Si fuera* primer ministro habría declarado, (...), al conjunto de los miembros del Gobierno turco personas non grata durante uno o dos meses (larioja.com, 05/03/2017). [COM-]**
- El ayuntamiento dispone de parcelas para este tipo de viviendas *si hubiera* demanda (levante.com, 05/03/2018). [NA]

4.2. Tiempos verbales en las apódosis condicionales

En las apódosis condicionales se han hallado nueve tiempos verbales. En la Tabla 3 se refleja la frecuencia en el corpus de cada tiempo (siendo el presente de indicativo el tiempo más veces identificado, 110, y el imperativo el que menos, 6) y los valores factuales en cada caso.

Tabla 3. Valores factuales de las apódosis condicionales.

TIEMPO VERBAL	VALOR FACTUAL VERBO (APÓDOSIS)	Nº DE EJEMPLOS / TOTAL TIEMPO APÓDOSIS
Presente de indicativo	[COM+]*	82/110 (74,55%)
	[COM-]	3/110 (2,73%)
	[NA]	25/110 (22,73%)
Pretérito imperfecto de indicativo	[NA]**	22/40 (55%)
	[COM-]	5/40 (12,5%)
	[COM+]	13/40 (32,5%)
<i>Pretérito perfecto simple</i>	[COM+]*	25/25 (100%)
<i>Pretérito perfecto compuesto</i>	[COM+]*	16/16 (100%)
<i>Futuro simple</i>	[NA]**	60/60 (100%)
Condicional simple	[NA]*	62/83 (74,70%)
	[COM-]	21/83 (25,30%)
Condicional compuesto	[COM-]**	47/56 (83,93%)
	[NA]	9/56 (16,07%)
Pretérito pluscuamperfecto subjuntivo	[COM-]*	40/44 (90,91%)
	[NA]	4/44 (9,09%)
<i>Imperativo</i>	[NA]*	6/6 (100%)

Como puede observarse, los únicos tiempos verbales que presentan un único valor factual en apódosis condicionales, es decir, que son concluyentes, son el pretérito perfecto simple (20a ahora recogido como 23a), el pretérito compuesto (23b), el imperativo (23c) y el futuro simple (1a ahora recogido como 23d) (marcados en cursiva en la Tabla 3). De estos, el único cuyo valor factual se contradice con el general es el futuro simple (acompañado de dos asteriscos), por lo que se requerirá formalizar una regla específica para este tiempo.

(23)

- Si Zapatero consiguió repetir la victoria *fue* gracias a los apoyos de los votantes nacionalistas y a la movilización de la izquierda (libertaddigital.com, 03/06/2018). [COM+]
- Si el importe de los cursos gestionados por el Ayuntamiento ha sido de 8,5 millones de euros entre 1994 y 1999, el coste de los cursos externos *ha superado* los 36 millones de euros (elpais.com, 09/01/2002). [COM+]
- «Si hubiera sido colocada en el sistema de aire acondicionado, *imaginen* el impacto», destacó Karnavian. (elmundo.es, 26/03/2015). [NA]
- Si gana hoy a el Getafe se *pondrá* con 40 puntos por los 35 de los colchoneros (...) (mundodeportivo.com, 06/01/2019). [NA]**

El resto de tiempos verbales de las apódosis condicionales, concretamente el presente (3a ahora recogido como 24a, y 24b) y el pretérito imperfecto de indicativo (12a ahora recogido como 25a, y 25b), el condicional simple (26a, y 10 ahora recogido como 26b) y compuesto (13 ahora recogido como 27a, y 27b), así como el pretérito pluscuamperfecto de subjuntivo (14a ahora recogido como 28a, y 28b), pueden expresar más de un valor factual según el contexto. Sin embargo, el valor más frecuente en el corpus coincide con su valor por defecto (v. Tabla 1), a excepción del pretérito imperfecto de indicativo y del condicional compuesto, acompañados con dos asteriscos. Así pues, para estos dos tiempos presentaremos también reglas específicas.

(24)

- El ayuntamiento *dispone* de parcelas para este tipo de viviendas si hubiera demanda (levante.com, 05/03/2018). [COM+]**
- Si hubiera mano es involuntaria, sólo veo que hay rodilla (mundodeportivo.com, 13/03/2012). [COM-]

- c. «Si hubiera arrancado el coche *hacemos* podio con bastante facilidad» (elconfidencial.com, 24/08/2014). [NA]
(25)
- a. «Si gobernara Vox, *vivíamos* sin autonomías (...)» (eldiario.es, 24/04/2019). [NA]**
- b. Si hubiera reventado, se *llevaba* por delante a todos los curiosos que los agentes han permitido estar a menos de 15 metros de dicha maleta (elnortedecastilla.es, 19/08/2016). [COM-]
(26)
- a. Si acaba la legislatura, el presidente ceutí *encadenaría* 22 años al frente de la ciudad (...) (elmundo.es, 15/06/2019). [NA]**
- b. «Si hubiera tanto paro nos *mataríamos* los unos a los otros» (abc.es, 27/01/2014). [COM-]
(27)
- a. Si mandara el pueblo *habrían metido* en cintura a los financieros hace tiempo (eldiario.es, 10/04/2013). [COM-]**
- b. (...) si hubiera entrado *habría sido* un gol para la historia del Real Madrid (...) (marca.com, 29/05/2014). [NA]
(28)
- a. Si hubiéramos tenido esta reforma laboral hace dos o tres años nos *hubiéramos ahorrado* un millón de parados (elpais.com, 01/01/2013). [COM-]**
- b. Si hubiéramos ganado las elecciones *hubiéramos hecho* lo mismo (elpais.com, 30/11/2014). [NA]

4.3. Correlación de tiempos verbales

En la Tabla 12 (v. Anexo) se muestran las distintas combinaciones temporales (prótasis-apódosis) identificadas, un total de 32, y el número de casos que presenta cada una. Cabe destacar que, de estas 32 combinaciones, 20 no se habían documentado todavía y cuyas casillas se destacan en gris. Como puede observarse, también se indican las combinaciones no identificadas en nuestro corpus (marcadas con -), que deducimos que no son posibles gramaticalmente o son poco frecuentes en el registro periodístico.

Una vez recopiladas las combinaciones temporales, se han analizado de forma detallada los casos de los tiempos no concluyentes en prótasis y apódosis. Recordamos que los tiempos no concluyentes son aquellos que pueden asociarse a más de un valor factual. El objetivo es identificar otros marcadores que permitan afinar la anotación de la factualidad en estas cláusulas en los casos de los valores menos frecuentes, ya que los más frecuentes ya han sido descritos. En estos casos hemos investigado si el tiempo verbal de la otra cláusula puede arrojar luz a este respecto. Solo se considerará que ello es posible si se cumple la siguiente condición: que el número de ejemplos en que el tiempo de dicha cláusula, en combinación con otro tiempo específico de la otra cláusula, sea superior al número de casos de esa misma combinación con el el valor por defecto (PD) (que puede ser 0).

Los resultados de este análisis relativo a la posible influencia del tiempo de una cláusula en la interpretación factual de la otra se presentan en las Tablas 13 y 14 (v. Anexo). Para facilitar la interpretación de dichas tablas, en la última fila del encabezado se indica cuáles son los valores generales habituales para cada tiempo (PD) y cuáles son otros posibles valores que dicho tiempo también puede adoptar en español («Otros»). En las filas de debajo se indican las cifras de los ejemplos encontrados para cada valor en el corpus diferenciando también el tiempo de la otra cláusula. Como se puede comprobar, en algunos casos se cumple la condición mencionada y se han marcado las celdas correspondientes en color gris. Se ha diferenciado entre el gris claro y el oscuro en función de si el número de casos es menor que 10 o igual o mayor a 10, respectivamente. Tal como se ha explicado (v. ap. 3), solo para estos últimos (igual o mayor a 10) se han formalizado las reglas correspondientes.

En primer lugar, en el caso de las apódosis (v. Tabla 13), tres de los cinco tiempos no concluyentes han dado algún resultado positivo respecto a la posible influencia del tiempo de la prótasis: presente y pretérito imperfecto de indicativo y condicional simple. Ahora bien, solo en este último el número de ejemplos recopilados es mayor que 10. Concretamente, se ha observado que cuando este tiempo de la apódosis aparece con prótasis con pretérito pluscuamperfecto de subjuntivo (14b ahora recogido como 29), el valor del condicional simple es, en su mayoría, [COM-], y no [NA], que es el valor PD.

(29) Si el Ayuntamiento *hubiera pagado* la resolución de 2005 en los seis primeros meses, el tema *estaría* resuelto (elpais.com, 08/03/2012).

En segundo lugar, en cuanto a las prótasis (v. Tabla 14), cuatro de los cinco tiempos no concluyentes han dado algún resultado positivo respecto a la posible influencia del tiempo de la apódosis: presente, pretérito imperfecto y pretérito perfecto compuesto de indicativo y pretérito imperfecto de subjuntivo. Como en el caso anterior, solo en uno de ellos los datos son mayores de 10. Concretamente, en las prótasis con pretérito imperfecto de subjuntivo, vemos que con presente (20b) o con futuro en apódosis (30), su valor más frecuente es [NA], y no [COM-], que es el valor PD.

(30) Si *hubiera* ilegalidad o infracción, la Junta *llevará* el tema a los tribunales (elpais.com, 26/11/2002).

5. Propuesta de formalización

Las reglas presentadas en este apartado son una propuesta de formalización para la anotación automática de la factualidad de las oraciones condicionales, diseñada para ser utilizada en el sistema desarrollado en el proyecto TAGFACT. Por ello, primero vamos a describir algunos aspectos básicos del mismo.

En primer lugar, el conjunto de etiquetas para la anotación de la factualidad usado en TAGFACT (Alonso y otros, 2018; Vázquez y otro, 2020) está inspirado en Saurí (2008), como la mayoría de trabajos en esta línea, pero es original. Lo más característico es que se usa la combinación de cuatro categorías: tiempo (presente-pasado-futuro), eventualidad (evento-mental-estado), polaridad (positiva-negativa), compromiso de certeza (compromiso-no_compromiso).

En segundo lugar, en TAGFACT, el sistema automático de anotación se ha diseñado de modo que, en un primer módulo, se decide qué oraciones expresan situaciones para las cuales no es aplicable el concepto de factualidad (como los deseos, entre otras). Estas son anotadas como NA (no aplica) a partir de unas reglas que determinan unos patrones tipo de esta clase de situaciones.

En un segundo módulo se etiquetan el resto de oraciones con las cuatro etiquetas, utilizando para todas ellas los valores más frecuentes en cada categoría (por ejemplo, es más habitual el compromiso que el no compromiso).

A continuación, se concatenan distintos módulos en los que se aplican reglas específicas de cada categoría. Las reglas más específicas se aplican antes que las menos restrictivas. Son reglas formuladas para los casos que no se corresponden con los valores más frecuentes. En estos casos los valores asignados por defecto son sustituidos por los nuevos. Algunas de estas reglas específicas están relacionadas con los tiempos verbales; por ejemplo, respecto a la categoría compromiso de certeza, todos los verbos usados en presente de subjuntivo pierden el valor de [COM+] asignado en el módulo 2 y se reetiquetan como [NA]. Hay otros tiempos que también son reetiquetados respecto a esta categoría, como el pretérito imperfecto y pluscuamperfecto de subjuntivo, el condicional simple y compuesto y el imperativo, en cada caso con los valores pertinentes (v. Tabla 1).

En este estudio las reglas creadas para la anotación de las oraciones condicionales se ciñen a las categorías del compromiso de certeza y polaridad, ya que en este tipo de oraciones son las únicas categorías donde se producen cambios respecto a los valores por defecto. Otra cuestión que debe tenerse en cuenta es que las reglas propuestas en este estudio se combinarán con otras en el sistema de TAGFACT, por lo que los valores asociados a los verbos de las oraciones condicionales después de la aplicación de estas reglas pueden ser modificados en algunos casos, según la etiqueta y el orden de aplicación de las reglas que se establezca. Por ejemplo, si un verbo de una prótasis de una oración condicional se etiqueta como [COM+] pero esta incluye un adverbio de probabilidad, ese valor será reetiquetado con [NO COM], que indica incerteza respecto a la realización de la situación (v. ap. 3).

Como se ha mencionado anteriormente, en el sistema TAGFACT se ha empleado el anotador morfológico y sintáctico Freeling 4.1 (Padró y Stanilovsky, 2012), el cual utiliza las etiquetas EAGLES (1994) para desambiguar los tiempos verbales. Así pues, como se verá en el apartado 5, en nuestras reglas se especifican los distintos tiempos verbales posibles en las prótasis y las apódosis de las oraciones usando dichas etiquetas.

Por último, todas las reglas que se presentan en este estudio contienen la conjunción *si*, puesto que todas ellas atañen a oraciones condicionales y solo se ha estudiado esta conjunción, y no otras posibles conjunciones condicionales, como ya se ha explicado. La partícula *si*, sin embargo, es polisémica, ya que puede ser una conjunción condicional o interrogativa. En Freeling, se identifica como CS (conjunción subordinante) en ambos casos. Para su desambiguación se ha creado una regla (v. Tabla 4) que nos permite determinar cuándo este elemento introduce una oración condicional y cuándo no.

Tabla 4. Representación de la regla de anotación para desambiguar *si* interrogativo indirecto.

VERBO		CONECTOR	ETIQUETA MORFOLÓGICA PARA EL CONECTOR
Verbo de comunicación / conocimiento {saber, preguntar, cuestionar, conocer...}	(.....)	<i>si</i>	CS-INT

Como puede observarse, en esta regla se discriminan los *si* interrogativos, que son los que están precedidos por verbos que pertenecen a la clase semántica de comunicación o conocimiento, *saber* (*sabrán*) en (31). Entre dichos verbos y la conjunción pueden haber otros elementos por lo que en la regla se añade (.....) entre ambos. Dicha regla debe aplicarse antes que las relativas a la factualidad. Con ello se conseguirá que las reglas propuestas en este estudio, aplicables solo a la conjunción *si* CS (no CS-INT), no actúen sobre oraciones interrogativas.

(31) Ellos *sabrán si* se sienten más cómodos con los socialdemócratas o con los ultraderechistas, ha dicho el ministro de Fomento y secretario de Organización, José Luis Ábalos (publico.es, 18/02/2019).

5.1. Tiempos verbales en las prótasis condicionales

Como hemos visto, de los cinco tiempos presentes en las prótasis, solo hemos determinado crear reglas específicas para el presente y el pretérito imperfecto de indicativo, así como para el pretérito imperfecto de subjuntivo (v. ap. 4.1).

En la Tabla 5 se refleja la configuración de tres reglas para estas prótasis y se numeran en la primera columna. Dichas reglas se han formulado para otorgar el valor más frecuente de estos tiempos en condicionales, que no coincide con el valor por defecto general de dichos tiempos. La segunda, tercera y cuarta columnas indican el patrón que se ha de localizar en el texto para que sea aplicable la regla y en la última columna de dicha tabla se especifica el valor factual que se asigna al verbo de la cuarta columna.

Respecto al patrón textual, primero se identifica el conector *si* para poder especificar que solo van a aplicarse a las condicionales, teniendo en cuenta que se ha aplicado la regla de desambiguación del *si* antes. Puede ser que después de la conjunción y antes del verbo aparezcan algunas palabras (...). En nuestro corpus se ha observado que mayoritariamente dicho conector se encuentra inmediatamente antes del verbo de la prótasis y en los pocos casos en que esto no sucede el máximo de vocablos establecido es tres. Concretamente, en el caso del presente, el pretérito imperfecto de indicativo y el pretérito imperfecto de subjuntivo, el 76,19%, el 84,91% y el 97,14% de los casos analizados, respectivamente, presentan el verbo de la prótasis tras el conector. En la cuarta columna, se indica que es necesaria la presencia de un verbo con un determinado tiempo verbal expresado con la etiqueta de Freeling. Para su descodificación hemos añadido entre // la denominación habitual de cada tiempo.

Tabla 5. Representación de las reglas de anotación para las prótasis condicionales no concluyentes.

	CONECTOR		VERBO (PRÓTASIS) TIEMPO VERBAL	VALOR FACTUAL VERBO (PRÓTASIS)
1	<i>si</i> (CS)	(...)	/Presente de indicativo/ (VMIP)	[NA]
2			/Pretérito imperfecto de indicativo/ (VMII)	
3			/Pretérito imperfecto de subjuntivo/ (VMSI)	[COM-]

A continuación (Tabla 6), presentamos otras reglas para la etiquetación del verbo de la prótasis en pretérito imperfecto de subjuntivo cuando no adopta su valor por defecto en condicionales, ya que este ha sido resuelto en la regla 3. Así, como se ha argumentado (v. ap. 4.3), hay dos casos en que el tiempo del verbo de la apódosis (cuando se usa el presente o el futuro) determina estos otros valores del imperfecto de subjuntivo en prótasis condicionales. Es por ello que ahora se añade otra columna para codificar la información sobre el verbo de la apódosis. Por último, en la última columna se especifican los valores factuales para el verbo de la prótasis. Hay que señalar que las reglas 4 y 5 se aplicarían cuando la prótasis es la primera cláusula de la oración condicional y la apódosis es la segunda.

Tabla 6. Representación de las reglas de anotación para las prótasis condicionales con pretérito imperfecto de subjuntivo y apódosis con presente y futuro.

	CONECTOR		VERBO (PRÓTASIS) TIEMPO VERBAL		VERBO (APÓDOSIS) TIEMPO VERBAL	VALOR FACTUAL VERBO (PRÓTASIS)
4	<i>si</i> (CS)	(...)	/Pretérito imperfecto de subjuntivo/ (VMSI)	(...)	/Presente/ (VMIP)	[NA]
5					/Futuro/ (VMIF)	[NA]

En el caso de que el orden de las cláusulas de los patrones de las reglas 4 y 5 fuera distinto, es decir, en el caso de oraciones condicionales con la misma combinación temporal para prótasis y apódosis pero con la apódosis ocupando el lugar de la primera cláusula, se aplicarían las reglas 6 y 7 de la Tabla 7.

Tabla 7. Representación de las reglas de anotación para las prótasis con pretérito imperfecto de subjuntivo y apódosis con presente y futuro.

	(...) VERBO (APÓDOSIS) (...) TIEMPO VERBAL	CONECTOR		VERBO (PRÓTASIS) TIEMPO VERBAL	VALOR FACTUAL VERBO (PRÓTASIS)
6	(...)/Presente/ (VMIP) (...)	<i>si</i> (CS)	(...)	Pretérito imperfecto de subjuntivo/ (VMSI)	[NA]
7	(...)/Futuro/ (VMIF) (...)			[NA]	

5.2. Tiempos verbales en las apódosis condicionales

En el caso de las apódosis de oraciones condicionales, en nuestro estudio los únicos tiempos verbales que presentan un valor distinto al prototípico son el pretérito imperfecto de indicativo, el futuro simple y el condicional compuesto (v. ap. 4.2).

En primer lugar, presentamos en la Tabla 8 cómo se configuran las reglas de anotación de la apódosis si esta se sitúa después de la prótasis. Respecto al patrón textual, primero se identifica el conector *si* usado como conjunción condicional (CS) y no como conjunción interrogativa (CS-INT). Ahora bien, esta conjunción se podría situar considerablemente lejos del verbo de la apódosis. En todo caso entre ambos se identificará como mínimo otro verbo, el de la prótasis. Así, la tercera columna se usa para representar un espacio de un número indefinido de palabras, que será mínimo de uno, el verbo de la prótasis. El tiempo de dicho verbo será alguno del modo indicativo o pretérito imperfecto o pluscuamperfecto de subjuntivo, que son los que pueden aparecer en estas prótasis condicionales, pero no es necesario especificarlo porque no modifica el valor factual de la apódosis. Así, en la regla aparece la etiqueta (V----), que indica que solo hay que comprobar el primer carácter de la serie de cinco de dicha etiqueta, o sea, que haya un verbo entre el *si* y el otro verbo.

Tabla 8. Representación de la regla de anotación para las apódosis condicionales con pretérito imperfecto de indicativo, futuro simple y condicional compuesto.

	CONECTOR	(...) VERBO (PRÓTASIS) (...) TIEMPO VERBAL	VERBO (APÓDOSIS) TIEMPO VERBAL	VALOR FACTUAL VERBO (APÓDOSIS)
8	<i>si</i> (CS)	(...) (V----) (...)	/Futuro simple/ (VMIF)	[NA]
9			/Pretérito imperfecto de indicativo/ (VMII)	[NA]
10			/Condicional compuesto/ (VAIC VMP)	[COM-]

A continuación (Tabla 9), presentamos una propuesta de regla en la que sí se tienen en cuenta el tiempo del verbo de la prótasis para asignar el valor factual (en este caso el que no es por defecto) al verbo de la apódosis cuando se conjugue en condicional simple, ya que en este caso sí es determinante el tiempo de la primera cláusula para determinar el valor factual de la segunda (v. ap. 4.3).

Tabla 9. Representación de la regla de anotación para las apódosis con condicional simple y prótasis con pretérito pluscuamperfecto de subjuntivo.

	CONECTOR		VERBO (PRÓTASIS) TIEMPO VERBAL	(...) VERBO (APÓDOSIS) (...) TIEMPO VERBAL	VALOR FACTUAL VERBO (APÓDOSIS)
11	<i>si</i> (CS)	(...)	/Pretérito pluscuamperfecto de subjuntivo/ (VASI VMP)	(...) /Condicional simple/ (VMIC) (...)	[COM-]

En segundo lugar, aunque son mucho menos frecuentes, cabe tratar el caso de que la apódosis se encuentre antes de la prótasis. En las tablas 10 y 11 presentamos la adaptación correspondiente de las reglas de las tablas 8 y 9.

Tabla 10. Representación de la regla de anotación para las apódosis condicionales con pretérito imperfecto, futuro y condicional compuesto.

	VERBO (APÓDOSIS) TIEMPO VERBAL	CONECTOR		VERBO (PRÓTASIS) TIEMPO VERBAL	VALOR FACTUAL VERBO (APÓDOSIS)
12	/Futuro simple/ (VMIF)	<i>si</i> (CS)	(...)	(V----)	[NA]
13	/Pretérito imperfecto de indicativo/ (VMII)				[COM-]
14	/Condicional compuesto/ (VAIC VMP)				[COM-]

Tabla 11. Representación de la regla de anotación para las apódosis con condicional simple y prótasis con pretérito pluscuamperfecto de subjuntivo.

	(...) VERBO (APÓDOSIS) (...) TIEMPO VERBAL	CONECTOR		VERBO (PRÓTASIS) TIEMPO VERBAL	VALOR FACTUAL VERBO (APÓDOSIS)
15	(...) /Condicional simple/ (VMIC) (...)	<i>si</i> (CS)	(...)	/Pretérito pluscuamperfecto de subjuntivo/ (VASI VMP)	[COM-]

6. Conclusiones

Este estudio aborda, en primer lugar, qué variedad de valores factuales expresan las oraciones condicionales con *si* en español y qué elementos lingüísticos son relevantes para su identificación automática en cada caso a partir del análisis de 440 ejemplos extraídos de textos periodísticos. En segundo lugar, se han propuesto reglas formales para la anotación automática de la factualidad de estas oraciones en el marco del proyecto TAGFACT. En dicho proyecto se ha desarrollado un anotador automático de la factualidad, cuyo funcionamiento se basa, primero, en la asignación de los valores factuales más comunes de forma masiva y, después, en su rectificación con reglas específicas basadas en la presencia de elementos o patrones concretos.

Dado que los textos procesados están anotados previamente morfológicamente y sintácticamente, para nuestro análisis y la posterior formalización, hemos tenido en cuenta, además de la presencia de la conjunción condicional *si*, los tiempos verbales usados en las prótasis y las apódosis por separado y en combinación. Respecto a este último aspecto, una primera aportación desde el punto de vista de la descripción lingüística es que se han identificado 20 nuevas combinaciones temporales para estas oraciones que no habían sido todavía documentadas en la bibliografía sobre el tema.

Otra aportación es que se ha podido comprobar empíricamente que las oraciones condicionales con *si* muestran una interesante complejidad en cuanto a la factualidad, ya que los valores factuales más frecuentes de los tiempos usados en ellas varían en un porcentaje no trivial respecto a los valores más frecuentes asociados a dichos tiempos en general para el español. Esta variación es, por ejemplo, bastante más elevada que la que se ha identificado en otros estudios similares realizados respecto a otro tipo de oraciones que también incluyen cláusulas interordinadas, como las concesivas (Barrios y Vázquez, 2020). Concretamente, en un 50% de los tiempos de las prótasis de las oraciones condicionales el valor más frecuente es distinto al valor más frecuente general de esos mismos tiempos. En el caso de las apódosis, esto ocurre para un 33,33%. Para todos estos tiempos usados en contextos de oraciones condicionales se han formulado reglas específicas para la adecuada anotación de la factualidad en estos casos.

Además, como los valores por defecto asociados a estos tiempos en su mayoría no son concluyentes, para los tiempos verbales usados en las oraciones condicionales que presentan más de un posible valor también se ha analizado qué papel tienen las correlaciones temporales a la hora de determinar los otros valores factuales, es decir, los que no son los usos más frecuentes en los contextos condicionales, con el objetivo de crear nuevas reglas y afinar más la anotación. Del estudio realizado se desprende que puede ser relevante esta información para determinar estos valores en diez casos (cinco de prótasis y cinco de apódosis) pero solo se han formulado tres reglas que incluyen esta información (una de prótasis y dos de apódosis). El motivo es que en los otros siete casos hemos considerado que el número de ejemplos obtenido de nuestro corpus es demasiado bajo (inferior a diez) como para establecer generalizaciones.

El resultado final ha sido un total de 15 reglas específicas para las oraciones condicionales del español. Concretamente, de estas, siete corresponden a tiempos verbales en prótasis y ocho en apódosis. En todas ellas se especifica un patrón que contiene los elementos que debe identificar secuencialmente el sistema en el texto y el valor factual que asignará al verbo de la cláusula. Este valor será el que substituirá el establecido por defecto de forma masiva inicialmente por el sistema.

Como trabajo futuro más importante, una vez implementadas las reglas en el anotador TAGFACT, se prevé evaluar el resultado para establecer las mejoras oportunas. Además, también se prevé ampliar el número de frases de combinaciones temporales poco representadas para determinar si sería adecuado llegar a formalizar esas reglas que hemos identificado solo como posibles tendencias. También habría que ampliar el estudio a oraciones condicionales con mayor complejidad sintáctica y con otras conjunciones distintas a *si* y comprobar si se pueden extrapolar los resultados.

Agradecimientos

Esta investigación se ha desarrollado en el proyecto TAGFACT: del texto al conocimiento (FFI2017-84008-P; Ministerio de Economía, Industria y Competitividad). Asimismo, Leyre Barrios ha contado con la financiación de la beca de ayuda UdL, Jade Plus y la Fundación Bancaria Caixa.

Contribución de autoría CREDIT

Ambas autoras han contribuido por igual a la elaboración de este artículo. Ambas se han encargado de la conceptualización, de la curación de datos, del análisis formal, de la investigación, de la metodología y de la redacción. No obstante, Glòria Vázquez ha llevado a cabo, además, la supervisión y ha llevado a cabo la adquisición de fondos como PI del proyecto en el que se inscribe este estudio.

Referencias bibliográficas

- Ahern, Aoife (2008). *El subjuntivo: Contextos y efectos*. (Vol. 104). Arco/Libros.
- Alarcos Llorach, Emilio (1994). *Gramática de la lengua española*. Espasa.
- Alonso, Laura, Castellón, Irene, Curell, Hortènsia, Fernández-Montraveta, Ana, Oliver, Sonia, y Vázquez, Gloria (2018). Proyecto TAGFACT: Del texto al conocimiento. Factualidad y grados de certeza en español. *Procesamiento del Lenguaje Natural*, 61, 151-154. <http://dx.doi.org/10.26342/2018-61-20>
- Banco de datos (CORPES XXI). Corpus del Español del Siglo XXI. <https://www.rae.es/banco-de-datos/corpes-xxi>

- Barrios, Leyre, y Vázquez, Gloria (2020). Las oraciones concesivas en español y la factualidad. *Estudios Filológicos*, 66, 151-183. <https://doi.org/10.4067/s0071-17132020000200151>
- Comajoan-Colomé, Llorenç, y Pérez Saldanya, Manuel (2018). *Los tiempos verbales del español. Descripción del sistema y su adquisición en segundas lenguas*. Octaedro Editorial.
- Dancygier, Barbara (1998). *Conditionals and prediction: Time, knowledge and causation in conditional constructions*. Cambridge University Press.
- Davies, Mark (2018). Corpus del Español NOW (News on the Web). <https://www.corpusdelespanol.org/now/>
- Declerck, Renaat, y Reed, Susan (2001). *Conditionals: A comprehensive empirical analysis*. Mouton de Gruyter.
- Diab, Mona. T., Levin, Lovin, Mitamura, Teruko, Rambow, Owen, Prabhakaran, Vinodkumar, y Guo, Weiwei (2009). Committed belief annotation and tagging. En *Proceedings of the Third Linguistic Annotation Workshop ACL-IJCNLP* (pp. 68-73). Singapore: Suntec. <https://aclanthology.org/W09-3012.pdf>
- Escandell, Victoria (2010). Futuro y evidencialidad. *Anuario de Lingüística Hispánica*, 26, 9-34. <http://uvadoc.uva.es/handle/10324/9480>
- Fernández-Montraveta, Ana, Curell, Hortènsia, Vázquez, Gloria, y Castellón, Irene (2020). The TAGFACT annotator and editor: A versatile tool. En *Research in Corpus Linguistics*, 8(1), 131-146. <https://doi.org/10.32714/ricl.08.01.08>
- García Berrio, Antonio (1970). Bosquejo para una descripción de la frase compuesta en español: El esquema tradicional a la luz de la moderna lingüística: tipos de construcción, transformación y funciones. En *Anales de la Universidad de Murcia. Filosofía y Letras*, 28 (3-4). Universidad de Murcia.
- Gili y Gaya, Samuel (1961). *Curso superior de sintaxis de la lengua española*. Bibliografía.
- Lee, Kenton, Artzi, Yoav, Choi, Yejin, y Zettlemoyer, Luke (2015). Event detection and factuality assessment with non-expert supervision. En Lluís Márquez, Chris Callison-Burch and Jian Su (Eds.), *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 1643-1648). Association for Computational Linguistics. <https://aclanthology.org/D15-1189.pdf>
- Montolío, Estrella (1999). Las construcciones condicionales. En I. Bosque y V. Demonte (Eds.), *Gramática descriptiva del español* (vol. 3, pp. 3653-3737). Espasa-Calpe.
- Moya Corral, Juan Antonio (2018). *Pautas para el análisis sintáctico funcional del español*. Universidad de Granada.
- Padró, Lluís, y Stanilovsky, Evgeny (2012). FreeLing 3.0: Towards wider multilinguality. En *International Conference on Language Resources and Evaluation (LREC2012)* (pp. 2473-2479.). Estambul. <http://hdl.handle.net/2117/15986>
- Porcar Miralles, Margarita (1993). *La oración condicional. La evolución de los esquemas verbales condicionales desde el latín al español actual*. Universitat Jaume I.
- Real Academia de la Lengua Española (2009). *Nueva gramática de la lengua española*. Espasa.
- Rodríguez Rosique, Susana (2008). *Pragmática y gramática: Condicionales concesivas en español*. Frankfurt am Main: Peter Lang.
- Rojo, Guillermo (1974). La temporalidad verbal en español. *Verba*, 1, 68-149.
- Rojo, Guillermo, y Montero, Emilio (1983). La evolución de los esquemas condicionales. (Potenciales e irreales desde el *Poema del Cid* hasta 1400). *Verba*. Anexo 22. Universidad de Santiago de Compostela.
- Saurí, Roser (2008). *A factuality profiler for eventualities in text* [Tesis doctoral. Brandeis University]. <https://www.proquest.com/openview/1743e640d05945237a0634656c2ebf4a/1?pq-origsite=gscholar&cbl=18750>
- Schwenter, Scott (1999). From hypothetical to factual and beyond: Refutational *si*-clauses in Spanish conversation. En Jean-Pierre Koenig (Ed.), *Discourse and cognition: Bridging the gap* (pp. 423-435). CSLI Publications.
- Sweetser, Eve (1990). *From etymology to pragmatics*. Cambridge University Press.
- Vázquez, Gloria, y Fernández-Montraveta, Ana (2020). Annotating factuality in the TAGFACT corpus. En M. Fuster-Márquez, C. Gregori-Signes, J. Santaemilia Ruiz (Eds.), *Multiperspectives in analysis and corpus design* (pp. 115-127). Comares. <http://hdl.handle.net/10459.1/69410>
- Veiga, Alexandre, y Mosteiro Louzao, Manuel (2006). *El modo verbal en cláusulas condicionales, causales, consecutivas, concesivas, finales y adverbiales de lugar, tiempo y modo*. Ediciones Universidad de Salamanca.
- Wonsever, Dina, Rosá, Aiala, y Malcuori, Marisa (2016). Factuality annotation and learning in Spanish texts. En *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*.

Tabla 12. Correlación de las prótasis y sus respectivas apódosis.

		PRÓTASIS							TOTAL TIEMPO APÓDOSIS
		Presente indicativo	Pretérito imperfecto indicativo	Pretérito perfecto simple	Pretérito perfecto compuesto	Pretérito imperfecto subjuntivo	Pretérito pluscuamperfecto subjuntivo		
APÓDOSIS	Presente indicativo	44	10	13	19	20	4	110	
	Pretérito imperfecto indicativo	–	22	1	1	9	7	40	
	Pretérito perfecto simple indicativo	1	5	18	–	–	1	25	
	Pretérito perfecto compuesto indicativo	–	6	–	10	–	–	16	
	Futuro simple	48	2	–	–	10	–	60	
	Condicional simple	7	8	–	–	50	18	83	
	Condicional compuesto	–	–	–	–	14	42	56	
	Pretérito pluscuamperfecto subjuntivo	1	–	–	–	2	41	44	
	Imperativo	4	–	–	1	–	1	6	
	TOTAL TIEMPO PRÓTASIS	105	53	32	31	105	114	440	

Tabla 13. Correlación de las apódosis no concluyentes y sus respectivas prótasis.

		APÓDOSIS											
		Presente indicativo		Pretérito imperfecto indicativo		Condicional simple		Condicional compuesto		Pretérito pluscuamperfecto de subjuntivo			
		PD	Otros	PD	Otros	PD	Otros	PD	Otros	PD	Otros		
PRÓTASIS	Presente indicativo	COM + 24	NA 20	-	-	NA 7	-	-	-	COM - 1	-		
	Pretérito imperfecto indicativo	COM + 10	-	NA 11	COM + 10	NA 8	-	-	-	-	-		
	Pretérito perfecto simple	COM + 13	-	-	COM + 1	-	-	-	-	-	-		
	Pretérito perfecto compuesto	COM + 17	NA 2	-	COM + 1	-	-	-	-	-	-		
	Pretérito imperfecto subjuntivo	COM + 17	NA 3	NA 9	-	NA 45	COM - 5	COM - 14	COM - 2	-	-		
	Pretérito pluscuamperfecto subjuntivo	COM + 1	COM - 3	NA 2	COM - 5	NA 2	COM - 16	COM - 33	COM - 37	NA 9	NA 4		
	Total PD/Otros apódosis	COM + 82	NA 25 / COM - 3	NA 22	COM + 12 / COM - 5	NA 62	COM - 21	COM - 47	COM - 40	NA 9	NA 4		
Total tiempo verbal apódosis	110/440 (25%)		39/440 (8,86%)		83/440 (18,86%)		56/440 (12,73%)		44/440 (9,77%)				

Tabla 14. Correlación de las prótasis no concluyentes y sus respectivas apódosis.

		PRÓTASIS											
		Presente indicativo		Pretérito imperfecto indicativo		Pretérito perfecto simple		Pretérito perfecto compuesto		Pretérito imperfecto subjuntivo			
		PD	Otros	PD	Otros	PD	Otros	PD	Otros	PD	Otros		
APÓDOSIS	Presente	NA 37	COM + 7	NA 3	COM + 7	COM + 9	NA 4	COM + 16	NA 3	COM - 4	NA 16		
	Pretérito imperfecto	-	-	NA 15	COM + 7	COM + 1	-	COM + 1	-	COM - 9	-		
	Pretérito perfecto simple	NA 1	-	NA 5	-	COM + 13	NA 5	-	-	-	-		
	Pretérito perfecto compuesto	-	-	NA 4	COM + 2	-	-	COM + 9	NA 1	-	-		
	Futuro	NA 48	-	NA 2	-	-	-	-	-	-	NA 10		
	Condicional	NA 7	-	NA 8	-	-	-	-	-	COM - 29	NA 21		
	Condicional compuesto	-	-	-	-	-	-	-	-	COM - 14	-		
	Pretérito pluscuamperfecto subjuntivo	-	COM - 1	-	-	-	-	-	-	COM - 2	-		
	Imperativo	NA 4	-	-	-	-	-	-	-	-	-		
	Total PD/Otros prótasis	NA 97	COM + 7 / COM - 1	NA 37	COM + 16	COM + 23	NA 9	COM + 26	COM - 58	NA 5	COM - 58	NA 47	
Total tiempo verbal prótasis	105/440 (23,86%)	53/440 (12,05%)	32/440 (7,27%)	31/440 (7,05%)	105/440 (23,87%)								