
Annotating factuality in the TAGFACT corpus

GLÒRIA VÁZQUEZ GARCÍA

Universitat de Lleida

ANA FERNÁNDEZ-MONTRAVETA

Universitat Autònoma de Barcelona

Abstract: The author of a text states, either implicitly or explicitly, his/her commitment with respect to the certainty of the situation being narrated (i.e. its factual status according to the author's stance). In the area of corpus linguistics and natural language processing (NLP), the annotation of factuality has experienced a growing interest since it simplifies complex processes such as information retrieval or fact checking. This article describes the labels used in the annotation scheme of the corpus TAGFACT. It presents some innovative aspects such as the labelling of situations according to their dynamic or non-dynamic nature. Non-dynamic situations are further classified as states, absolute truths and properties of an eventive nature. Even though this scheme has been created for the annotation of a Spanish corpus, it is applicable to other languages.

Keywords: factuality, certainty, commitment, corpus annotation.

1. INTRODUCTION

In recent years there has been a growing interest in the annotation of the degree of certainty with which events are presented in texts, that is, their factual status. This topic is especially relevant in the area of NLP and, more specifically, in the field of information retrieval.

The objective of this paper is to provide an account of the annotation scheme created in the TAGFACT¹ project. In this project, we aim to create an automatic tool for the annotation of the factual status of situations as presented in different newspapers written in Spanish. This tool will be solely based on linguistic information. In a first phase, a part of the corpus (Gold Standard) is being manually annotated in order to study the textual elements that need to be taken into consideration. We would like to

¹ This project is developed by the research group GRIAL and funded by the Ministerio de Economía, Industria y Competitividad - FFI2017-84008-P.

highlight that this project represents an important contribution for Spanish, since very little work has been done in this area for this language.

Next section is dedicated to review other projects that deal with the annotation of factuality. Section 3 presents some crucial decisions taken before the establishment of the scheme and Section 4 defines the tagset used in the annotation process, describing the different attributes and their values.

2. FACTUALITY²

The simplest treatment of factuality is based on two options: either the speaker shows commitment to the certainty of what is being narrated (presenting it as either true or false) or presents the situation as doubtful. Thus, in this annotation scheme, used by Diab et al. (2009) and projects derived (Prabhakaran et al. 2015; Colomer et al. 2016), there are two labels: *commitment* and *non commitment*. The applications of this basic analysis are scarce since it does not distinguish between facts and counterfactuals.

Other models distinguish three possible views of a situation: positive certainty (fact), negative certainty (counterfact) and uncertainty. Tonelli et al. (2014), Minard et al. (2015, 2016a, 2016b) and Wonsever et al. (2016) follow this tripartite model. However, not all these projects include the equivalent cases under the same label, especially as for future situations. Thus, Tonelli et al. (2014) and Wonsever et al. (2016) consider uncertain future situations and all those present and past situations not presented with commitment, but, whereas Tonelli et al. (2014) use the label *nonfactual*, Wonsever et al. (2016) use *IND* (indefinido ‘undefined’). In our opinion, future situation should be dealt separately from uncertain past and present situations because only in the first case uncertainty is absolute. A different approach is presented in Minard et al. (2015, 2016a, 2016b), where future situations are split into all three categories. Thus, they differentiate between true situations (*factual*), situations presented as not true (*counterfactual*) and uncertain situations (*nonfactual*), regardless of the temporal reference. From our point of view, if an event has not happened yet (future situation), it can never be presented as certain (factual or counterfactual).

Van Son et al. (2014) propose a set of four labels. Hence, future situations are dealt differently from all the other situations and tagged with a specific tag *future*. Then, past and present situations are tagged as *certain*, *not true* or *uncertain*. From our perspective, this proposal can be improved by further differentiating between positive or negative future situations, which would allow us to study the reliability of predictions *a posteriori*.

² We would like to thank Leyre Barrios for her help in developing this review.

In Narita et al. (2013), Matsuyoshi et al. (2010), Soni et al. (2014) and Tianxiong et al. (2017)³, the labels *probable* and *possible* are also considered as subclasses of uncertainty. However, as Velupillai (2011) shows, the use of different categories in this area causes problems since the boundary between them is unclear in some cases.

Light et al. (2004) and Medlock and Briscoe (2007) present some of the pioneering automatic systems created for the annotation of factuality. Nevertheless, Saurí (2008) and Pustejovsky and Saurí (2009) (Factbank project) are considered the seminal papers in this field. All the authors and projects mentioned above (with the exception of Diab et al. [2009]) are based on their proposal. FactBank considers polarity not only for certainty (*CT +*, *CT-*), but also for probability (*PR +*, *PR-*) and possibility (*PS +*, *PS-*).

Finally, Wonsever et al.'s (2008) also propose six labels, but from a slightly different perspective. The labels proposed are: positive (*R*: performed) and negative (*NR*: unrealized) certainty, possible and probable situations (*POS*), programmed future (*FP*: very likely to happen), indefinite future (*IND*: actions which are probable or possible) and future denied (*FN*: a slight possibility of something happening). In subsequent proposals (Wonsever et al. 2016), the model was considerably simplified.

All the above mentioned works are concerned with the annotation of English texts, with the exception of Matsuyoshi et al. (2010) and Narita et al. (2013), who work with Japanese, Minard et al. (2015, 2016a) with Italian and Wonsever et al. (2008, 2016) with Spanish. In addition, there are several multilingual projects; for example, Colomer et al. (2016) study English and Chinese whereas Minard et al. (2016b) deal with English, Italian and German.

3. PREANNOTATION CONSIDERATIONS

One of the main problems encountered when one sets to annotate factuality is the definition of the term *fact*. A fact is an event that happens, i.e. it can be proved to be true (or false) (Wittgenstein 1921/1994). According to this definition, those situations describing or assigning properties to an entity (states) should, in fact, be excluded from the annotation. Since states are also considered in our project, we prefer to talk about situations.

Situations may be interpreted in a different way by different people (depending on different authors, perceptions, life experiences or beliefs). Moreover, situations could also be a consciously manipulated view of one's interpretation (lies). In our approach, partially following the "possible worlds semantics" (Montague 1968), when we annotate the certainty of a situation, factual values are always assigned to a voice (an author) and to a particular time. Because of limits of space, in this work we do not provide further information about these aspects.

³ Actually, there are 5 labels in this proposal since some cases cannot be labelled with regard to factuality.

Next, we present other decisions made in the annotation task: in Section 3.1, the rationale behind the selection of the predicates to be annotated, and in Section 3.2, the linguistic foundation of our proposal.

3.1. Selection of predicates

As a starting point, all situations that are expressed through verbs⁴ are considered for annotation. In the manual phase, the predicate's arguments (including the subject) and the adjuncts were delimited so that the certainty value is assigned not only to the verb but to all the elements in the scope of the predication. This procedure is expected to be maintained in the automatic phase, but it will obviously depend on the quality of the parser available.

Whenever a compound tense, a passive construction or a verb periphrasis is found, the element annotated is the verbal set. Modality expressions such as *es probable que* ('it is probable that') have also been treated as a single unit.

Predicates that will not be dealt with describe unreal situations and are marked with the label *NA* (*no aplica* 'it does not apply')⁵. In order to make this distinction, we take into account the type of world represented in the statement. Two types of worlds are distinguished. The first one (A) describes reality. We further contemplate three subtypes that we present next. All situations that belong to this world will be annotated in relation to factuality (see Section 4).

- i. Situations presented by the author as certain (facts); for example, present or past situations (1).
 - (1) Los resultados de este estudio *se presentaron* en la reunión anual⁶.
 "The results of this study *were presented* at the annual conference".
- ii. Situations whose truth has been agreed upon in science or by a particular culture. These are the so-called absolute truths (2) (see Section 4.1). Even though some authors claimed the convenience of identifying these situations independently, to our knowledge, only Tonelli et al. (2014) annotate them with a specific tag.
 - (2) ...el planeta Tierra *es* redondo...
 "...planet Earth *is* round..."
- iii. Counterfacts, that is, situations that have not happened (3a), and negative absolute truths (3b) are also part of this world.
 - (3) a. La presidenta madrileña (...) *no ha comparecido* ante los medios.
 "The president of Madrid (...) *has not appeared* before the media".
 - b. Es importante marcar un límite, ya que *no es sano* vivir en constante presión...

⁴ Deverbal nouns are not considered for annotation (*La elección fue buena* 'The election was correct').

⁵ This solution is also adopted in other projects, for example in FactBank.

⁶ Examples come from the TAGFACT corpus. We underline the predicate in each sentence but we do not mark the scope of the predication. If there is a word in italics, it is a linguistic clue (see Section 3.2).

“It is important to set a limit, since it *is* not healthy to live in constant pressure...”

The second type of world (B) refers to situations presented by the author as uncertain. In (B), we also contemplate three subtypes that we develop below. Only the first two will be annotated about factuality and the third is classified as *NA*.

- i. Present or past situations presented as uncertain because the author does not know if they have happened, regardless their polarity. These situations can be real or not.

(4) *Puede que haya sido* objeto de una injusticia histórica.

“He/she *may have been* the subject of historical injustice”.

- ii. Future situations expressed as scheduled (5a), intended or planned (5b), promised (5c) or predicted (5d). The author considers that these situations will take place.

(5) a. El COI reveló este lunes (...) que los Juegos Olímpicos de la Juventud de verano de 2022 *se celebrarán* en Dakar ...

“The IOC revealed on Monday (...) that the 2022 Summer Olympic Games *will be held* in Dakar...”

b. La segunda planta (...) *estará* dedicada a bar de copas...

“The second floor (...) *will be* dedicated to a cocktail bar...”

c. ...el también vicepresidente del Gobierno italiano (...) ha dicho: ‘Las zonas francas (...) *ya no serán toleradas*.’

“...the also vice-president of the Italian Government (...) has said: Free zones (...) *will no longer be tolerated*”.

d. El PP *no va a ganar* las elecciones....

“PP *will not win* the elections...”

- iii. Situations, placed at any temporal moment, which are uncertain because they express wishes (6a), needs (6b), permission (6c), obligation (6d), prohibition (6e), attempts (6f), suggestions or advice (6g) or future possibility (6h). These situations belong to an unreal world⁷.

(6) a. ...son muchos los que *desean entrar* al país...

“...there are many who *wish to enter* the country...”

b. Señaló que el sector *necesita desarrollar* trabajo...

“He pointed out that the sector *needs to develop* work...”

c. El grupo *tenía permiso* para *escalar* el pico oriental de Nanda Devi.

“The group *was allowed to climb* the eastern peak of Nanda Devi”.

d. Concluyó el chef añadiendo que “*tenemos que estar* mucho más pendientes del tema de salud...”.

⁷ Some of the examples below contain more than one verb but, sometimes, not all of them are part of the same annotation set: *desean* in (6a) will be annotated with a factual value since it describes a real situation (world A) whereas the infinitive will be discarded (NA) because belongs to world B-iii.

“The chef concluded adding that “we *have to be* much more aware of health issues...”

- e. El conductor, en principio, se negó porque la empresa *prohíbe* que vayan acompañados de personal ajeno a el servicio, pero luego cambio de criterio.

“The driver, in principle, refused because the company *prohibits* them from *being* accompanied by non-service personnel, but then changed his mind”.

- f. Estos videos *tratan* de *mostrar* evidencias que demuestren que la tierra no es redonda.

“These videos *try to show* evidence that proves that the earth is not round”.

- g. Las asociaciones de consumidores *plantan* a los usuarios bancarios que *actúen* con precaución.

“Consumer associations *are asking* bank users to *act* with caution”.

- h. Si *no recibe* respuesta inmediata, debe acudir a los tribunales.

“If you *do not receive* an immediate answer, you must go to court”.

3.2. Linguistic clues

This project aims to automatically tag the degree of certainty with which a situation is narrated. The program will be based on linguistic knowledge, that is, the presence or absence of some linguistic clues (called triggers) will be used to determine the factual status of a clause.

In the TAGFACT project, we only use the information contained in the sentence (not in the rest of the text or the context). More specifically, in simple sentences, only the linguistic information found in these sentences is used to identify the clues taken into account to decide the factual annotation. If it is a complex sentence (coordinated or juxtaposed sentence), each clause is analyzed in isolation, since, in general, the linguistic clues of one sentence should not affect the other. In compound sentences, when there is a dependent clause, the clues can also be found in the main sentence.

For example, in relative clauses the factual status of the verb is completely independent of the main clause. In other cases, the influence of the main verb is clear, as in infinitive clauses: if the main verb belongs to the semantic class of verbs of desire the subordinated clause is not annotated (6a), but if the subordinating clause is a temporal infinitive clause (*antes de ir a Japón militó en el Liverpool* - “before going to Japan, he played in the Liverpool football team”) it is annotated.

Once the boundaries to identify the linguistic clues have been established, an important task is how to differentiate the predicates that will be annotated from those that will not, since, as mentioned above, unreal situations (B-iii), e.g. orders, wishes or hypothetical conditions, are discarded. Thus, in order to detect sentences that describe unreal situations, it will be necessary to identify certain classes of verbs, such

as desire verbs (6a) or modal verbs, especially deontic modals like *tener que* ('have to' + INF) (6d)⁸.

In addition, some subordinate clauses, such as purpose or conditional clauses, are also mostly associated with hypothetical or unreal situations. In both cases, the presence of certain linguistic elements combined with the use of specific tenses (*para* 'to' + INF / *que* 'that' in the case of purpose clauses and *si* 'if' + PRES IND / IMP SUBJ in the case of conditionals) will be considered the elements triggering the interpretation.

For the situations in which factuality is relevant (real world —A— and uncertain world —Bi, Bii—), they are annotated with respect to reference time, polarity (positive or negative) and the degree of certainty expressed (certain or uncertain). For the annotation of temporal information, the tense is the basic clue, but other aspects, such as temporal phrases, must also be considered since under some conditions tenses do not refer to the corresponding reference time (in (10a), the present tense is used for future time, and *esta noche* 'tonight' is the clue).

Regarding negative polarity, the adverb *no* (see example (2) above), but also pronouns or indefinite adjectives or adverbs of negation (e.g. *ningún* 'any', *nunca* 'never'), must be considered. In addition, some verbs lexically imply negative polarity of the situation described in the subordinate clause (*evitar* 'avoid' (7)).

- (7) 43 personas que pasaron por aquí *evitaron desarrollar* un cáncer de piel.
 "43 people that came by *avoided to develop* skin cancer".

Other elements are used to identify the degree of certainty, such as modal elements. Epistemic modality presents a direct relationship with factuality since it implicitly shows the absence of complete trust in the truth value of the statement. Generally speaking, in Spanish possibility or probability can be expressed through verbal periphrases (*poder*+INF 'can+INF') or adverbs or expressions such as *quizás*, *puede que*, *seguro que*... maybe, it might be the case that, to be sure that" (see (4) above). On the contrary, other types of verbal periphrases, such as non-epistemic periphrases, are usually associated with certainty (*acabar de*+INF 'have just+PP', as in (8a), *llegar a*+INF 'even +VERB', as in (8b)).

- (8) a. El grupo Flat Waves *acaba de publicar* el videoclip de la canción 200 veces.
 "The group Flat Waves *has just released* the music video for the song 200 times".
 b. El propio Pacheta *llegó a decir* que se asemejaba al Messi de Segunda División.
 "Pacheta himself *even said* that he resembled the Second Division Messi".

As said above, some verbs determine the factual value of their subordinate clause. For example, factive verbs, such as *alegrarse* 'be glad' (9a) o *lamentar* 'regret' (9b), determine the value of the situation over which they predicate. Opinion verbs, on the

⁸ Sometimes modal verbs are polysemous, for example *deber* 'must' + INF that expresses obligation or probability. The automatic identification of these senses is not trivial.

other hand, always introduce non-factual situations since they express opinions, beliefs, etc. (see (9c)).

- (9) a. Este mismo miércoles cinco inquilinas *se alegraban* de *disponer* ya de luz.
 “This Wednesday, five tenants *were happy* to have light”.
- b. ...el ultraderechista *lamentó* nuevamente que el militar no *hubiese sido capturado* en Indonesia.
 “...the far-right (politician) *regretted* again that the military *had not been captured* in Indonesia”.
- c. ...su familia *cree* que *salió* con un saco de dormir.
 “...his family *believe* that she *took* a sleeping bag with her”.

When operators, such as negative or epistemic elements, are used in the main clause, the scope of such elements may vary in relation to subordinating clauses depending on their position and combination.

All these formal markers help the reader to understand the author’s perspective regarding factuality. Nevertheless, exceptionally world knowledge has been used to identify absolute truths. On this matter, lexical or morphological analysis of verb tenses, for example, is not always effective since a tense can denote more than one constructional meaning. For instance, in Spanish present simple can be used to describe a future event, as in (10a), or an absolute truth, as in (10b).

- (10) a. ...De hecho, Panamá *juega esta noche* contra Jamaica.
 “...In fact, Panama *is playing* Jamaica tonight”.
- b. El niño *juega* y, mientras *juega*, no se hace grandes reflexiones filosóficas y sociales.
 “The child *plays* and, while *playing*, he doesn’t make great philosophical and social reflections”.

4. LABELS IN THE ANNOTATION SCHEME

Our model takes into consideration four aspects: eventual types, the writer commitment to the certainty of the assertion, polarity and reference time. The first two concepts are discussed in Sections 4.1 and 4.2, respectively. The other two concepts go beyond the scope of the present work. In any case, the tags used in TAGFACT for polarity are POS, for affirmative sentences, and NEG, for negative sentences. Regarding reference time, in our project we differentiate between present (PRES), past (PAST) and future (FUT) situations.

4.1. Dynamism and types of situations

In our model, we differentiate two eventual types distinguishing between dynamic situations and non-dynamic situations, following Vendler (1967). From this first classification, we further distinguish other relevant features, as shown below.

Dynamic situations are characterized because they progress over time. Under this label, both events and processes are considered. We have further distinguished between

physical (11) and mental situations (12), also including in this latter group the semantic class of perception verbs, since this difference has proven to be relevant to the assignment of factual information. The former type is annotated with the tag EVENT and the latter with the tag (suggestion;or check sentence) MENTAL.

- (11) Tres bomberos (...) *fueron arrestados* cuando *iban* en la barca de una ONG...
 “Three firefighters (...) *were arrested* while on the boat of an NGO...”
- (12) el hecho de que haya llegado a juicio les *preocupa* ...
 “...the fact that he’s come to trial *worries* them...”

Non-dynamic (stative) situations cover those statements in which a property of an entity is described. Following Fernández et al. (2002), further subclassifications have been made. First, we have considered those states that relate a property to an entity (13) (PROPERTY). In this group, we have included sentences with lexically stative verbs (13a), copulative (13b) or pseudocopulative verbs (13c) and participle clauses that denote a state (13d).

- (13) a. ...dijo que este país del oeste de África *tiene* un “plan social y económico emergente”.
 “...he said that this West African country *has* an “emerging social and economic plan”.
- b. Pensaban que todo *era* un error, pero la denuncia siguió adelante.
 “They thought it *was* all a mistake, but the complaint went ahead”.
- c. ... *acabaron* convencidos gracias a los argumentos presentados en Youtube.
 “...they *ended up* convinced by the arguments presented on Youtube”.
- d. Carles Font, (...), *emocionado* ante la comisión de investigación del Parlamento, ...
 “Carles Font, (...), *moved* before the Parliament’s committee of inquiry...”

Another type of state that has been identified in the scheme is *absolute truths* (ABS-TRUTH PROPERTY). Understandably, this class of properties is not annotated with respect to the author’s commitment (see Section 4.2). Strong scientific claims (2) are found within this subclass, together with general advice or wisdom (14a), cultural beliefs (14b), proverbs and fixed expressions (14c).

- (14) a. ...es bueno llevar en el coche algún tipo de tela que puedas tirar al suelo...
 “...it *is* a good idea to carry in the car some kind of fabric that you can throw to the ground...”
- b. Pocas veces me afecta, porque uno *se acostumbra* a todo.
 “It rarely affects me, because you *get used* to everything”.
- c. ...en mi tierra natal hay un refrán que dice: “sarna con gusto no *pica* ”...
 “...in my homeland there is a saying: “A burden of one’s own choice *is not felt* ”...”

Lastly, other non-dynamic situations identified in our corpus describe properties that could be seen as eventive, since they refer to series of actions repeated over time

or to an action susceptible of being repeated in the future (EVENT PROPERTY). Within this type of repeated eventive properties we have included, first, programmed repeated (habitual) actions, traditions or rules (15) and actions considered characteristic of an entity (middle constructions), being it an individual (16a), an object (16b)⁹ or a group or society (16c)¹⁰.

- (15) ...a casa de Lidia *acude* una persona del Ayuntamiento *todos los lunes* ...
 "...A person from the City Council *comes* to Lidia's house *every Monday*..."
- (16) a. ...el chico es guapo, *canta* bien e hizo bailar y cantar a los espectadores.
 "...the boy is handsome, *sings* well and made the spectators dance and sing".
- b. ...este material *se estropea fácilmente* si no se cuida de forma adecuada.
 "...this material *is easily damaged* if not properly taken care of".
- c. Las ruedas de prensa son parte del trabajo del presidente y en EEUU *se toman muy en serio*.
 "Press conferences are part of the president's work and in America they *are taken very seriously*".

4.2. The degree of commitment with the situation

Regarding the labels used to annotate situations, we have opted to follow Diab et al.'s (2009) terminology (COMMITMENT vs. NON COMMITMENT), since we believe it better reflects the fact that the evaluation of a situation is not objective but always reveals the author's stance. In contrast, the terms *certainty / non-certainty*, proposed by Saurí (2008), could be interpreted as the representation of the truthfulness of the narrated facts. Something similar could be said about the use of the labels *factuality - non factuality* (Van Son et al. 2014), or *performed vs. non performed* (Wonsever et al. 2008, 2016).

The tag COMMITMENT is used for situations related to the real world (see (1) and (3) above), and also future situations presented as certain (predictions and future events planned or subjected to a schedule) (5). On the other hand, situations annotated as NON COMMITMENT correspond to present or past sentences depicted as uncertain by the author (4), and also to those describing future events that are also presented with uncertainty (17).

- (17) ...*creo que* el mercado *probablemente seguirá bajando*...
 "...I *think that* the market *will probably continue* to fall..."

In our model we have also established the use of another label, QUALIFIED COMMITMENT, for those situations in which the author emphasizes his/her commitment towards the certainty of what is being said. We use this label for present (18a), past and

⁹ This construction is called *construcción media* (middle construction), following Mendikoetxea (1999).

¹⁰ This construction, in this case, is known as *pasiva refleja* (reflexive passive).

future situations (18b). We believe that this type of commitment has to be categorized differently because if the author has chosen to emphasize a particular event there must be a reason. We understand that this emphasis, somehow, hides the recognition of an event being arguable and, therefore, in some way, it could be considered an opinion, probably a very strong opinion:

- (18) a. *...me apuesto lo que quieras a que el porcentaje de muertos es muchíííí-simo mayor en alpinismo que en carreras de motos.*
 "...I bet whatever you want that the percentage of the dead is a lot higher in mountaineering than in motorcycle racing".
- b. *...y seguro que agradecerán el tiempo de descanso que les otorga el pase directo a cuartos de final.*
 "...and I am pretty sure they will appreciate the rest time that the direct access to quarterfinals gives them".

5. CONCLUSIONS

In this paper, we have presented the scheme defined within the project TAGFACT for the annotation of factuality, understanding it as the degree of commitment of an author regarding the certainty of the facts he or she is narrating. This annotation is conceived compositionally, based on the combination of four aspects: reference time, polarity, the eventual type of the situation and the degree of commitment. Although it is a model created for the Spanish language, we believe it can be extrapolated to other languages.

Our proposal integrates a new aspect in the annotation of factuality since it incorporates information about event types. More specifically, it contributes with three innovations: the distinction of dynamic situations (processes-events) from non-dynamic ones (properties), the identification and annotation of absolute truths and the distinction between prototypical and eventual properties (habitual and middle constructions among others). Furthermore, all the predicates that are not considered for the annotation of their factual status are also specified.

In order to carry out the annotation, contextual linguistic information has been used and the boundaries considered in this work have been established. The next step in the project is to create an automatic labeling tool based on these criteria and all the information collected in the manual annotation of the Gold Standard. To our knowledge this resource will be the only tool based exclusively on linguistic knowledge for Spanish.

REFERENCES

- COLOMER, J. P., K. LAI and O. RAMBOW (2016) "Detecting level of belief in Chinese and Spanish". In E. BLANCO, R. MORANTE and R. SAURÍ (eds.) *Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics (ExProM)*, Osaka: The COLING 2016 Organizing Committee, 22-30.
- DIAB, M. T., L. LEVIN, T. MITAMURA, O. RAMBOW, V. PRABHAKARAN and W. GUO (2009) "Committed belief annotation and tagging". In *Proceedings of the Third Linguistic Anno-*

- tation Workshop Suntec, Singapore: Association for Computational Linguistics, 68-73
- FERNÁNDEZ, A., G. VÁZQUEZ and M.A. MARTÍ (2002) "Alternancias diatéticas relacionadas con el aspecto", *Verba*, 29, 389-402.
- LEE, K., Y. ARTZI, Y. CHOI and L. ZETTEMAYER (2015) "Event detection and factuality assessment with non-expert supervision". In L. MÁRQUEZ, C. CALLISON-BURCH and J. SU (eds.) *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon: Association for Computational Linguistics, 1643-1648.
- LIGHT, M., X. YING QIU and P. SRINIVASAN (2004) "The language of bioscience: Facts, speculations, and statements in between". In L. HIRSCHMAN and J. PUSTEJOVSKY (eds.) *HLT-NAACL 2004 Workshop: BioLINK 2004, Linking Biological Literature, Ontologies and Databases*. Boston: Association for Computational Linguistics, 17-24.
- MATSUYOSHI, S., M. EGUCHI, C. SAO, K. MURAKAMI, K. INUI and Y. MATSUMOTO (2010) "Annotating event mentions in text with modality, focus, and source information". In N. CALZOLARI, K. CHOUKRI, B. MAEGAARD, J., MARIANI, J. ODIJK, S. PIPERIDIS, M. ROSNER and D. TAPIAS (eds.) *Proceedings of the International Conference on Language Resources and Evaluation*, LREC 2010. Valetta: ELRA, 1456-1463.
- MEDLOCK, B. AND T. BRISCOE (2007) "Weakly supervised learning for hedge classification in scientific literature". In *Proceedings of the 45th Association for Computational Linguistics*. Praga: Association for Computational Linguistics, 992-999.
- MENDIKOETXEA, A. (1999) "Construcciones con se: Medias, pasivas e impersonales". In I. BOSQUE and V. DEMONTE (dirs) *Gramática descriptiva de lengua española*, vol. 2. Madrid: Espasa-Calpe, 1631-1721.
- MINARD, A., SPERANZA, M., SPRUGNOLI, R., and CASELLI, T. (2015) "FactA: Evaluation of Event Factuality and Temporal Anchoring". In C. BOSCO, S. TONELLI AND F. MASSIMO ZANZOTTO (eds.) *Proceedings of the Second Italian Conference on Computational Linguistics*, Academia University Press, Turin: Accademia University Press, 187-192.
- MINARD, A. L., SPERANZA, M., and T. CASELLI (2016a) "The EVALITA 2016. Event Factuality Annotation Task (FactA)". In P. BASILE, F. CUTUGNO, M. NISSIM, V. PATTI and R. SPRUGNOLI (eds.) *EVALITA. Evaluation of NLP and Speech Tools for Italian*. Torino: Academia University Press, 32-39.
- MINARD, A., M. SPERANZA, R. URIZAR, B. ALTUNA, M. VAN ERP, A. SCHOEN and C. VAN SON (2016b) "MEANTIME, the NewsReader Multilingual Event and Time Corpus". In N. CALZOLARI, K., CHOUKRI, T. DECLERCK, S. GOGGI, M. GROBELNIK, B., MAEGAARD, J. MARIANI, H. MAZO, A. MORENO, J. ODIJK and S. PIPERIDIS (eds.). *Proceedings of LREC 2016*. Portoroz: ELRA, 4417-4422.
- MONTAGUE, R. (1968) "Pragmatics". In R. KLIBANSKI (ed.) *Contemporary Philosophy*. Florencia: La Nuova Italia Editrice, 102-121.
- NARITA, K., J. MIZUNO and K. INUI (2013) "A lexicon-based investigation of research issues in Japanese factuality analysis". In R. MITKOV Y J. C. PARK (eds.) *Proceedings of the Sixth International Joint Conference on Natural Language Processing*. Nagoya: Asian Federation of Natural Language Processing, 587-595.
- PRABHAKARAN, V., T. BY, J. HIRSCHBERG, O. RAMBOW, S. SHAIKH, T. STRZALKOWSKI, J. TRACEY, M. ARRIGO, R. BASU, M. CLARK, A. DALTON, M. DIAB, L. GUTHRIE, A. PROKOFIEVA, S. STRASSEL, G. WERNER, J. WIEBE and Y. WILKS (2015) "A new dataset and evaluation for belief/factuality". In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*. Denver, Colorado: Association for Computational Linguistics, 82-91.
- SAURÍ, R. (2008) *A Factuality Profiler for Eventualities in Text*. PhD dissertation. Brandeis University.
- SAURÍ, R., and J. PUSTEJOVSKY (2009) "FactBank: A corpus annotated with event factuality",

- Language Resources and Evaluation*, 43(3). 227-268.
- SONI, S., MITRA, T., GILBERT, E., and EISENSTEIN, J. (2014) "Modeling factuality judgments in social media text". In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Baltimore, Maryland: Association for Computational Linguistics, 415-420.
- TIANXIONG, H., L. PEIFENG and Z. QIAOMING (2017) "Identifying Chinese Event Factuality with Convolutional Neural Networks". In Y. WU, J-F HONG and Q. SU (eds.) *Chinese lexical semantics*. Lecture Notes in Artificial Intelligence. 18th Workshop, CLSW, Leshan, China. Springer, 284-292
- TONELLI, S., SPRUGNOLI, R., and M. SPERANZA (2014) "NewsReader Guidelines for Annotation at Document Level". In *Extension of Deliverable D3. Technical Report NWR-2014-2*. Trento.
- VAN SON, C., VAN ERP, M., FOKKENS, A., and P. VOSSEN (2014) "Hope and fear: Interpreting perspectives by integrating sentiment and event factuality". In N. CALZOLARI et al. (eds.) *Proceedings of the 9th Language Resources and Evaluation Conference*. Reykjavik: ELRA, 26-31.
- VELUPILLAI, S. (2011) "Automatic classification of factuality levels. A case study on Swedish diagnoses and the impact of local context". In A MOEN, S.K. ANDERSEN, J. AARTS and P. HURLEN (eds.) / *User Centred Networked Health Care Proceedings of MIE 2011*. / Amsterdam, Berlin, Tokyo, Washington D.C.: IOS Press, 559-563.
- VENDLER, Z. (1967) *Linguistics in Philosophy*. Ithaca, NY: Cornell University Press.
- WITTGENSTEIN, L. (1994) *Tractatus Logico-philosophicus* (E. T. GALVÁN, Trad.). Madrid: Alianza.
- WONSEVER, D., MALCUORI, M. and ROSÁ, A. (2008) *SIBILA: Esquema de anotación de eventos*. Reportes Técnicos RT 08-11. Universidad de la República, Uruguay.
- WONSEVER, D., A. ROSÁ and M. MALCUORI (2016) "Factuality annotation and learning in Spanish texts". In N. Calzolari, et al. (eds.) *Proceedings of LREC 2016*, Portoroz: ELRA, 2076-2080.