

1 Ana Fernández and Gloria Vázquez*

2 **The SenSem Corpus: an annotated corpus**
3
4 **for Spanish and Catalan with information**
5 **about aspectuality, modality, polarity and**
6 **factuality**
7

8
9 **Abstract:** In this paper, we present the annotation scheme used in the SenSem¹
10 corpora (SSC), for Spanish and Catalan, to codify information regarding aspectual-
11 ity, modality, polarity and factuality. As regards aspectuality, the most relevant
12 contribution is the codification of information about dynamicity, telicity and iter-
13 ativity. Regarding factuality, we present a more fine-grained annotation of uncer-
14 tainty as applied to the identification of impossible events, completely uncertain
15 events and neutral uncertain events. Although information about factuality in
16 Spanish has been provided elsewhere, the Catalan SSC is the only corpus to do so
17 for Catalan.
18

19 **Keywords:** corpus annotation, aspectuality, modality, polarity, factuality, asser-
20 tivity, certainty, impossibility, dynamicity, telicity
21

22
23 ***Corresponding author: Gloria Vázquez:** Universitat de Lleida, English and Linguistics,
24 Lleida 25003, Spain. E-mail: gvazquez@dal.udl.cat

25 **Ana Fernández:** Universitat Autònoma de Barcelona, English and German Barcelona, Spain.
26

27
28 **1 Introduction**
29

30 The Spanish and Catalan SenSem Corpora (SSC)² are made up of approximately
31 half a million annotated words: 455,905 and 391,267, respectively. The Spanish
32 corpus (Fernández et al. 2006) contains 30,365 sentences (25,075 extracted from
33 the journalistic register and 5,299 from the literary). The Catalan corpus (Vázquez
34 et al. 2013) does not contain sentences from the literary register.
35

36
37 **1** Acronym for Sentence Semantics.

38 **2** These corpora are the result of the work carried out by the research team members for the last
39 nine years in five different projects (2004–2012). The last is Standardization and Transference of
40 Lexical and Textual Resources – Ministerio de Ciencia e Innovación – FFI2011-27774.

These two SSC have been annotated with linguistic information regarding the lexical, phrasal and sentential levels. First, words were annotated with morpho-syntactical tags using FreeNet (Atseries et al. 2006).³ This was the only completely automatic annotation process applied. Verbs were also disambiguated with respect to the sense they exemplify. This was accomplished by assigning each verb a sense from our lexical database, the corresponding WordNet sense and information about their *Aktionsart*. The next level of annotation was the phrasal level. Each phrase was marked with respect to first its syntactic category, and then its semantic and syntactic function. At sentential level, taking into account the type and order of participants in each clause, we created a hierarchy of constructions related to argument structure. This hierarchy includes agentive, causative, passive, anticausative, reflexive and reciprocal constructions, among other constructions. At sentential level, we also annotated some other information related to aspectuality as well as modality, polarity and factuality. In this paper we will deal with the annotation scheme applied to these last four features.

During the years we have been working in this project, 5 trained linguists have manually annotated the sentences following the guidelines (Vázquez et al. 2005) proposed. In (Alonso et al. 2007) the score for the evaluation of inter-annotator agreement is presented.

The corpora themselves are available online.⁴ The online interface allows users to browse several linguistic phenomena at once in a significant number of sentences for just one verb or a set of verbs. The corpora can also be downloaded in XML format.

The main contributions of this proposal are, for aspectuality, the codification of information about dynamicity, telicity and iterativity, and regarding factuality, a more fine-grained annotation of uncertainty as regards the identification of impossible events, completely uncertain events and neutral uncertain events. The Catalan SSC is the only corpus covering information about factuality. Prior to this work, only the SIBILA corpus (Wonsever et al. 2008) covered factuality for Spanish.

Next, we present the state of the art with regard to corpora containing information on aspectuality, modality, polarity or factuality. In Section 3, we present a general description of how aspectual information has previously been annotated. In Section 4, we present the annotation of modality, polarity and factuality as applied to the SSC. Section 5 presents some results and Section 6 is devoted to

³ Currently, this information is not visualized in the user's interface.

⁴ <http://grial.uab.es/sensem/corpus>

1 conclusions. Finally, in the appendix we present a summary of the tags used to
2 annotate information regarding aspectuality, modality, polarity and factuality.

3

4

5

6 **2 Related works**

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

2.1 Corpora annotated with temporal and aspectual information

As highlighted in Feldman and Arshavskaya (2007), it is common to find corpora annotated with information concerning verb tense and mood in relation to morphological information. However, with the exception of English corpora, there are a few corpora annotated with more specific temporal and aspectual information.

Interest in annotating temporal information in corpus linguistics began in the last decade with the objective of improving results in the field of information extraction. Nowadays the most comprehensive and ambitious scheme annotation is TimeML (Saurí et al. 2006), which includes dates, times, temporal relations and some aspectual information such as phases, mode, progressivity or (im)perfectivity. However, it does not include information about dynamicity, telicity and iterativity. This annotation scheme also covers aspects related to the annotation of modality and polarity. Two corpora have been annotated with this language: TimeBank 1.2 and AQUAINT TimeML (Pustejovsky et al. 2006). Versions of TimeBank in both Spanish and Catalan are currently available (Saurí 2010). TimeBank covers a total of 68,000 words.

2.2 Corpora annotated with information about modality, polarity and factuality

Several projects in which aspectuality and temporality are accounted for also take an interest in questions related to the description of modality, since all these theoretical aspects of language description are closely related. Similarly, modality and polarity are also related, so it is also common that a corpus that seeks to address one of these issues also needs to address the other. By modality we refer to the point of view of the speaker in relation to the degree of certainty about the events described (i.e., factuality). This type of annotation is currently one of the most innovative issues in the field of corpus linguistics.

The distinction between factive and non-factive events is related to aspectuality, polarity and modality and it is crucial to extract “real” facts from texts (Saurí et al. 2006, Baker et al. 2010 and Hendrickx et al. 2012). Nevertheless, the direct annotation of factuality is not common.

In TimeBank, the annotation of polarity is very basic, being either positive or negative. Modality is not specified comprehensively either and, in many cases, factuality is not indicated at all. FactBank was therefore created to make up for these shortcomings (Saurí and Pustejovsky 2009). In FactBank, about 10,000 events (belonging to the same 208 documents annotated in TimeBank) were manually annotated with information about factuality such that an event can be true (or not), possible (or not) or probable (or not). In addition, the type of source is indicated, that is, whether the narrator was directly the source or instead the event is being reported indirectly by another source (another person, a newspaper, etc.). Another corpus project in which exceptionality factuality is directly tagged, also manually, is SIBILA.

In addition, Vincze et al. (2008) and Hendrickx et al. (2012) have recently put forward new proposals in the field of the annotation of polarity and modality. Thanks to the existence of corpora already annotated with this information, automatic annotation of these phenomena has recently started. Results in this area are still not very good but efforts are being made in this direction (Morante and Daelemans 2012).

3 Aktionsart, aspect and aspectuality in SenSem

3.1 Lexical and phrasal levels

With regard to lexical aspect (*Aktionsart*), the traditional categories (Vendler 1957) are *state* (for non-dynamic situations) and *process*, *accomplishment* and *achievement* (for dynamic situations). In our proposal (Coll 2007), we use 4 tags: we keep the labels for *states* and *processes*; accomplishments and achievements are grouped into just one general category, *event*; and finally, we have added a fourth type, *process-event*, to indicate those cases where telicity is not lexically defined, following the commonly accepted idea that telicity is a lexical property that may be modified by certain elements in the sentence, specifically the type of object (phrasal level) or the verb tense.

An example would be *analizar* (1) (‘analyze’): in (1a) the object (underlined) triggers a bounded interpretation of the event whereas (1b) is an unbounded process.

- 1 (1) a. Se *analizaron* uno por uno los candidatos posibles.
 2 ‘All possible candidates were analyzed one by one.’
 3
 4 b. (...) los ingenieros (...) analizarán todas las semanas programas y soft-
 5 wares existentes en la red.
 6 ‘The engineers (...) will analyze programs and software on the net every
 7 week.’

8 Thus, the codification of *Aktionsart* is inherited from our lexical database and
 9 maintained through the other phases of annotation, except in the case of verbs
 10 annotated as *process-event*. These verbs, like *analizar* ‘analyze’ are disambiguated
 11 at phrasal level as being either bounded (1a) or unbounded (1b).

12 Other predicates that are also specified lexically as “process-event” are cer-
 13 tain verbs of movement, such as *viajar* (‘travel’) or *perseguir* (‘follow’). When they
 14 are used with a PP that limits the end of the path, they are interpreted as events
 15 (2a). If they are used without this PP they are interpreted as processes (2b).
 16

- 17 (2) a. (...) un ciudadano que *viaje* a diario desde Terrassa hasta Plaza Catalunya
 18 de Barcelona (...).
 19 ‘(...) a citizen who travels daily from Terrassa to Plaza Catalunya in Barcelona
 20 (...)’
 21
 22 b. La policía encontró en el coche que *perseguían* un saco (...).
 23 ‘The police found in the car they were following a bag (...).’
 24
 25

26 3.2 Sentential level

27
 28 At sentential level we also codify three other pieces of information: a)
 29 im(perfectivity), b) the temporary/permanent nature of stativity and c) iterativity.
 30 The labels used are:

- 31 a) *Perfective vs. Imperfective*
 32 b) *Permanent State vs. Temporary State*
 33 c) *Habitual*
 34

35 First, the element that contributes to determine the (im)perfective nature of a
 36 construction in Spanish/Catalan is verb morphology, since, in these languages,
 37 verb tense, mode and aspect are expressed by verbal inflection (2b). Certain aux-
 38 iliary verbs may express (im)perfectivity too.

39 Second, we distinguish those states that are permanent (3a) from those that
 40 are temporary and, therefore, denote a reversible state (3b).

- (3) a. Entre 40 y 45 personas (. . .) *cabén* en una patera tradicional. 1
 ‘Between 40 and 45 people (. . .) *fit* in a traditional fishing boat.’ 2
 b. La señora *se encontraba* mejor (. . .) 3
 ‘The lady of the house *was feeling* better (. . .).’ 4
 5

As regards the third element, iterativity, we annotate as *habitual* those sentences that denote repeated actions and are, therefore, not related to *specific* time-space coordinates. There are numerous linguistic elements that may contribute to this interpretation. For instance, in (4) there are two habitual events (*gastar* ‘consume’ and *crecer* ‘grow’). The key elements that point to a habitual interpretation for this sentence are the subject (*Cada catalán* ‘Every Catalan’, with distributive interpretation), the use of an imperfective tense (present), and the reference to the repetition of years (*al año* ‘annual’).

- (4) *Cada catalán* *gasta* 3,3 toneladas de petróleo *al año* y la demanda *crece* un 3 % *anual*. 15
 ‘*Every Catalan* *consumes* 3.3 tons of petroleum a year and the demand is *growing* by 3% *every year*’. 16
 17
 18
 19
 20
 21

4 Modality, polarity and factuality in SenSem 22

There are four basic types of epistemic modality described in the literature: (a) 24
 certainty, (b) non-certainty or counterfactuality, (c) possibility or probability, and 25
 (d) impossibility or improbability. The former two categories refer to what is cer- 26
 tain (what really happens/has happened or does not happen/has not happened, 27
 4.1), while the latter two refer to what is uncertain (facts are presented as uncer- 28
 tain, 4.2). In SenSem we identify these two categories with the tags *assertive* and 29
non-assertive. Deontic modality has also been tagged as *non-assertive*. Sentences 30
 have also been annotated with information about polarity, *positive* or *negative*. 31
 The annotation of factuality derives from the information about modality and po- 32
 larity, as we will explain below. 33
 34
 35

4.1 Factuality vs. non factuality/counterfactuality 36

Within the group of assertive sentences we further differentiate positive assertive 38
 constructions that express certain or factive events (5) from assertive events with 39
 negative polarity, that is, counterfactuals (6): 40

- 1 (5) (...) Oftaláser (...) *ha invertido* 110 millones en un nuevo servicio de cirugía
 2 refractiva.
 3 ‘(...) Oftaláser (...) *has invested* 110 million [euros] in a new refractive sur-
 4 gery service.’
 5 (6) El Athletic quedó noqueado y no se recuperó tras el descanso.
 6 ‘Athletic [Football Club] was knocked down and did not recover after the
 7 break.’
 8
 9

10 4.2 Annotation of non-factuality

11
 12 In contrast to assertivity, polarity does not affect non-assertive sentences. In the
 13 literature, we find different subcategories within the field of non-assertivity that
 14 indicate degrees of uncertainty. In SenSem we have made three distinctions,
 15 some of which are new in the field of corpora annotation.

16 First, following proposals such as Ahern (2008) and Wonsever et al. (2008),
 17 we further differentiate two sub-categories within the field of doubt (i.e., non-
 18 assertivity). We have annotated the cases of non-assertivity when it refers to *past*
 19 *or present* events (7) differently than when it refers to *future* events (8):
 20

- 21 (7) Es posible que, inicialmente, *hayan beneficiado* a los países desarrollados.
 22 ‘They may initially *have benefited* the developed countries.’
 23
 24 (8) Quizá me *destinarán* fuera de aquí.
 25 ‘Maybe I *will be sent away*.’
 26

27 This distinction, not present in FactBank but present in SIBILA, is important in
 28 order to make inferences. In (7), uncertainty is based on ignorance regarding
 29 whether the facts are certain or not; that is, facts are presented as uncertain, re-
 30 gardless of whether they might have truly occurred or might not have, so that
 31 their degree of certainty is actually unknown. By contrast, the second type of
 32 uncertainty (8) might be called absolute uncertainty because we are describing
 33 something that belongs to the future.

34 Second, we have annotated *impossible events*, a tag used in neither FactBank
 35 nor SIBILA, to label past or present situations presented as completely unreal (9),
 36 because the interpretation is that the situation described never happened and
 37 therefore we consider them intrinsically impossible (Morante and Daelemans 2012).⁵
 38

39 _____
 40 ⁵ In terms of factivity, these are counterfactuals.

This is not the case for future events, because, even when they are presented as virtually impossible (10), they are not intrinsically impossible since they are still outside the scope of what is real. Again, this distinction is important when it comes to making inferences.

(9) Si el Barcelona *hubiera tenido* diez palcos, los *habría llenado* hasta los topes
 ‘If Barcelona [Football Club] *had had* ten boxes, they *would have been completely full*.’

(10) Ocasiones que no se consiguen, pero que de conseguirse nos *transformarían* en ángeles, evitarían el que siguiéramos enfangados en el crimen y el pecado (. . .).
 ‘Opportunities which do not materialize but which, if they did, would transform us into angels, would prevent us from becoming mired down in crime and sin (. . .).’

Finally, the distinction between probable and possible events, on the one hand, and improbable and impossible events, on the other, is represented in FactBank, thus allowing us to express the full range of degrees of uncertainty. In SSC we feel that it is important to differentiate between when a statement is presented neutrally as uncertain (10) and when it is presented as uncertain by the use of an epistemic expression denoting probability *-por supuesto* ‘of course’- or possibility *-quizás* ‘maybe’ (8).

(11) Banderas aseguró que “en España nunca se hubiera concedido a un actor” un reconocimiento de este tipo.
 ‘Banderas declared that ‘recognition of this kind would never have been granted to an actor in Spain’.

Had we used the adverb *probablemente* ‘probably’ with *llenar* ‘be full’ in sentence (9), the interpretation of uncertainty would be more clear-cut, as is the case of example (8). In this respect, the uncertainty is more ‘neutral’ since a subjective opinion is not directly identified in (9). Furthermore, in SSC, whenever we come across an uncertainty marker, we further annotate the sentence as *epistemic*.

The tagging of modality, polarity and factuality takes place at the sentential level. Occasionally, we have been able to automatically pre-assign some of these values. For example, it is simple to detect the presence of words such as adverbs, determiners and pronouns that make statement *negative*. All remaining statements are therefore labeled *positive*. Thus, the annotation of polarity becomes a trivial task that can be automatized. Regarding modality, the pre-assignment of

1 values has sometimes been possible for non-assertive sentences when formal
 2 marks, such as specific verb tenses, like subjunctive, future or conditional
 3 (*non-assertive* general tag), are used. However, this automatic annotation must
 4 be double-checked manually given the fact that we can find instances like (12)
 5 where the simple past conditional expresses assertivity:

6

7 (12) A lo largo de esa angustiada madrugada, Juan me *aclararía* que no había
 8 ningún secreto (. . .).

9 ‘In the course of those distressing early morning hours, Juan *would explain*
 10 that there was no secret (. . .).’

11

12 In appendix 1 we present a summary of the tags used in relation to polarity and
 13 assertivity (scheme 2) and also show how these tags are combined to express fac-
 14 tuality (scheme 3).

15

16

17 5 Results⁶

18

19 We present a graphical example of the annotation (Figure 1), in this case taken
 20 from the Spanish SSC. It corresponds to the annotation of sentence (4), which is
 21 associated with sense 1 of the verb *gastar* (‘consume’). The sentence excerpted
 22 from the corpus (‘Cada Catalán gasta 3,3 toneladas de petróleo al año’) is given in
 23 the middle bar. The three bands 2a–2c above the sentence are used to represent
 24 information about sentence semantics: semantic construction at argument-
 25 structure level is labeled in (2a), factivity in (2b) and aspectuality in (2c). For fac-
 26 tuality, the combination of assertive and positive polarity indicates that we are
 27 describing a “real” event. For aspectuality, on the one hand, the sentence is
 28 annotated as an event, in this case denoting an iterative situation (“habitual”
 29 tag). The tag “event” has been inherited from the lexicon. Information about
 30 argument structure is provided in bands 1a–1c under the sentence (orange-shaded
 31 boxes) with syntactic functions appearing in (1a), syntagmatic categories in (1b)
 32 and semantic roles in (1c).

33 In Figure 2 we present some information in the lexical entry for *gastar 1*,
 34 where we see, among other information, that this verb is classified lexically as
 35 bounded (event).

36

37

38 ⁶ We only present data for Spanish because, except for the literary subcorpus, Catalan sentences
 39 are the same as Spanish sentences and, consequently, the annotation of aspectuality, modality,
 40 polarity and factuality is coincident. Furthermore, we do not differentiate according to register
 for Spanish SSC because differences do not appear to be significant.

gastar (1):

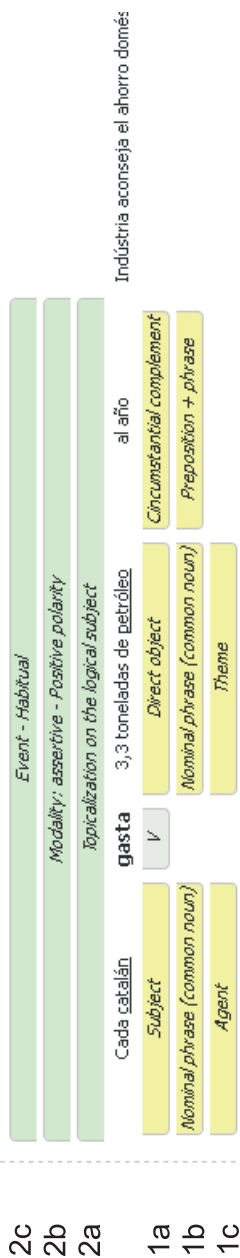


Fig. 1: Example of annotation

Definition:	Hacer uso de dinero u otro bien o producto que pueda consumirse.
Semantic roles:	Agent, Theme, Purpose
Aspectual class:	Event
Wordnet:	0066.1955v

Fig. 2: Lexical entry for *gastar_1* in SenSem lexicon

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

1 We present the figures for modality and polarity in the Spanish corpus in Tables 1
 2 and 2. As can be seen, the vast majority of sentences are assertive and positive.
 3 The difference between positive and negative statements is highly significant.
 4 Negative statements represent merely 7.6% of the sentences in the corpus.

5
 6
 7 **Table 1:** Modality in the Spanish SSC

8 Assertive [5/6] ⁷	23,451 (77.21%)
9 Non-assertive sentences [14]	6,923 (22.79%)

10
 11
 12
 13 **Table 2:** Polarity in the Spanish SSC

14 Positive [5/7a]	27,987 (92.14%)
15 Negative [6/7b]	2,387 (7.85%)

16
 17
 18
 19 In Table 3 we present the figures for each subtype of non-assertivity in the Spanish
 20 corpus and note that future non-assertive events clearly predominate (they con-
 21 stitute more than half). Impossible sentences are poorly represented (less than
 22 1%). Also, there are very few sentences in which a direct mark for epistemicity is
 23 found (1.40%).

24
 25
 26 **Table 3:** Non-assertivity in the Spanish SSC

27 Future [8]	4,521 (65.30%)
28 Past and present [11]	1,070 (15.46%)
29 Impossible [10]	65 (0.93%)
30 Epistemic [8]	97 (1.40%)

31
 32
 33
 34 Dynamic situations are the most common type of action found in the corpus
 35 (86.34%) (Table 4) and events practically double the number of processes. As for
 36

37
 38
 39 ⁷ Numbers between square brackets in this section refer to the sentences in the paper used to
 40 exemplify this kind of meaning.

Table 4: Aspectual information level in the Spanish SSC

States	4,149 (13.66%)	3
<i>Temporary states</i> [3b]	525 (12.65%)	4
<i>Permanent states</i> [3a]	3,624 (87.35%)	5
Events [1]	17,037 (56.09%)	6
Processes [2b]	9,188 (30.25%)	7
Perfective [1]	10,360 (48.74%)	8
Imperfective [2b]	10,894 (51.25%)	9
Habitual [4]	1,121 (10.29%)	10

states, a small number of *temporary stative constructions* have been found in the corpus. At the other extreme, the difference between the number of perfective and imperfective situations is not significant.⁸ Finally, only 10% of the actions are habitual.

6 Conclusions

The SSC are quite large in terms of the diversity of information reported. Furthermore, they constitute important resources in that they incorporate information about high-level sentence semantics, which has been the focus of this paper, namely aspectuality, modality and polarity, which together provide information concerning factuality. It should be emphasized that the Spanish SSC is one of only two corpora with information regarding factuality in Spanish. As for Catalan, this is the first corpus to annotate this kind of information.

As indicated by Morante and Daelemans (2012), “factuality involves polarity, epistemic modality, evidentiality and mood” (p. 3). We would add that aspectuality also plays a central role, as in the case of habitual interpretation. This is one key contribution of this project, since there are no other corpora where habitual actions are annotated. Also, the SSC are the first corpora in which reversibility of states is represented. Another contribution is the annotation of the degrees of uncertainty, which is not considered in similar projects. The identification of

⁸ States and future or imperative actions (30% of the total) have not been annotated regarding (im)perfectivity.

1 these nuances in texts are very important when it comes to extracting inferences
 2 from them, which can be very useful in specific tasks of natural language process-
 3 ing (NLP).

4 These two SSC are also the first corpora where dynamicity and telicity is rep-
 5 resented. Though this kind of information has no direct applicability in NLP, it
 6 does hold a certain value from the point of view of descriptive linguistics.

7 As regards results, we can claim that, both in Spanish and Catalan, the most
 8 frequent situations found are positive assertive and dynamic actions. Imperfec-
 9 tive and perfective situations are almost equally represented. As for non-assertive
 10 situations, they are clearly more represented when dealing with the future than
 11 with the present or the past.

12

13

14

15 Appendix

16

17

18 Scheme 1. Aspectuality tags

19

20

I. Lexical level

21

➔ ***Inheritance of tags from the lexical database:***

22

Dynamicity:

23

Neutral with respect to telicity:

24

➤ *State* (Non dynamic)

25

➤ *Process-event* (Dynamic)

26

Defined with respect to telicity:

27

➤ *Event* (Dynamic)

28

➤ *Process* (Dynamic)

29

30

II. Phrasal level

31

➔ ***Modification of process-event tag:***

32

Telicity:

33

➤ *Event*

34

➤ *Process*

35

36

III. Sentential level

37

➔ ***Subspecification of state tag:***

38

Duration:

39

➤ *Permanent state*

40

➤ *Temporary state*

→ Addition of tags:	1
Perfectivity:	2
➤ <i>Imperfective</i>	3
➤ <i>Perfective</i>	4
Iterativity:	5
➤ <i>Habitual</i>	6
	7
	8
	8
Scheme 2. Combination of polarity and modality tags to express factuality	9
	10
<u>Sentential level</u>	11
– Factuality/Certainty: <i>Positive polarity + Assertive</i>	12
– Counter-factuality/Non certainty:	13
○ <i>Negative polarity + Assertive</i>	14
○ <i>Positive/Negative polarity + Non-assertive + Past-present + Impossible</i>	15
– Non-factuality/Uncertainty: <i>Positive/Negative polarity + Non-assertive</i>	16
○ Scope of uncertainty:	17
■ Unknown uncertainty: <i>Past/present</i>	18
■ Absolute uncertainty: <i>Future</i>	19
○ Degrees of uncertainty:	20
■ Epistemicity: <i>Epistemic</i>	21
	22
	23
	24
	25
	26
	27
	28
	29
	30
	31
	32
	33
	34
	35
	36
	37
	38
	39
	40

References

- Ahern, Aoife. 2008. *El subjuntivo: contextos y efectos*. Madrid: Arco Libros. Cuadernos de Lengua Española.
- Alonso, Laura, Josep Antoni Capilla, Irene Castellón, Ana Fernández, Gloria Vázquez. 2007. The Sensem Project: Syntactico-Semantic Annotation of Sentences in Spanish. Angelova, Galia, Kalina Bontcheva, Ruslan Mitkov & Nicolas Nicolov. (eds.), *Recent Advances in Natural Language Processing IV. Current issues in linguistic theory 292*. John Benjamins Publishing Co, p. 89–98.
- Atserias, Jordi, Bernardino Casas, Elisabet Comelles, Meritxell González, Lluís Padró & Montse Padró. 2006. FreeLing 1.3: Syntactic and semantic services in an open-source NLP library. En *Proceedings of the fifth international conference on Language Resources and Evaluation (LREC)*. Genoa, Italy, p. 2281–2286.
- Baker, Kathryn, Michael Bloodgood, Bonnie J. Dorr, Nathaniel W. Filardo, Lori Levin & Christine Piatko. 2010. A modality lexicon and its use in automatic tagging. In *Proceedings of the Seventh International Conference on Language, Resources and Evaluation (LREC)*. Valletta, p. 1805–1812.
- Coll, Mercè. 2007. *Propuesta de anotación del aspecto en el corpus SenSem*. Oral Communication. II Seres Workshop, Barcelona, Universitat de Barcelona.

- 1 Feldman, Anna & Katya Arshavskaya. 2007. English and Russian event annotation: A pilot study.
 2 In Meurman-Solin, Anneli & Arja Nurmi (eds.), *Studies in variation, contacts and change in*
 3 *English*, vol. 1. http://www.helsinki.fi/varieng/series/volumes/01/feldman_arshavskaya/
 4 (last accessed Nov. 11th).
- 5 Fernández, Ana, Gloria Vázquez, Irene Castellón. 2006. SenSem: a Databank for Spanish Verbs.
 6 In *Proceedings of the X Ibero-American workshop on artificial intelligence, IBERAMIA*.
 7 Ribeirão Preto, Brasil. <http://grial.uab.es/archivos/SenSem%20databank-5-def.pdf> (last
 8 accessed Nov. 11th).
- 9 Hendrickx, Iris, Amália Mendes & Silvia Mencarelli. 2012. Modality in text: A proposal for
 10 corpus annotation. In *Proceedings of the Eighth International Conference on Language,*
 11 *Resources and Evaluation (LREC)*. Istanbul, Turkey, p. 1805–1812.
- 12 Morante, Roser & Walter Daelemans. 2012. Annotating Modality and Negation for a Machine
 13 Reading Evaluation. In *CLEF 2012 Evaluation Labs and Workshop Online Working Notes*.
 14 <http://www.clips.ua.ac.be/~walter/papers/2011/md11.pdf> (last accessed Nov. 11th).
- 15 Pustejovsky, James, Marc Verhagen, Roser Saurí, Jessica Littman, Robert Gaizauskas, Graham
 16 Katz, Inderjeet Mani, Robert Knippen & Andrea Setzer. 2006. *TimeBank 1.2*. Linguistic Data
 17 Consortium (LDC). Philadelphia, PA.
- 18 Saurí, Roser. 2010. *Annotating temporal relations in Catalan and Spanish*. *TimeML annotation*
 19 *guidelines*. Barcelona Media Technical Report, BM 2010-04.
- 20 Saurí, Roser & James Pustejovsky. 2009. Factbank: A corpus annotated with event factuality.
 21 *Language Resources and Evaluation*, 43(3). 227–268.
- 22 Saurí, Roser, Marc Verhagen & James Pustejovsky. 2006. Annotating and Recognizing Event
 23 Modality in Text. In *Proceedings of FLAIRS Conference*. Melbourne Beach, Florida,
 24 p. 333–339.
- 25 Vázquez, Gloria, Ana Fernández, & Laura Alonso. 2005. Description of the guidelines for the
 26 syntactico-semantic annotations of a corpus in Spanish. In Angelova, Galia, Kalina
 27 Bontcheva, Ruslan Mitkov & Nicolas Nicolov (eds.), *International Conference: recent*
 28 *advances in natural language*. Shoumen (Bulgaria), p. 603–607.
- 29 Vázquez, Gloria, Ana Fernández, Elena Beà. 2013. SenSemCat: Corpus de la lengua catalana
 30 anotado con información morfológica, sintáctica y semántica. Emili Casanova & Cesàreo
 31 Calvo (eds.), *Actes del 26è Congrés internacional de lingüística i filologia romàniques*
 32 (2010) De Gruyter, VIII, 159–170.
- 33 Vendler, Zeno. 1957. Verbs and Times. *Philosophical Review*, 56. 143–160.
- 34 Vincze, Veronika, György Szarvas, Richárd Farkas, György Móra & Janos Csirik. 2008. The
 35 BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes.
 36 *BMC Bioinformatics*, 9:11.
- 37 Wonsever, Dina, Marisa Malcuori & Aila Rosá. 2008. *SIBILA: Esquema de anotación de eventos*.
 38 Instituto de Computación. Universidad de la República Montevideo. Reporte técnico RT
 39 08-11.

Bionotes

40 Gloria Vázquez ■■■■

Ana Fernández ■■■■

Please
supply
data

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40