

Spanish FrameNet and FrameSQL¹

Carlos Subirats and Hiroaki Sato

Autonomous University of Barcelona (Spain) and Senshu University (Japan)

Carlos.Subirats@uab.es, hiroaki@isc.senshu-u.ac.jp

Abstract

In this paper, we introduce the Spanish FrameNet Project which is creating an online lexical resource for Spanish, based on Frame Semantics and supported by corpus evidence. Spanish and English lexical units in the Emotion and Motion frames are compared and differences between the lexicalization patterns and constructions in both languages are described. This paper also introduces FrameSQL which is a web-based application to search and view the Spanish and Berkeley's FrameNet data on the web browser. The application handles both seamlessly, showing on the same window the Spanish and English lexical units belonging to the same frame. It makes it easier to compare semantic structures of the two lexicons.

Background to Spanish FrameNet

The Spanish FrameNet Project (<http://gemini.uab.es/SFN>) is creating an online lexical resource for Spanish, based on Frame Semantics (Fillmore, 1982, 1985) and supported by corpus evidence. The "starter lexicon" will be available to the public by January 2006, and will contain at least 1000 lexical items –predicative verbs, nouns and adjectives– representative of a wide range of semantic domains. The aim is to document the range of semantic and syntactic combinatory possibilities (valences) of predicates in specific senses, through:

- human approved and automatic annotated example sentences and
- automatic capture and organization of the annotation results.

The Spanish FrameNet (SFN) database will be in a platform-independent format, and it will be able to be displayed and queried via the web and other interfaces. The SFN database will act both as a dictionary and a thesaurus. The dictionary features include:

- definitions, tables showing how frame elements are syntactically expressed in sentences containing each word,
- annotated examples from the corpus: human approved and automatically annotated, and an alphabetical index.

Like a thesaurus, words are linked to the semantic frames in which they participate, and frames, in turn, are linked to wordlists and to related frames. The basic assumption of Frame Semantics is that each word evokes a particular frame and possibly profiles some element or aspect of that frame. Semantic frames are schematic representations of situations involving various participants, props, and other conceptual roles, each of which is called a frame element (FE). The semantic arguments of a predicating word correspond to the frame

elements of the frame (or frames) associated to that word. A frame semantic description of a lexical unit identifies the frames which underlie a given meaning and specifies the ways in which frame elements are realized in structures headed by the word (See Johnson et al., 2002). For example, consider the *Judgement_communication* frame which deals with communicating a positive or negative judgment of an Evaluatee to an Addressee, e.g. *alabar* (praise) or *criticar* (criticize). This frame minimally includes the FEs Communicator, Evaluatee and Addressee. Sentence (1) below is a canonical example of a verb in the *Judgement_communication* frame.

- (1) Max **elogió** a Eva ante los
Max praised to Eva in-front-of the
directivos de la empresa.
directors of the company
Max praised Eva before the company directors.

Here, *Max* fills the role of Communicator; *Eva* is the Evaluatee; and *los directivos de la empresa* is the Addressee. Note that the Addressee is expressed in (1) above, but it may not be realized in other sentences, as shown in sentence (2) below, where *la actuación de la empresa* is the Evaluatee, and the Addressee is not instantiated.

- (2) Sara **denunció** la actuación de la empresa.
Sara reported the performance of the company
Sara reported the company performance.

Each frame element tag is part of a set of three tags, consisting of the frame element (Communicator, Evaluatee, etc.), the grammatical function and the phrase type of the annotated constituent. The mappings between the semantic and syntactic information constitutes its valence. This information is given in the triples of annotation for the set of sentence types in which a given lexical unit

¹ Spanish FrameNet is a research project which is sponsored by the Department of Science and Technology of Spain (Grant No. TIC2002-01338), and is developed at the Autonomous University of Barcelona in cooperation with the FrameNet Project, administered at the International Computer Science Institute in Berkeley, CA. The authors would like to thank their colleagues, Rocío Donés, Jordi Duran, Mercedes García, Lidia Moya, and Marc Ortega for their assistance.

occurs. The goal of Spanish FrameNet is to annotate corpus citations and to discover the valence patterns for a large number of words showing how those valence patterns are instantiated in actual sentences. Each Spanish FrameNet entry will provide links to other lexical resources, including Spanish EuroWordNet synsets and syntactic subcategorization frames. The project's deliverables will consist of the SFN database itself: lexical entries for individual word senses, frame descriptions, and annotated subcorpora.

The SFN project is based on the evidence offered by a 330 million-word corpus which includes both New World (60%) and European Spanish (40%). The corpus is POS tagged and lemmatized with a tool that uses an electronic dictionary of Spanish of 600,000 forms, both single (92%), and multi-word lexical units (8%), basically multi-word nouns (85%), like *bomba atómica* (atomic bomb), *carga de profundidad* (depth charge), and multi-word adverbs (9%) like *a ciegas* (unknowingly), *por ahora* (by now), etc. Multi-word verbs like *tener en cuenta* (to take into account) and lexicalized prepositional phrases with support verbs like *estar de moda* (to be in fashion) are tagged and lemmatized with transducers. SFN uses the Corpus Workbench software from the Institut für Maschinelle Sprachverarbeitung of the University of Stuttgart² for searching the corpus. The semantic and syntactic annotation is carried out by using the FNDesktop, the system developed by the Berkeley FrameNet Project. The input of the FNDesktop is composed by sentences that have been automatically extracted from the corpus, and then POS tagged and lemmatized (Subirats and Ortega, 2000). The extraction of subcorpora where predicates appear in all their relevant constructions provide annotators with examples of each possible syntactic configuration in which a given lexical item can occur. Annotators then select sentences for annotation that best illustrate the ways in which frame elements are realized syntactically. Figure 1 shows an actual sentence from the database annotated with the FNDesktop.

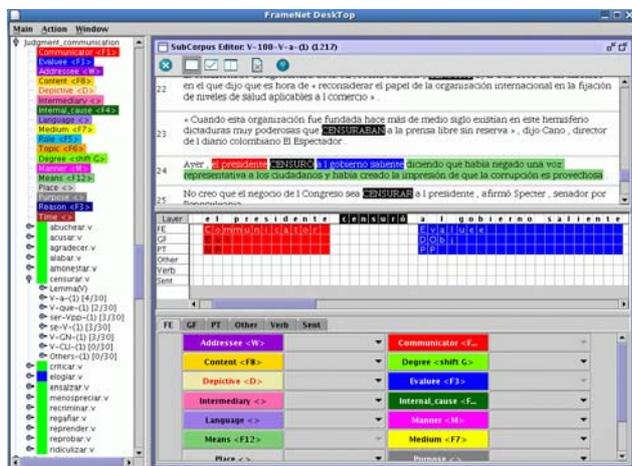


Figure 1: Annotation of a sentence in the Judgement_communication frame

Different lexicalization patterns in English and Spanish emotion predicates

SFN is studying areas of the lexicon that parallel existing English FrameNet descriptions. Most of the frames defined so far are valid cross-linguistically, because frames are meant to characterize conceptual structures at a basic level of description.

Valence descriptions provided by SFN and FN can be used to study different lexicalization patterns in English and Spanish. Thus, for instance, *sorprender* (to surprise) in (3) is a Cause_emotion verb characterizing an event, in which an agent seeks to cause an emotion on an Experiencer.

- (3) Juan **sorprendió** a María al contarle la verdad.
 Juan surprised to Maria on explaining-her the truth
 Juan surprised María by telling her the truth.

The reflexive verb *sorprenderse* (to get surprised) in (4) and the adjectival past participle *sorprendido* (surprised) with the support verb *estar* (to be) in (5) are two Experiencer_objet predicates in which the Experiencer is the subject and the Stimulus is the object.

- (4) María **se sorprendió** de que Juan cantase.
 María REFL surprised of that Juan sang
 María got surprised when Juan sang.
- (5) María está **sorprendida** de que Juan cante.
 María is surprised of that Juan sang
 María is surprised that Juan sang

Both *sorprenderse* in (4) and *sorprendido* in (5) express parts of the complex event characterized by *sorprender* in (3): *sorprenderse* is an inchoative verb which characterizes the beginning of an event and *sorprendido* expresses the ongoing state which occurs after the above mentioned beginning. Therefore, *sorprenderse* and *sorprendido* are simpler parts of the complex event *sorprender* (cf. Subirats and Petruck, 2003).

This analysis allows us to study the lexicalization pattern differences among English and Spanish emotion predicates. Both Spanish and English lexicalize the causative meaning with two verbs, namely, *surprise* and *sorprender*. On the contrary, there exists a difference in the lexicalization of the inchoative meaning: Spanish uses the reflexive verb *sorprenderse*, whereas English uses the construction made by *get* and the adjectival past participle *surprised*. As a result, English only has the LU *surprised* in the Experiencer_subject frame and Spanish has two LUs, that is, the reflexive verb *sorprenderse*, and the adjective *sorprendido* (cf. Figure 2). These differences can be verified thanks to FrameSQL, an application which allows to compare predicates or predicate-related

² <http://www.ims.uni-stuttgart.de/>

constructions in the frames which share the same name, and therefore the same characteristics in English and Spanish.

| | Stative being in a state | Inchoative entering into a state | Causative putting into a state |
|---------|--------------------------------|--|-----------------------------------|
| Spanish | Experiencer_subject | | Cause_emotion |
| | estar V-PP | V REFL | V |
| | estar sorprendido | sorprenderse | sorprender |
| English | Experiencer_subject | | Cause_to_experience |
| | be V-PP | get V-PP | V |
| | be surprised | get surprised | surprise |

Figure 2: Different lexicalization patterns in Spanish and English emotion predicates (cf. Subirats and Petruck, 2003).

Different constructions in English and Spanish motion predicates

Comparative valence descriptions between SFN and FN have still shown other differences. For instance, motion predicates in Spanish, like the majority of predicates from other frames, accept Purpose FEs, such as *para pedirle dinero a un amigo* in (6) below.

- (6) **Voy** a San Francisco para pedirle dinero.
 Go to San Francisco To ask-him money
 a un amigo.
 to a friend
 I go to San Francisco to ask a friend for money.

However, many motion predicates in Spanish accept an Intentional FE, such as *a ver un amigo* in (7), which expresses the intention of the motion event, which is semantically different from the purpose.

- (7) **Voy** a San Francisco a ver a un amigo.
 Go to San Francisco to see to a friend
 I go to San Francisco to see a friend.

Intentional FEs not only have a different meaning from Purpose FEs, but they are also syntactically different. In this way, the Intentional FE *a ver un amigo* in (7) is a prepositional object and, therefore, it is not only a conceptual argument of the target, but also a syntactic argument. On the contrary, the extrathematic Purpose FE *para pedirle dinero a un amigo* in (6) is an adjunct which is not a syntactic argument of the target. The semantic difference between Intention and Purpose allows both FEs to be present in the same sentence, such as in (8), acting as different conceptual arguments of the same target.

- (8) Juan **fue** a San Francisco a visitar a
 John went to San Francisco to visit a
 un amigo para pedirle dinero.
 a friend to ask-him money
 John went to san Francisco to visit a friend and ask
 him for money.

There is a clear difference between Spanish and

English. While in Spanish, there are two conceptual and syntactic arguments attached to the same target, in English there are two coordinated sentences with two different targets. Thus English uses another construction to express the same meaning.

FrameSQL

FrameSQL is a web-based application to search and view the Berkeley FN data on the web browser (Sato, 2003). Since its data structure is basically the same as that of SFN, FrameSQL can handle the SFN data with a little modification. The application stores the FrameNet data in an MySQL database, and executes various searches in the SQL language, when users select search parameters on the web browser. The application handles both of the FN data seamlessly, showing Spanish and English lexical units belonging to the same frame on the same window. It makes it easier to compare semantic structures of the two lexicons.

FrameSQL has several search modes. Figure 3 shows a basic menu for searching the Spanish lexical unit *elogiar* of the Judgement_communication frame. The search menu consists of four panes: the upper one for selecting search modes, the middle-left for specifying frames and lexical units to search and view, the middle-right for setting search parameters, and the bottom for showing help files and search results.

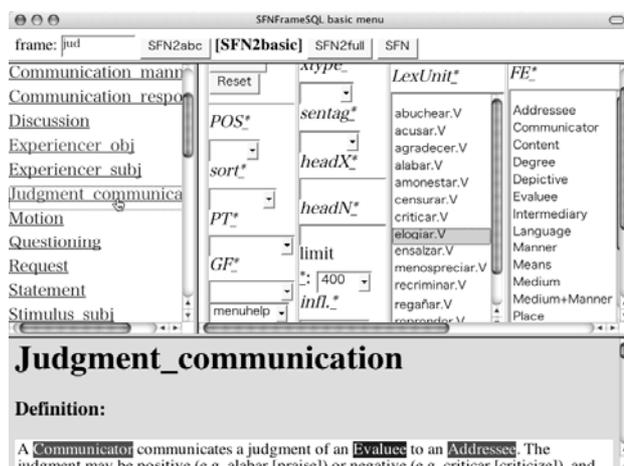


Figure 3: Basic search menu of FrameSQL

Figure 4 shows search results of the lexical unit *elogiar*. The bottom pane summarizes how each of FEs are used in annotated examples. Each line in the bottom pane consists of the number of annotated examples (Num), two hyperlinks to English FrameNet (Sloppy, Exact), and a set of FEs and LU used in annotated examples (FE/LUset). The left numbers are hyperlinked to annotated examples. For example, when a user clicks on the hyperlink *01* of the last line which have the FE/LU set *Communicator+elogiar.V+Evaluate+Role*, annotated examples with this set appear on the middle-left pane of Figure 4.

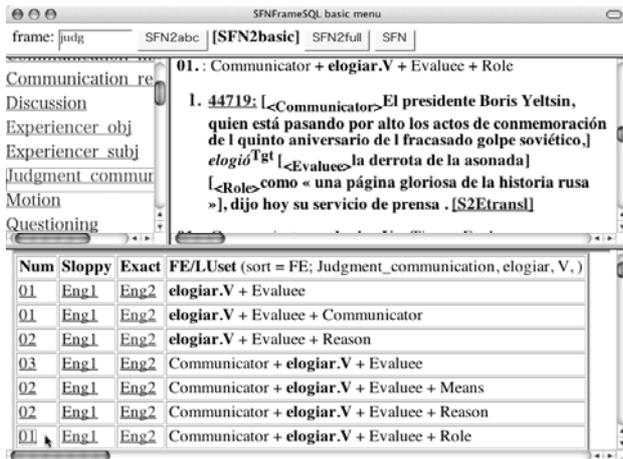


Figure 4: Search results of *elogiar*

The hyperlinks to English FrameNet lead to English annotated examples which have similar FE sets to Spanish ones. For example, when a user click on the Eng1 link of the FE/LUset *Communicator+elogiar.V+Evaluatee+Role*, English annotated examples of the *Judgement_communication* frame appear in the middle-left pane that contain this FE set, as in Figure 5.

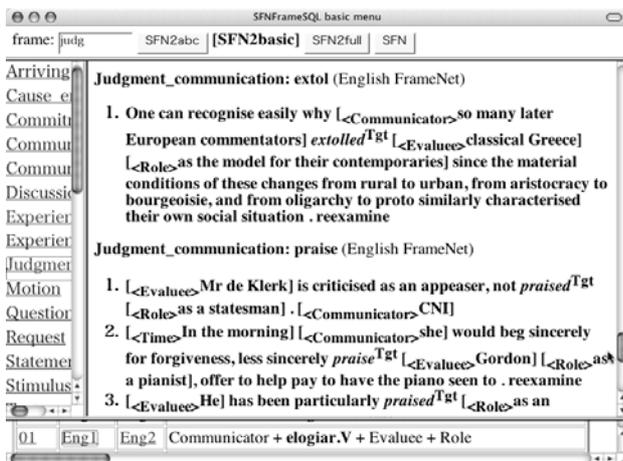


Figure 5: Accessing English FrameNet

References

Baker, C.F., C.J. Fillmore and B. Cronin. (2003). The Structure of the FrameNet Database, *International Journal of Lexicography*, Volume 16(3), 281--296.

Baker, C.F. and J. Ruppenhofer (2002) FrameNet's Frames vs. Levin's Verb Classes. In J. Larson and M. Paster (Eds.) In *Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society*. (pp.27--38).

Fillmore, C.J. (1985). Frames and the semantics of understanding. In *Quaderni di Semantica* 6(2), 222-254.

Fillmore, C.J. (1982). Frame Semantics. In *Linguistics in the Morning Calm* (pp.111--137). Seoul: Hanshin Publishing Co.

Fillmore, C.J., C.R. Johnson and M.R.L. Petruck. (2003). Background to FrameNet. *International Journal of Lexicography* 16(3), 235--250.

Fillmore, C.J. and H. Sato. (2002). Transparency and Building Lexical Dependency Graphs. In J. Larson and M. Paster (eds.) *Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society* (87--99).

Johnson, C.R., M.R.L. Petruck, C.F. Baker, M. Ellsworth, J. Ruppenhofer, and C.J. Fillmore. (2002). *FrameNet: Theory and Practice*. Berkeley, CA: International Computer Science Institute, Technical Report-02009.

Sato, H. (2003). *FrameSQL: A Software Tool for FrameNet*. In *ASIALEX '03 Tokyo Proceedings* (pp. 251--258), Asian Association of Lexicography.

Subirats, C. and M. Ortega. (2000). Tratamiento automático de la información textual en español mediante bases de información lingüística y transductores. *Estudios de Lingüística del Español* 10: <http://elies.rediris.es/elies10/>

Subirats, C. and M. R. L. Petruck. (2003). Surprise: Spanish FrameNet! Presentation at Workshop on Frame Semantics. Prague (Czech Republic): International Congress of Linguists. July 29, 2003.