

New Functions of FrameSQL for Multilingual FrameNets

Hiroaki SATO

Senshu University

Office #8411, 2-1-1 Higashimita, Tama-ku, Kawasaki-shi, Kanagawa-ken, 214-8580, JAPAN

E-mail: hiroaki@isc.senshu-u.ac.jp

Abstract

The Berkeley FrameNet Project (BFN) is making an English lexical database called FrameNet, which describes syntactic and semantic properties of an English lexicon extracted from large electronic text corpora (Baker et al., 1998). Other projects dealing with Spanish, German and Japanese follow a similar approach and annotate large corpora. FrameSQL is a web-based application developed by the author, and it allows the user to search the BFN database in a variety of ways (Sato, 2003). FrameSQL shows a clear view of the headword's grammar and combinatorial properties offered by the FrameNet database. FrameSQL has been developing and new functions were implemented for processing the Spanish FrameNet data (Subirats and Sato, 2004). FrameSQL is also in the process of incorporating the data of the Japanese FrameNet Project (Ohara et al., 2003) and that of the Saarbrücken Lexical Semantics Acquisition Project (Erk et al., 2003) into the database and will offer the same user-interface for searching these lexical data. This paper describes new functions of FrameSQL, showing how FrameSQL deals with the lexical data of English, Spanish, Japanese and German seamlessly.

1. Introduction

FrameSQL is a web-based application which I have developed and been maintaining for BFN since 1999. The application is designed to search the FrameNet database, a lexicon of contemporary written English that the project is producing for both computational and lexicographic interests. Other projects dealing with Spanish, German and Japanese follow a similar approach and annotate large corpora. The Spanish FrameNet Project (SFN) and the Japanese FrameNet Project (JFN) have been working within the framework of **Frame Semantics** (Fillmore, 1976) which BFN is based on. The Saarbrücken Lexical Semantics Acquisition Project (SALSA) also adopts the paradigm of Frame Semantics and created a large German annotated corpus.

FrameSQL can search and view the BFN data release 1.3 on a standard web browser. You do not need to install any additional software tools to use FrameSQL, nor do you even need to download the BFN data to your local computer, because FrameSQL accesses the database of the server computer in Japan, and executes searches. The server computer handles the BFN data with MySQL, a popular relational database application that can execute complex searches. Directly searching a MySQL database requires learning the database query language SQL, but with FrameSQL, you do not have to learn SQL at all. You only select several search parameters and a form of the output search results on the web browser, and FrameSQL translates these parameters into their corresponding SQL command and executes the search.

2. Basic Searches

FrameNet is based on the theory of Frame Semantics. The central idea of the theory is that word meanings must be described in semantic **frames** which schematically represent conceptual structures. Each semantic frame has a group of **lexical units (LUs)**. A lexical unit is a pairing of a word with a sense (Cruse, 1986) whose semantic properties are described with semantic roles called **frame elements (FEs)**. For example, the *Arriving* frame has a

list of LUs such as *approach.N*, *approach.V*, *arrive.N* and *arrive.V*, and semantic properties of example sentences in this frame are described with its own set of FEs such as *Theme* (the object that moves) and *Goal* (the place where *Theme* ends up).

There are several search modes in FrameSQL. Figure 1 shows the **LexUnit search mode** (http://sato.fm.senshu-u.ac.jp/fn2_13/notes/index2.html) for the BFN data, where you can search and view annotated sentences of a single LU at a time.

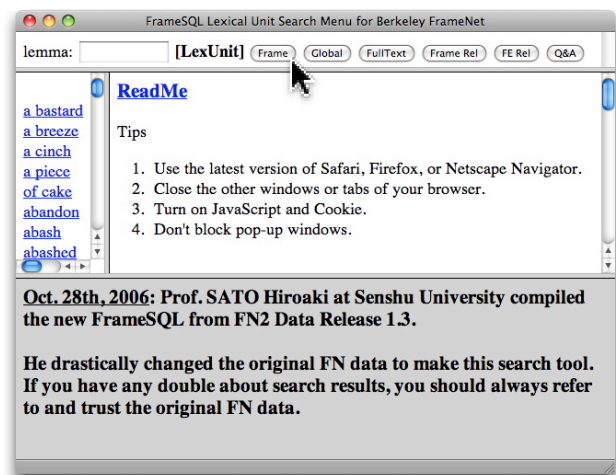


Figure 1: FrameSQL LexUnit Search Mode

You can select another search mode by clicking a button on the upper pane. When you click on the [Frame] button on the upper pane, you can select the **Frame search mode**, where you can search and view annotated sentences of a single frame at a time.

FrameSQL uses lists of frames, LUs and FEs as search parameters. Figure 2 shows the Frame search mode where a user selected *Arriving* in the frame list of the middle-left pane, *approach.V* in the LU list and *Source* (a general direction from which *Theme* moves) in the FE list of the middle-right pane. When the user clicks on the [Search] button, FrameSQL translates these search

parameters into SQL command (1) and executes the search.

```
(1) SELECT bncid, word0.sentid, sentence, list.frameID,
list.frame, list.lemma, list.pos, sensedesc, word0.fe, pt,
gf, headX, headN FROM list, word0, sentence
WHERE word0.fe = 'Source' AND pos = 'V' AND
lemma = 'approach' AND list.frame = 'Arriving' AND
list.listid = sentence.listid AND word0.sentid =
sentence.sentid LIMIT 400;
```

In the lower pane of Figure 2, FrameSQL shows search results which include all the example sentences with the FE *Source* of the LU *approach.V* in the *Arriving* frame of the BFN data.

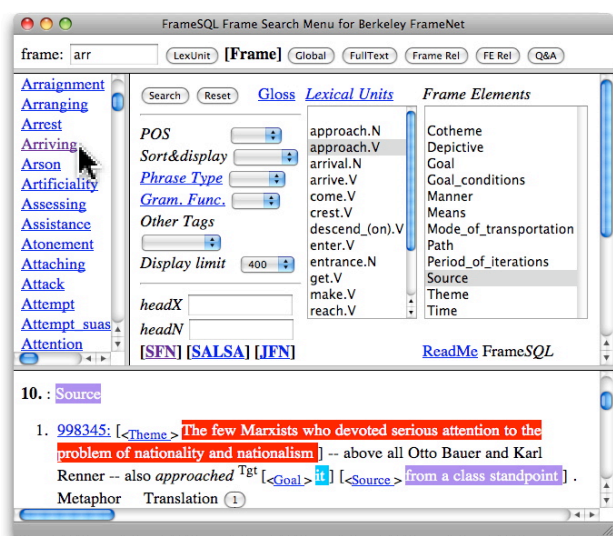


Figure 2: BFN Frame Search Mode

3. Mutual Hyperlinks of the Same Frame

At the left-bottom of the middle-right pane in Figure 2, there are three hyperlinks: [SFN] [SALSA] [JFN]. They are linked to FramesQL search menus of SFN, SALSA and JFN of the Arriving frame respectively.

Since all the three projects of SFN, JFN and SALSA adopt the basic notions of BFN, FramesQL search menus for these projects are basically the same. Once you get accustomed to one of those search menus, you can search those lexical databases in the same manner, and compare LUs of the same frame among those languages.

For example, when you click on [SFN] of Figure 2, the corresponding SFN search menu of the Arriving frame appears in the middle-right pane, as in Figure 3. The user-interface in Figure 3 is basically the same as that of Figure 2. You can execute a search command similar to that in Figure 2 by selecting the Spanish LU *aproximarse.V* (to approach) and the FE *Source* of this search menu. FrameSQL shows search results in the lower pane where all the example sentences for the LU *aproximarse.V* with the FE *Source* in the Arriving frame of SFN are displayed.

FrameSQL is Unicode compliant, and it can process

Japanese Unicode data of JFN as well as English and Spanish data. Figure 4 shows the search menu of Arriving frame for the JFN data. A user executed a search after selecting the Japanese LU 着く.V (to arrive) of the LU list and the FE *Source* in the FE list in the middle-right pane. Search results in the lower pane display a Japanese annotated sentence with the FE *Source* of the Arriving frame.

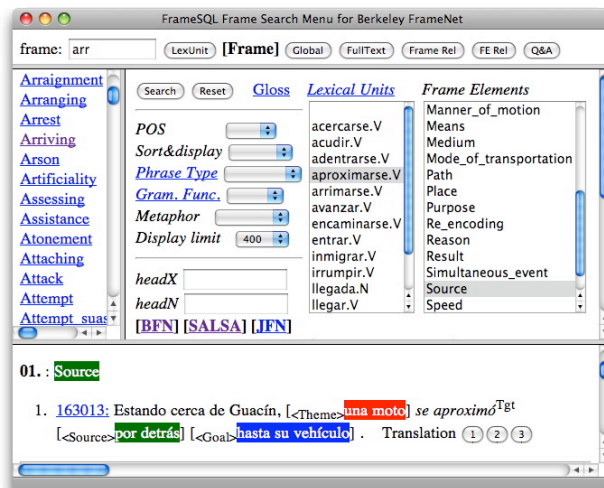


Figure 3: SFN Frame Search Mode



Figure 4: JFN Frame Search Mode

SFN and JFN use the same XML format as BFN, but SALSA uses a different XML format for the German database (Burchardt et al, 2006). I wrote a Perl script to convert the SALSA XML format into that used in BFN, and imported the SALSA data into FramesQL. Figure 5 shows the search menu of Arriving frame of the SALSA data, which corresponds to Figures 2, 3 and 4. Those pages are mutually linked to each other with the hyperlinks [BFN] [SFN] [SALSA] and [JFN].

FrameSQL adds translation buttons to the end of each annotated sentence for translating the Spanish sentence into English. When you click on the translation

button [1] in Figure 3, FrameSQL sends the Spanish sentence to the Spanish-to-English machine translation website (Instituto Cervantes, http://62.97.114.150/ats_0001.aspx) and shows the translation as in Figure 6. FrameSQL provides similar translation buttons for the JFN and SALSA data.

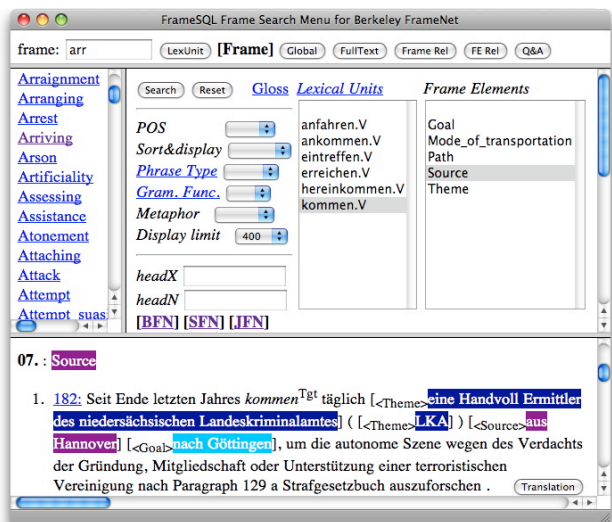


Figure 5: SALSA Frame Search Mode

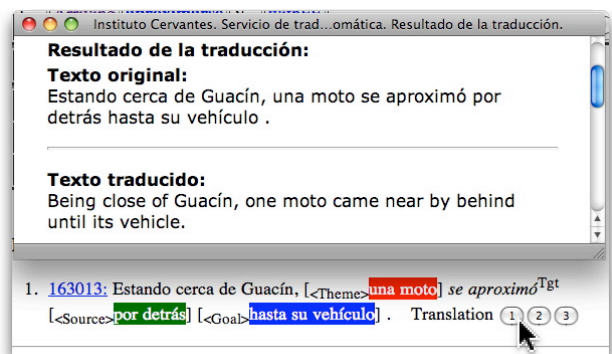


Figure 6: Translating Spanish to English

4. Comparing Corresponding LUs and Semantic Structures

Another search mode of FrameSQL makes it possible to compare semantic structures of corresponding LUs of two languages. In this mode, FrameSQL shows a table of FEs assigned to annotated sentences. For example, Figure 7 shows how FEs of the Arriving frame are assigned to annotated sentences of the Spanish LU *aproximarse.V* and its English translation equivalent *approach.V*.

The numbers on the right in the table show occurrences of example sentences annotated with those FEs on the left. For example, the third row (*Goal*) + *Manner* + *Theme* 01 01 shows that the LU *aproximarse.V* has one annotated Spanish sentence with the FEs (*Goal*) + *Manner* + *Theme*, and *approach.V* has one English sentence with the same FEs. The numbers on the right are linked with the corresponding annotated sentences, and you can access English and Spanish annotated sentences by clicking on the hyperlinked numbers. When you click

on the hyperlinked number 01 of *approach* beside (*Goal*) + *Manner* + *Theme*, FrameSQL will display on your web-browser sentence (2) which contains this FE set.

- (2) Guided by the agent [_{<Theme>}the pair] *approached*^{Tgt} [_{<Manner>}with extreme caution], using the trees for cover until they emerged from the tree-line above the valley . [_{<Goal>}DNI]

The symbol [_{<Goal>}DNI] in (2) corresponds to the parenthesized (*Goal*) of Figure 7. The parentheses of (*Goal*) signify that the FE *Goal* is conceptually salient, but does not appear in the sentence. BFN calls this type of FE **null instantiation**. There are three types of null instantiation and one of them is DNI (Definite Null Instantiation), in which the missing element is already understood in the linguistic or discourse context. FrameSQL parenthesizes all null instantiated FEs.

	aproximarse	approach
Goal + Manner + Theme		01
(Goal) + Manner + Theme	01	01
Depictive + Goal + Place + Theme	01	
Depictive + Goal + Manner + (Theme)	01	
(Goal) + Source + Theme		04
(Goal) + Manner + (Theme)		01
Degree + Goal + Theme	01	
Goal + (Theme)	02	
Goal + Theme	07	12
Goal + Theme + Time	01	
Goal + Means + Theme		01
(Goal) + Result + Theme	01	
Goal + Manner + Manner_of_motion + Theme	01	
Goal + Path + Theme		05
Depictive + (Goal) + Manner + Theme	01	
Cotheme + Goal + Theme	01	
Goal + Source + Theme	01	06
Circumstances + (Goal) + Manner + (Theme)	01	
(Goal) + Path + Theme		01
Goal + Speed + Theme + Time	01	
(Goal) + Theme		02

Figure 7: Spanish *aproximarse* and English *approach*

FrameSQL uses sense description in a Spanish-to-English electronic dictionary for making a list of corresponding LUs. For example, the Spanish Adorning frame, which involves a static relationship between a Location and a Theme, has the LU *adornar.V* whose sense is described in the dictionary as (3).

- (3) tr. to adorn, beautify, grace, decorate, embellish, deck, garnish.

English LUs in the same frame are as follows:

- (4) adorn.V, blanket.V, cloak.V, coat.V, cover.V, deck.V, decorate.V, dot.V, encircle.V, envelop.V, festoon.V, fill.V, film.V, line.V, pave.V, stud.V, wreath.V

When you extract English words which occur both in (3) and (4), you get the English translation equivalents *adorn*, *decorate* and *deck* for the Spanish LU *adornar*. I wrote a Perl script to make English translation equivalents for every Spanish LU in this manner, and made such a table as Table 1, which is used for comparing semantic structures of corresponding LUs as in Figure 7. FrameSQL provides similar bilingual tables for the JFN and SALSA data.

Frames	Spanish LUs	English LUs
Adorning	abarrotar	∅
Adorning	adornar	adorn,decorate,deck
Adorning	atestar	∅
Adorning	cubrir	cover,envelop,coat
Adorning	decorar	decorate,adorn
Adorning	envolver	∅
Adorning	impregnar	cover,envelop

Table 1: Corresponding Spanish and English LUs

5. Gaps of LUs and frames

Table 1 shows correspondences between Spanish LUs and English LUs in the Adorning frame. The symbol ∅ means that the Perl script cannot find English LUs corresponding to Spanish ones, when it compares sense description in the electronic dictionary and the English LUs of the Adorning frame of the BFN data. These gaps are common not only in the Adorning frame but also in other frames. Human translators could fill appropriate English LUs in some of the gaps, when there are corresponding LUs.

There are also gaps in frame correspondences. BFN makes an inventory of frames and FEs for annotating English sentences. Most of them work for annotating sentences of other languages, because frames are meant to characterize conceptual structures at a basic level of description. Some frames, however, seem to be language-specific. For example, SFN created new frames which are not defined in BFN, such as *Become_different*, *Becoming_wet* and *Being_influenced*, because these are needed for annotating Spanish sentences.

FrameSQL does not show hyperlinks in search menus of different languages explained in Section 3, when there are not corresponding search menus of the same frame. In other words, the presence of the hyperlinks in Figures 2, 3, 4 and 5 shows that the Arriving frames exist not only in BFN, but also in SFN, JFN and SALSA.

6. Training Annotators

The process of annotation is more complicated in FrameNet than in most of other lexical projects, because the number of frames and FEs is large. Although BFN

compiled a user manual which explains every concept of BFN in detail, it is still difficult to annotate sentences with rich semantic information. It is useful for annotators to exercise annotation on training materials before they start real annotation.

FrameSQL has quiz pages in which you can try “quasi” annotation. You are asked to select the appropriate FE in the pull-down menu for each of phrases. Figure 8 shows the quiz page for the LU *approach.V* of the Arriving frame of BFN, in which a user is about to select *Path* in the pull-down menu as the FE assigned to the word *it* in the sentence. After the selection, FrameSQL automatically marks the user’s choice as in Figure 9, where the striked-through *Path* signifies that the choice is the wrong answer. The quiz pages also show the definition of the FE which the user selects in the lower pane, as in Figure 9, when the user makes a choice in the pull-down menu.

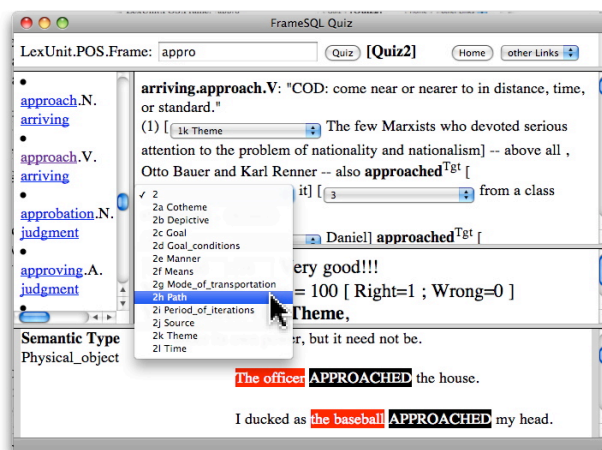


Figure 8: Quiz for training annotators

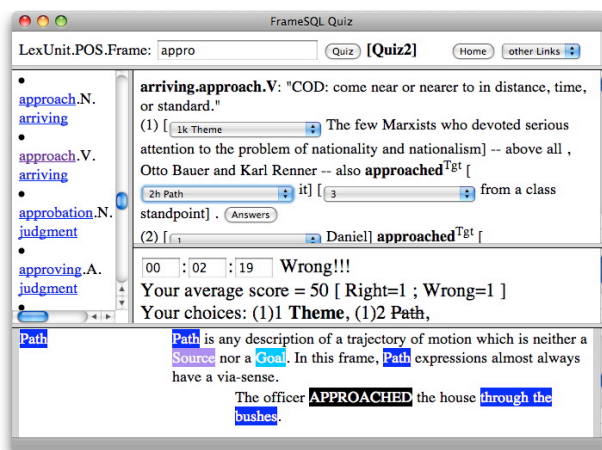


Figure 9: Quiz for training annotators

I wrote a Perl script to make the quiz pages from the BFN XML data. In other words, sentences in the quiz pages are actually used in the BFN corpus.

There are hundreds of FE names used in the BFN corpus, and showing all the FEs in the pull-down menu would confuse the trainee. The quiz pages only show those FEs which are used in annotated sentences of a

single frame. For example, each pull-down menu in Figure 8 has 12 FEs: *Cotheme*, *Depictive*, *Goal*, *Goal_conditions*, *Manner*, *Means*, *Mode_of_transportation*, *Path*, *Period_of_iterations*, *Source*, *Theme* and *Time*. These FEs are used in annotated sentences of the Arriving frame, and their definitions appear in the lower pane, when you select them in the pull-down menu.

FrameSQL has similar quiz pages for the SFN, JFN and SALSA data.

7. Conclusion

There are lots of search tools to process XML data, but only FrameSQL can investigate BFN, SFN, JFN and SALSA databases seamlessly. Mutual hyperlinks among these databases and the new search mode make it easy to compare the semantic structures of corresponding LUs across languages, and it could be useful for building multilingual lexical resources, or more generally for multilingual studies. Training good annotators for FrameNet is difficult, and the quiz pages of FrameSQL will be useful for this purpose.

In the poster presentation with demo session, I will show how the functions of FrameSQL mentioned in this paper really work.

References

- Baker, C.F., Fillmore, C.J., and Lowe, J.B. (1998). The Berkeley FrameNet project. In *Proceedings of the COLING-ACL*, pp. 86--90, Montreal, Canada.
- Burchardt, A., Erk, K., Frank, A., Kowalski, A., Pado, S., Pinkal, M. (2006). The SALSA Corpus: a German Corpus Resource for Lexical Semantics. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, pp. 969--974, European Language Resources Association.
- Cruse, D.A. (1986). *Lexical Semantics*. Cambridge: Cambridge University Press.
- Erk, K., Kowalski, A., Pinkal, M. (2003). A Corpus Resource for Lexical Semantics. In *Proceedings of the Fifth International Workshop on Computational Semantics (IWCS)*, pp. 106--121, Tilburg, The Netherlands.
- Fillmore, C.J. (1976). Frame Semantics and the Nature of Language. In *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, Volume 280, pp. 20--32.
- Fillmore, C.J., Baker, C.F., Sato, H. (2002). The FrameNet Database and Software Tools. In *Proceedings of the Third International Conference on Language Resources and Evaluation*, pp. 1157--1160, European Language Resources Association.
- Ohara, K.H., Fujii, S., Saito, H., Ishizaki, S., Ohori, T., Suzuki, R. (2003). The Japanese FrameNet Project: A Preliminary Report. In *Proceedings of Pacific Association for Computational Linguistics (PACLING'03)*, pp. 249--254. Halifax, Canada.
- Sato, H. (2003). FrameSQL: A Software Tool for FrameNet. In *ASIALEX '03 Tokyo Proceedings*, pp. 251--258, Asian Association of Lexicography, Tokyo, Japan.
- Subirats, C., Sato, H. (2004). Spanish FrameNet and FrameSQL. In *Proceedings of Building Lexical Resources from Semantically Annotated Corpora*, pp. 13--16, European Language Resources Association.