A Search Tool for FrameNet Constructicon

Hiroaki Sato

School of Commerce, Senshu University

Office #9513, 2-1-1 Higashimita, Tamaku, Kawasaki, Kanagawa, JAPAN

E-mail: thc0408@gmail.com

Abstract

The Berkeley FrameNet Project (BFN, https://framenet.icsi.berkeley.edu/fndrupal/) created descriptions of 73 "non-core" grammatical constructions, annotation of 50 of these constructions and about 1500 example sentences in its one year project "Beyond the Core: A Pilot Project on Cataloging Grammatical Constructions and Multiword Expressions in English" supported by the National Science Foundation. The project did not aim at building a full-fledged Construction Grammar, but the registry of English constructions created by this project, which is called Constructicon, provides a representative sample of the current coverage of English constructions (Lee-Goldman & Rhodes 2009). CxN Viewer is a search tool which I have developed for Constructicon and the tool shows its typical English constructions on the web browser. CxN Viewer is a web application consisting of HTML files and JavaScript codes. The tool is a useful program that will benefit researchers working with the data annotated within the framework of BFN. CxN Viewer is a unicode-compliant application, and it can deal with constructions of other languages such as Spanish.

Keywords: Construction Grammar, FrameNet, grammatical constructions

1. Introduction

The Berkeley FrameNet Project (BFN) is building a lexical database of English that is both human- and machine-readable, based on annotating examples of how words are used in actual texts (https://framenet.icsi.berkeley.edu/fndrupal/). BFN started with annotating simple examples, avoiding complex ones. But when it started annotating continuous text, BFN was faced with difficulties of how to deal with "non-core" grammatical constructions in the text. BFN changed the structure of the FrameNet database and created a number of adaptations to the annotation software to deal with these constructions.

In its one year project "Beyond the Core: A Pilot Project on Cataloging Grammatical Constructions and Multiword Expressions in English" supported by the National Science Foundation, BFN created descriptions of **73 constructions**, annotation of 50 of these constructions and about **1500 example sentences** (Lee-Goldman & Rhodes 2009). The project did not aim at building a full-fledged Construction Grammar, but the registry of English constructions created by this project, which is called **Constructicon**, provides a representative sample of the current coverage of English constructions.

CxN Viewer is a search tool which I have developed for Constructicon and the tool shows its typical English constructions on the web browser. CxN Viewer is a web application consisting of HTML files and JavaScript codes. BFN stores the Constructicon data in the MySQL database. The data can be exported into XML files, and I made perl scripts to convert the XML files into the HTML files used by CxN Viewer.

2. How CxN Viewer Shows XML Annotation Data

Among the constructions registered in the above project is the *Way* construction, which has been investigated by Goldberg (1995) among others. (1a) is an example sentence of the *Way_manner* construction in which the verb 'whistled' exceptionally takes 'her way' as a direct object. This construction indicates that some entity moves while performing the action indicated by the manner verb. CxN Viewer shows its annotation like (1b) on the web browser.

(1) a. She whistled her way down the lane to the silo.
   b. [$_{\text{Theme}}$ She] {[$_{\text{Transitive\_manner\_verb}}$ whistled] <$_{\text{Construction-evoking element}}$ her way>} [$_{\text{Path}}$ down the lane] [$_{\text{Goal}}$ to the silo].

Annotation practices in grammatical constructions are similar to those in FrameNet lexical units. *Theme*, *Transitive_manner_verb*, *Path* and *Goal* in (1b) are **construct elements** (CE's), which correspond to frame elements in the FrameNet lexical annotation. **Construction-evoking element** (CEE) in (1b) is the noun phrase 'her way', where 'her' is coindexed with the *Theme* 'She'. The CEE is analogous to the target lexical unit in FrameNet lexical annotation, although not all constructions have the CEE.

The phrases bracketed in [...] represent CE's, the one bracketed in < ... > represents the CEE, and the one bracketed in { ... } represents the **construct**, which licenses the rules of a construction. (1b) is made from the original XML annotation data of BFN, which is "represented in a way compatible with the development of full grammar of the language of the sort presented elsewhere [such as Sag (2012)]" (Fillmore, Lee-Goldman & Rhodes 2012).

3. Browsing the Data with CxN Viewer

3.1 Three-paned Interface

The BFN database for Constructicon consists of three

parts: the list of constructions, their definitions and annotation data of example sentences illustrating the constructions. (2) is a list of construction names in the database.

(2) *Absolute_clause, Adjective_as_nominal.abstract, Adjective_as_nominal.definite, Adjective_as_nominal.people, As.role, Attributive_degree_modification, Bare_argument_ellipsis, Bare_noun_phrase.role, Be_present-participle, Be_recip, Comparison, Comparison_equality, ... , Tautology.deep_tautology, Tautology.nominal, Tautology.nostalgia, Test, There.presentational, There_be_a_time_when, Uniqueness, Unitary_event_gerundive_construction, Valence_sharing.raising, Way_manner, Way_means, Way_neutral, What-with_absolute, With_absolute*

There three parts are stored in different tables of the original BFN database, and users would need to execute SQL commands to search the database, if they did not use CxN Viewer.

CxN Viewer is user-friendly and the user can access these three parts of data by selecting menus on the web browser. CxN Viewer has three panes for representing the above three parts. On the upper-left is the list of constructions. When the user clicks on a construction in the list, the viewer shows its definition on the upper-right pane and its annotation of example sentences on the lower pane. In Figure 1, the user clicks on the *Adjective_as_nominal.definite* construction on the upper-left list, and CxN Viewer shows its definition on the upper-right pane and its set of annotated examples on the lower pane. The three-paned interface makes it easier for users to browse annotated examples, because they can readily access definitions of constructions and CE's on the upper-right pane.
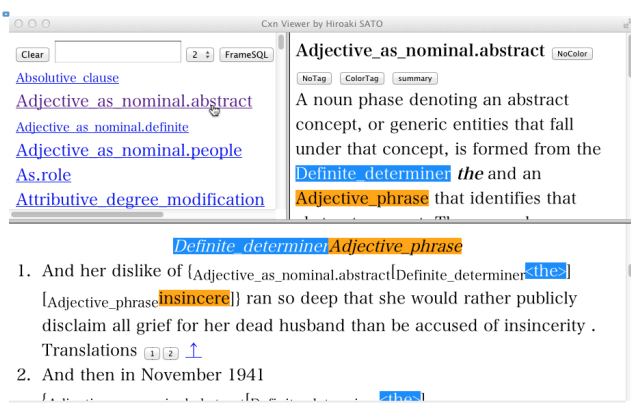

Figure 1: Three panes of CxN Viewer

3.2 Valence Table

When BFN annotators annotate example sentences, they input CE's with their phrase types and grammatical functions in the constructicon database. CxN Viewer shows valences of each construction in the upper-right pane. Figure 2 shows the valence table of the

Way_manner construction. The first line of the table "01 Theme Transitive_manner_verb CEE Direction" reads that there is one example sentence registered in the *Way_manner* construction that has the elements *Theme*, *Transitive_manner_verb*, *CEE* and *Direction*. The second line "01 NP.Ext VPing._ NP._ AVP.Dep" shows phrasal types and grammatical functions of these elements.

Figure 2: Valence table of the construction

The hyper-linked numbers on the left of the valences in Figure 2 are linked with example sentences with the valences, so that the user can easily access example sentences with a particular valence by clicking on its link.

4. Search Functions

There are many tools that can process linguistic data in the XML format. CxN Viewer is tailored for the BFN database and its advantage is the ability to search different fields of different tables of the database with a simple search.

4.1 Searching Construction Names

There is a text box for searching in the upper-left pane of CxN Viewer. The user can type a word in the box, and CxN Viewer searches **construction names**, **definitions of constructions** and **contents of example sentences** for the search string. In Figure 3, the user inputs 'subject' in the search box and below the box appear search results of construction names that match the search string.
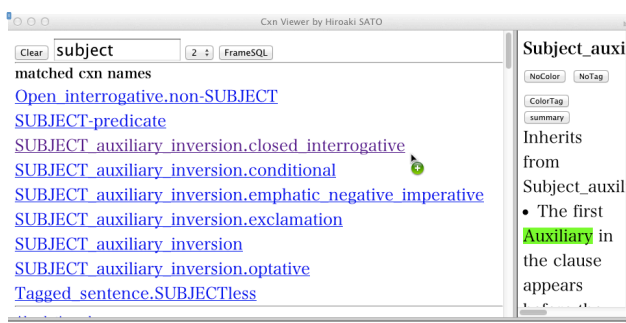

Figure 3: Search results of the string 'subject'

The five *Subject Auxiliary Inversion* constructions in Figure 3 are derived from the abstract *Subject Auxiliary Inversion* construction, which creates signs that can serve as components of a large variety of specific constructions with meanings like question (3a), condition (3b), negative imperative (3c), exclamation (3d), wish (3e), etc. (Fillmore 1999).

3 a. [**Auxiliary** Do] [**Subject** I] [**Remainder** need any training]?
Subject_auxiliary_inversion.closed_interrogative

b. [**Auxiliary** Were] [**Subject** it] [**Remainder** an elephant], our feet would be in its midriff, our heads just under its rib-cage. *Subject_auxiliary_inversion.conditional*

c. [**Auxiliary** Don't] [**Subject** you] [**Remainder** be patronizing with me]. *Subject_auxiliary_inversion. emphatic_negative_imperative*

d. [**Conventional_exclamation** Boy], [**Auxiliary** do] [**Subject** they] [**Remainder** have a surprise coming their way].
Subject_auxiliary_inversion.exclamation

e. [**Auxiliary** May] [**Subject** the Force] [**Remainder** be with you]. *Subject_auxiliary_inversion.optative*

None of the *Subject Auxiliary Inversion* constructions has the construction-evoking element, and thus there is no CEE in (3). The user can access the annotated sentences in (3) by clicking on the hyper-linked construction names in Figure 3.

4.2 Searching Construction Definitions

CxN Viewer can search keywords in definitions of constructions, and it helps the user to understand syntactic and semantic properties of each construction. There are many parts in definitions of constructions which refer to *frame*. When the user searches for the string 'frame' with CxN Viewer, it shows a list of definitions which contain the search string, as in Figure 4.
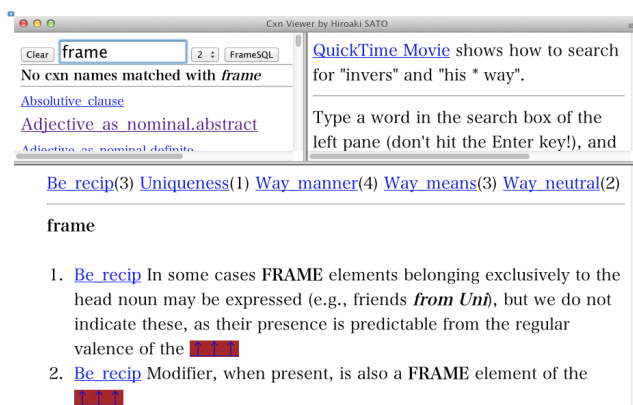

Figure 4: Search results of the string 'frame'

The lower pane of Figure 4 shows that the search string 'frame' is used in the definition of the *Be_recip* construction three time, in that of the *Uniqueness* construction once, in that of the *Way_manner* construction four times, in that of the *Way_means* construction three times, and the *Way_neutral*

construction twice. Each name of these constructions is hyper-linked with its part of the definition which contains the search string. For example, when the user clicks on Way_manner(4), the viewer shows excerpts of the definition, as in Figure 5.
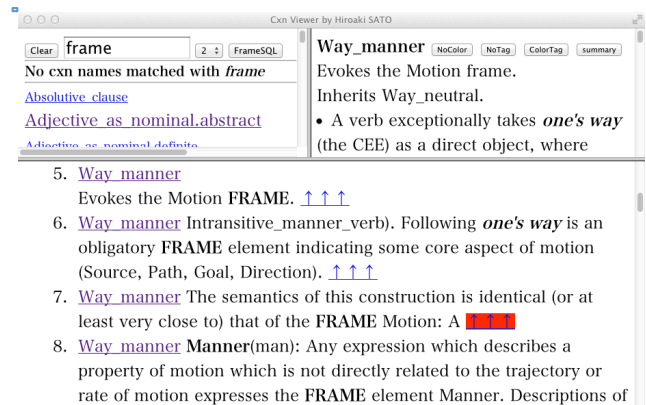

Figure 5: Excerpts of the definition containning 'frame'

4.3 Searching Example Sentences

Sometimes CE's that are conceptually salient do not show up in the example sentence. They are annotated in Constructicon as null instantiation in the same way as in FrameNet lexical annotation. The definite null instantiation (DNI) is the one whose missing element must be something that is already understood in the linguistic or discourse context (Ruppenhofer et al. 2006: Appendix A). When the user searches for 'cni', CxN Viewer shows a list of constructions whose example sentences contain 'cni', as in Figure 6.
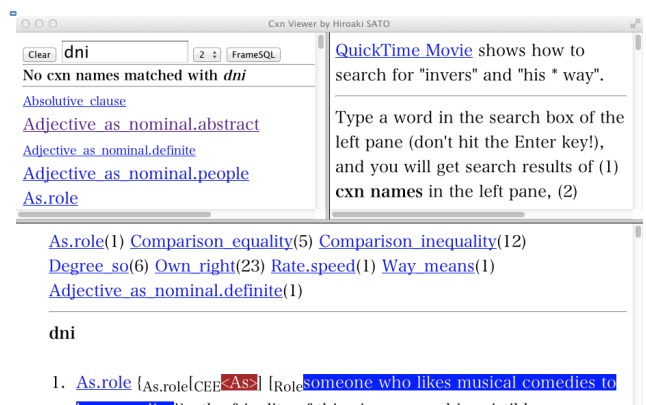

Figure 6: Search results of 'dni'

Each name of the constructions in Figure 6 is hyper-linked with example sentences which contain the search string. The user can also search other elements such as 'supp' (support verb), 'gov' (governor), 'CNI' (Constructional Null Instantiation) which are also used in Constructicon annotation, as well as in FrameNet lexical annotation.

5. Parallelism between Constructicon and FrameNet lexical annotation

Constructicon has many parallels to FrameNet lexical

annotation. Construction definitions, which describe constructions and their components, correspond to frame definitions in FrameNet lexical annotation, which describe frames and their components. As I mentioned in section 2, CE's of Constructicon correspond to frame elements of FrameNet lexical annotation, and the construction-evoking element of Constructicon corresponds to the lexical unit in FrameNet lexical annotation.

The annotation processes of Constructicon and FrameNet lexical units are also similar. Both of them use the same software tools for annotation. Annotators mainly search the British National Corpus for illustrative sentences for annotation, except for annotating continuous text.

Figure 2 in section 3.2 shows how CxN Viewer displays valence patterns of a construction, and FrameSQL (Sato 2008), a search tool for FrameNet lexical annotation, displays valence patterns of FrameNet lexical annotation in a similar manner, as in Figure 7.

Figure 7: Valence table of FrameSQL

FrameSQL links each valence pattern to its set of annotated sentences. The user-interface of FrameSQL is similar to that of CxN Viewer, and there are mutual links between these two search tools.

Frame-to-frame relations in FrameNet lexical annotation, which capture directed relations between two frames, correspond to constructicon-to-constructicon relations. For example, the *Self_motion* frame, which is evoked by such a lexical unit as the verb 'walk', inherits some of its basic properties from its parental and more abstract *Motion* frame, which is evoked by such a lexical unit as the verb 'move'. In a similar manner, the *Way_manner* construction, which I mentioned in section 2 and 3.2, inherits some of its basic properties from its parental and more abstract the *Way_neutral* construction as in (4).

(4) a. She carefully made her way through the cluttered room to the altar at the north end.
   b. [Theme She] [manner carefully] {[Make-verb made] <Construction-evoking element her way>} [Path through the cluttered room] [Goal to the altar at the north end].

6. Conclusion

CxN Viewer is a useful program that will benefit researchers working with the data annotated within the framework of BFN. CxN Viewer is a unicode-compliant application, and it can deal with constructions of other languages. The Spanish FrameNet Project (http://gemini.uab.es:9080/SFNsite/) annotates several Spanish constructions and the same CxN Viewer can search and display its annotation in the same way on the web browser, as in Figure 8.
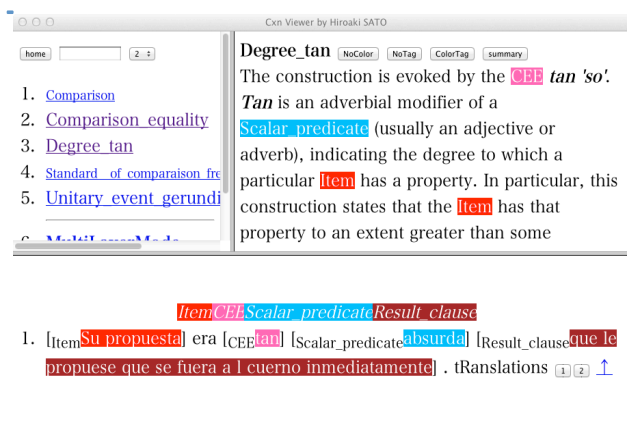

Figure 8: CxN Viewer for Spanish FrameNet Project

7. References

Fillmore, Ch. J. (1999). Inversion and Constructional Inheritance. In G. Webelhuth, J.P. Koenig and A. Kathol (Eds.), *Lexical and Constructional Aspects of Linguistic Explanation*, Stanford: CSLI, pp. 113--128.

Fillmore, Ch. J., Lee-Goldman, R. R., and Rhodes, R. (2012). The FrameNet Constructicon. To appear in Boas, H.C. and Sag, I.A. (Eds.), *Sign-based Construction Grammar*, Stanford: CSLI.

Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: Chicago University Press.

Lee-Goldman, R. R., and Rhodes, R. (2009). Corpus-based Analysis and Annotation of Constructions (Slides presented at Fillmore FEST: Conference in honor of Charles Fillmore) 31 July-2 August 2009. University of California, Berkeley.

Ruppenhofer, J. K., Ellsworth, M., Petruck, M. R. L. and Johnson, C. R. (2006). *FrameNet: Theory and Practice*. Retrieved February 22nd, 2006 from http://framenet.icsi.berkeley.edu/.

Sag, I. A. (2012). Sign-Based Construction Grammar: An Informal Synopsis. To appear in Boas, H.C. and Sag, I.A. (Eds.), *Sign-based Construction Grammar*, Stanford: CSLI.

Sato, H. (2008). New Functions of FrameSQL for Multilingual FrameNets. In P*roceedings of the Sixth Language Resources and Evaluation Conference*, Marrakech Morocco: European Language Resources Association.