

JFNWAT Web Report: A New Japanese FrameNet Web Application Tool

Lidia Pierre

IMT Atlantique

lidia.pierre@imt-atlantique.net

Kyoko Ohara

Keio University

ohara@hc.st.keio.ac.jp

Abstract

This paper introduces the JFNWAT Web Report, a web-based tool to display and query annotated data from the Japanese FrameNet database. It is part of the Japanese FrameNet Web Application Tool (JFNWAT), which ultimately aims to replace the existing JFN production and reporting tools. JFNWAT is designed to be language independent, meaning that the Web Report module could be used for non-Japanese FrameNet projects.

1 Introduction

The Japanese FrameNet project aims at creating a lexicon of Japanese language, supported by annotated corpus data (Ohara, 2016; Ohara et al., 2004). It complies with the Frame Semantics theory and works in collaboration with the English-based FrameNet project (Baker et al., 1998). The produced data follows the structure of FrameNet data (Baker et al., 2003) and contains a collection of annotated sentences extracted from the Balanced Corpus of Contemporary Written Japanese (BCCWJ) core data. A *frame* refers to a schematic representation of a situation, each of its participants being a *frame element*. A frame element along with its grammatical function and phrase type in a sentence forms a *valence*, and the *valence pattern* is the set of valence groups realized in one sentence. Finally, frame-evoking words are called *lexical units*.

In this paper we present the JFNWAT Web Report, a new web-based application to browse JFN data, reflecting the content of the underlying database in real-time. Based on the existing FrameNet Web Report, it is designed to be a more user-friendly tool, and includes additional features to facilitate the search. It supports look-up by *frame*, *lexical unit* and *full-text* searches.

2 FrameNet Data

The FrameNet data includes frames, with their respective frame elements, and eventual frame relations. Frame elements can be ‘core’, meaning that they are essential to the meaning of a frame, or not. The annotation data is organized in annotation sets, which links a sentence and a lexical unit.

While the Japanese FrameNet Production tools (such as the JFNDesktop, the annotation software) are based on the FrameNet data model, JFNWAT and the Web Report module rely on a MySQL database with a slightly different structure (Kabach et al., 2015). The new design addresses limitations of the existing model such as the absence of foreign keys, and keeps high constraints on data integrity. A simple SQL script can migrate data from the original data model to the new one.

3 JFNWAT overview

3.1 Architecture

JFNWAT is a Java web application featuring a user-friendly web interface and designed as a modular unified tool for the Japanese FrameNet tasks. The Web Report module connects to a MySQL database to run queries and display annotation data.

3.2 Underlying technologies

JFNWAT has been created with the Spring Boot framework, allowing for a lightweight application with minimal configuration and easy deployment. The data access layer of the application is managed by the Spring Data JPA, which automatically creates the implementation of the methods performing queries on the data, directly from the repository interfaces. Object-relational mapping to convert MySQL data to Java objects is performed with Hibernate. The controller layer uses the JavaServer Faces specification, which processes the requests, loads the appropriate views,

builds and renders the response. Finally, the library for the user interface is PrimeFaces, a set of JavaServer Faces components used with the Adamantium layout.

The Web Report module has support for both English and Japanese language and it can be run as an independent application.

3.3 Authentication

Authentication for the JFNWAT is managed with the Spring Security framework, user names and encrypted passwords (standard MD5 algorithm) are stored in the database.

4 Web Report Interface

4.1 Home Page

The top page of the Web Report gives access to its three main sections: the Frame Index, the Lexical Unit Index and the Document Index and allows users to switch language.

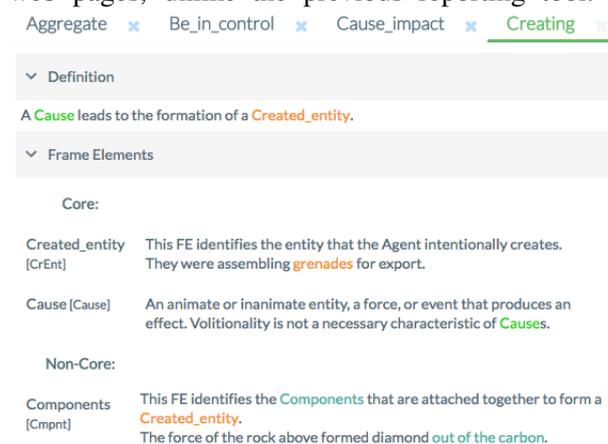


The search bar provides a convenient way to look for existing frames, lexical units or documents, and the search is facilitated by an auto-completion function. After entering a query, the user is redirected to the result page, which presents an appropriate list of frames, lexical units and documents.

4.2 Frame Index

The Frame Index provides complete description of frames: name, definition, frame elements (divided into core and non-core elements), frame relations and the lexical units associated with the frame. The frame related names and the lexical units can be clicked to quickly jump to their associated entry reports. The left panel displays the list of existing frames. It supports a new feature for real-time filtering based on the frame names. Every frame description belongs to a tab in the main panel, from which

users can switch without having to open new web pages, unlike the previous reporting tool.



4.3 Lexical Unit Index

The structure of the lexical unit index is comparable to the frame index. In the left-side panel a list of all existing lexical units can be filtered based on the lexical unit lexeme but also on the frame it belongs to. The main panel is organized in tabs, each tab displaying details of one lexical unit. Instead of automatically pulling all the annotation available, the user can make a selection based on the frame element, their specific realizations and valence patterns.

As a first step, the user chooses between getting all frame elements, core only or non-core only. Then an additional filter enables to keep only specific realizations (grammatical function and phrase types) of the frame elements. This results in a list of possible valence patterns.

	Theme	Manner	Goal
2	NP	AVP	NP
1	Ext ガ	Dep へ	Dep へ

When selecting a valence pattern, a new panel appears at the bottom of the page, containing the annotated corresponding sentences. Particular attention has been paid to improve the appearance of the annotated sentences. In the new panel, frame elements names are placed under their corresponding text segment while the target lexical unit is highlighted in a box. Explanations concerning the frame elements can be seen when clicking on the names.



Sentences can be added/removed from the panel or downloaded in a .txt file.

The complementary panel is not linked to a unique lexical unit report but stays open in the other tabs of the index as well, meaning that a user can display annotated sentences for different LUs at the same time which makes it easier for comparisons.

4.4 Document Index

The document index reflects the annotation of continuous texts extracted from BCCWJ and is designed to replace the existing JFN Full Text application.

In a similar way, the index is split into a left panel to filter documents on their name and the corpus they belong to, and a main panel divided into tabs for every document selected. A document is in fact an ordered list of sentences and each sentence occupies a new row. If an annotation is available, the target lexical unit is highlighted with a bold font. Users can turn on/off the annotation for a LU by clicking on it, but only one annotation at a time can be displayed for one sentence.



The annotation style is consistent with the lexical unit index, with frame elements name appearing under their associated text segment and target LUs in a green box. The evoked frame is also displayed and links to the matching frame index report.

5 Conclusion and Future Work

In this paper we presented a new reporting tool to browse Japanese FrameNet data, with additional features and low maintenance and deployment costs. The JFNWAT Web Report is designed as a module of a larger web-based application, JFNWAT, that aims to merge and replace the existing JFN tools for annotation creation and display. Relying on an improved data model, the ultimate goal of JFNWAT is to facilitate the annotation process thanks to a user-friendly interface while keeping the highest quality of annotated data. Besides the Web Report, JFNWAT currently includes a concordancer module (KWIC) and the annotation tool is now under development. In the future, the

JFNWAT application and its independent modules could be used by non-Japanese FrameNet projects, given a database migration script.

References

- Baker, Collin F.; Fillmore, Charles J. and Lowe, John B. 1998. *The Berkeley FrameNet Project*. Proceedings of ACL-COLING.
- Baker, Collin F.; Fillmore, Charles J. and Cronin, Beau. 2003. *The Structure of the FrameNet Database*. International Journal of Lexicography.
- Kabbach, Alexandre and Ohara, Kyoko Hirose. 2015. *Japanese FrameNet Data Quality Overview: Toward Full Data Integrity*. MAPLEX2015. Yamagata, Japan.
- Ohara, Kyoko Hirose. 2016. *Japanese FrameNet. Paper given at the Pre-conference Tutorial Multilingual FrameNet: Linguistic Insights, Computational Challenges, and Applications* 10th Language Resources and Evaluation Conference (LREC 2016). Portoroz, Slovenia
- Ohara, Kyoko Hirose; Fujii, Seiko; Ohori, Toshio; Suzuki, Ryoko; Saito, Hiroaki and Ishizaki, Shun. 2004. *Japanese FrameNet Project: An Introduction*. Proceedings of the Workshop on Building Lexical Resources from Semantically Annotated Corpora at LREC.