

About FrameNet

The FrameNet project 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.

Contents

Vision.....	P4
Mission	P4
Values	P4
Word Usages.....	P5
1. Dictionary.....	P5
2. Semantic Roles	P5
3. Valence.....	P5
Administrative Information	P5

FrameNet (FN)

Stakeholder(s):

International Computer Science Institute :

The project has been in operation at the International Computer Science Institute in Berkeley since 1997, supported primarily by the National Science Foundation, and the data is freely available for download.

National Science Foundation

Researchers :

It has been downloaded and used by researchers around the world for a wide variety of purposes (see FrameNet downloaders). FrameNet-like databases have been built for a number of languages (see FrameNets in other languages) and a new project is working on aligning the FrameNets across languages.

French FrameNet :

The ASFALDA French FrameNet was built within the ASFALDA project (2013-2016), funded by ANR and the Empirical Foundations of Linguistics Labex. The French FrameNet focuses on 4 notional domains (verbal communication, commercial transactions, cognitive stances and causality). The objective of the project was to exhaustively cover these four domains, in terms of pertaining frames, lexical units and annotations on pre-existing syntactic treebanks. The 1.1 release contains 105 frames, mainly adapted from Berkeley FrameNet 1.5, a lexicon of 1100 lexical units, and about 16000 annotation sets with frames and core frame elements.

Chinese FrameNet :

Chinese FrameNet (CFN) is a lexical database comprising frames, lexical units, and annotated sentences. It is based on the theory of Frame Semantics, making reference to the English FrameNet work in Berkeley, and supported by evidence from a large Chinese corpus. CFN currently contains 323 semantic frames, 3947 lexical units, more than 18,000 sentences annotated with both syntactic and frame-semantic information, covering both the common core of the language and the more specialized domains of tourism, on-line book sales, and law, and 200 annotated discourses. In addition to building the CFN database, they are studying the theory of frame semantics as it relates to the Chinese language and researching techniques for building applications based on CFN. They have developed frame semantic role labeling systems for both individual sentences and discourses.

FrameNet Brasil :

A project at the Universidade Federal de Juiz de Fora (UFJF) in Brazil has been working to create a FrameNet-style lexical database for Brazilian Portuguese, in collaboration with the FrameNet team at ICSI. They have created a corpus of roughly 104 million words of Brazilian Portuguese, comprising written text, transcribed speech, and movie subtitles. They have also created their own annotation software. Two M.A. theses have already been written in connection with the project and two more M.A. theses and two Ph.D. dissertations are in progress. The first Portuguese data, some 32 frames and 38 lexical units, are now available for the general public through the link "Data" in the Main Menu of their website, FrameNet Brasil. In addition to building the FN Brasil database, two related projects are being carried out by the Brazilian team: 1. "Frames

and Constructions", aimed at annotating grammatical constructions in FrameNet Brasil, the beginning of a Brazilian Portuguese Constructicon. 2. "Copa 2014", focused on the development of a trilingual electronic dictionary for use during next FIFA World Cup, to be held in Brazil in 2014. This is joint work with the FrameCorp team at UNISINOS and the Computer Sciences Department at UFJF in consultation with the FrameNet team at ICSI. The World Cup Dictionary is now available online at www.framenetbrazildictionary.com

German FrameNet(s) :

Currently three research groups are collaboratively investigating FrameNet for German, with different foci:

- *The largest of these is the SALSA Project in Saarbrücken. Their goals are to provide a large, frame-based lexicon for German, with rich semantic and syntactic properties, as a resource for linguistic and computational linguistic research, to investigate probabilistic and hybrid methods for wide-coverage semantic annotation, and to explore the use of frame semantic annotations for dynamic semantic analysis in practical NLP applications, especially information access. For more information, see their overview page.*
- *In Stuttgart, the group headed by Uli Heid is working on corpus tools and extraction techniques, particularly for the investigation of collocations and nominalizations.*
- *The German FrameNet at Austin, under the direction of Prof. Hans C. Boas, uses the Saarbrücken data as a starting point and is creating a detailed German FrameNet database that employs the Berkeley software and methodology and is based on a much larger corpus. That database will be used in Saarbrücken to study automatic lexicon acquisition, and in Stuttgart to investigate morphological productivity. Since the project will involve teams in the U.S. and Germany working together, German FrameNet will create web-based tools that support the linguistic work in an internationally distributed environment, including utilities for linguists as well as project management utilities as required.*

Spanish FrameNet :

Spanish FrameNet, headquartered at the Department of Linguistics of the Autonomous University of Barcelona, includes researchers from several Spanish universities working in cooperation with the Berkeley FrameNet Project. It is financed by the Dept. of Science and Technology of Spain for the period September 2002 - September 2005. They have constructed a 300 million-word corpus which they can search for examples and are working on a chunker.

Japanese FrameNet :

Japanese FrameNet aims at building a lexicon that records the valence descriptions of Japanese words, based on Frame Semantics and corpus data. After a preliminary study funded by a grant from Keio University between November 2000 and

— continued next page

Stakeholders (continued)

March 2002, the Japanese FrameNet project was launched in July 2002 by a grant from the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) and Keio University. It is currently supported by a grants from MEXT. JFN works in collaboration with the English-based FrameNet project, an ongoing project undertaken at the International Computer Science Institute in Berkeley, California (<http://framenet.icsi.berkeley.edu/>). The ultimate goal of JFN is to produce a FrameNet-style database of Japanese lexical units. The resulting database will thus contain valence descriptions of Japanese lexical units and a collection of annotated corpus attestations. Important research questions being asked by JFN are to what extent the Frame-semantic approach is suitable for analyzing Japanese lexicon and also to what extent the existing English-based semantic frames are applicable to characterizing Japanese lexical units. Also, while purporting to retain the richness of semantic information in FN, JFN pays close attention to typological differences in lexicalization patterns between Japanese and English. JFN is currently analyzing and annotating basic content words in Japanese as well as annotating texts from the Balanced Corpus of Contemporary Written Japanese (BCCWJ) core data. Also, JFN has started a pilot study of "constructicon" building of Japanese. (cf. <http://framenet.icsi.berkeley.edu/>).

Swedish FrameNet :

Researchers at Gothenburg University are building a Swedish FrameNet mainly based on existing Swedish lexica, but using Berkeley FrameNet frames and frame elements. The data is freely downloadable in several formats; as of 2010.03.16, they have more than 2,300 LUs in 51 frames. The lexicon comes with example (Swedish) sentences for each core FE of the frames which they have done.

Korean FrameNet :

This project is being carried out by KAIST, and has a very clear website displaying their data, including annotated sentences. They have several thousand LUs in frames corresponding to English FN frames. They write, "The aim of this website is to thoroughly demonstrate the step-by-step procedures involved in the process of developing Korean FrameNet that is of great potential use as training and testing data for improved semantic analysis of natural language texts in Korean. The website explicitly explores the course of our approach to manually constructing FrameNet for Korean through translation of the English FrameNet, followed by making suggestions for further extensive studies on the application of Korean FrameNet to various natural language processing (NLP) tasks."

FrameNet Users :

https://framenet.icsi.berkeley.edu/fndrupal/framenet_users

Vision

The form, meaning, and use of words is well documented and understood

Mission

To build a lexical database of English that is both human- and machine-readable

Values

Lexicology

Human-Readability

Machine-Readability

Word Usages

Annotate examples of how words are used in actual texts

1. Dictionary

Provide a dictionary of word senses

Stakeholder(s):

Students :

From the student's point of view, it is a dictionary of more than 13,000 word senses, most of them with

annotated examples that show the meaning and usage.

2. Semantic Roles

Provide a training dataset for semantic role labeling

Stakeholder(s):

NLP Researchers :

For the researcher in Natural Language Processing, the more than 200,000 manually annotated sentences linked to more than 1,200 semantic frames provide a unique training dataset for semantic role labeling,

used in applications such as information extraction, machine translation, event recognition, sentiment analysis, etc.

3. Valence

Serve as a valence dictionary

Valence ~ the number of grammatical elements with which a particular word, especially a verb, combines in a sentence.

Stakeholder(s):

Linguistics Teachers :

For students and teachers of linguistics it serves as a valence dictionary, with uniquely detailed evidence

for the combinatorial properties of a core set of the English vocabulary.

Linguistics Students

Administrative Information

Start Date:

End Date:

Publication Date: 2021-09-06

Source: <https://framenet.icsi.berkeley.edu/fndrupal/about>

Submitter:

Given Name: Owen

Surname: Ambur

Email: Owen.Ambur@verizon.net

Phone: