The FrameNet Database and Software Tools

Josef Ruppenhofer, Collin F. Baker, and Charles J. Fillmore

International Computer Science Institute
1947 Center St.
Berkeley, CA 94704-1198, USA
{josef, collinb, fillmore}@icsi.berkeley.edu
website: http://framenet.icsi.berkeley.edu/~framenet

Abstract

The FrameNet Project is producing a lexicon of English for both human use and NLP applications, based on the principles of Frame Semantics, in which sentences are described on the basis of predicators which evoke semantic **frames** and other constituents which express the participants (**frame elements**) in these frames. Our lexicon contains detailed information about the possible syntactic realizations of frame elements, derived from annotated corpus examples. In the process, we have developed a suite of tools for the definition of semantic frames, for annotating sentences, for searching the results, and for creating a variety of reports. We will discuss the conceptual basis of our work and demonstrate the tools we work with, the results we produce, and how they may be of use to other NLP projects.

Introduction

The FrameNet (FN) research project [Fillmore and Baker 2001, Baker et al. 1998, Lowe et al. 1997][1] is building a lexical resource that aims to provide, for a significant portion of the vocabulary of contemporary English, a body of semantically and syntactically annotated sentences from which reliable information can be reported on the **valences** (combinatorial possibilities) of each item to be analyzed.

The project is committed to a descriptive framework based on semantic frames [Fillmore 1985, Fillmore and Atkins 1988] and to documenting its observations on the basis of carefully annotated attestations taken from corpora. A semantic **frame**, (henceforth simply frame) is a script-like structure of inferences, linked by linguistic convention to the meanings of linguistic units -- in our case, lexical items. Each frame identifies a set of **frame elements** (FEs) -- participants and props in the frame. A frame semantic description of a lexical item identifies the frames which underlie a given meaning and specifies the ways in which FEs, and constellations of FEs, are realized in structures headed by the word. Generalizations about frame structure and grammatical organization are derived automatically from a large body of annotated sentences, each of these annotated to show one combinatory arrangement for the particular targeted word.

Corpora and Software

For the first part of the project, The British National Corpus (BNC, http://info.ox.ac.uk/bnc) was used, courtesy of Oxford University Press. For our continuing work, we are depending on both the BNC and the corpora of English news texts provided by the LDC; eventually we hope to be able to add the full resources of the American National Corpus (http://www.cs.vassar.edu/ide/anc/). The project has used an in-house user interface to run the Corpus Workbench software from Institut für Maschinelle Sprachverarbeitung of the

University of Stuttgart for searching the corpora and creating subcorpora of sentences for annotation. In the first phase of the project, Annotation was done using the Alembic Workbench tools from MITRE; for the second phase, FN staff have written a custom Java GUI annotation tool as a front-end to the MySQL database. Java GUIs have also been written for frame and frame element

What is to be annotated

Because FN is primarily lexicographic, we are not attempting to annotate whole texts or even a random sample of sentences which include each lemma. Rather, we want to annotate a set of sentences which exemplify the range of combinatorial possibilities of a lexical unit, including all the types of syntactic constituents which can embody the frame elements. We regard the following constituents as worthy of annotation for our purposes:

- For verbs, nouns, adjectives and prepositions: their post-head complements, i.e., constituents of the phrase headed by the target (within the VP, NP, AP or PP) which amplify our understanding of the frame evoked by the head.
- For verbs, constituents external to the VP which instantiate an FE of the verb, either directly (as the verb's subject) or indirectly (by being a direct argument of a predicate which governs the VP).
- For nouns, we annotate frame-relevant possessive determiners (*[Roger's] decision to join the party*), relational adjectives (*[environmental] protection*) and modifier nouns in compounds (*[environment] protection*).
- For nouns which occur with support verbs, we annotate constituents which are complements of the noun or the verb.

Implicit FEs

Certain FEs are regarded as core FEs for a particular frame, in that they are always conceptually present; yet they are not always expressed in each occurrence of a predicator which evokes the frame. We find it useful to distinguish three conditions under which such arguments can be omitted:

- Constructional, where the absence of a constituent representing a particular FE was authorized by the grammar of the language: e.g. the missing subject of an imperative sentence (*Say something!*).
- Existential, where the missing element could be given a generic or existential interpretation, such as when the objects of certain common verbs are not mentioned: sew, eat, bake, etc.
- Anaphoric, in which the missing element has to be "understood" or "given" in the discourse context.

Description of the FrameNet Database

The FrameNet database is distributed in two parts, the frame database, covering approximately 300 semantic frames, and the lexical database, comprising roughly 5,000 lexical units.

The frame database contains, for each frame, its name and description, a list of frame elements, each with a description and examples, and information about relations among them. The most important relations include frame inheritance (ISA, with inheritance of FEs from parent to child) and frame composition (PART-OF, with optional bindings between FEs of subframes and those of the complex frame they constitute).

The lexical database consists of a lexicon with entries for nouns, verbs, and adjectives. Each entry represents a **lexical unit**, a pairing of a lemma with a semantic frame (i.e. one sense of a word). Each entry details the FEs that can occur with a particular lexical unit and the syntactic patterns in which they can occur, in terms of phrase type and grammatical function. Every such pattern is supported by annotated examples from a corpus (averaging more than 20 examples per lexical unit).

This section of the demo will describe the database in some detail, including internal database structure (part of which is shown in Fig. 1), the format of the XML files used for the distribution, and give instructions for obtaining copies. Because the data for the second phase of the FrameNet project are much more complex than those for the first phase, a new XML format has been used, allowing the representation of multiple layers, overlapping labels, frame inheritance, etc. We are also interested in making the FN data accessible through the "Smart Web"; to this end we are in the process of adding RDF using DAML+OIL to our XML representation. Some preliminary versions of this format will be displayed.

Software to be Demonstrated

Web-based Report System
Reports are generated from the database providing various views of the data. Of particular interest is the Lexical Entry report, which concisely shows the definition, the FEs, and the valence patterns (FEs in particular combinations of phrase type and grammatical function), with links to the annotated sentences supporting each line in these tables.

Other reports give the complete description of a frame and its FEs, and provide convenient ways to look up frames from lemmas and lemmas from frames.

Frame Editing Tools
We will demonstrate the frame editor, which uses a GUI written in Java to edit the tables in a MySQL database which represent the frames, frame elements, lemmas and lexical units being described. The editor not only facilitates creating these units, but also establishing relations among them, i.e. frame inheritance and frame composition as mentioned above.

Demonstration of Manual Frame Element Annotation
We will show the process of annotation used in the daily work of the project, using a different GUI which adds data to different tables in the MySQL database, representing sentences and labels attached to them. Information regarding POS tags, location of the target lemma, and FEs is represented by labels in several layers associated with each sentence. The annotation software uses multiple layers that allow not only overlapping FE labels and

discrepancies between syntactic and semantic constituents, but also multiple targets lemmas within a sentence, each associated with a separate set of annotation layers.
We will briefly discuss the policies for choosing appropriate sentences, delimiting frame elements, defining grammatical functions, etc.

The FrameSQL Tools for Searching the FN Database

FrameSQL is a web interface written by Hiroaki Sato of Senshu University., Japan, which allows the user to search the FN database in a variety of ways; there are two levels of complexity, depending on the needs and sophistication of the user. Search parameters include the frame, the lemma, the FEs, specific phrase types and grammatical functions of FEs, the head noun of a particular FE, etc. We will demonstrate how to use FrameSQL for queries such as "find all example sentences containing verbs in the Communication frame whose Addressees are expressed as direct objects".

Automatic Frame Element Recognition

In an effort to speed up the annotation process, we are developing a system for recognition of frame elements using rules specified a priori by lexicographers, based partially on introspection and partially on preliminary corpus searching. For example, in annotating the lemma *tell*, with sentences such as
1. The president told the reporters the answer to the question.
2. The president told the story to the Cabinet later that morning.
the lexicographer recognizes two possible valence patterns, V NP NP (ditransitive) and V NP to NP; in Ex. 2, the preposition to marks the Addressee, while in Ex. 1, the presence of two NPs signals that the first must be the addressee. A set of simple rules, functioning as a cascaded filter, can be written to automatically mark frame elements which match portions of the rules. The annotation task can then be reduced (in most instances) to approving, disapproving, or editing the pre-labeled sentences. Results from this rule-based system will be compared to those from a different system, created by Dan Gildea [Gildea and Jurafsky 200, Gildea 2001], using an algorithm that learns from lexical units that have already been annotated.

Acknowledgements
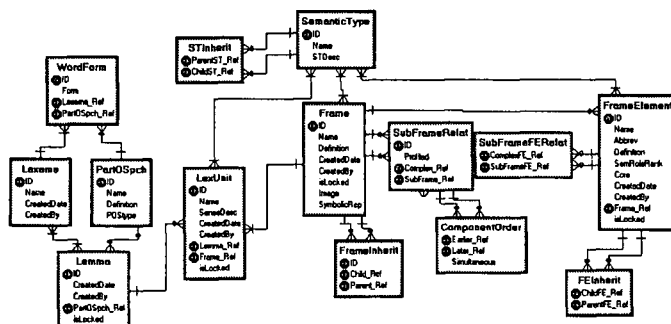
Fig. 1 Structure of the FrameNet Database (partial)

References

[Baker et al. 1998] Baker, C.F., C.J. Fillmore & J.B. Lowe, 1998. The Berkeley FrameNet Project, in: *COLING-ACL Í98: Proceedings of the Conference, held at the University of Montreal*, pp. 86-90, Association for Computational Linguistics, Montreal.

[Fillmore 1985] Fillmore, C.J. 1985, Frames and the Semantics of Understanding, in: *Quaderni di Semantica* VI.2

[Fillmore & Atkins 1998] Fillmore, C.J. & B.T.S. Atkins 1998, FrameNet and Lexicographic Relevance, in: *Proceedings of the First International Conference on Language Resources And Evaluation*. Granada, Spain.

[Fillmore & Baker 2001] Fillmore, C.J. & C.F. Baker 2001, Frame Semantics for Text Understanding, in: *Proceedings of WordNet and Other Lexical Resources Workshop*, held at North American Association for Computational Linguistics, Pittsburgh.

[Fillmore et al. 2001] Fillmore, C.J., C. Wooters & C.F. Baker 2001, Building a Large Lexical Database Which Provides Deep Semantics, in: B. Tsou & O. Kwong (eds.), *Proceedings of the 15th Pacific Asia Conference on Language, Information and Computation*. Hong Kong.

[Gildea & Jurafsky 2000] Gildea, Daniel and Daniel Jurafsky. 2000, Automatic Labeling of Semantic Roles, In Proceedings of the ACL 2000, Hong Kong.

[Lowe et al. 1997] Lowe, John B. and Collin F. Baker and Charles J. Fillmore (1997) A frame-semantic approach to semantic annotation, in: Marc Light (ed.), *Tagging Text with Lexical Semantics: Why, What and How?* Special Interest Group on the Lexicon, Association for Computational Linguistics.

Endnotes

[1] The full texts of most references in this paper are available at http://framenet.icsi.berkeley.edu/~framenet/Papers.html.