

Welcome to Release 1.3 of the FrameNet data

Thank you for your interest in FrameNet; we hope that you will find it both interesting and useful in your work.

R1.3 contains a number of changes that bring the data closer in line with the full range of frame semantics, and should also make it more suitable for training machine learning algorithms and providing data for inferencing.

	R1.2	R1.3	Change
frames	609	795	31%
(non-lexical)	58	74	28%
FEs	4909	7124	45%
FE/lexical frame	8.91	9.88	11%
pct. non-lexical	9.5%	9.3%	-2%
frame relations	914	1723	89%
FE relations	4182	8486	103%
Lexical Units	8869	10195	15%
LU/lexical frame	16.1	14.14	-12%
LUs w/ lexicog anno	6642	6815	3%
prop. LUs w/ lexicog. anno	74.9%	66.8%	-11%
AnnoSets in lexicog anno	133846	139439	4%
lexicog AnnoSets/annotated LU	20.2	20.5	2%
AnnoSets in full text anno	0	11671	
Total AnnoSets	133846	151110	13%

Table 1: Some statistics comparing Releases 1.2 and 1.3

1 More lexical units, frames, and FEs.

As shown in the summary table above, the number of lexical units has increased to more than 10,000, while the numbers of frames and FEs have increased by more than 30% to 795 and 7125 respectively. Some frames are marked as “non-lexical”, meaning that we have created them because they are logically necessary in the frame hierarchy, even though they do not contain lexical units. The proportion of non-lexical frames has remained roughly constant, at a little less than 10%.

The number of LUs per lexical frame has decreased from about 16 to about 14. This may be due in part to the gradual extension of FrameNet to more specialized terminology (although we continue to work in many domains simultaneously, so that FN remains as domain-independent as possible). The percentage of LUs with lexicographic annotation has also decreased, primarily because, as we annotate full texts, we need to create new LUs to cover vocabulary found in them, and may not have the resources to keep up the lexicographic annotation at the same pace. For those LUs that do receive full lexicographic treatment, the number of annotated sentences has remained roughly constant at 20 per lexical unit.

2 Greatly expanded frame-to-frame (and FE-to-FE) relations

Over the last year, we have made a major effort to link all the frames to one another, as reflected in the summary table. In fact, we have linked more than more than 750 of the frames into a single connected graph via these relations, with the remainder falling into roughly 20 smaller groupings.

Not all of these relations are hierarchical; this is consistent with our conception of frame semantics, in which two frames A and B may be related in a variety of ways. For example B and C may both be subframes (subevents) of a complex frame A (a complex event), and B may further connected to C by a Precedes relation, indicating their temporal order. (For full details of frame-frame and FE-FE relations, please see Chapter 6 of “FrameNet Theory and Practice” 2007, [book.pdf](#) in this directory, and hereafter referred to as “the Book”).

3 Major improvements in consistency and completeness

We have also devoted a lot of attention to ensuring that the data in the database is consistent and complete in a number of aspects, and this is reflected in this data release. We are using two different approaches for this checking: for the annotation data, we use a suite of checking programs, but for the frames, FEs and LUs themselves, we use a single program with a long list of declarative rules. Further details about the principles and implementation of the consistency checking can be found in the Appendix to the Book, and in two other documents in this directory, [fnDbConsistencyChecking.html](#) and [annotationCheckers.pdf](#).

4 Full text annotation

We have been working for more than two years on annotating continuous texts. This sort of annotation is quite different from the annotation of lexicographic examples. In lexicographic annotation, we are interested in selecting clear, easily understandable sentences that exemplify the valences of the predicator and do not involve lexicographically irrelevant complexity. For example, we tend to avoid passive sentences, since the ability to be passivized is true of (almost) all transitive verbs, and reveals nothing that is frame- or verb-specific. We also tend to prefer sentences which are not too long, and where the predicator in question is in a main clause, rather than deeply embedded. Since the sentences are annotated as examples of a single predicator, they are annotated only with respect to that one, even though they may contain other predicators which are also FrameNet LUs.

In full-text annotation, we attempt to annotate **all** the LUs of the sentence, so a single sentence typically contains several annotation sets.

These full-text annotations have been viewable on our website for some time, but we are now able to release the full annotations in XML format:

We are releasing 5 files which were also annotated as part of the PropBank project (<http://www.cis.upe\nn.edu/~ace/>), made available courtesy of Martha Palmer (U Colorado, Boulder) and the Linguistic Data Consortium at University of Pennsylvania. The texts are from the Wall Street Journal, but none have to do directly with business; in fact the subjects are quite varied:

1. Election Victory (Editorial on Felipe Gonzalez’s victory in Spanish elections)
2. Loma Prieta (Feature story on the 1989 Loma Prieta earthquake in N. California)
3. Polemic Progressive Education (Opinion piece against “progressive” education by Irving Kristol)
4. Ticket Splitting (The history of ticket splitting in U.S. politics)
5. Bell Ringing (Feature on the decline in the art of bell ringing in English country churches)

We are also releasing 12 files that come from the Nuclear Threat Initiative website (<http://www.nti.org>) and were made available courtesy of Gary Ackerman, of the Center for Nonproliferation Studies on Monterey, California. These texts represent intelligence analyses of the status of various countries with regard to the development of nuclear, chemical and biological weapons and missile systems. The titles are:

1. Libya Country
2. North Korea Introduction
3. Kazakhstan
4. Iran Nuclear
5. Iran Missile
6. Iran Biological
7. Iran Chemical
8. Iran Introduction
9. North Korea Nuclear Capabilities
10. North Korea Nuclear Overview
11. Russia Introduction
12. Taiwan Introduction

5 Semantic Types and Connections to SUMO

Since Release 1.2, we have completely revised our set of semantic types, so that there are now 72 semantic types within the FN database, and they are organized into a reasonable, if simple, taxonomy. This is contained in its own XML file, `semtypes.xml`. We distinguish three large categories of semantic type: true ontological types (applied primarily to frame elements), frame types (applied to frames, which are somewhat theory-internal) and lexical types (applied mainly to lexical units). We hope that the ontological types will be of use in research on semantic parsing and NLU. More detail on this topic can be found in the Book Chapter 6.

Although FrameNet has not had the ambition of building a formal ontology, we find ourselves doing something that looks suspiciously like it. A visiting post-doc, Jan Scheffczyk, has been the driving force behind efforts to link FrameNet to the Suggested Upper Merged Ontology (SUMO), in collaboration with Adam Pease and Michael Ellsworth. So far, (1) the FN semantic types have been linked with SUMO nodes, giving an external reference semantics for those types, and (2) a program has been developed for computer-assisted linking of FN FEs to SUMO nodes. For more details see the materials in the `ontologies/fn_sumo` directory.

6 OWL DL representation of data

Also through the efforts of Jan Scheffczyk, building on work done for Release 1.2 by Srin Narayanan and Matt Gerber, we now have a complete representation of the frame database in OWL DL. (This includes the frames, frame elements, lexical units, frame relations and frame element relations.) We chose OWL DL because it is more expressive than OWL lite, but more computationally tractable than OWL full.

It is possible to convert all of the annotation data to OWL DL as well, but it produces very large files. Rather than do all the conversion and distribute the results, Jan has produced a java program to read the annotation XML and produce the OWL output, so that users can convert only the part that they need. All of this, along with some papers describing the principles behind it are in the directory `ontologies/fn_owl`.

7 Minor changes in XML format

We have changed the leXML files so that the lexemes appear there, inside an element `<lexemes>`, in order, with one marked as the head. This is only informative in the case of multi-lexeme lemmas (i.e. multi-word expressions), but we have done it for all lemmas for the sake of consistency.

(See the file “XMLDocumentation” in this directory for details.)

8 Change in licensing policy

We have decided to make the FrameNet data available without charge for research and development purposes to everyone, including for-profit companies. Release 1.2 was only free to academic researchers. We continue to hold the copyright on the data and require a commercial license from anyone intending to sell a product or service derived from or based on FrameNet. For details, please see the license agreements on the FrameNet website, <http://framenet.icsi.berkeley.edu>

Collin Baker
Project Manager
2006.08.22