Welcome to Release 1.6 of the FrameNet data

We are very happy to be offering a new data release after a long period during which we updated the public website with new data, but the official release remained the same. We hope that the new release will make it unnecessary for users to depend on the website for current data. Users familiar with the FrameNet data will probably be pleased to hear that we have not made substantial changes to the data format; we are including the XML documentation from Release 1.5, since nearly all of it is still applicable, and will give details of the changes in Sec. 8 below. The good news is that the continuity of format has made it easy to produce a detailed report of the differences between releases 1.5 and 1.6, included here as "DifferencesR1.5-R1.6.xml".

We have been pleased at the increasing use of FrameNet as a resource for semantic analysis both by academic and commercial users. We are continuing to license the FrameNet data under a Creative Commons Attribution-Only license. This license allows use for any purpose, including commercial use, and requires public acknowledgment of such use. In general, academic users have included clear acknowledgments in published papers, but we are aware of a number of commercial users who have failed to do so. If you produce a product or provide a service based on the use of the FrameNet data, you are required to provide your users with a written reference to FrameNet, including the URL of the FrameNet website (http://framenet.icsi.berkeley.edu) and reference to a current FrameNet publication. (At this time, the most complete, reasonably current published summary of work on FrameNet is Fillmore and Baker (2010) "A Frame Semantic Approach to Linguistic Analysis" in Heine and Narrog (eds.) *Oxford Handbook of Linguistic Analysis*.)

We have made substantial changes in the frame-to-frame relations and semantic type system (discussed below), and believe that the data in the new release is considerably more regular and logically consistent. We have gotten some way through the backlog of changes we have wanted to make for a long time, and welcome your comments and new suggestions for improvement.

"The Book"

We have not revised the documentation in *FrameNet II: Extended Theory and Practice*. Therefore, we are continuing to distribute the version dated Sept. 14, 2010 (by Ruppenhofer, Ellsworth, Petruck, Johnson, and Scheffczyk) for the time being. We hope to produce a new version of "the Book" soon, which will include all the changes in data release 1.6.

1 Growth of the FrameNet database

In all, we have added 191 frames, deleted 5, and renamed 27. At the Frame Element (role) level, we have added 1,683 Frame Elements (FEs), deleted 47, and renamed 514. In the DifferencesR1.5-R1.6.xml file, we have indicated which of these FE changes are due to adding or deleting frames, and which are additions to or deletions from existing frames. For example, we created a new frame **Reserving**, along with six new Frame Elements as part of that process: Booker, Client, Beneficiary, Services, Organization, and Time. We also added five new Frame Elements (Configuration, Goal, Imposed purpose, Circumstances, and Containing event) to the existing frame **Reshaping**. These are marked in the XML with the attribute `added_to_frame = "y"`.

	R1.2	R1.3	R1.5	R1.6	Change R1.5→R1.6
Frames	609	795	1019	1205	18%
(non-lexical)	58	74	111	134	21%
FEs in lexical frames	4909	7124	8884	10333	16%
FE/lexical frame	8.91	9.88	9.78	9.65	-1.3%
Pct. non-lexical	9.5%	9.3%	10.9%	11.1%	0.2%
Frame relations	550	1152	1507	1805	20%
FE relations	2770	6311	8252	11791	43%
Lexical Units	8869	10195	11829	13312	13%
LUs/lexical frame	16.1	14.14	13.03	12.43	-4.6%
LUs w/ lexicog anno	6642	6815	7711	8313	7.8%
Pct. LUs w/ lexicog. anno	74.9%	66.8%	65.2%	62.5	-2.7%
AnnoSets in lexicog anno	133846	139439	149931	157739	5.2%
Lexicog AnnoSets/annotated LU	20.2	20.5	19.5	19.0	-2.6%
AnnoSets in full text anno	0	11671	23087	43955	90%
Total AnnoSets	133846	151110	173018	201695	17%
Full Text anno/Total anno	0%	7.7%	13%	22%	

Table 1: Some statistics comparing Releases 1.2, 1.3, 1.5 and 1.6

2 New semantic domains

The reasons for decisions as to which new frames to add to FrameNet may not be obvious to the outside observer, but in fact, the FrameNet team follows several consistent approaches in initiating new frames.

1. In the process of creating frame A, we include a Lexical Unit (LU) (based, as usual, on a combination of native-speaker recollection, corpus search, and consultation with other dictionaries and thesauri). We find that this LU has another sense that does not fit into any existing frame, and we consider what other words should be LUs in a new frame with that sense; let us call this frame B. If we go on to work on frame B, some of those LUs may also have other senses which would require the creation of still more frames. Thus, the network grows by polysemy relations of lemmas.

2. As part of our collaboration with other researchers or in response to the needs of our funders, we create frames in their domains of interest. In this release, this process has resulted in new frames related to fire-fighting (e.g. the **Fire_emergency_scenario** and its subframes), military combat (e.g. **Suicide_attack**), frames for spatial relations such as **Adjacency** and **Surrounding**, and linguistic phenomena such as negation, conditionals, and functions of modals (discussed below).

3. We sometimes add lemmas simply because we notice (or colleagues elsewhere point out to us) conceptually important and/or high-frequency lemmas that FrameNet lacks; often creating these lemmas involves creating new frames, as well. Among the new frames created in this way in R1.6 are **Making_arrangements, Sex, Business_closure** and **Product_development**.

3 Regularization of names

We have made all the naming much more regular with regard to spelling, capitalization, hyphens vs. underscores vs. spaces. As a general policy, we use spaces between the words of a multiword lexical unit, so that it will look like it does in text, but underscores between the words of frame names that contain more than one word. Thus the frame **Adding_up** contains the lexical units *add up.v, tally.v*, and *total.v*. We use hyphens only for words conventionally written with a hyphen; for example, the **Firing** frame contains the lexical units *lay off.v* (*She was laid off a month ago*) and *lay-off.n* (*There were a lot of lay-offs after the merger*). The addition of more lexical units of more varied parts of speech means that we now have names of lexical units ending in .scon (subordination conjunction) and .c (coordinating conjunction).

Some of the changes of names are not simply matters of form. In particular, Release 1.5, contained many Frame Elements across many frames named PATIENT and 18 Frame Elements named UNDERGOER. We came to the conclusion that this difference did not represent a distinction that we had made consistently, so we renamed the UNDERGOER Frame Elements to PATIENT across the board. Likewise, we renamed all the FEs called REASON to EXPLANATION for the sake of consistency. (We do continue to maintain a distinction between PATIENTS and THEMES, which occur in frames connected with motion and location.)

4 Frame-to-frame relations

Frame-to-frame relations have been added in many places where they were missing, again partially as a result of our own review of the data and partially in response to suggestions from researchers not directly connected with the project. As a result, we have reduced the number of "orphan" frames (those with no frame-to-frame relations connecting them to the other frames) from 45 to 20. We also ran checks to find accidental overlaps within the frame hierarchy, e.g. places where frame A both Inherits from and Uses Frame B. We removed such overlaps, except for a very small number of unusual cases, in which we judged them to be correct. All this has made the frame lattice much more complete and logical.

The intended interpretations of the frame-to-frame (and FE-to-FE) relations are explained in some detail in the chapter on frame relations in the book, so we will not discuss them further here. However, please note the importance of the "See also" relation: This is not a strictly logical relation, but a suggestion to the reader to look at another frame that might seem very similar to the one under consideration. The FrameNet team has done its best to write clear explanations of the differences among sets of closely related frames, and these explanations are included in the definition of just **one** of the frames in the set. The others point to that frame via "See also" relations. Following these links and perusing those explanations should help to clarify many difficult frame distinctions.

5 Modals, Negation, Conditionals, etc.

We have recently begun exploring certain linguistic phenomena that might not usually be thought of as related to semantic frames. One of these involves the meanings of modals in English, including the usual list of *might, may, should, can, could, would, shall*, and *will* and closely related expressions such as *have to (hafta), got to (gotta), ought to (oughtta)*, and *used to*. We have created frames for these LUs based on their use in discourse, e.g. *should* is

an LU in the **Desirable_event** frame, which has only one core FE, the STATE_OF_AFFAIRS that the speaker regards as desirable, e.g.

[STATE_OF_AFFAIRS They] SHOULD [STATE_OF_AFFAIRS see to it that their kids don't play truant].

which can be interpreted as "It is desirable that the following event should occur: their seeing to it that their kids don't play truant."

Altogether, we have identified nine semantic functions of these modals; so far we have created frames and done some annotation for five of these. Our work on Negation is still at an early stage, but we have created a frame for **Alternativity**; the frames **Negation**, **Conditionality_scenario** and **Negative_conditional** all inherit from it, since both negation and conditionals imply the existence of two alternative realities. Please see the definitions of these frames for further explanation. Some of these functions interact in complex ways with tense; we are just beginning our research on how to represent that interaction.

We notei that some of these LUs have multiple pragmatic functions, and we have marked some of the annotation sets with labels on the "Sent" layer to indicate which function they are performing, e.g.

Content conditional: If it rains, the ceremony will be under the tent.

Epistemic: If the lights are on, then somebody is home.

Speech act: If you need anything, my name is Bob.

Metalinguistic: If that's not success, I don't know what is.

These labels do not appear in the HTML in a browser window, but are found as extra attributes in the XML of the relevant sentence annotations. We have also performed a small survey to get an idea of the relative frequency of these uses; the results are available on the FrameNet public website.

6 Spatial relations

In Release 1.6, we have also added many frames for static spatial relations, all of which inherit from **Locative_relation**; most of the LUs in these frames are either prepositions or adjectives. For example, the **Gradable_proximity** frame contains the LUs *close.a, distant.a, far.a, near.a*, and *proximity.n*, Inherits from **Locative_relation** and has a Perspective_on relation to **Proximity_image_schema**. The theory behind the relations and image schemas is closely related to the work of Len Talmy (1983, 2000). As part of our work on this domain, we have also added more than two dozen semantic types, in a subtree under "LU_with_FE_specified". We plan to work out the metaphorical extensions to time and some other domains in the next phase of this project.

7 Word form mappings

We have added two supplementary XML files containing all the mappings of word forms to lexemes and lemmas in our database. For the most part, these mappings are identical with the equivalent data from an earlier version of the CELEX database which we used to seed these tables initially, but we have added lemmas and word forms as needed since then, and in particular, linked British spellings with American spellings of the lemma. There are two

files, one giving lexeme-to-word form mappings and the other giving lemma-to-word form mappings. The latter include multi-lexeme lemmas, which is the approach used in FrameNet to represent multiword lexical units. These files may be useful for lemmatization and query expansion, quite apart from any use of the main FrameNet data.

8 Changes in XML format from R1.5

As will be apparent from the Differences file, both frame names and frame element names are subject to change, either because we have found a more perspicuous name for them or in order to convey a change in the scope of the frame or FE. The ID numbers of frames and FEs, however, are a reliable means of identifying them, from one version of the data to the next (and in some cases, across FrameNets in different languages). In processing the FrameNet data, therefore, we urge users to supplement frame and frame element names with ID numbers wherever possible. ID numbers will be still more important when comparing frames across languages, now that projects for frame semantic annotation are underway in more than 10 languages.

In keeping with this suggestion, we have revised the XML generation to include frame IDs as well as names of frames: In the XML for each frame, after the last FE definition, you will find a list of frame relations in which the current frame participates; in R1.5, those other frames are identified only by frame name. The XML report has been revised so that the <`relatedFrame`> elements now include the ID of the related frame, as well as the name. We have also decided to change the ''Created by" field in the Frame definition so that it now displays the annotator's initials rather than a numeric code.

Most of the changes in XML format are of this sort, either adding more information in additional attributes on an element, or making something required that was previously optional. We believe that nothing has been removed from the XML specification, and no existing XML element names have been changed; current software designed for Release 1.5 that uses standard XML parsing algorithms should have no problems in reading the new data.