Prepublication version of:
Ruppenhofer, Josef, Hans C. Boas, and Collin Baker. 2013. The FrameNet approach to relating syntax and semantics. In: R.H. Gouws, U. Heid, W. Schweickard, and H.E. Wiegand (eds.), Dictionaries. An International Encyclopedia of Lexicography, 1320–1329. Berlin/New York: De Gruyter/Mouton.

The FrameNet Approach to Relating Syntax and Semantics

Abstract:
FrameNet is a research project that analyzes the vocabulary of English in terms of frame semantics based on corpus evidence. This article describes how the work proceeds, what the resulting product looks like and how the resource is used. It also draws comparisons to other electronic resources and commercial lexicography to highlight FrameNet's unique features.

1.Introduction

FrameNet is a research project that seeks to put into practice the principles of Frame Semantics as proposed by Charles J. Fillmore (1977, 1985). The main idea is that the

Prepublication version of:
Ruppenhofer, Josef, Hans C. Boas, and Collin Baker. 2013. The FrameNet approach to relating syntax and semantics. In: R.H. Gouws, U. Heid, W. Schweickard, and H.E. Wiegand (eds.), Dictionaries. An International Encyclopedia of Lexicography, 1320–1329. Berlin/New York: De Gruyter/Mouton.

meanings of words are best characterized in terms of experience-based schematizations of events and objects in the speaker's world. Such schematizations concern particular types of events and the participants and circumstances involved in them. The schematizations are referred to, in Frame Semantics, as semantic *'frames*'. Individual word senses are called *'lexical units*' (LUs). When a lexical unit is related to a given frame, Frame Semantics says that the LU *'evokes*' that frame. Typically, senses of different words share the same schematization with other word senses, that is, groups of word senses evoke the same frame. The roles associated with the event are referred to as '*frame elements*' (FEs). The same system of analysis applies not only to events but also to relations, states, and objects; the frame-evoking expressions may be single words or multi-word expressions, and they may be of any syntactic category (Fillmore, Johnson and Petruck 2003). For example, in (1) the verb *comply* (the '*target'* LU) '*evokes'* the `Compliance` frame: several semantically related words such as *adhere, adherence, comply, compliant*, and *violate*, among others, evoke the same frame.

(1) *The appellant failed to comply with the rules of the hostel*.

Prepublication version of:
Ruppenhofer, Josef, Hans C. Boas, and Collin Baker. 2013. The FrameNet approach to relating syntax and semantics. In: R.H. Gouws, U. Heid, W. Schweickard, and H.E. Wiegand (eds.), Dictionaries. An International Encyclopedia of Lexicography, 1320–1329. Berlin/New York: De Gruyter/Mouton.

The Compliance frame is characterized by relationships holding between different FEs, such as *Acts* and *States_of_Affairs* for which *Protagonists* are responsible; the frame has to do with whether the *Acts* or *States_of_Affairs* are or are not in accord with some *Norm*.

The FE *Act* identifies the act that is judged to be in or out of compliance with the norm(s). The FE *Norm* identifies the norms that ought to guide a person's behavior. The FE *Protagonist* refers to a person whose behavior is in or out of compliance with norms. Finally, the FE *State_of_Affairs* refers to a situation that may violate a norm. Besides compiling frame descriptions and lists of frame-evoking LUs, frame semanticists are also interested in determining how FEs are realized syntactically (Atkins, Fillmore and Johnson 2003, Fillmore 2007).

The remainder of this article deals with the practical implementation of Frame Semantics in the FrameNet project (http://framenet.icsi.berkeley.edu; Ruppenhofer et al. 2010), focusing on the work-flow of the project and its different uses as a lexical resource, its relation to other electronic resources, and its impact on linguistic research. For a detailed glossary of FrameNet (FN) terminology, see

Prepublication version of:
Ruppenhofer, Josef, Hans C. Boas, and Collin Baker. 2013. The FrameNet approach to relating syntax and semantics. In: R.H. Gouws, U. Heid, W. Schweickard, and H.E. Wiegand (eds.), Dictionaries. An International Encyclopedia of Lexicography, 1320–1329. Berlin/New York: De Gruyter/Mouton.

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

2. Creation and contents of the FrameNet resource

At the core of FrameNet's activities since the late 1990s is the compilation of a lexicographic database, which currently contains more than 1,000 frame descriptions together with lexical entries for more than 12,000 LUs and more than 185,000 annotated examples (as of November 2011). The FrameNet database is freely available for any purpose, and is used by hundreds of research groups worldwide.

At the beginning of the lexicographic process, the FN staff proposes a frame description, including what sorts of FEs the frame needs and what LUs might evoke it. During this stage, lexicographers depend both on their intuitions and on careful searches of electronic corpora. The basic criterion for defining the boundaries of a frame is that all LUs should evoke the same type of event and share the same inventory and configuration of FEs. That is, the FEs should have the same semantic types; be of the same relative importance (traditionally, this is thought of in terms of the argument-adjunct distinction); and take the same pragmatic perspective (e.g. in active-form clauses, the same FE should be the

Prepublication version of:
Ruppenhofer, Josef, Hans C. Boas, and Collin Baker. 2013. The FrameNet approach to relating syntax and semantics. In: R.H. Gouws, U. Heid, W. Schweickard, and H.E. Wiegand (eds.), Dictionaries. An International Encyclopedia of Lexicography, 1320–1329. Berlin/New York: De Gruyter/Mouton.

subject of all the verbs in the frame). For more discussion of these criteria, see Petruck et al. (2004).

Next, example sentences are extracted from electronic corpora for each target LU; trained annotators then use the FN desktop software to manually annotate roughly 10-20 sentences per LU.
As the annotators apply labels showing which parts of the sentence fill which FE role, labels for their phrase types (PTs), and their grammatical functions (GFs) vis-à-vis the target word are added automatically. (The GF and PT can be corrected by the annotator if necessary.) Two example annotations are given in Figs. 1 and 2.

	I	**substitute**	tea	for	coffee
FE	Agent		New	Old	
PT	NP		NP	PP	
GF	Ext		Obj	Dep	

Fig. 1: *substitute* NP PP_for

	I	**substitute**	coffee	with	tea
FE	Agent		Old	New	
PT	NP		NP	PP	
GF	Ext		Obj	Dep	

Fig. 2: *substitute* NP PP_with

Prepublication version of:
Ruppenhofer, Josef, Hans C. Boas, and Collin Baker. 2013. The FrameNet approach to relating syntax and semantics. In: R.H. Gouws, U. Heid, W. Schweickard, and H.E. Wiegand (eds.), Dictionaries. An International Encyclopedia of Lexicography, 1320–1329. Berlin/New York: De Gruyter/Mouton.

The goal is to arrive at an annotation set for every possible combination of FE, GF, and PT. For instance, for the verb *substitute* in the frame `Replacing`, the two different mappings of the FEs *`Old`* and *`New`* to grammatical functions in active-form clauses, as shown in Figs. 1 and 2, are documented. When the annotation is completed, the valence of each LU is automatically summarized by abstracting over the annotated examples. Finally, the lexical entry is "produced" as a report from the FrameNet database. It includes a brief definition of the LU together with the frame it evokes and valence tables providing exhaustive information about every attested combinatorial possibility of FEs and their syntactic realizations. Depending on the LU, it may also contain information such as the semantic type of the LU and a list of annotated support or controller verbs, etc. For an in-depth description of the FN work-flow, see Fillmore et al. (2003), Ruppenhofer et al. (2010), and Fillmore and Baker (2010). Baker, Fillmore and Cronin (2003) provide a technical description of the FN database.

Since 2004, FN staff have also been annotating continuous texts, thereby creating lexical entries and a continuous corpus of annotated sentences. This full-text annotation differs from FN's established lexicographic research in several ways:

Prepublication version of:
Ruppenhofer, Josef, Hans C. Boas, and Collin Baker. 2013. The FrameNet approach to relating syntax and semantics. In: R.H. Gouws, U. Heid, W. Schweickard, and H.E. Wiegand (eds.), Dictionaries. An International Encyclopedia of Lexicography, 1320–1329. Berlin/New York: De Gruyter/Mouton.

- Typically, between two and ten LUs are annotated per sentence, in as many different frames. By contrast, in the lexicographic work, only one LU per sentence is annotated.
- With full texts, annotators must label whatever the text contains, regardless of syntactic complexity, ambiguity, rhetorical infelicity, etc., whereas lexicographers can choose the clearest and simplest examples of a given LU to annotate.
- Full-text annotation drives the discovery of new frames, that is, if there are no existing frames they have to be created "on the fly".

After more than a decade, the number of semantic frames has grown to more than 1,000. To show that frames are not simply isolated entities, FN has developed an elaborate system of frame-to-frame relations representing a hierarchy of frames in which some frames are instances of others, some are components of others, etc. The most commonly found relations include those representing generalizations (INHERITANCE, USING, PERSPECTIVE ON), complex events (SUBFRAME, PRECEDES), and "systematic" relations (CAUSATIVE OF, INCHOATIVE OF); an additional relation, called "SEE ALSO" serves as a cross-reference between frames (see Fillmore and Baker (2010) and

Prepublication version of:
Ruppenhofer, Josef, Hans C. Boas, and Collin Baker. 2013. The FrameNet approach to relating syntax and semantics. In: R.H. Gouws, U. Heid, W. Schweickard, and H.E. Wiegand (eds.), Dictionaries. An International Encyclopedia of Lexicography, 1320–1329. Berlin/New York: De Gruyter/Mouton.

Ruppenhofer et al. (2010) for more details on frame-to-frame relations).

3. Human access to FrameNet data

Although snapshot data releases are available for download, human users mainly access FN on the web, where the data is updated frequently. The top-level page of FN's website (http://framenet.icsi.berkeley.edu) provides several access paths to the data. The information in FN can be accessed from two major perspectives: '*onomasiologically'*, beginning with a concept (frame) to be encoded, or '*semasiologically'*, beginning with a form (lemma) to be decoded. The first perspective is realized by the Frame Index, which leads to a page with an alphabetized list of frames in a navigation pane on the left and a larger right panel where descriptions of selected frames are displayed. The frame descriptions include, in order:

- a definition;
- a list of FEs, in sub-groups reflecting grammatical prominence, where each item is accompanied by a definition and one or more examples;
- a list of frame-to-frame relations;
- and a list of LUs that can evoke the frame.

In the frame and FE definitions, the FEs used in the example sentences are identified for

Prepublication version of:
Ruppenhofer, Josef, Hans C. Boas, and Collin Baker. 2013. The FrameNet approach to relating syntax and semantics. In: R.H. Gouws, U. Heid, W. Schweickard, and H.E. Wiegand (eds.), Dictionaries. An International Encyclopedia of Lexicography, 1320–1329. Berlin/New York: De Gruyter/Mouton.

users by mark-up. In the frame-to-frame relations section, hyper-links are provided to related frames. Likewise, in the list of associated LUs, two links each are provided per LU: one to an Annotation Report and another to a Lexical Entry Report.

The Annotation Report consists of a listing of the FEs for the associated frame and then a display of annotated instances of the LU, grouped into subsets ('*sub-corpora*') according to the syntactic patterns used to extract them. The syntactic patterns usually serve to target particular syntactic complementation types. E.g for the noun *oath* in the `Commitment` frame, there are sub-corpora named *vpto* and *sthat* for uses where the noun takes infinitival and finite-clause complements, respectively. The Lexical Entry Report contains two types of summary information. First, it provides a list of the individual FEs and the ways in which they are realized in the annotated instances. Fig. 3 shows the syntactic realizations found with the verb *substitute* for the three core frame elements in the `Replacing` frame. Note that the notion of "realization" used by FrameNet covers cases in which the Frame Element is not overtly realized at all but instead "null instantiated". This can happen through the idiosyncratic licensing of a lexical unit or the licensing of a syntactic construction. FrameNet distinguishes three types of missing elements: DNI ("definite null instantiation")

Prepublication version of:
Ruppenhofer, Josef, Hans C. Boas, and Collin Baker. 2013. The FrameNet approach to relating syntax and semantics. In: R.H. Gouws, U. Heid, W. Schweickard, and H.E. Wiegand (eds.), Dictionaries. An International Encyclopedia of Lexicography, 1320–1329. Berlin/New York: De Gruyter/Mouton.

marks FEs that are unrealized but which have to be recoverable from the context. An example is the FE *Old* in the Replacing frame in a sentence such as *We'll substitute sugar*. Fig. 3 shows that there are 9 such cases of DNI for the FE Old in the FrameNet annotations for *substitute*. INI ("indefinite null instantiation") marks FEs that are merely existentially bound. An example is the FE *New* in *I have to replace sugar in my diet*. CNI ("constructional null instantiation") marks all omissions licensed by a syntactic construction. A typical case is the omission of agentive FEs in the passive construction, as in *The tire needs to be replaced*.

Frame Element	Number Annotated	Realizations
Agent	57	CNI.-- (18) NP.Ext (35) PP[by].Dep (3) Poss.Ext (1)
New	57	NP.Ext (16) INI.-- (5) NP.Dep (1) NP.Obj (30) PP[with].Dep (5)
Old	57	DNI.-- (9) NP.Ext (5) PP[for].Dep (32)

		PP[in].Dep (4) PP[as].Dep (1) PP[of].Dep (1) NP.Obj (5)

Fig. 3: Partial Syntactic Realization Table for *substitute* in `Replacing`

The Lexical Entry report also contains a list that shows all the attested combinations in which groups of FEs were co-realized. Some of the attestations found with *substitute* in the Replacing frame are shown in Fig. 4.

Number Annotated		Patterns		
1 Total	Agent	Manner	New	Old
(1)	PP[by].Dep	PP[without].Dep	INI	NP.Ext
1 Total	Agent	Means	New	Old
(1)	CNI	PP[by].Dep	NP.Ext	DNI

Fig. 4: Partial Valence Pattern Table for *substitute* in `Replacing`

The second row in the table describes the pattern of FEs found in the following sentence:

Prepublication version of:
Ruppenhofer, Josef, Hans C. Boas, and Collin Baker. 2013. The FrameNet approach to relating syntax and semantics. In: R.H. Gouws, U. Heid, W. Schweickard, and H.E. Wiegand (eds.), Dictionaries. An International Encyclopedia of Lexicography, 1320–1329. Berlin/New York: De Gruyter/Mouton.

(2) … *some alternative can be substituted by electoral decision.*

Both types of tables contain hyper-linked counts for the number of annotated instances exemplifying the FE (or FE combination), which can be displayed by following the links. The two report pages for a given lexical unit are also interlinked.

The decoding perspective is represented by the alphabetized Index of LUs. Users can simply browse this index or use a search box to find LUs. Note that there is no direct representation of lemmas: the different LUs (senses) connected to the same lemma simply follow each other. Unlike on the semantic side, where one can easily follow links from frame to related frame, on the form side there are no direct links between the LUs associated with the same lemma, or between LUs belonging to morphologically related lemmas. Each entry for a LU comes with information on the work-flow status of the LU, and with three links: one to the Frame description, one to the Annotation report, and one to the Lexical Entry Report (as described above).

A third type of access point is the index of Full-Text annotations. A Full-Text Annotation Report pairs a display of the running text of a

Prepublication version of:
Ruppenhofer, Josef, Hans C. Boas, and Collin Baker. 2013. The FrameNet approach to relating syntax and semantics. In: R.H. Gouws, U. Heid, W. Schweickard, and H.E. Wiegand (eds.), Dictionaries. An International Encyclopedia of Lexicography, 1320–1329. Berlin/New York: De Gruyter/Mouton.

document with a display of the annotations of specific user-selectable items from the running text. Although the annotated instances in full-text documents are also part of the Annotation and Lexical Entry Reports for the relevant lexical units, they can be viewed in their document-context only via the Full-text Annotation Report.

A fourth kind of data access is given by the FrameGrapher visualization tool (https://framenet.icsi.berkeley.edu/fndrupal/FrameGrapher). While a frame's frame-to-frame relations are also listed in the Frame reports, the FrameGrapher is particularly suited to interactively exploring the topology of relations that exist in a set of related frames. For example, users can choose a focal frame on which the display is to be centered and a set of frame relation types they are interested in. In order to manage the size and layout of the resulting display, users can specify the number of levels of the (sub-)hierarchy and the number of leaf nodes to be displayed. Both the frames and the links in the graph can be clicked by the user, resulting, respectively, in a re-centered graph with a new focal frame, or in an expanded graph, where the frame element-to-frame element relations that go along with the chosen frame-to-frame relationship are displayed in detail (see Fillmore and Baker 2010: 332-333 for details).

Prepublication version of:
Ruppenhofer, Josef, Hans C. Boas, and Collin Baker. 2013. The FrameNet approach to relating syntax and semantics. In: R.H. Gouws, U. Heid, W. Schweickard, and H.E. Wiegand (eds.), Dictionaries. An International Encyclopedia of Lexicography, 1320–1329. Berlin/New York: De Gruyter/Mouton.

Fig. 5 shows a set of frames related to the idea of employment as portrayed in the FrameGrapher. The dashed lines indicate `SUBFRAME` relations, the solid lines indicate `PERSPECTIVE ON` relations.

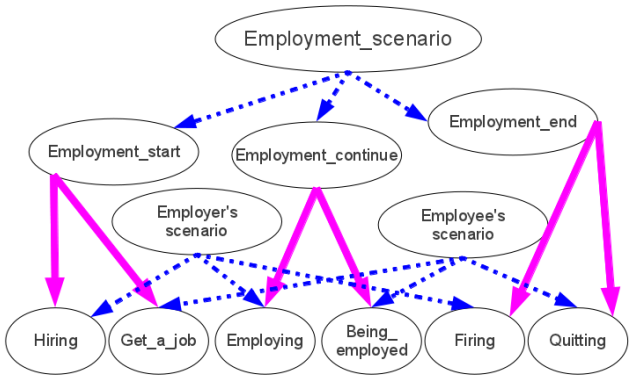

Fig. 5: FrameGrapher display for employment-related frames

The final view on the FrameNet data is provided by the FrameSQL search tool (http://framenet2.icsi.berkeley.edu/frameSQL/fn2_15/notes/), which is developed and maintained independently by Prof. Hiroaki Sato of Senshu University, Japan. In addition to the English-language FrameNet, FrameSQL can search the frame semantic annotations of FrameNet's German, Japanese, and Spanish sister projects. Furthermore, it also integrates information from all four resources. As a result, users can access the information from the different

Prepublication version of:
Ruppenhofer, Josef, Hans C. Boas, and Collin Baker. 2013. The FrameNet approach to relating syntax and semantics. In: R.H. Gouws, U. Heid, W. Schweickard, and H.E. Wiegand (eds.), Dictionaries. An International Encyclopedia of Lexicography, 1320–1329. Berlin/New York: De Gruyter/Mouton.

sources aligned seamlessly, as if they were a single database. Hyper-links among the databases and accompanying search capabilities make it easy to compare the semantic structures of corresponding frames or lexical units across the languages (see Sato et al. 2008).

Besides the FN website, human users can also access the first three types of information (Frame, LU-specific Lexical Entry and Annotation Reports, and Full-text annotation reports) in the FN data release, which consists of a set of static interlinked XML pages that can also be viewed in a web browser. The data release does not, however, include the FrameGrapher or the FrameSQL tools. Morphological information on inflectional forms of lemmas is not provided online nor in the releases.

4. Automatic access to FrameNet data

As used in the work of the FN project, the physical database is implemented as a relational database (MySQL), whose contents lexicographers and annotators access and edit via graphical user interfaces, controlled by a client-server architecture that ensures data integrity and consistency (see Baker, Fillmore and Cronin 2003). For purposes of distribution, the contents of the database are converted into several sets of XML

documents. For instance, the release includes one file per frame, one file per LU annotated in the lexicographic mode, one file per full-text document annotated, etc. The types of XML files distributed in the release do not mirror the database tables one-to-one. For instance, the descriptions of the FEs are simply included in the frames files of the release. Along with the data files, the FN data release includes XML schema description files, but no ready application programming interface (API). However, several user-contributed APIs are available for different versions of FN (http://www.cl.uni-heidelberg.de/trac/FrameNetAPI, https://ptl.sys.virginia.edu/msg8u/NLP/Source/ResourceAPIs/FrameNet/).

5. Relation to other electronic resources

5.1. WordNet

WordNet (WN; Miller et al. 1990) is a far more extensive lexical resource than FrameNet, with more than 200,000 word senses, (i.e. lemma-synset pairs), compared to FrameNet's 11,000; furthermore, WN has more than 117,000 synsets to FrameNet's 1,000 frames. Indeed, although synsets and frames both represent groupings of words, they are fundamentally different conceptually: Words in a synset are intended to be synonymous, and

Prepublication version of:
Ruppenhofer, Josef, Hans C. Boas, and Collin Baker. 2013. The FrameNet approach to relating syntax and semantics. In: R.H. Gouws, U. Heid, W. Schweickard, and H.E. Wiegand (eds.), Dictionaries. An International Encyclopedia of Lexicography, 1320–1329. Berlin/New York: De Gruyter/Mouton.

thus substitutable in a least in some context. Words in a frame, on the other hand, are supposed to be related to the same type of event, but that does not necessarily imply synonymy (see Boas 2005). For example, antonyms such as the verbs *praise* and *criticize* are both in the FN `Judgment_Communication` frame, as they involve the same sorts of participants; aside from the differences in their definitions, they are also distinguished with regard to semantic type: *praise* is marked as "positive judgment", while *criticize* is neutral. In WordNet, *praise* is in one synset and *criticize* in two, where one is defined as 'to act as a critic', and the other ('to find fault with') is linked to the *praise* synset by an antonymy relation.

When FrameNet first began its lexicographic work, there was an expectation that WN synsets would guide the lexicographers in deciding which word senses go into which frames. In practice, although FN lexicographers often consult WN along with other machine-readable dictionaries in the course of defining frames and LUs, the set of LUs in a FN frame and the set of lemmas in a WN synset often do not overlap that closely. The two resources are constructed on quite different principles and this is reflected in their different treatments of individual LUs. In fact, there is an emerging consensus that the two resources should be aligned with each other

to the extent possible, but that they cannot be merged. For studies of parallel WordNet-FrameNet annotation, see Fellbaum and Baker (*submitted*) and DeMelo et al. (2012).

5.2. Valence-focused resources (Levin, PropBank, NomBank)

Levin's (1993) linguistic study classifies more than 3,000 English verbs on the basis of their diathesis alternations; verbs which share the same subset of possible syntactic alternations form a group which is also likely to be semantically similar in important respects. FN's work differs from Levin's in two key respects. First, it covers open lexical classes beyond verbs. Second, while it acknowledges that syntactic alternations can be the basis for groupings of verbs that make some semantic sense, it considers the nexus between shared syntactic alternations and common semantic features to be much less tight than does Levin's work (Baker and Ruppenhofer 2002, Boas 2011).

VerbNet (Kipper et al. 2006) is a large on-line verb lexicon for English. It is a hierarchical domain-independent, broad-coverage verb lexicon, built on the basis of Levin-style alternations, and has mappings to other lexical resources, among them WordNet and FrameNet.

Prepublication version of:
Ruppenhofer, Josef, Hans C. Boas, and Collin Baker. 2013. The FrameNet approach to relating syntax and semantics. In: R.H. Gouws, U. Heid, W. Schweickard, and H.E. Wiegand (eds.), Dictionaries. An International Encyclopedia of Lexicography, 1320–1329. Berlin/New York: De Gruyter/Mouton.

The primary goal in creating the PropBank resource (Palmer, Gildea and Kingsbury 2005) was to provide an analysis of all instances of every verb in the Penn Treebank II corpus (Marcus et al. 1994). PropBank (http://verbs.colorado.edu/~mpalmer/projects/ace.html) provides no analyses for other parts-of-speech. The NomBank sister project (Meyers et al. 2004) began later to extend parallel analyses to nouns. The sense analyses that PropBank produces (called '*framesets*') are verb-specific. Although sometimes connections to VerbNet are pointed out in the descriptions of PropBank's framesets, as a whole PropBank does not readily allow for the kinds of generalizations and inferences that FrameNet's frames and their hierarchy allow. Further, PropBank overall seems to make fewer fine-grained distinctions than FrameNet does. For instance, causative-inchoative pairs are not distinguished and PropBank has coarser role analyses for cases where FrameNet avails itself of role-to-role relations such as Excludes, Requires, and CoreSet. For instance, while FrameNet distinguishes asymmetric and symmetric uses of verbs like *meet* and *resemble* by using different roles, PropBank does not explicate the difference.

For example, in *She and I first met in college* (an instance of *meet* in the `Making acquaintance` frame), the conjoined NP *She and I* is labeled with the FE *`Individuals`*; conversely, in *I first met her in college*, *I* is labeled with the FE *`Individual_1`* and *her* with *`Individual_2`*. The difference is subtle, but the latter wording emphasizes the role and point of view of the speaker; the former is more reciprocal.

6. FrameNet work-flow and coverage

An idealized presentation of FrameNet's lexicographic work-flow may suggest a fairly strict, linear approach beginning with frame creation, continuing with FE definition and LU selection and ending with annotation. In reality, FN's work proceeds in a more interdependent and iterative fashion. Frame definition already takes into account possible LUs, and facts discovered during annotation may lead to modifications of a frame's definition or its set of frame elements.

It is also worth emphasizing that FN's progression through the vocabulary is "organic" rather than linear. In lexicographic work, the choice of which new frame to define is often influenced by the frames most recently

defined: patterns of polysemy exhibited by the lemmas that have lexical units associated with Frame A may raise awareness of the existence of Frame B. In the more recent work on full-text annotation, the frames to work on are determined by the texts chosen. The "organic" growth of FrameNet, driven either by patterns of polysemy or domain coherence, means that, contrary to conventional lexicography, FrameNet is completed frame by frame. As a result, some of the senses of polysemous words, perhaps even frequent ones, will be analyzed by FN much later than others.

7. Impact

7.1. Linguistics and Lexicography

The field of "traditional" linguistics in the U.S. is divided into a variety of schools with quite different approaches, so it is natural that frame semantics and FrameNet will have more impact on some schools than others. Those linguists who have explicitly disavowed any interest in the communicative use of language would not be expected to follow FN research, with its focus on documenting patterns of use LU by LU. The major influence of the FrameNet project has probably been to draw more attention to frame semantics, particularly in the fields of Cognitive Linguistics and lexical semantics. A new journal, *Constructions and*

Frames (http://benjamins.com/#catalog/journals/cf), has begun publishing articles on construction grammar and frame semantics, many of which use FrameNet data.

FrameNet is also a resource that lexicographers increasingly pay attention to as discussed in Atkins, Rundell and Sato's (2003) article in the special issue of the International Journal of Lexicography devoted to FrameNet. Atkins was in fact one of the founders of the FrameNet project, and she and her collaborators routinely teach beginning lexicographers how to apply the principles of frame semantics in their daily work (Lexicography Master Class, cf. http://www.lexmasterclass.com). Over the past ten years, FrameNet has also been taken as a basis for building FrameNets for other languages such as Spanish, German, Japanese, and French (see Boas 2009 for an overview).

7.2. Computational linguistics

FrameNet has had a large impact on the field of computational linguistics. First and foremost, it has paved the way for the task of automatic semantic role labeling (ASRL) introduced by the seminal work of Gildea and Jurafsky (2002). Several evaluations for automatic role labeling systems have since

taken place under the umbrella of the SemEval (http://aclweb.org/aclwiki/index.php?title=SemEval_Portal) series of evaluation campaigns (2004, 2007, 2010). Furthermore, systems that automatically identify frames and assign semantic roles have also been incorporated into various other natural language processing tasks such as Question Answering (Shen and Lapata 2007), Sentiment Analysis (Bethard et al 2004), or Entailment Recognition (Burchardt 2008).

Apart from analysis systems, FN has also inspired interest as a model for encoding certain kinds of domain knowledge, e.g. Dolbey (2009) for the bio-medical domain, Borin, Gronostaj and Kokkinakis (2007) for medical knowledge, and Schmidt (2008) for football vocabulary (www.kicktionary.de/). FN's data has also come to serve as a gold standard data set in computational work seeking to automatically cluster the English vocabulary into semantically coherent groups (Green, Dorr and Resnik 2004, Schulte im Walde 2008). It has been used similarly in research on computational psycho-linguistics (Pado 2007).

The main factors limiting FN's use in natural language processing systems are the following. First, the organic growth of FrameNet presents a problem for word sense

disambiguation or role labeling systems (as well as for human users) that are interested in full coverage of the general language. Without the full set of frames that a lemma can evoke, it is not easy to determine where the boundaries of the defined senses are and to know which as yet undefined senses might exist. In domain-specific applications, this problem may not loom as large, however. Second, the focus on prototypical examples and the lack of statistical representativeness in the lexicographic annotation make it difficult to train supervised statistical analysis systems on FN's annotated data and to apply them to new texts. Third, FN, like other resources such as WordNet, is sometimes said to make too fine-grained semantic distinctions that current NLP systems cannot learn robustly (Palmer et al. 2007).

8. Recent developments and outlook

8.1 Grammatical Constructions

According to the theory of construction grammar, there is only one kind of linguistic object that speakers of a language have to learn: the construction, a pairing of a form and a meaning. The lexical units of FN are simply constructions whose form pole is one or more word-forms, and whose meaning pole is partially represented as a specific semantic frame; but in general, the form pole of a

Prepublication version of:
Ruppenhofer, Josef, Hans C. Boas, and Collin Baker. 2013. The FrameNet approach to relating syntax and semantics. In: R.H. Gouws, U. Heid, W. Schweickard, and H.E. Wiegand (eds.), Dictionaries. An International Encyclopedia of Lexicography, 1320–1329. Berlin/New York: De Gruyter/Mouton.

construction can also be purely syntactic, independent of any LU. Various degrees of specificity are possible on each side of the construction. Many of the ``interesting'' constructions are partially, but not entirely, lexical and are precisely what causes conventional parsers to fail or give incomplete analyses of sentences such as the following:

(2) *I can't stand to see, let alone touch, boa constrictors.*
(3) *The gifted have a duty to help the less fortunate.*
(4) *What's this scratch doing on the table top?*
(cf. Kay and Fillmore 1999)

In 2011, the FrameNet team completed a pilot project to document non-lexical constructions, just as the current FrameNet documents the lexical constructions by manually annotating examples drawn from corpora, using a set of '*construction elements'* (CEs), analogously to annotation with frame elements, using an enhanced version of the FN annotation software. Roughly 75 constructions were described in the pilot study, and most were documented by annotating representative examples drawn from corpora or the web. Some of these also evoke frames already described in FN, and so are annotated with regard to both their syntax and their frame semantics. A particular area of research is constructions related to rates of various kinds,

Prepublication version of:
Ruppenhofer, Josef, Hans C. Boas, and Collin Baker. 2013. The FrameNet approach to relating syntax and semantics. In: R.H. Gouws, U. Heid, W. Schweickard, and H.E. Wiegand (eds.), Dictionaries. An International Encyclopedia of Lexicography, 1320–1329. Berlin/New York: De Gruyter/Mouton.

such as *ten dollars an hour*, *30 m.p.g.,10 m./s^2*, and *1.3 hectares per family*. Standard parsers simply analyze such phrases as two adjacent NPs, yet they are clearly members of a family of regular constructions based on the underlying notion of a fraction with a numerator and a denominator.
The goal of this project is to create a ``Constructicon'' and a set of gold standard annotation data which can be used to train automatic recognizers for all sorts of constructions, including the ``interesting ones'', in other words, to produce construction-aware parsers (Fillmore, Lee-Goldman and Rhodes (in press)).

8.2. Enrichment

Since the creation of lexical resources such as FrameNet is labor-intensive, it is natural that developers and users explore various ways to maximize the benefit of FrameNet data or reduce the effort involved in their creation. One strategy, as discussed in section 5.1. using the example of the WordNet-FrameNet alignment, is to link FrameNet to other resources in order to combine knowledge. A more ambitious approach is embodied by the Swedish FrameNet++ project (http://spraakbanken.gu.se/swefn/), which from the start has sought to recycle existing resources in building up a Swedish frame semantic resource. A third, recently emerging

strategy is to semi-automatically add certain kinds of information to the database as a whole based on manual seed data. For instance, information related to positive and negative associations of lexical units for use in Sentiment Analysis might be propagated throughout FrameNet's frames using frame relations and information gleaned from definitions and annotated instances. Finally, another avenue for expanding FrameNet is to explore so-called crowd-sourcing approaches to the acquisition of lexical knowledge (Hong and Baker 2011).

9. References

9.1 Printed References

Atkins, S., C.J. Fillmore, and C.R. Johnson 2003 Lexicographic relevance: selecting information from corpus evidence. International Journal of Lexicography.16.3: 251--280.

Atkins, S., M. Rundell and H. Sato 2003 The contribution of Framenet to practical lexicography. International Journal of Lexicography. 16.3:333--357

Baker, C.F., C.J. Fillmore, and B. Cronin 2003 The Structure of the FrameNet

Prepublication version of:
Ruppenhofer, Josef, Hans C. Boas, and Collin Baker. 2013. The FrameNet approach to relating syntax and semantics. In: R.H. Gouws, U. Heid, W. Schweickard, and H.E. Wiegand (eds.), Dictionaries. An International Encyclopedia of Lexicography, 1320–1329. Berlin/New York: De Gruyter/Mouton.

Database. International Journal of Lexicography. 16.3: 281--296.

Baker, C.F. and J. Ruppenhofer 2002 FrameNet's Frames vs. Levin's Verb Classes. In: J. Larson and M. Paster (eds), Proceedings of 28th Annual Meeting of the Berkeley Linguistics Society, 27--38.

Bethard, S., H. Yu, A. Thornton, V. Hatzivassiloglou, and D. Jurafsky 2004 Automatic extraction of opinion propositions and their holders. In: 2004 AAAI Spring Symposium on Exploring Attitude and Affect in Text.

Boas, H.C. 2005 From Theory to Practice: Frame Semantics and the Design of FrameNet. In: S. Langer and D. Schnorbusch (eds.), Semantisches Wissen im Lexikon,129--160. Tübingen: Gunter Narr.

Boas, H.C. (ed.) 2009 Multilingual FrameNets in Computational Lexicography. Berlin/New York: Mouton de Gruyter.

Boas, H.C. 2011 A frame-semantic approach to syntactic alternations with build-verbs. In: P. Guerrero Medina (ed.), Morphosyntactic alternations in English, 207--234. London: Equinox.

Borin, L., M. Toporowska Gronostaj and D. Kokkinakis 2007 Medical frames as target and tool. In: R. Johannson and P. Nugues (eds.), Frame 2007: Building frame semantics resources for Scandinavian and Baltic languages, 11--18.

Burchardt, A. 2008 Modeling Textual Entailment with Role-Semantic Information. Ph.D. Thesis, Computational linguistics and phonetics, Saarland University.

de Melo, G., C.F. Baker, R. Passonneau, C. Fellbaum, and N. Ide. 2012 Empirical Comparisons of MASC Word Sense Annotations: a First Look. *LREC 2012.*

Dolbey, A. 2009 BioFrameNet: a FrameNet Extension to the Domain of Molecular Biology. Ph.D. thesis, Department of Linguistics, UC Berkeley.

Fellbaum, C. and C.F. Baker submitted Aligning Verbs in WordNet and FrameNet. Linguistics.

Fillmore, C. J. 1977 Scenes-and-frames semantics. In: A. Zampolli (ed.), Linguistic Structures Processing, 55—88. (Fundamental Studies in Computer Science.) Amsterdam: North Holland Publishing.

Fillmore, C. J. 1985 Frames and the semantics of understanding. Quaderni di Semantica. 6: 222--254.

Fillmore, C. J. 2007 Valency issues in FrameNet. In: T. Herbst and K. Goetz-Votteler (eds.), Valency: theoretical, descriptive, and cognitive issues.

Fillmore, C.J. and C. F. Baker 2010 A Frame Approach to Semantic Analysis. In: B. Heine and H. Narrog, (eds.), Oxford Handbook of Linguistic Analysis, 313--341. Oxford: Oxford University Press.

Fillmore, C.J., C.R. Johnson, and M.R.L. Petruck 2003 Background to FrameNet. International Journal of Lexicography 16.3: 235--250.

Fillmore, C.J., M.R.L. Petruck, J. Ruppenhofer, and A. Wright 2003 FrameNet in Action: The Case of Attaching. International Journal of Lexicography 16: 297--332.

Fillmore, C.J., R. Lee-Goldman, and R. Rhodes 2012 The FrameNet Constructicon. In H.C. Boas and I. Sag (eds.), Sign-based Construction Grammar. Stanford: CSLI Publications.

Gildea, D. and D. Jurafsky 2002 Automatic labeling of semantic roles. Computational Linguistics 28: 245--288.

Green, R., B.J. Dorr and P. Resnik 2004 Inducing frame semantic verb classes from WordNet and LDOCE. In: D. Scott, W. Daelemans, M.A. Walker (eds.), Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, 375--382.

Hong, J. and C.F. Baker 2011 How good is the crowd at "real" WSD? In *Proceedings of the 5th Linguistic Annotation Workshop* (LAW V '11). Association for Computational Linguistics, Stroudsburg, PA, USA, 30-37.

Kay, P. and C.J. Fillmore 1999 Grammatical constructions and linguistic generalizations: the What's X doing Y? construction. Language. 75.1--33.

Kipper, K., A. Korhonen, N. Ryant, and M. Palmer 2006 A large-scale classification of English verbs. Language Resources and Evaluation 41.1:21--40.

Levin, B. 1993 English verb classes and alternations: A preliminary investigation. Chicago, IL: University of Chicago press.

Marcus, M., G. Kim, M.A. Marcinkiewicz , R.Macintyre, A. Bies , M. Ferguson , K. Katz ,

B. Schasberger 1994 The Penn treebank: Annotating predicate argument structure. Proceedings of the workshop on Human Language Technology. 114--119.

Meyers, A., R. Reeves, C. Macleod, R. Szekely, V. Zielinska, B. Young, and R. Grishman 2004 Annotating Noun Argument Structure for NomBank. In: M.T. Lino, M.F. Xavier, F. Ferreira, R. Costa, and R. Silva (eds), Proceedings of LREC-2004, held at Lisbon, Portugal, 803--806.

Miller, G. A., R. Beckwith, C. Fellbaum, D. Gross and K. Miller 1990 Introduction to wordnet: An on-line lexical database.International journal of lexicography 3.4:235--244.

Pado, U. 2007 The integration of syntax and semantic plausibility in a wide-coverage model of human sentence processing. Ph.D. thesis, Computational linguistics and phonetics, Saarland University.

Palmer, M., D. Gildea and P. Kingsbury 2005 The Proposition Bank: An Annotated Corpus of Semantic Roles. Computational Linguistics 31:71-106.

Palmer, M., H. Trang Dang and C. Fellbaum 2007 Making Fine-grained and Coarse-grained sense distinctions, both

manually and automatically. Natural Language Engineering 13: 137--163.

Petruck, M.R.L., C.J. Fillmore, C.F. Baker, M. Ellsworth, and J. Ruppenhofer 2004 Reframing FrameNet Data. In: G. Williams and S. Vessier (eds.), Proceedings of The 11th EURALEX International Congress, held at Lorient, France 2004, 405--416.

Ruppenhofer, J., M. Ellsworth, M.R.L. Petruck, C.R. Johnson and J. Scheffczyk 2010 FrameNet II: Extended Theory and Practice. Technical Report. Berkeley, California: International Computer Science Institute.

Sato, H., S. Kuboya, T. Sone, H. Tagami and K. Ohara 2008 The Japanese FrameNet Software Tools. In: European Language Resources Association (ed.), Proceedings of the 6th International Conference on Language Resources and Evaluation, held at Marrakech, Morocco, 2008.

Schmidt, T. 2008 The Kicktionary revisited In: A. Rothkegel and J. Laffling (eds), Text Resources and Lexical Knowledge. Berlin: Mouton de Gruyter, 239-251.

Schulte im Walde, S. 2008 Human Associations and the Choice of Features for

Semantic Verb Cassification. Research on Language and Computation 6.1:79--111.

Shen, D. and M. Lapata 2007 Using Semantic Roles to Improve Question Answering. In: J. Eisner (ed.), Proceedings of the Conference on Empirical Methods in Natural Language Processing and on Computational Natural Language Learning, 12--21.

9.2 Electronic resources

FrameGrapher visualization tool
https://framenet.icsi.berkeley.edu/fndrupal/FrameGrapher Last access: Jan 31, 2012

FrameNet API for C/.Net
https://ptl.sys.virginia.edu/msg8u/NLP/Source/ResourceAPIs/FrameNet/ Last access: Jan 31, 2012

FrameNet API for JAVA www.cl.uni-heidelberg.de/trac/FrameNetAPI Last access: Jan 31, 2012

FrameNet Glossary
https://framenet.icsi.berkeley.edu/fndrupal/glossary Last access: Jan 31, 2012

FrameNet Website
https://framenet.icsi.berkeley.edu/fndrupal/ Last access: Jan 31, 2012

Prepublication version of:
Ruppenhofer, Josef, Hans C. Boas, and Collin Baker. 2013. The FrameNet approach to relating syntax and semantics. In: R.H. Gouws, U. Heid, W. Schweickard, and H.E. Wiegand (eds.), Dictionaries. An International Encyclopedia of Lexicography, 1320–1329. Berlin/New York: De Gruyter/Mouton.

FrameSQL
http://framenet2.icsi.berkeley.edu/frameSQL/fn2_15/notes/ Last access: Jan 31, 2012

Kicktionary www.kicktionary.de Last access: Jan 31, 2012

Lexicography Master Class
http://www.lexmasterclass.com Last access: Jan 31, 2012

PropBank Website
http://verbs.colorado.edu/~mpalmer/projects/ace.html Last access: Jan 31, 2012

SemEval
http://aclweb.org/aclwiki/index.php?title=SemEval_Portal Last access: Jan 31, 2012

Swedish FrameNet++
http://spraakbanken.gu.se/swefn/ Last access: Jan 31, 2012