
BACKGROUND TO FRAMENET

Charles J. Fillmore: *International Computer Science Institute and University of California, Berkeley, California, USA* (fillmore@icsi.berkeley.edu),

Christopher R. Johnson: *University of Chicago, USA* (crj@uchicago.edu), and

Miriam R.L. Petrucc: *International Computer Science Institute and University of California, Berkeley, California, USA* (miriamp@icsi.berkeley.edu)

Abstract

This article presents general background information about the FrameNet project, including an introduction to its basic assumptions and goals, a description of its precursors, and information about its evolution during the six years of the project. The companion articles in this special issue of IJL describe various aspects of the project in greater detail.

I. What is FrameNet?

FrameNet is a computational lexicography project that extracts information about the linked semantic and syntactic properties of English words from large electronic text corpora, using both manual and automatic procedures, and presents this information in a variety of web-based reports. The name ‘FrameNet’, inspired by ‘WordNet’ (Fellbaum 1998), reflects the fact that the project is based on the theory of *Frame Semantics*, and that it is concerned with *networks* of meaning in which words participate.¹

Frame Semantics is an approach to the study of lexical meaning based on work by Fillmore and his collaborators over the past thirty-odd years (1975, 1977a, 1977b, 1982, 1985, among others). The central idea of Frame Semantics is that word meanings must be described in relation to *semantic frames* – schematic representations of the conceptual structures and patterns of beliefs, practices, institutions, images, etc. that provide a foundation for meaningful interaction in a given speech community. FrameNet identifies and describes semantic frames, and analyzes the meanings of words by directly appealing to the frames that underlie their meanings and studying the syntactic properties of words by asking how their semantic properties are given syntactic form.

The primary units of lexical analysis in FrameNet are the frame and the *lexical unit* (LU: Cruse 1986), defined as a pairing of a word with a sense (e.g.,

the *hot* of temperature and the *hot* of taste experiences are two among the many lexical units that use the adjective *hot*). Generally speaking, the separate senses of a word correspond to the different semantic frames that the word can participate in (or, as we will see below, different sets of frames). When a word's sense is based on a particular frame, we say that the word *evokes* the frame: thus, the word *hot* is capable of evoking a temperature scale frame in some contexts and a particular taste experience frame in others. Interpreting a sentence containing this word requires assumptions about which frame is relevant in the given context.

1.1. Valence

Some words play a special role in the structure of sentences, because their meanings are intrinsically relational, and require the support of other phrases in a sentence to be coherent. This property is typically illustrated with verbs, and shows up in perhaps its simplest form in the phenomenon of transitivity. Thus, some verbs must combine with direct objects in order to make sense and result in grammatical sentences, as shown below in (1).

- (1) a. The player *hit* the ball.
- b. *The player *hit*.

The asterisk indicates that (b) is an unacceptable sentence of English. We can explain the unacceptability by recognizing that the meaning of the verb *hit* involves an entity doing the hitting (expressed by the subject in a simple active sentence) and another entity with which the first comes into contact (expressed by the direct object). Unless a sentence with the verb *hit* mentions both entities, it is neither grammatically nor semantically acceptable. We can say the verb *hit*, in a normal active sentence, 'requires' a subject that expresses the hitter, and a direct object that expresses the *hit* entity. The requirement that a word combine with particular kinds of phrases in a sentence is often referred to as the *valence* of the word, by analogy to the word as used in chemistry to refer to the combining power of atoms. The phrases with which a word combines are often called *complements*, because they complement or complete the word, both semantically and syntactically. They are also sometimes called *arguments*, by analogy to predicate-argument structure in symbolic logic. Valence goes far beyond transitivity. In FrameNet, information about valence must be specified in both semantic and syntactic terms; the semantic roles that complements play with respect to the meaning of the word must be accounted for, and the grammatical properties of the possible complements of a word must be identified. Semantic valence information is often recorded in a notation that is similar to logic, and referred to as *argument structure*. Syntactic valence information is usually specified in terms of the phrase types (e.g. noun phrase, prepositional phrase, etc.) of the possible complements, and in terms of the grammatical functions (e.g. subject, object, etc.) that the complements bear with

respect to the word. Syntactic valence information is sometimes expressed in terms of *subcategorization frames*, these referring to the kinds of sets of grammatical phrase types that accompany a verb in predicate phrases: such sets are seen as classifying verbs into such categories as transitive, intransitive - and a great number of other types.

In FrameNet, the semantic valence properties of a word are expressed in terms of the kinds of entities that can participate in frames of the type evoked by the word. We call these roles *frame elements* (FEs). FEs bear some resemblance to the argument variables used in first-order predicate logic, but have important differences deriving from the fact that frames are much more complex than logical predicates. Here is a description of a simple frame and its FEs:

Frame: Transfer

FEs: DONOR, THEME, RECIPIENT

Description: Someone (the DONOR) is in possession of something (the THEME), and then causes someone else (the RECIPIENT) to be in possession of the THEME, perhaps by causing the THEME to move to the RECIPIENT.

We use font differences for the names of frames and FEs to show that we are not simply using the English words on which the names are based. The FE name DONOR, for example, does not indicate that the being who transfers a THEME is a donor in the normal English sense of the word (i.e. someone who makes a charitable contribution). Rather, all the important properties of the DONOR *role* are characterized by the description of the frame. The particular names of the FEs simply serve a mnemonic purpose.

FrameNet uses frames and FEs to describe the valence properties of predicates. For example, the verb *give* can be described in terms of the Transfer frame. We find *give*, in its transfer sense, occurring in two basic complementation patterns:

- (2) a. The teacher **gave** the student a book.
- b. The teacher **gave** a book to the student.

The major constituents of both of these sentences can be understood in terms of the semantic and grammatical roles that they play with respect to the verb *give*, which we refer to as the *target* and indicate in examples with boldface. The semantic roles are characterized by the FEs of the Transfer frame, which is evoked by *give*. In both of these sentences, the DONOR role is expressed by the noun phrase (NP) *the teacher*, and the THEME role is expressed by the NP *a book*. The RECIPIENT is expressed by the NP *the student* in the first sentence, and by the prepositional phrase (PP) *to the student* in the second sentence.

Grammatically, *the teacher* is the subject of *gave* in both sentences (we call subjects *external arguments* (Ext)). In the first sentence, *the student* is the grammatical object (Obj) of *gave*, and *a book* is what we call an NP *complement* (Comp). In the second sentence, *a book* is the Obj of *gave*, and *to the student* is

a PP Comp. These patterns of grammatical realization are represented as triples of FE names, phrase types (PTs) and grammatical functions (GFs), as seen in the columns in (3):

(3)	give	<i>FEs:</i>	Donor	Theme	Recipient
		<i>PTs:</i>	NP	NP	NP
		<i>GFs:</i>	Ext	Comp	Obj
	give	<i>FEs:</i>	Donor	Theme	Recipient
		<i>PTs:</i>	NP	NP	PP-to
		<i>GFs:</i>	Ext	Obj	Comp

We refer to patterns like these as *valence patterns*. One of the main purposes of FrameNet is to identify valence patterns for a large number of English verbs, nouns, adjectives, adverbs, and prepositions, and to annotate corpus citations to show how those valence patterns are instantiated in actual sentences.

1.2. Relations between words in a frame

Semantic frames provide a way to characterize semantic relations between words. For example, the verbs *give* and *receive* are related by the fact that they both evoke the Transfer frame. This meaning relation is apparent from inferences or lexical entailments associated with these words. If we are told that the teacher gave a book to the student, we can infer that the student received a book from the teacher. The verbs differ in the *perspectives* that they impose on transfer events, and on the attendant difference in the relative salience of the DONOR and RECIPIENT, which shows up in the syntactic expression of the FEs. In contrast to *give*, *receive* expresses the RECIPIENT as the Ext, and expresses the DONOR optionally in a PP Comp headed by *from*:

- (4) The student **received** a book (from the teacher).

receive	Donor	Theme	Recipient
	(PP-from)	NP	NP
	(Comp) Obj		Ext

Differences like these between words in a frame show that there is a complicated relation between FEs and the more general *thematic roles* that linguists often use to describe argument structure. Thematic roles are meant to capture regularities about *linking*, i.e. the relation between semantic participant roles and the grammatical functions of constituents that express them. A thematic role analysis might claim that the verbs *give* and *receive* have different sets of thematic roles:

(5)	give	Agent	Theme	Recipient
	receive	Recipient	Theme	Source

This analysis preserves generalizations about linking – in particular, that Agents tend to be expressed as subjects (Exts), and that Recipients are more

likely to be subjects than are Sources. However, in doing so it misses the meaning relation that is captured in the frame semantic analysis. In the Transfer frame, it is apparent from the description that the DONOR has properties of both an Agent and a Source, emphasized by the verbs *give* and *receive* respectively.

1.3. Relations between frames

In addition to relations between words within a frame, FrameNet is also interested in various relations between frames.

Sometimes one frame has all the properties of another frame and elaborates on it, adding some detail. In this case we say that the elaborating frame *inherits* the simpler, less detailed frame. For example, there is an elaboration of the Transfer frame in which transferring is done via some institution, channel, or medium. This frame is evoked by verbs like *mail*, *wire* and *fax*, which have the syntactic properties of *give*:

- (6) a. The company **mailed** me the books.
b. The company **mailed** them to me.
- (7) a. Pat **wired** Lou the money.
b. Pat **wired** the money to Lou.
- (8) a. The reporter **faxed** the editor the story.
b. The reporter **faxed** the story to the editor.

Sometimes one frame is a part of another more complex frame. For example, the Transfer frame plays a role in the more complex Commercial Transaction frame, which has the frame elements BUYER, SELLER, GOODS, and MONEY. We do not think of the Commercial Transaction frame as an elaboration of the Transfer frame, because a Commercial Transaction event is not just a more specific type of Transfer event. Rather, it is a complex event that consists of two Transfer sub-events: the BUYER gives the SELLER some MONEY, and the SELLER gives the BUYER the GOODS. The verbs *pay* and *sell*, respectively, give special salience to these two Transfer sub-events. Consequently, they also have syntactic properties of *give*, though they also have more roles.

- (9) a. The student **paid** me \$2000 (for the car).
b. I **paid** \$5000 to a dealership (for that car).
- (10) a. The dealership **sold** me the car (for \$5000).
b. I **sold** the car to the student (for \$2000).

1.4. The Database

The kind of information in the FrameNet database is not expressed in the same level of depth in any existing print dictionary or computational lexical resource. By recognizing conceptual schemas that mediate the semantic relations between

words, FrameNet provides the means to link arguments of predicating words with the semantic roles they express.

FrameNet data is stored in a relational database that reflects, insofar as possible, the theoretical basis of the project. Because of the different kinds of information that is represented in the FrameNet database, it is convenient to characterize it in terms of two parts: the *lexical database* and the *annotation database*.

1.4.1 *Lexical database*. The lexical database holds information about frames and frame elements, as well as lemmas, lexemes, word forms, and parts of speech. In other words, it contains all that is necessary to characterize LUs. Frame-to-frame relations, such as the Inheritance and SubFrame relations, discussed briefly below, and FE-to-FE relations within related frames, are also indicated in the lexical database.

1.4.2 *Annotation database*. The annotation database stores the annotated sentences, along with the subcorpora from which they were selected for annotation. For each target word in respect to which the annotation of exemplifying sentences is done, there is a set of annotation layers for the frame elements, phrase types, and grammatical functions.

2. Precursors to FrameNet

2.1. Semantic Roles

The linguistic basis for FrameNet can be traced back to Fillmore's theory of Case Grammar, beginning with work in the late 60's (Fillmore 1968). It was offered as contribution to generative-transformational grammar, and consisted of the proposal that syntactic deep structures are best expressed as configurations of 'deep cases', these given general semantic-role names such as Agent, Patient, Goal, etc. As applied to verbs, the idea was that a verb had to be described first in terms of the assembly of semantic roles forming an essential part of its meaning, and second in terms of the rules that were needed for converting these into grammatically realized constituents – for example, as subjects, objects, and oblique phrases. From the start, however, there were questions about the correct set of semantic role labels, or in fact about whether it was possible to characterize the predicates of natural languages using a small set of such labels. Fillmore's later work on lexical semantics (Fillmore 1976, 1977a, 1982, 1985) led to the conviction that a small fixed set of 'deep case' roles was not sufficient to characterize the complementation properties of lexical items.² Indeed, frame elements are designated in terms of frame specific situational roles, rather than semantic roles as articulated in Case Grammar.

2.2. RISK Study

As outlined above, frame semantics is based on the idea that word meanings are organized around schematic conceptual scenarios, or *frames*, that underlie the use and interpretation of the lexical items and their general complementation and modification properties. Fillmore's (1978) characterization of the frame as the most central and powerful kind of domain structure paved the way for a frame-based organization of the lexicon, where the frame provides the conceptual underpinnings for related senses of a single word and semantically related words. With this approach to lexical semantic analysis and description, it is possible to characterize all categories of words, as well as phrases and expressions, using the same apparatus – the frame.

Combining notions from frame semantics with corpus-based lexicographic practice, Fillmore and Atkins conducted a large-scale study of the English word *risk* using dictionary descriptions and corpora. Their work resulted in two papers (Fillmore and Atkins 1992, 1994) which suggested the possibilities of designing a dictionary that aimed at the coverage provided by exploiting a corpus and interpretations shaped by frame-semantic notions. The *Risk* frame was characterized with two sub-frames, *Chance* and *Harm*, and a description of the categories needed for the word's valence description. The lexico-syntactic patterns in which *risk* occurs, as a verb and as a noun, were presented in detail, along with the polysemy system of the verb. The different uses of the verb were explained in terms of the different syntactic realizations of the elements in their common semantic frames. Compare the following sentences, each of which illustrates a use of the verb *risk*.

- | | | |
|------|--|-----------------------|
| (11) | a. Fred risked [his relationship | ASSET]. |
| | b. Fred risked [excommunication | BAD_OUTCOME]. |
| | c. Fred risked [telling the truth | RISK_ACTION]. |

Notice that in the first two sentences, it is an entity that is risked, which can be an **ASSET**, as in (a), or a **BAD_OUTCOME**, as in (b); furthermore, that entity is realized as the grammatical object of the verb. In (c), however, we characterize that which is risked as a **RISK_ACTION**, and note that the constituent is a VP-ing Comp. The work on *risk* described in Fillmore and Atkins (1992, 1994) constitutes the first attempt at demonstrating the applicability of frame semantic theory to lexicographic practice, as it instantiated a frame-based lexicon. It served as a model for the DELIS project.

2.3. The DELIS Project

In 1993, Heid, Atkins and an international group of computational linguists including Nicoletta Calzolari (ILC Pisa), Anna Braasch (CST Copenhagen), the late Ole Norling Christensen (Danish National Dictionary), Nicholas Ostler (Linguacubun Ltd.), Annie Zaenen (Xerox) and Willy Martin (Free University, Amsterdam), together with Fillmore as consultant, undertook the European-

Union sponsored DELIS project (DEscriptive Linguistic Specifications).³ This research - the first to be grounded in frame semantics – produced a contrastive lexicon for English, Dutch, French, Italian, and Danish verbs of communication (*say, talk, speak*, and their near-equivalents in the other languages) and verbs of perception (*hear, listen, see, look, watch, feel, taste, smell*, etc.), with the frame element configurations serving as an interlingua. The first FrameNet project came out of the experience gained in DELIS, including a keen appreciation of many problems of linguistic description, which led to fresh approaches, such as frame inheritance and multiple frame inheritance, in the context of FrameNet.

2.4. *Proposals*

Fillmore, with Atkins and Heid and J. B. Lowe, wrote a proposal for a frame-based lexicon that would cover the general vocabulary of English, which was funded in 1997. The goal of the first phase of the work was to develop a set of tools for lexicon building, and begin building the FrameNet database. During the second phase of the project, which began in 2000, work continued on building the database by increasing word coverage and annotating example sentences. In addition, the goal was to demonstrate the usefulness of the database as a lexical resource for its application to speech and language technology.

2.5. *Principals*

During the first phase of the project Fillmore was the principal investigator, with Dan Jurafsky (University of Colorado) as co-principal investigator; Atkins served as lexicographic consultant and Heid served as program consultant. J.B. Lowe was the technical director and Hiraoki Sato developed the FrameSQL graphical user interface (GUI). For the second phase of the project, three co-principal investigators participated, each using FrameNet data in a natural language processing task: Jurafsky, studying question-answering; J. Mark Gawron working on machine translation; and Srini Narayanan investigating information extraction, further details of which are given in section 6.2. Atkins and Sato continued in their original capacities; Charles Wooters designed the new database (Fillmore, Wooters, and Baker 2001) and Beau Cronin became the head programmer. Throughout the six-year period of FrameNet, many researchers from academia and industry visited the FrameNet worksite to learn more about the project and discuss possibilities of collaboration.

3. Evolution of the Project

In line with its emphasis on empirical evidence, FrameNet was designed to allow its picture of lexical organization to emerge organically, from the bottom up, rather than being fixed from the start by static preconceptions. And in fact,

a number of changes have been introduced since the start of the project. These are discussed briefly here.

3.1. Domains vs. Inheritance

In the first phase of FrameNet, we grouped words into *domains*, partly as a matter of convenience, but also to ensure coverage of different areas of the general vocabulary. At the time we were not explicit about the theoretical basis of groupings of frames into domains, and we came to understand that we needed to represent the different ways in which frames could be related to each other. In the second phase of the project, we defined several frame-to-frame relations and began to implement them in the database. One such relationship is that of inheritance, where more specific frames inherit all the features of a more general frame. This means that all of the frame elements, subframes, and semantic types of the parent have equally or more specific correspondents in the child frame. To illustrate, the *Communication_means* frame, evoked by the verbs *mail*, *wire* and *fax*, inherits from the more general *Communication* frame, and thus all of the FEs of *Communication* have correspondents in the *Communication_means* frame as follows:

(12)	Communication	Communication_means
	ADDRESSEE	ADDRESSEE
	COMMUNICATOR	COMMUNICATOR
	MEDIUM	DOCUMENT
	MESSAGE	MESSAGE
	TOPIC	TOPIC

Notice that the FE *MEDIUM* from the *Communication* frame corresponds to the FE *DOCUMENT* in the *Communication_means* frame, illustrating that the FE in the child frame is more specific than its correspondent in the parent frame.⁴ Moreover, the semantic typing, for instance, of the FEs *ADDRESSEE* and *COMMUNICATOR* (both of which are +Human) carries over from the *Communication* frame to *Communication_means*.

3.2. Support Verbs and Transparent Nouns

Recognizing the discrepancies between syntactic and semantic structure required that we recognize the role of *support verbs* and various classes of *transparent nouns*. While our thinking about the importance of support verbs began rather early on, greater interest in transparent nouns developed as we started to articulate the unique quality of the FrameNet database, specifically that of providing semantic information about LUs.

3.2.1 *Support Verbs*. Certain semantically neutral verbs can turn an event noun or a state noun into a verb phrase-like predicate and allow for the expression of a frame element as their subjects. We call such verbs support verbs.⁵ For example, both sentences in (13) report on the same event, that of deciding something and (13)(b) is not about an event of making. We want to record the fact that the noun phrase *the committee* instantiates the same frame element in both sentences, and recognizing the role of the support verb *make* allows us to do so.

- (13) a. The committee **decided** to convene again next month.
b. The committee made a **decision** to convene again next month.

Support verbs also provide lexicographically relevant information since they are selected by the noun rather than the other way around. Consider the following:

- (14) a. say a prayer/pray
b. *give a prayer
c. give a speech/speak
d. *say a speech

The support verb for the noun *prayer* is *say*; and the support verb for *speech* is *give*. Reversing the support verbs yields ungrammatical collocations, as shown in 14(b) and 14(d).

Finally, support verbs have other properties that we want to record in our database. For instance, different support verbs can determine the semantic role instantiated by a given constituent in a sentence, as shown below in (15). In the first sentence, the syntactic subject is the EXPERIENCER, while in the second sentence the syntactic subject is the AGENT.

- (15) a. Sean underwent a surgical **procedure**.
b. Sean performed a surgical **procedure**.

3.2.2 *Transparent Nouns*. A transparent noun is one which can appear as the first noun in N₁-of-N₂ constructions in contexts where the governing verb actually selects N₂ semantically rather than N₁, the syntactic head (see also Fillmore and Sato 2002 and Fontenelle 1999). Some examples include *glass*, *piece*, and *bunch*, as seen here:

- (16) a. Maria sipped a **glass** of dry white wine.
b. Rafael fingered a **piece** of sheer black silk.
c. Jenn bought a **bunch** of carrots at the produce market.

Nouns that can be transparent fall into several semantic classes, including aggregates (e.g. *bunch*, *group*), quantities (e.g. *gaggle*, *flood*), types (e.g. *kind*, *ilk*), portions and parts (e.g. *top*, *piece*), unitizers (*container*, *flask*), and evaluations (e.g. *gem*, *idiot*). While these words were annotated in appropriate frames such as Unitizers for *glass*, annotators added a special kind of tagging,

which we call Gov-X Annotation, for those sentences where the target noun appears as a transparent N₁, as illustrated in (17). This indicated that the governing verb ‘saw through’ the transparent noun syntactic head of the construction and selected the semantic head. In (17), we understand that Maria drank some dry white wine, and not a glass.

- (17) Maria [sipped GOV] [a **glass** of dry white wine X]

3.3. Semantic Types

We introduced the concept of *semantic type* in order to capture semantic facts about frames, FEs, or LUs that didn’t necessarily fit into our developing hierarchy of frames. For example, lexical units across a range of frames may incorporate positive versus negative evaluation. Consider the Judgment verbs *praise* and *criticize*, the Experiencer_subject verbs *like* and *hate*, and the Frugality adjectives *generous* and *stingy*. We can indicate the semantic type ‘positive evaluation’ on the first word of each pair, and ‘negative evaluation’ on the second, thus recording the information in the database. In addition, for some frames, while we wanted to include a full listing of lexical units, we decided not to annotate in respect to all of them. For example, the Calendric_units frame includes the names for the twelve months of the year, but we only annotated in respect to two of them. Using the semantic type ‘month name’ on the other ten provides a means of capturing the fact that these lexical units behave in the same way as the two that were annotated. Other uses of semantic type include basic typing of frame elements, for example, ‘sentient’ for the FE COGNIZER, and functional marking of frames, such as ‘non-lexical’ on frames that are present to participate in Inheritance, Subframe, or Use relations with other frames.⁶

3.4. Null Instantiation

In the first phase of the project, the software that we used required that annotation labels be tagged on a piece of text. We created a dummy symbol on which annotators tagged missing frame elements, that is, FEs that are conceptually necessary but do not show up as lexical or phrasal material in the sentence chosen for annotation. We wanted to indicate their absence because it provides lexicographically relevant information about omissibility conditions. In the second phase of FrameNet, we used in-house developed software that precluded the need for the dummy symbol.

We recognize three types of *null instantiations* – *constructional*, *definite*, and *indefinite*, the latter two of which are lexically specific. The omission of constructionally omitted constituents is licensed by a grammatical construction and is of little interest for lexicographic purposes: examples are the omitted subjects in imperative sentences (*Leave now, please!*), omitted agents in passive sentences (*We were robbed.*) and the like.

With definite null instantiation (DNI), the omission of an element is only licensed by an agreement that it is understood in the linguistic or discourse context: one doesn't say *I object!* unless both speaker and hearer are aware of the proposition currently being opposed. This is sometimes referred to as *anaphoric zero*. The concept Indefinite Null Instantiation (INI) covers mainly implicit arguments of certain transitive verbs that are characterized as used intransitively, for example with such verbs as *eat*, *bake*, *drink*, and so on, we know that something is involved as the second participant in the associated acts, but its nature is left unspecified.⁷

3.5. Grammatical Functions and Phrase Types

During the second phase of the project, we came to understand that we needed to expand our list of grammatical functions and phrase types to cover the facts that we found in the data. For example, in measure phrases (e.g. *two cups of coffee*), we wanted to characterize the quantifier or number that precedes the noun as bearing a special grammatical function in relation to the target noun that it modifies. In these cases, we introduced the GF Quant(ifier) and the PT Num(ber), used as illustrated in (18).

- (18) Sarah poured [two Quant/Num] **cups** of coffee.

4. How the Work is Done

4.1. Empirical Work

Our work proceeds empirically, allowing us to discover things not known intuitively or found in dictionaries. For instance, as discussed further in Fillmore et al.'s paper on 'FrameNet in Action' in the present volume, the LU *tie.v* in the Attaching frame is used to talk about an event in which somebody causes one thing to be physically connected to something else, or causes two things to be connected to each other. Examination of the data from our corpus determined that the 'connected-to' entity is larger and more stable than the 'connected' entity. The sentence in (19) illustrates the point. Such information is not intuitively apparent.

- (19) He **tied** the driving wheel to Pete's cardboard box with string.

Furthermore, the empirical work required that we distinguish (at least) between *tie.v* in the sense of tying one thing to another with a rope-like entity, which we characterize in terms of an Attaching frame, and *tie.v* in the sense of tying a knot, which we characterize in terms of the Knot_creation frame. In our understanding, Knot_creation differs from Attaching in a number of important ways, but crucially in that it essentially involves creating a knot, rather than attaching one entity to another.

Finally, the empirical work allows us to develop a perspective on the lexicon, specifically one that is based on the uncontroversial assumption that to understand word meaning we must first have knowledge of the conceptual structures, or semantic frames, which provide the background and motivation for their existence in the language and their use in discourse.

4.2. Assumptions

4.2.1 *Corpus assumption.* One of the commitments of the FrameNet project is to make generalizations based on attested corpus sentences. This is motivated by the belief that there are almost always interesting generalizations to make about lexical items and their complementation properties that elude introspection. For most of the project, the British National Corpus was our only source of data;⁸ during the last year, we also began to use the American Newswire Corpus.⁹

4.2.2 *Theory-neutral grammatical assumption.* Another commitment of the FrameNet project has been to use relatively theory-neutral grammatical descriptions. As a practical matter, this has typically meant limiting the grammatical analysis of sentences to the recognition and labeling of the major constituents expressing frame elements. The motivation for this commitment is a practical one: we would like the FrameNet database to be useful to researchers from a variety of theoretical backgrounds, as well as those whose goals are practical rather than theoretical.

4.3. The Work

The daily work of FrameNet includes semantic frame development, corpus extraction, and annotation of example sentences. Semantic frame development involves an initial informal characterization of the kind of entity or situation represented by the frame, selecting names for labeling the entities or components of the frame (the FEs), and constructing lists of words that appear to belong to the frame. Corpus extraction requires examining the use of a central member of the frame in the corpus, first to verify our understanding of the syntax and semantics of the words in the frame and then to determine the syntactic and collocational contexts for the sense we have in mind. We have developed an automatic process to make extracted subcorpora ready for annotation. Human annotators choose representative instances of each LU and add frame relevant labels to whole constituents in sentences; automatic processes add grammatical function (GF) and phrase type (PT) information to the annotation. Another set of automatic processes uses the resulting annotations to produce corpus-based formal descriptions of the valences of each LU. A variety of reports and viewing tools are used to display the results of the analysis and the annotation internally and on the web.

5. Types of Information in the Database

5.1. *Things Not Included*

There are a number of important types of information about lexical units that are not included in the FrameNet database. First of all, there is no phonological (i.e. how to pronounce words), morphological (i.e. the structure and formation), or etymological (i.e. the origin and historical development) information about the words in the database. Moreover, we do not provide information about lexical relations, such as synonymy, antonymy, or hyponymy, other than that which could be deduced from studying the LUs in a given frame or examining the frame-to-frame relations among a set of frames. Finally, although we aim to provide three to five annotated example sentences for each syntactic pattern in which a given LU is used, we do not offer any statistical information about frequency of occurrence of syntactic patterns or about LUs.

5.2. *Things Included*

The FrameNet database contains lexical entries for individual words, descriptions of frames and frame elements, and annotated subcorpora. Each lexical entry includes a sense description, either from the Concise Oxford Dictionary (Pearsall 1999, COD) or one created by a FrameNet lexicographer, and provides summaries of the syntactic realization of the frame elements and the valence patterns, discussed in Section 1.1, above (see Section 8 of Fillmore et al.'s article on 'FrameNet in Action' in this volume).

6. What is it Good For?

6.1. *Lexicography*

We believe that lexicographers starting a new dictionary entry from scratch or revising an existing one, will welcome FrameNet's complete characterization of the headword's grammar and combinatorial properties, with corpus-derived example sentences showing authentic contexts. The word senses in our lexicon already offer an accelerated launch pad for conventional lexicography, although we recognize that at present our coverage is limited. In addition, the paper by Atkins, Rundell and Sato in this volume shows one way that FrameNet data can be used in lexicography, specifically how polysemy structures can be studied with our results.

6.2. NLP Applications

As mentioned above (section 2.4), FrameNet's goal is to show the utility of its data in natural language processing. To that end, various applications are currently being tested. Research in progress at the University of Colorado focuses on creating a semantic parser that produces FrameNet roles for input sentences, building on the work of Gildea (Gildea and Jurafsky 2000). The FrameNet data augmented parser is being used as part of a project on question-answering. At the International Computer Science Institute and the University of California, Berkeley, researchers are investigating the use of the FrameNet semantic resource in Information Extraction (IE) and Natural Language Understanding (NLU) (Mohit and Narayanan 2003). Finally, work at San Diego State University addresses the relation between the frames developed for English and the frames needed to represent Japanese texts (specifically in the crime domain) with a view to using frame semantics in a machine translation system.

6.3. Semantic Research

The FrameNet database provides a wealth of material for research in semantics. Aside from valence information for each lexical unit, it is possible to use the database to study polysemy (e.g. Boas 2001), collocation, the discrepancies between syntactic and semantic structure, as demonstrated by work on support verbs and transparent nouns (e.g. Dodge and Wright 2002, Fillmore and Sato 2002, Fillmore, Baker, and Sato 2002), lexically licensed omissibility (Ruppenhofer and Baker 2003), and inferencing (Chang et al. 2002).

6.4. FrameNet Abroad

At present, there are three projects under way that make use of both the theoretical principles and the unique methodology of FrameNet: Spanish FrameNet, under the direction of Carlos Subirats of the Department of Linguistics of the Autonomous University of Barcelona; Japanese FrameNet, organized by Kyoko Ohara, at Keio University; and German FrameNet in a three-way collaboration between Uli Heid at Stuttgart, Manfred Pinkal at Saarbrücken, and Hans C. Boas at the University of Texas, Austin. In addition, interest has been expressed by researchers for Chinese, French, Italian, and Swedish.

Concluding Remarks

In this paper, we have contextualized the FrameNet project historically, provided a brief description of the key concepts, operating assumptions, and working procedures of the project. The remaining articles in the present volume provide

a more in-depth look at theoretical considerations, the structure of the database, the daily workings of the project, and FrameNet's contribution to the practice of lexicography.

Notes

¹ Unlike WordNet, however, FrameNet does not have the means of directly displaying the 'network' connections: we see lexical units as related to each other by virtue of belonging to the same frame, and through frame-to-frame relations identified in the database, belonging to words in related frames.

² Thematic roles have continued to play an important part in a number of theoretical frameworks even as generative grammar has moved away from the notion of transformational derivations.

³ See <http://www.hltcentral.org/projects/DELIS> for more details.

⁴ Consider 'He faxed me the income tax forms', where faxed is the target in the Communications_Means frame and the phrase the *income tax forms* is the FE DOCUMENT.

⁵ Our notion of *support verb* is broader than the traditional notion of *light verb* but is narrower than Igor Mel'čuk's (1996) *lexical functions*.

⁶ Other frame-to-frame relations introduced in the course of the project are discussed in Baker, Fillmore, and Cronin of the present volume.

⁷ In some uses specific types of entities are assumed: with *Uncle Harry drinks too much* the assumed object is something alcoholic; with *I hate baking* the assumed object tends to be grain-based products rather than, say, ham or potatoes.

⁸ Information about the British National Corpus is available at <http://www.hcu.ox.ac.uk/BNC/>.

⁹ Information about the American National Corpus is available at <http://www.cs.vassar.edu/~ide/anc>.