Border Conflicts:
FrameNet Meets Construction Grammar

Charles J. Fillmore
University of California, Berkeley, CA
and
International Computer Science Institute, Berkeley, CA

1. The Problem

I count myself among the linguists who believe in a continuity between grammar and lexicon (Fillmore et al. 1988, Joshi 1985), and I entertain the common image that each lexical item carries with it instructions on how it fits into a larger semantic-syntactic structure, or, alternatively, on how semantic-syntactic structures are to be built around it. My remarks here specifically concern an ongoing effort to describe and to annotate instances of, *non-core* syntactic structures, and to see how the products of this work can be integrated with the existing lexical resource, called FrameNet (FN), which is a set of procedures, and a growing database for recording the meanings and the semantic and syntactic combinatorial properties of lexical units. The FrameNet project, which I have directed since 1997, has recently begun exploring ways of creating a *constructicon*, a record of English grammatical *constructions*, annotating sentences by noting which parts of them are licensed by which specific constructions.

 The grammatical constructions that belong in the larger constructicon - that is, in a construction-based grammar - include those that cover the basic and familiar patterns of predication, modification, complementation, and determination, but the new project is concentrating on constructions that ordinary parsers are not likely to notice, or that grammar checkers are likely to question. Some of them involve purely grammatical patterns with no reference to any lexical items that participate in them, some involve descriptions of enhanced demands that certain lexical units make on their surroundings, and some are mixtures of the two.

2. The Work, the Product, and the Limitations of FrameNet

Since many features of the new resource are modeled on FrameNet, I think it useful to review FN's goals and activities, and the features of its database (Baker et al. 2003, Fillmore et al. 2003). FrameNet research amounts to

1. describing *lexical units* (LUs) in terms of the *semantic frames* they evoke, and describing those frames (i.e., the situation types, etc., knowledge of which is necessary for interpreting utterances in the language),
2. defining the *frame elements* (FEs) of each frame that are essential for a full understanding of the associated situation type (the frame elements are the props, participants, situation features that need to be identified or taken for granted in sentences for which the frame is relevant),
3. extracting from a very large *corpus* example sentences which contain each LU targeted for analysis (FN has worked mainly with the British National Corpus),
4. selecting from the extracted sentences representative samples that cover the range of combinatorial possibilities, and preparing *annotations* of them as layered segmentation of the sentences, where the segments are labeled according to the FEs they express, as well as the basic syntactic properties of the phrases bearing the FE,

5. displaying the results in *lexical entries* which summarize the discovered *combinatorial affordances*, both semantic and syntactic, as *valence patterns*, and creating links from these patterns to the annotated sentences that evidence them, and
6. defining a network of *frame-to-frame relations* and the graphical means of displaying these, that will show how some frames depend on or are elaborations of other frames.

2.1 The frames

The frames developed in FrameNet are the conceptual structures against which the LUs in the FN lexicon are understood and defined (Fillmore 1982, Fillmore & Atkins 1992, 1994). These can be as general as the location of some entity in an enclosure, or as specific as interest on investment.

One FN frame that is simple enough to describe completely, and just complex enough to be interesting, is the so-called Revenge frame, the nature of which requires understanding a kind of history. In that history, one person (we call him the OFFENDER) did something to harm another person (what he did we call the OFFENSE and his victim we call the INJURED_PARTY); reacting to that act, someone (the AVENGER, possibly the same individual as the INJURED_PARTY) acts so as to do harm to the OFFENDER, and what he does we call the PUNISHMENT. Thus, we have the frame Revenge, and the frame elements AVENGER, OFFENDER, OFFENSE, INJURED_PARTY, and PUNISHMENT. Other features of the Revenge frame include the fact that this kind of pay-back is independent of any judicial system. There is a very large set of verbs, adjectives and nouns that *evoke* this frame, by which we mean that when users of the language understand these words, their understanding includes all of the elements of that scenario. Among the verbs that evoke this frame are *avenge* and *revenge*, the nouns include *vengeance* and *retribution*, there are phrasal verbs like *pay back* and *get even*, adjectives like *vengeful* and *vindictive*, support constructions like *take revenge on*, *wreak vengeance on*, and *exact retribution against*, plus prepositional adverbials like *in retribution*, or *in revenge*.

FrameNet has developed descriptions of over 800 frames to date, and nobody is ready to estimate how many there are altogether. The list from the time of the last official release can be found at {http://framenet.icsi.berkeley.edu}.

2.2 The frame elements

The frame elements (FEs) are somewhat analogous to the *deep cases* of early Fillmore (Fillmore 1968, 1971), *thematic roles* in various generativist writings (Jackendoff 1990), *actants* and *circonstants* in the Tesnière tradition (Tesnière 1959). There are good reasons for not tying the frame elements into any of the familiar lists of semantic roles (agent, patient, theme, experiencer, instrument, etc.). Since annotators are asked to find expressors of frame elements in actual sentences, FE names that are memorable in respect to the frame itself will facilitate such identifications. Thus to take the case of the arguments of *replace* in a sentence like

> *[I] replaced [my stolen bicycle] [with a much cheaper one]*,

it makes more sense to refer to the phrases introducing the two bicycles as the OLD and the NEW than to try to figure out how well these roles can be accommodated in the "standard" lists. (The missing bicycle, in fact, is not a participant in the event described by the sentence but is a necessary element of its meaning.) The recognition of FE commonalities across frames is made possibly by the system of frame-to-frame relations.

We wanted to think of the frame elements as representing the kinds of information that could be expressed in the sentences and phrases in which the frame is "active", and we wanted to

be able to discover which parts of a sentence reveal information about which frame element. There is an important constraint on this task, distinguishing it from annotation practices that seek to learn everything about each event in a continuous text. Since the information we record is supposed to be relevant to the syntactic description of a given lexical unit, we require that the frame elements we attend to are *in grammatical construction with* the lexical unit being described. Annotators will ignore event-relevant information elsewhere in the text.

We make a distinction between *core* and *peripheral* FEs. The core FEs are those that are conceptually necessary in any realization of the frame by the nature of that frame; the peripheral frame elements are the adjuncts that fit the familiar description 'time, place, and manner, etc', especially the 'etc.'. (The core/periphery distinction can vary across frames; for verbs like *reside*, *elapse*, and *behave*, the locative, temporal and manner components, respectively, are not peripheral.) A characteristic of the peripheral FEs is that they have essentially the same meaning and the same syntactic marking wherever they appear; whatever distributional limitations they have are explained by the fact that frames about *happenings* can take time and place modification, frames about *intentional acts* can take instrument and purpose modification, and so on. A third kind of frame element is what we refer to as *extrathematic*: these are expressions (like benefactives or phrases like *in revenge* or *in return*) that have the effect of situating the event signaled by the target's frame in some larger or coterminous situation.

The goal of FrameNet lexical descriptions is, for each frame-bearing word, to match the word's semantic combinatorial requirements with the manner of their syntactic realization. Reversing the point of view, we seek to recognize in the syntactic nature of the phrases around a given frame-bearing lexical unit, information about the participants in situation that is an instance of the frame. The resulting pairing of semantic and syntactic roles constitutes the *valence* description of the item.

2.3 Example sentences
The goal in providing examples was to have, for each lexical unit, a full set of illustrations of its basic combinatorial properties, and we preferred sentences whose content was clearly relevant to the meaning of the word being exhibited. If we were looking for an illustration of *knife*, we would prefer *the butcher sharpened his knife* than *the poet photographed a knife*. These example-selecting decisions were made in resistance to several kinds of pressure. Some members of the research community wanted to see sentences of the most frequent type; but for many verbs, the most frequent examples had mainly pronouns (*I risked it*). Some wanted us to include complex and distorted sentences as well as the simplest type; some wanted us to make sure we include creative uses of a word wherever we found them, scolding us for neglecting metaphor and other figurative uses: our view echoes that of Patrick Hanks (MS), namely, that we had the obligation to produce clear descriptions of the *norm*, leaving it to some auxiliary research to explore the ways in which speakers *exploit* the norm for creative expression. Where a metaphorical use was *lexicalized*, the LU resulting from that lexicalization was included in its appropriate frame.

2.4 The annotation
The original mission of FN was purely lexicographic: to annotate a variety of typical uses of each target LU and to seek to cover a wide range of relevant contexts for the LU (i.e., all of its *valence* possibilities and representative samples of its semantic collocates), and this meant

creating a collection of sentences in which each was annotated with respect to one word in it. Thus a sentence like

> *She smiled when we told her that her daughter had been*
> *nominated to receive an important award.*

might be annotated for the verb *smile* alone, as a member of the Make_faces frame, where it belongs in the set *frown, grimace, grin, pout, scowl, smile, smirk.*

As the size of the lexicon increased, it became clear that there were sentences for which FN was prepared to describe many of the words in it, and ultimately we received a subcontract to look into the possibility of producing full text annotations. That meant annotating each word in the sentence - that is, each frame-evoking word. For the above example, that would mean showing the frame structure of the words *smile, tell, daughter, nominate, receive, important* and *award*. For our purely lexicographic purposes, we would have no reason to annotate the word *told* in this sentence - we already have more than enough examples of the lemma - but it would have to be done here again in order to prepare the semantic structure of the sentence as a whole. Obviously this need increased our eagerness to find ways of automating parts of the annotation process.

FrameNet has to date annotated a growing number of texts, some of them viewable on the FN website. Most of them are only partially annotated, partly because they contain lexical material FN has not yet worked through, and partly because they contain meaningful grammatical patterns that FN annotation has not been prepared to capture.[1]

The annotations themselves are presented in layered stand-off representation in multiple layers. For lexicographic annotations, one layer identified the target LU and its frame; another represented the FEs in the phrases that serve as its valents; one indicated the phrase types of the constituents so identified; one indicated the grammatical function of each valent; and a few other layers were dedicated to special features associated with individual parts of speech. The FEs were annotated manually, the GF and the PT labels were attached automatically and checked manually. Annotations viewable on the FrameNet website show only the frame element labeling, as in Figure 1.

Figure 1: FE annotation of a sentence

[Fluid The River Liffey] *FLOWS*Target [Source from west] [Goal to east] [Area through the center of the city] [Goal to Dublin Bay] .

Full text annotations consist of sets of layers, each corresponding to one target LU. It is virtually impossible to get a view of the full annotation of a long sentence, but there is some experimental work being done to derive dependency trees from these, with the nodes indicating lexical heads and their frames, the branches labeled according to the frame element represented by the dependent nodes.

One special feature of FN annotation is the recording of FEs that are conceptually present but syntactically missing. These are sorted into *constructional null*, such as the missing subject of an imperative sentence; *indefinite null*, such as the object of intransitivized *eat, sew, bake,* etc.; and *definite nulls* (*zero anaphora*), entailing that the missing element has to be recoverable in the context, such as the missing object of *we won* (what is understood but unexpressed is the

---

[1] The texts - chosen because other researchers are examining them as well - were taken from the Wall Street Journal section of the Penn TreeBank, the Nuclear Text Initiative website, and a selection of Berlitz Travel Guides that have been made available to the American National Corpus.

contest - not the prize), the missing preposition phrase in *she arrived* (where the destination has to be known) or *mine is similar* (where the unexpressed comparand has to be part of the conversation), and so on. The last of these plays an important role in construction annotation as well. Such information is associated with the annotation of the LU that licenses the omission.
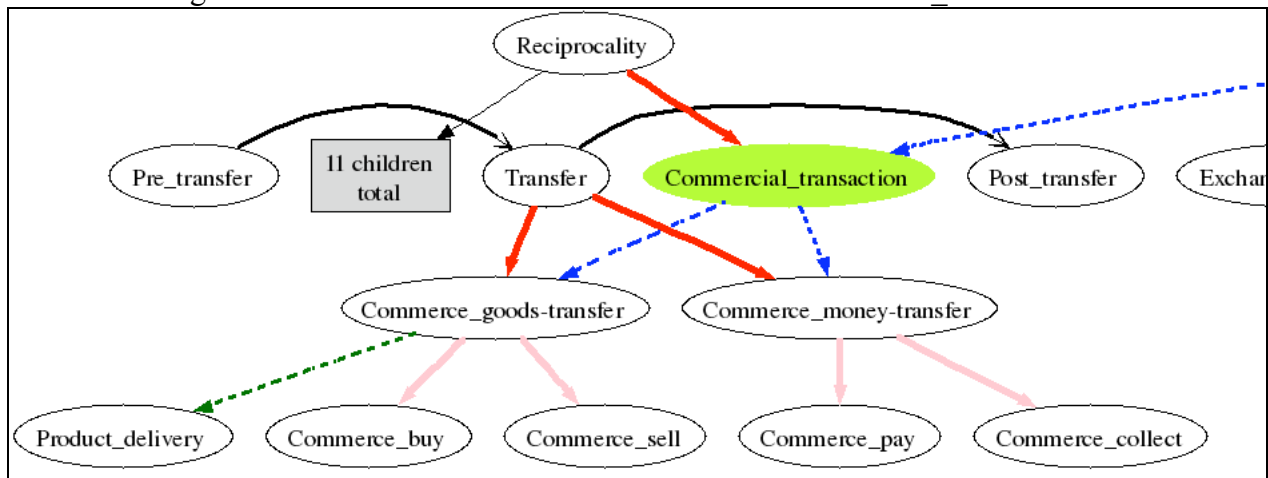
2.6 The entries

Each LU is identified by lemma, part of speech, and frame name. The LUs were chosen because of their membership in one of the frames being covered by FrameNet, and what that means is that in many cases the most common use of a lemma is not to be found: FN researchers have not reached that frame yet. Almost all features of the lexical entry are produced automatically: handmade features include a simple definition[2]. For valence-bearing words, the entry contains a table showing the ways in which each frame element can match a phrase type, and a separate table showing the variety of ways in which combinations of FEs and PTs make up the valence exhibited by individual sentences. Viewers of the valence descriptions can toggle between core FEs only, or all FEs found in the sentences - core, peripheral, and extrathematic.

The entries for nouns that designate events or states of affairs also include information about the existence of *support verbs* and *support prepositions*; access to the sentences will reveal which FEs are represented among the arguments of the LU's verbal or prepositional support[3].

2.6 Frame-to-frame relations

Since frames can differ from each other in granularity, and some frames are clearly related to other frames, it has proved necessary to create an ontology of frames, linked to each other by several kinds of relations. Figure 2 is a display of the frame relations centered on Commercial_transaction:

Figure 2: Frame-to-frame relations centered on Commercial_Transaction

---

Several different kinds of relations can be seen in this diagram. Commercial_transaction has two components (related to the mother node by a *Part_of* relation as indicated by the broken line), and these are Commerce_goods_transfer and Commerce_money_transfer. Each of these is a type of (=has an *Inherits* relation to) the frame Transfer. The low frames Commerce_buy and Commerce_sell have separate *Perspective_on* relations to Commerce_goods_transfer, and the frames Commerce_pay and Commerce_collect have *Perspective_on* relations to Commerce_money_transfer. Thus, a commercial transaction is an instance of Reciprocality, involving two co-occurring reciprocal transfers, one of goods and one of money. Buying and Selling are perspective-varying instances of goods-transfer, differing from the point of view of the buyer and the seller; and similarly with paying and collecting (=charging) and their relation to money-transfer.

3. FrameNet Treatment of Multiwords So Far

The constructicon-building work concerns itself with linguistic knowledge that goes beyond simple grammar and simple words, and hence it will include various kinds of idioms and other multiwords. There are many kinds of multiwords that already fall within the scope of FrameNet work[4]. Among the multiwords covered by current FrameNet[5] we find

1. phrasal verbs, with particles, which are simply treated as two-part verbs that take a specific particle as a syntactic valent; the particle is more or less motivated, but can't be understood as simply contributing its own meaning
   a. Intransitive: pick up (increase), take off (start flying)
   b. Transitive: take up (consider), take off (remove)
2. words with selected prepositional complements, listed with preposition, syntactically selects P-headed phrase
   a. Verbs: depend on, object to, cope with
   b. Adjectives: fond of, proud of, interested in
   c. Nouns: fondness for, pride in, interest in
3. support constructions - syntactically separate, treated as evoking a frame linked to the noun rather than the verb
   a. Verbal heads: take comfort in, take pride in, put emphasis on
   b. Prepositional heads: at risk, in danger, under arrest
4. combinations - combining selected prepositional complement with particle or noun
   a. put up with (tolerate), break in on (interrupt)
   b. take comfort in, place emphasis on
   c. take into possession, take under consideration
5. transparent nouns - the first noun in [N *of* N] structures signifying types, aggregates, portions, units, measures, epithets, etc.; the motivation for recording these is to be able to recognize selectional or collocational relations between the context and the <u>second</u> noun
   a. <u>my</u> gem of a <u>wife</u>, <u>in</u> a part of the <u>room</u>, <u>on</u> this part of the <u>shelf</u>, <u>wreak</u> this kind of <u>havoc</u>

---

[4] Josef Ruppenhofer delivered a paper on this topic at an earlier Euralex meeting (Ruppenhofer et al. 2002).

[5] FN treatment of compound words has more or less awaited the capability of constructional annotation. In the current databases, there are compounds that are simply treated as single unanalyzed units, and there are others in which the head is a frame-bearing word and the modifier is labeled as an FE in the head's frame. FN has lacked the means of describing a compound word both as a unit on its own and as having an internal structure.

4. Full-Text Annotation and the Confrontation with Constructions
In carrying out full-text annotation the goal was to end up with structures which could be the basis of the semantic integration of the whole sentence. Working with one of those linguist-invented sentences like
The Secretary ordered the Committee to consider selling its holdings to the members.
we should be able to identify straightforwardly the participants in the *ordering* event: *the Secretary* gave the order, *the Committee* received the order, and *to consider selling its holdings to the members*, specifies the order. For the verb *consider*, the entity that was to do the considering was *the Committee*, and *selling its holdings to the members* was to be the content of such considerations; and the three participants in the *selling* event are to be *the Committee* as seller, *the members* as buyer, and *the holdings* as the asset destined to change ownership. The words *Secretary*, *Committee* and *members* are all relational nouns used without any indication of what the other term of the relation is, and that's possible if that other entity is understood in the context. A simple frame-annotated dependency tree will fairly well capture the meaning of the whole, with word-frame pairs making up the node labels, the branches labeled according to the semantic role, and with the missing entities in the relational nouns marked with the possibility of indexing them to contextually given entities.

One doesn't have to look far to find sentences containing structures that do *not* lend themselves to such simple treatment. Here are the first three sentences of a leader from the Economist newspaper of June 17, 2007, with comments on those features that go beyond simple lexicon and simple grammar.

> *For all the disappointments, posterity will look more kindly on*
> *Tony Blair than Britons do today. Few Britons, it seems, will shed*
> *a tear when Tony Blair leaves the stage on June 27th after a*
> *decade as prime minister, as he finally announced this week he*
> *would do. Opinion polls have long suggested that he is unpopular.*

1. *for all the disappointments*:
   for all X is a concessive structure with a meaning like 'in spite of X'; seems to be restricted to definite objects; not best treated as a complex preposition
2. *look kindly on*:
   a phrasal verb with the meaning 'judge positively'
3. *[posterity] will look more kindly on Tony Blair than [Britons] do [today]*:
   a comparative structure with a double-focus comparand - [Britons] [today], each accented, requiring the semantic unpacking of *posterity* as something like [the world] [in the future] (a contestable interpretation)
4. *few Britons*:
   not a vague indication of cardinality like *a few Britons*, semantically a negator (= 'not many'), creating a *negative polarity* context (see item 6)
5. *it seems*:
   an epistemic parenthesis, bearing no structural relation to the rest of the sentence but limited in the positions that would welcome it

6. *shed a tear*:
   a VP collocation of the *minimizer* type, appropriate to the negative polarity context created by *few*; similar in this respect to *drink a drop, lift a finger, give a damn, eat a bite*
7. *leave the stage*:
   metaphor, referring here to leaving the PM-ship
8. *on June 17th*:
   use of the preposition *on* with day-level temporal units (cf. *in March, at noon, in the morning*)
9. *June 27th*:
   one of various ways of pairing a date with a month name
10. *as prime minister*:
    as selecting 'role' name; requires context implying service in a role
11. *as he announced he would do*:
    relativizer *as* (consider replacing *as* with *which*)
12. *would do*:
    the form of VP ellipsis (including *do* after a modal) found in BrE missing or rare in AmE (*as he announced he would*)
13. *this week*:
    an expression in which the first element is taken from the list *this/next/last* and the second is a calendric unit name like *week, month, year*, but not *day*
14. *have long suggested*:
    the use of *long* in the meaning 'for a long time' has numerous contextual constraints, difficult to pin down; here both (a) the position between *have* and the participle and (b) restriction to certain classes of verb meanings seem necessary (compare *I have long known that ...* with **I long knew that ...* and **I have long lived in California*.)

5. Constructions and the New Constructicon
Section 3 offered a number of ways in which the behavior of multiword expressions can be incorporated into the FN lexicon and into FN-style annotations, that is, where the information recorded is mainly limited to a small numbr of requirements that lexical items impose on their immediate grammatical environment. Stepping outside of that is a definite new challenge.

5.1 The Annotation Challenge
How did FrameNet become concerned with such matters? First, with our efforts in full text annotation, we became interested in the possibilities of making better coverage of all of the linguistic properties of texts, not just those involving simple predicates and their valence structures. Second, it seems clear that while with support constructions we moved slightly beyond "standard" valence projections, the view of syntactic structure within which we explained the syntactic concomitants of lexical selection needs to be expanded. Third, the community in Berkeley that got started with FrameNet is also a community that has an interest in the broader theory of grammatical constructions. Fourth, and most importantly, it seemed likely that *the same data structure and annotation software devised for lexical annotation could be assigned to the treatment of constructions*.

In 2007 FrameNet received a small grant for doing exploratory research on designing a constructicon, an inventory of "minor" grammatical constructions, and to demonstrate a means of

annotatng instances of them. The parallels to ordinary FN lexical annotation were triking, as can be seen in Table 1

Table 1: Lexical and Constructional Description and Annotation Compared

Lexical FrameNet	Constructicon
Frame descriptions describe the frames and their components, set up FE names for annotation, and specify frame-to-frame relations; lexical entries are linked to frames, valence descriptions show combinatory possibilities, entries link valence patterns to sets of annotated sentences.	Constructicon entries describe the constructions and their components, set up construction elements (CEs, the syntactic elements that make up a construct), explain the semantic contribution of the construction, specify construction-to-construction relations, and link construction descriptions with annotated sentences that exhibit their type.
The FEs are given names according to their role in the frame, and provide labels for the phrases in the annotations that give information about the FE.	The CEs are named according to their function in the constructs, they provide the labels on words and phrases in annotated sentences.
The syntactic properties - grammatical functions and phrase types - are identified for all constituents that realize frame elements.	Phrase types are identified for constituents that serve as CEs in a construct; for constructions that are headed by lexical units, grammatical function labels will also be relevant.
Example sentences are selected that illustrate the use of the lexical units described.	Example sentences are selected and annotated for the ways they illustrate the use of the construction.
Annotations identify the LU, the FEs, and the GFs and PTs of the segments marked off,	Annotations contain labels for the CEs and identify, for lexically marked constructions, the relevant lexical material.
Valence patterns are identified, and linked to the annotations.	Varieties of construct patterns are identified and linked to the annotations.
Frame-to-frame relationships are documented and displayed in a separate resource.	Construction-to-construction relationships are identified and (will eventually be) displayed

The questions to ask for setting up an annotation system for constructions include: What is the constituent (the *construct*) within which a construction operates? What needs to be tagged within a construct? What are the functions of the elements of the construction? What if anything reveals to the reader/listener that there's anything special about the sentence?

In FN lexicographic annotation, we describe a frame and its components or participants, we annotate sentences by identifying the target lexical item and bracketing off the valents and labeling them with frame element names. In constructional annotation, then, we should be able to describe a construction and name the parts of sentences that are the constituents of the constructs licensed by the construction, and then to bracket off those components and assign them labels assigned to the elements of the construction. One important difference is that often there is no *target* LU to link the construction to.

Figures 3 and 4 show the similarity of lexical and constructional annotations, as they appear in the annotation tool. The lexical example represents the clause *one of them accused Mr Wisson of kidnapping*; the constructional example represents the sentence *None of these arguments is notably strong, let alone conclusive*. The list of labels at the bottom of each is the list appropriate to a single level: the FE level in the lexical example, the CE level in the construction example.

Figure 3: Lexical annotation of the verb accuse in the Judgment_Communication frame
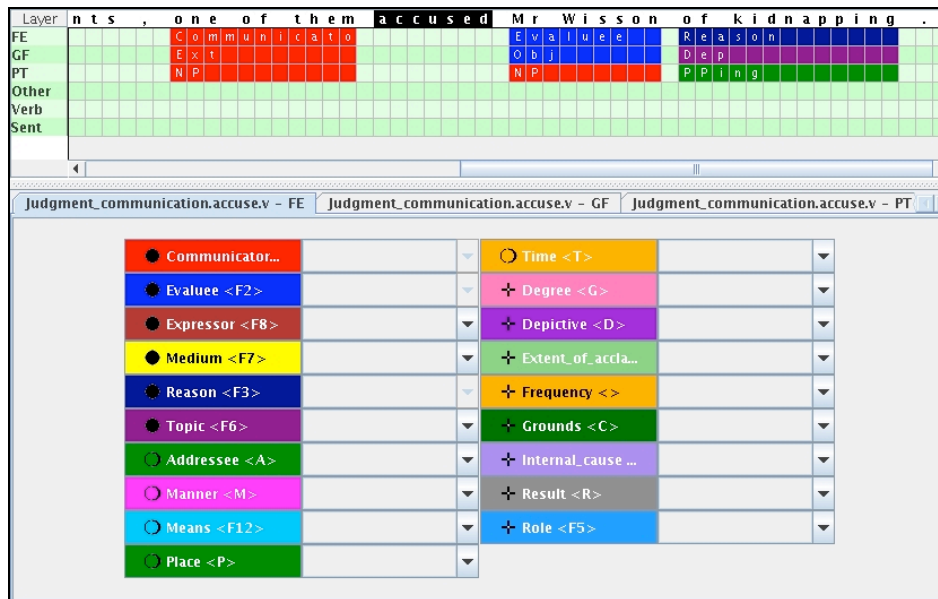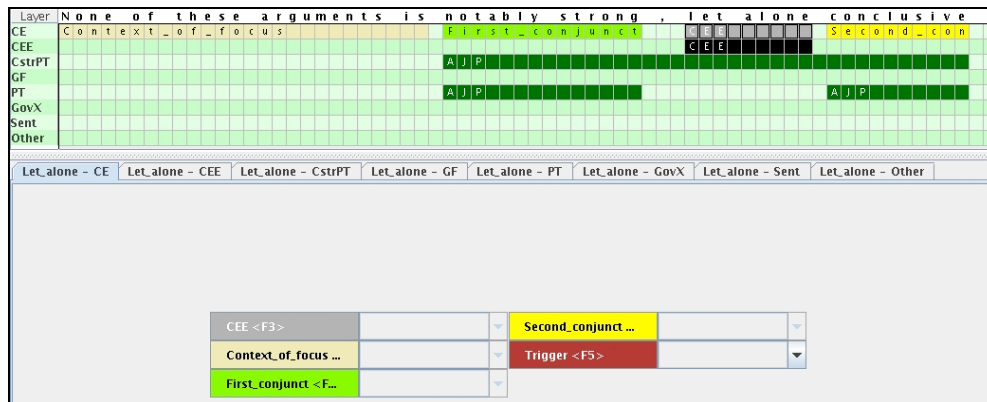

Figure 4: Constructional annotation of a phrase built around the conjunction *let alone*

5.2 The Varieties of Constructions Needing Annotation

The assumption that it would be easy to adapt the FrameNet annotation tool to construction annotation turned out to be false. Essentially the first half of the year of this grant passed by before a proper annotation tool was ready. Finally, in the spring semester, there are two graduate students working on the project, Russell Lee-Goldman and Russell Rhodes, with strong backup

by Michael Ellsworth and Project Manager Collin Baker. By the time of the Euralex meeting, I expect to be able to give a coherent report on our accomplishments and their significance. In the meantime, however, I offer some hastily gathered notes on the types of constructions we need to cover. In the final report almost all of the construction descriptions will include references to the relevant literature, omitted here with apologies, including names like Boas, Borsley, Croft, Goldberg, Jackendoff, Kay, Lakoff, Lambrecht, McCawley, Michaelis, O'Connor, Pullum, Pustejovsky, Sag, Wierzbicka, Zwicky.

5.2.1 Lexical constructions. For an important class of cases, the grammar allows words with one meaning to be paired with the *combinatory affordances* that are common to a semantically defined class of words. (In the case of verbs, this amounts to valence patterns; for nouns, the difference between proper and common nouns, or that between count and non-count nouns; for adjectives the difference between scalar and non-scalar adjectives.) The word *coercion* is sometimes used to cover such relationship.

We can distinguish the words that are "at home" with these affordances from the words that are their "guests". There is an obvious problem for a corpus-based lexicon-building effort like FrameNet, since there is no automatic way of telling the difference: should the derived behavior of "frequent guests" be listed in the lexicon or merely recognized in context as an instance of the construction? It's a problem for lexicography in general, since the decisions that need to be made one way or another are not always clearcut.

EXAMPLES include the phenomena in much of the literature on Argument Structure Constructions, especially in the work of Adele Goldberg. The meanings created by these constructions involve specified relations between the meaning of the "guest" and the semantic expectations of the "host" pattern: *slipping someone a banknote* is using a slipping action to give someone a banknote, *wriggling into the swimsuit* is "entering" the swimsuit (putting it on) with a wriggling motion; an event of *sneezing the napkin off the table* is one in which the air current created by a sneeze has motive force. With nouns, examples like *we had beaver for dinner* show the use of the name of an animal with the grammar of a mass noun, coercing a construal as the flesh of the animal prepared for human consumption[6].

5.2.2 Verbs with contextual requirements outside of their phrasal projection. For the kinds of examples we have in mind under this category it should be possible simply to specify the greater context as part of the combinatory affordances - but there is no familiar formal way to do this within theories of valence. The most common cases are words that fit negative polarity contexts, contexts including negation straight on or other sources of general *irrealis* contexts, like questions, conditional clauses, and dozens of others. (Since we are mainly interested in identifying cases and annotating them, the kinds of careful formulation that a true grammar would need can be glossed over.) Verbs that require contexts that involve both ability and negation allow various ways of expressing those contexts.

EXAMPLES include *can't stand, can't afford, can't tell, can't seem to ..., can't help.* The contexts can be expressed in different ways: in *were you ever able to afford such luxuries?* the polarity is not triggered by a negative morpheme, and the ability is expressed by an adjective rather than a modal. In *it's too dark to tell what they're doing*, the semantics of 'not + able' is entailed in the meaning of *too*. In the case of the verb *brook* a first impression might be that its

---

[6] The construction does not merely convert the animal name into the name of a continuous substance. A sentence like the neighborhood fox likes beaver is not licensed by this construction.

required negation is "local" - i.e., in the determiner of the direct object - but the negation can be presented by an external negation with *any* replacing the *no* in the determiner position: *I will brook <u>no</u> interruption*, *I am <u>too</u> busy to brook <u>any</u> distraction*.

5.2.3 Templatic constructions. Some constructions seem to require a pattern of fixed positions with strict requirements on what can fill those positions: such is the case of the linguistic way of expressing proportions of the kind A:B=C:D; it is sufficient to think of the sentences as providing ways of pronouncing the symbols in such a representation.

      EXAMPLES are often found in lower-grades test questions: *Six is to three as four is to two; blood is to red as snow is to white.*[7]

5.2.4 A mere five dollars. There is a phrasing of numerical expressions that requires (a) the singular indefinite determiner, (b) an adjective that qualifies a number, and (c) a number, such that the combination demands a noun head that matches the number and can contradict the singularity of the article *a*. That is, for something like *a mere five dollars*, all three elements are required: **a five dollars* doesn't work, **mere five dollars* doesn't work, **a mere dollars* doesn't work. We see the construction as determining the prenominal phrase only: in the manner of an ordinary cardinal number, the noun can be deleted if its nature is understood in the context: - as people or dollars, for example, in *a mere two million*.

      EXAMPLES show adjectives with minimizing, neutral and maximizing senses: *a paltry twenty cents, an additional thirty pages, a whopping seven billion dollars*. An expression like *another $200* is a disguised instance of this construction, where *an+other* is analogous to *an+additional*, and *$200* is shown as *two-hundred +dollars*. The modifying adjectives that appear in constructs that instance this construction make up an interesting class.

5.2.5 Presentative constructions. George Lakoff has discussed a family of constructions using *here* and *there* which have important communicative functions. Formally, they begin with *here* or *there*, they have a verb which most typically is *be, come, go, sit, stand*, or *lie*, with the restriction that if the subject is a pronoun it precedes the verb but if it is a lexical NP it follows the verb, and utterances of them have the function of announcing something about the appearance or presence of something. In the complete version, they include some kind of secondary predicate, that can be an adjective, a preposition phrase, a participial phrase, or a with(out) clause.

      EXAMPLES include *here comes that old fool; there she stood, with her hands on her hips; here comes Billy, crawling on his hands and knees; here I am, ready to serve.*

5.2.6 Wherewithal. There is a construction which uses the determiner *the* and a noun construed as naming a resource; it is followed by an indication of what the resource could be used for, expressed as an infinitival VP or a *for*-PP; and its governing context identifies someone as a POSSESSOR (or not) of a sufficient supply of the resources to carry out the purpose represented by the noun's complement. A parallel construction exists with the word *enough* in place of *the*. The name it's been given is due to the fact that the noun *wherewithal* occurs *only* in this construction!

---

[7] These sentences could be given a somewhat tortured parse, involving the extraposition of the *as*-phrase: if we think of *as four is to two* as identical to *what four is to two*, and as naming a particular relation, then we can see the pattern by putting things "back": *Six is [what four is to two] to three.*

EXAMPLES with physical resources include *I don't have the resources to landscape the garden, we lack the staff for such a project, who will provide me the wherewithal to accomplish this, they denied me the funds to complete the job, do we have the fuel to make it to the next town?* Nouns that designate spiritual resources that fit the same construction include *courage, spirit, will, guts, balls*, and several others. Arguments that this construction is needed include the observation that the combination of the nominal and the complement cannot serve as a self-standing NP: **we spilled the fuel to make it to the next town*. The purpose complement can be omitted in contexts where it is understood: A sentence like *where did you find the cash?* can be an instance of this construction, addressed to someone who had just bought an expensive car, or it can be used simply to refer to some until-now misplaced amount of money. The existence of the *Wherewithal* construction explains that ambiguity.

5.2.7 Gapping and Right Node Raising. Some constructions are purely organizational, and have no lexical components beyond conjunctions or words that can function as conjunctions. Those referred to as Gapping and Right Node Raising (RNR) omit phrases whose meaning is shared against elements that are in focal contrast.

EXAMPLES of RNR include *John loves, but Mary hates, rock music*, where comma intonation separates the two truncated conjuncts from their common completion; Gapping is seen when the shared element is between the focal elements: *John loves peaches and Mary apples*. Those are obviously made-up sentences, chosen for their brevity. An attested sentence that exemplifies both of these constructions simultaneously is *Bears have become largely, and pandas entirely, noncarnivorous*.

5.2.8 Let Alone. *Let alone* is a conjunction whose combinatory potential and semantic-pragmatic interpretation are discussed in Fillmore-Kay-O'Connor 1988 and some discussions following that. Briefly, the pieces that are in focal contrast can be[8] assembled with their surrounding contexts to form two propositions, one of these propositions is responsive to the context (i.e., to some assumed or expressed *context proposition*), the other is strongly asserted by the speaker, and it contextually entails the first.

EXAMPLES include the sentence in Figure x, *None of the argments is notably strong, let alone conclusive*. Numerous examples of multiple foci are found in the FKO article. *Let alone* sentences frequently exemplify RNR: *I wouldn't touch, let alone eat, anything that ugly*. (Made-up sentence.)

5.2.9 Verb one's way. A much-studied construction is a way of providing motion verbs by inserting a verb that indicates an action by which someone is able to move, or a path through which the mover moves, or an activity on the mover's part during which they moved. The structure is (a) verb plus (b) possessive pronoun coreferential to the moving entity plus (c) the

---

[8] For example:

   Context proposition spoken by interlocutor: *Can you give me a dollar?*
   Direct response to the context proposition: *I won't give you a dollar*.
   Response that strongly entails the context-relevant response: *I wouldn't lend my mother a nickel*.
   Result: *I wouldn't <u>lend</u> my <u>mother</u> a <u>nickel</u>, let alone <u>give</u> <u>you</u> a <u>dollar</u>*.
Relevant scales for the triple contrasting foci: I'm more likely to lend money to someone than to give it away; I'd be more generous to my mother than to you; a dollar is a lot more than a nickel.

word way: VERB one's WAY. The most neutral verb that is "at home" in this construction is make (*Let's start making our way home.*) The verb *wend* exists <u>only</u> in this construction.

EXAMPLES that show the variety include *She pushed her way through the crowd, the river winds its way through the prairie, we dined our way through the south of France.*

5.2.10 In one's own right. A number of constructions depend on the extended reflexive possessive pronoun *one's own*: *he finally has a room of his own*, *you're on your own now*, but one we have examined is the adjunct *in one's own right*. A typical background assumption for its use is something like this: A is affiliated with B in some way (a relative, an assistant), B is already known for some property or accomplishment, the sentence asserts that same property or accomplishment of A, and the construction conveys the assumption that A's accomplishments are not due to the affiliation with B. The son of a poet can be *a fine poet in his own right*, the husband of a famous chemist can be *an accomplished chemist in his own right*. It would sound odd to say of the wife of right-wing radio commentator Rush Limbaugh that *she is a major intellectual in her own right*, without invoking a belief that Mr. Limbaugh is a major intellectual. (I don't even know if he's married - this is just an example.)

5.2.11 Rate Phrases. The concept of rate is expressed in English with two adjacent NPs in which the first identifies a quantity of units of some type and the second introduces a unit of a different type across which the measurement applies, more or less as numerator to denominator. Typically the second NP is marked with *a* or *per*, but other types occur as well. These expressions express such notions as growth rate, frequency, fuel efficiency, speed, and the like.

EXAMPLES include *it grows four inches a day,* but also *four inches every three days; my Hummer gets seven miles a gallon; our committee meets twice a week; we were moving at 150 km per hour.* The type of rate can be calculated by comparing the two kinds of units, and can be supported by making note of aspects of the governing context, such as the items *grow, meet, gets,* and *at* of the examples.

5.2.12 Measurement Phrases. Some scalar adjectives, but not all, support measurement qualifiers that indicate a quantity of units used for values on the scale.

EXAMPLES include *five meters long/wide/tall/thick*, and *seventeen years old*. Weight and cost values are expressed verbally, with the verbs *weigh* and *cost*; there is no **twenty pounds heavy* or **twenty dollars expensive*. Comparative expressions, however, can have measured "gaps" across the board: *twenty pound heavier, twenty dollars cheaper, three years older*, etc.

5.2.13 Deictically Anchored Calendar Units. The lexical set *this-next-last* occurs in several constructions dedicated to locating a reference time to the present moment - the *temporal deictic center* - with respect to *calendric* time periods like *week*, *month*, and *year*. *This* makes reference to the period containing 'now'; *next* refers to the period following the period containing 'now'; and *last* refers to the period preceding the period containing 'now'. These patterns do not apply to days, however: at the day level the same functions are served by the lexical items *today, yesterday, tomorrow*.

EXAMPLES illustrating one of the constructions, simply identifying a period, are *next year, last month, this week*; a second construction uses these words to mark a recurring point or subdivision of a larger unit and locates the event within the lower unit with respect to whether the larger period is current, past, or future to 'now': *next Wednesday, last summer, this August*;

the third construction uses *next* and *last* in a fixed pattern where the word is understood as picking up the immediately preceding mention of the time entity: *the week after next, the month before last*, and *the summer after next, the Christmas before last*.

5.2.14 THE + Adjective. Expressions like *the rich* and *the poor* are usually thought of as showing these adjectives being "used as a noun". Instead of attributing a part-of-speech change to the adjective, it would seem that a better analysis is that the combination THE + Adjective-Phrase behaves like a full NP. How else could we understand *the very rich, the very young*? Not as *very* modifying a noun, presumably. The constraints seem to be that the adjectives designate some categorizing property of humans; the resulting phrase is human, generic, and plural. Certain adjectives - *poor, rich, young, old* - are "frequent guests" of this construction, but the lexicographers' decision to identify them as actual nouns in those contexts does not seem helpful.

5.2.15 Adjective + AND + Adjective. These same adjectives can be used, in roughly the same meaning, when they surround and, as in *he was beloved of rich and poor alike*. In this case the definite article is not needed, but the conjunction is necessary: **he was beloved of poor* does not work.

5.2.16 Degree Modifiers of Adjectives. It's difficult to decide how many constructions are needed for the intended family of constructions, perhaps several, with constructional inheritance connecting them. Some examples communicating sufficiency or excess have extraposable complements: *too* and *enough* go with an accompanying infinitival VP, *so* goes with a *that*-clause. Others question a scalar value posed in the context, require negative polarity, are accented, and do not have an extraposed complement.

   EXAMPLES include *she's not <u>that</u> young, you can't be <u>too</u> hungry or you'd help us get dinner ready, you're too young to understand, he's so senile that he can't follow the conversation, I am hungry enough to eat a horse*. For *too* and *enough*, the complement can be omitted when the idea is contextually given: *she's too young, she's not old enough*.

5.2.17 Adjective Comparison. Comparison makes up a huge topic, that will not be conquered during the time of this pilot study, but they're included here because of some further constructions that will include them. The comparative markers also carry extraposable complements: *more/er-* and *less → than*; *[not...] so* and *as → as*.
   EXAMPLES include *She's much more intelligent than you said, are you as angry as you seem, it's less warm today than it was yesterday*.

5.2.18 Comparative Negation with NO rather than NOT. If I say that *you're not more qualified for the job than I am*, I could believe that we are both well qualified, and that I should certainly be included among the candidates. On the other hand, if I say that *you're no more qualified for the job than I am*, it's assumed that we're both barely qualified, and (say) I'm complaining that they had no right to give you the job. Using this construction seems to suggest that both of the things being compared are at the low end of the scale. *Your puppy is no bigger than a mouse!*

5.2.19 NP-Internal Degree-Modified Adjectives. All of the adjective modifiers we've just reviewed can be used predicatively, but there is a construction that allows them to be used attributively, *but only in the case of a singular indefinite count noun*. Those that have extraposed

complements allow them to be extraposed after the noun. The adjectival part precedes the indefinite article. (Compare *[an] [intelligent] man* with *[too intelligent] [a] man*.) A variant of the construction has an intrusive *of* which sounds more natural in some contexts than others. We have nothing to say about that just now.

EXAMPLES include *you're too intelligent a man to act like that, that's much bigger of a house than we need, that's as sensible a solution as we can expect, is it really that big of a problem, that's no bigger a problem than others we had in the past, that's so big a problem that we'll never be able to deal with it, is this big enough of a box?* The limitation to indefinite singular count nouns is striking: **it's not that hot of soup, *they're no older of people than my parents*.

5.2.20 One's Every Something. I once proposed that a particular expression with *every* was dedicated to talk about indulgence fantasies, but have learned from corpus data that it is also frequent in paranoid talk.

EXAMPLES of the former kind include *we are here to meet your every need, you will obey my every command, my every dream has been fulfilled, I've satisfied my every wish*; but the other kinds include *why are you dogging my every step, they watch my every move. he records my every gesture*. And there are neutral expressions as well, so it probably requires no more than a sense of extreme attentiveness. Whatever it is, the relationship between the POSSESSOR and the noun has to be agentive in some way - it cannot be one of simple possession: **they stole my every donut* doesn't seem to work.

5.2.21 Plural-Noun Reciprocals as Predicates. Some plural undetermined nominals can occur as predicates indicating a symmetrical social relation between two people. *We were best friends in high school* can be expressed from one member's point of view: *I was best friends with him in high school.* . If the subject is singular, a *with* is needed to identify the other member of the relationship. This only works with nominals that indicate some kind of social relation that inherently is (like *cousin* or *friend*) or can be (like *brother* or *sister*) symmetrical: *we're siblings* can stand alone as a predicate, *we're sons* requires mention of the second term of the relationship, **I was foreigners with him in Japan* doesn't work: *foreigner* isn't a relation between two people

EXAMPLES include *we were colleagues in the post office, she is cousins with a very rich man*, and, from the web, *my theory is that Harry's mother is siblings with Voldemort*.

6. Opportunities for a Construction-Expanded FrameNet

The decision to enter constructional information and lexical information in the same database turns out to have many advantages. In particular, it's seldom necessary to worry about whether we're dealing with a lexical or a grammatical structure. Some products of a construction are simply lexical units in essentially every way, except in that they are "generated" rather than requiring individual listing in a dictionary's wordlist: this is true of the products of argument structure constructions as well as a number of derivational patterns, morphological or "zero" derivation. The lexicographer might now have a principled way of deciding whether a "frequent guest" deserves inclusion in the lexicon's standing wordlist. Some constructs behave like ordinary lexical items in their external environment, and can then be annotated as equivalent to single LUs in their own right: the reciprocal *best friends* can be annotated as an ordinary symmetric predicate of the kind that permits both joint and disjoint expression of the paired participants. The phrase to *push one's way* in its external syntax works just like an ordinary

motion verb and acquires the valence expectations shared by ordinary motion verbs and can be annotated as such. Many of the constructions produce constituents that fit their environment in normal ways requiring nothing special: a rate expression classified as indicating FREQUENCY, or SPEED, or UNIT_PRICE, or WAGES, can combine with whatever marking goes with the governing predicate and find its place in the annotations for that predicate. The zero anaphora facts that FrameNet has encountered in preparing lexical descriptions are similar to those that occur with constructions as well, and pose similar challenges to theories of anaphora. Thus, to take a sentence like *otherwise most members wouldn't have the funds*, a search for cohesion with preceding texts would have to include the condition implied by *otherwise*, the organization presupposed by *members*, and the purpose-indicating complement of the Wherewithal construction that *the funds* are needed for.

Whether parsers can recognize (and interpret) instances of special constructions will remain to be seen. It's possible that a very large sample of construction-annotated texts could provide the learning corpus for statistics-based parsers. An apparent number agreement failure could lead to interpretations that permit such possibilities: *she is friends with the president, a mere twenty pages*. In many cases there are overt markers of a construction that could initiate specific steps to find the components (the phrase *let alone*). A comma before a conjunction in will trigger a search for discontinuities permitted by RNR and Gapping structures. And in some cases the failure to find, in the immediate context, a needed valent of a verb or head of a modifier should guide the search for explanations: the hanging *largely* in the sentence *bears have become largely and pandas entirely noncarnivorous* should serve as a clue.

References:

Fillmore, Charles J. (1982):Frame semantics. In Linguistics in the Morning Calm, Seoul, Hanshin Publishing Co., 111-137.

Fillmore, Charles J. and B. T. S. Atkins (1992): Towards a frame-based organization of the lexicon: The semantics of RISK and its neighbors. In Lehrer, A and E. Kittay (Eds.) Frames, Fields, and Contrast: New Essays in Semantics and Lexical Organization. Hillsdale: Lawrence Erlbaum Associates, 75-102.

Fillmore, Charles J. and B. T. S. Atkins (1994): Starting where the dictionaries stop: The challenge for computational lexicography, In Atkins, B. T. S. and A. Zampolli (Eds.) Computational Approaches to the Lexicon. Oxford: Oxford University Press, 349-393.

Fillmore, Charles J., Kay, Paul and O'Connor, Mary Catherine (1988). Regularity and idiomaticity in grammatical constructions. Language, 64(3): 501-538.

Goldberg, Adele . 1995. Constructions. A Construction Grammar approach to argument structure. Chicago: University of Chicago Press.

Jackendoff, Ray (1990). Semantic Structures. Cambridge, Mass.: MIT Press, 322

Joshi A. K. (1985). Tree-adjoining grammars: How much context sensitivity is required to provide reasonable structural descriptions? In D. Dowty, L. Karttunen, and A. Zwicky, editors, Natural Language Parsing, 206-250, Cambridge University Press.

Lakoff, George (1987). Women, Fire, and Dangerous Things. Chicago, IL: University of Chicago Press.

Tesnière, Lucien (1959), Elements de syntaxe structurale. Paris. Klingksieck.