Corpus-based analysis and annotation of constructions

We report on a year-long project to extend FrameNet methodologies (Fontenelle, 2003) to the analysis of linguistic constructions. The database constructed is a natural extension of FrameNet's, allowing for both ease of use and integration of the two resources.

A construction is understood to be a specification of how to build simple and complex linguistic **constructs**: bundles of phonological, morphosyntactic, semantic, and pragmatic information. For the present project, constructs are described in terms of their internal and external morphosyntax, and their semantics and pragmatics (Fillmore, 1988; Sag 2007). The semantic and pragmatic aspects of a construction are captured by specifying a frame evoked by the construction.

The building of a **constructicon** (cf. lexicon) proceeds in parallel to FrameNet's process of defining frames, frame elements (FEs), and frame-evoking lexical units, and annotating sentences exemplifying these. Each construction is associated with a set of Construct Elements (CEs), its constituent parts. Just as with Frames and FEs, constructions and CEs have their own morphosyntactic and semantic specifications, and both constructions and CEs form a complex inheritance hierarchy. Sentences are then collected from corpora such as the British National Corpus and annotated to show the arrangement and variety of CEs realization. Two examples follow.

(1) [The car]$_{\text{Theme}}$ was moving *at*$_{\text{support}}$ {$_{\text{NP}}$ [$_{\text{NP}}$ 50 miles] [$_{\text{NP}}$ an hour] }.
(2) [Charles]$_{\text{Theme}}$ {$_{\text{V}}$ [bulldozed]$_{\text{V}}$ [his way]$_{\text{NP}}$ } [to the front]$_{\text{Goal}}$.

An annotation of the Rate.speed construction (a subtype of Rate) is in (1). Rate.speed licenses the combination of two CEs—named Distance and Time, with syntactic and semantic restrictions, not showable here—to create an NP that designates a speed, which may appear in constructions calling for such an expression: *at a high speed*, *at 50 miles an hour*. The curly braces indicate the constuctionally-licensed construct; in Sag's (2007) terms, *50 miles an hour* is the Mother, formed of two Daughters (our CEs). As the construction evokes the Speed frame, it licenses a Theme argument (*the car*). (2) illustrates the Way-means construction (Goldberg 1995), which calls for a verb to take a possessed *way*-headed NP, creating a complex verb. This construct evokes the Motion frame (cf *move*, *go*), and so the construct licenses the appearance of FEs: here, Theme and Goal.

The project has to date analyzed over 75 constructions and annotated over 1,500 sentences exemplifying most of the constructions. Among those annotated are subject-auxiliary inversion (Fillmore 1999), conjunction-related constructions (*let alone*, gapping, right node raising), and numerous other non-core expressions: *the morbidly obese*, *so big a problem*, *Wednesday next*, and so on. We have also engaged in analysis of running text, a prerequisite to full constructional (and thus semantic) analysis of a document. We envision the project to be a resource for natural language processing, in particular semantic role labeling and information extraction. Because so much of the meaning of a text is dependent upon non-core constructions, their collection, analysis, and incorporation with a lexical resource like FrameNet is a prerequisite.

Fillmore, C. J. 1988. The Mechanisms of "Construction Grammar." In Axmaker, Jasser, and Singmaster (eds), *Proeedings of the 14th Annual Meeting of BLS*, 35-55. **Fillmore, C. J.** 1999. Inversion and constructional inheritance. In Webelhuth, Koenig, and Kathol (eds) *Lexical and constructional aspects of linguistic explanation*. Stanford: CSLI. 113-128. **Fontenelle, T.** (ed). 2003. Special Issue: FrameNet and Frame Semantics. *International Journal of Lexicography*. Vol.16, Special Issue 3. **Goldberg, A. E**. 1995. *Constructions. A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press. **Sag, I. A**. 2007. Sign-Based Construction Grammar: An informal synopsis. Manuscript, Stanford University.