Ambitions for FrameNet[1]

Miriam R. L. Petruck (FrameNet / International Computer Sciences Institute)

Introduction

The title of the present work comes from Fillmore's impromptu talk during an international workshop about FrameNet, held in Berkeley on April 19, 2013.[2] The event brought together developers of FrameNet style resources (FONTENELLE, 2003; RUPPENHOFER ET AL., 2010) and constructicons (FILLMORE ET AL., 2012) for Brazilian Portuguese, English, Japanese, and Swedish as well as computer scientists and software developers who exploit FrameNet data for natural language processing applications. This paper, initially only intended to introduce the collection of articles in the first part of this special issue of *Veredas* dedicated to Frame Semantics, also provides the opportunity to share Fillmore's wisdom and recommendations about future directions for FrameNet, the research project that he founded in 1997. In so doing, the present paper also seeks to demonstrate that aspects of each article in this special issue start to fulfill some of Fillmore's *Ambitions for FrameNet*.[3]

The rest of this introductory invited essay is organized as follows. Section 1 presents Fillmore's ambitions for FrameNet, quoting directly from the talk where possible, or summarizing the essence of Fillmore's contributions, and section 2 describes the articles in the first part of the current special issue of *Veredas*, also indicating (where appropriate) how each one addresses these ambitions. Section 3 summarizes another of Fillmore's ambitions for FrameNet, offering a suggestion about how FrameNet might handle the concept of TIME.

1. Ambitions for FrameNet

High on Fillmore's list of ambitions for FrameNet is responding to questions that surface continually from lexicographers and theoreticians interested in understanding Frame Semantics and (potentially) using FrameNet data. The first few questions listed here come with attribution.

1. What can FrameNet contribute to lexicography? (Patrick Hanks)
2. Are there words that Frame Semantics cannot handle? (BUSSE, 2012)
3. Why does Frame Semantics concentrate on verbs and verbal nouns, ignoring other forms? (BUSSE 2012)

---

[1] A significant portion of this paper derives from Fillmore's contributions, some of which were recorded and transcribed for the purpose of writing this essay. Clearly, the authorship of the paper belongs to Fillmore. Since circumstances precluded Fillmore's active involvement in the writing of the paper, his name does not appear as its author. Any misrepresentation of Fillmore's ideas is solely the responsibility of the author, who was funded through an International Guest Researcher Grant from the Faculty of Arts at Gothenburg University for part of the writing of this paper.

[2] Funding to the Swedish FrameNet++ project (grant agreement 2010-6013) by the Swedish Research Council supported the workshop.

[3] Note that the authors of the papers in this issue of *Veredas* did not know explicitly about Fillmore's *ambitions for FrameNet* when they wrote their papers.

Fillmore characterized the next three questions as *OAQs — Occasionally Asked Questions.*

4. Is FrameNet useful for practical lexicography?
5. Why do frames and constructions belong together?
6. Clarify the distinction between *evoking* and *invoking*.

2. Frame Semantics, FrameNet and FrameNet Constructicons

The four papers included here concern research and development that the original FrameNet project (framenet.icsi.berkeley.edu) inspired. Torrent and Ellsworth's work addresses the need to define new analytical categories that FrameNet-Brasil (www.ufjf.br/framenetbr) requires for its lexicographic annotation. Ruppenhofer's paper proposes developing SentiFrameNet, which employs Frame Semantics to create an information-rich resource for sentiment analysis. Ohara discusses the challenges of leveraging the information that FrameNet provides about the mapping of meaning to form in English to create resources for Japanese, specifically for developing a Japanese constructicon as part of the Japanese FrameNet effort (jfn.st.hc.keio.ac.jp). Borin, Forsberg, and Lyngfelt's paper presents the issues that surface in work to develop resources for Swedish, specifically Swedish FrameNet (spraakbanken.gu.se/swe/swefn).[4]

FrameNet Brasil has been developing the Brazilian Portuguese analog of the original (English-language) FrameNet, producing analyses of Brazilian Portuguese based on the principles of Frame Semantics and supported by corpus evidence. Torrent and Ellsworth report on the creation of new analytical categories to support that project's lexicographic annotation, specifically discussing the addition of new phrase type and grammatical function labels for the Brazilian Portuguese data. The authors compare the sets of labels that each language requires and motivate the decisions of the teams developing the different knowledge bases. Although not dedicated explicitly to Fillmore's ambitions for FrameNet, this work nevertheless contributes to the larger effort of demonstrating the efficacy of Frame Semantics for the characterization of the lexicon. Motivating the language specific annotation requirements of Brazilian Portuguese (compared to those of English), the paper supports the theory-grounded and data-driven approach that characterizes FrameNet methodology and practice.

Ruppenhofer's paper constitutes a contribution about the potential use of Frame Semantics in general as well as FrameNet conceptual and computational infrastructure and apparatus in particular for sentiment analysis. Current work in sentiment analysis faces several challenges including the use of shallow approaches, pragmatic focus, as well as the ad-hoc creation of data sets and methods. The author argues that progressing towards deep analysis (of text) depends on taking two critical steps: (1) enriching shallow representations with linguistically motivated, rich information; and (2) focusing different branches of research and combining resources and efforts to collaborate with those engaged in related NLP work. Ultimately, Ruppenhofer proposes developing SentiFrameNet, an extension to FrameNet, as a novel representation for sentiment analysis tailored to these goals. Insofar as SentiFrameNet would enrich representations in FrameNet, using an appropriately extended model of Frame Semantic representation, Ruppenhofer's work can be viewed as beginning to answer Hank's question about FrameNet's contribution for practical lexicography. To illustrate, SentiFrameNet proposes creating new lexically specific frames for lexical units that include a negative or positive semantic type. In addition, all inherently evaluative lexical units would be associated with opinion frames that facilitate identifying the semantic role of the opinion holder. For instance, the adjective *profligate* comes with a negative evaluation of the

---

[4] Distinguish between Swedish FrameNet and Swedish FrameNet++.

individual whom the speaker of an utterance characterizes with that adjective. Ordinary dictionaries simply do not specify that information for *profligate*. In keeping with Atkins and Rundell's (2008) suggestion to compile lexical analyses before crafting a dictionary entry, the work of SentiFN provides both the analysis and (to some extent) the compilation.

Ohara's "Toward Constructicon Building for Japanese in Japanese FrameNet," presents Japanese FrameNet, focusing on the kinds of problems that arise in building a FrameNet-style lexical resource for a language with no genetic or typological relation to English. The author discusses similarities and differences in both frames and constructions between English and Japanese. The present work suggests that the organization of frames for Japanese may be different from that of English. For instance, noting that the original FrameNet frames defined for the semantics of English "involve the transitive perspective rather than the intransitive perspective," Ohara suggests the need for defining frames from the "intransitive perspective" to provide frames for intransitive/inchoative and transitive verbs in Japanese, many of which are morphologically related. In addition Ohara intimates the need for defining an Intransitive_of frame-to-frame relationship, to handle differences in perspectives and lexical aspects (*aktionsart*) between English and Japanese. Based on its lexicographic and full-text annotation work, the Japanese FrameNet project has characterized the relations between sentential, lexical, and constructional meanings. An interest in a range of multiword expression types brought out the inadequacy of Japanese FrameNet's practices for the analysis of complex particles, complex auxiliaries and clause-level constructions, thereby prompting the project's fairly recent effort to build a Japanese constructicon. Modeling its development on the FrameNet constructicon (FILLMORE ET AL., 2012), Ohara provides analyses for a number of Japanese constructions, all of which also evoke frames. These analyses serve as a reminder of the close connection between frames and constructions, as well as between FrameNet-style lexica and constructicons, although not all constructions will include a frame evoking element (FEE). The analyses begin to demonstrate why frames and constructions belong together, specifically illustrating one of Fillmore's earlier characterizations of lexicon and grammar:

> [a] **lexicon** should specify the grammatical affordances of its entries; a **grammar** should specify the kinds of lexical units capable of occurring in specifiable positions within grammatical constructions. The most consistent way to represent such mutual dependencies would be to provide both kinds of information in a single well-articulated **grammar** + **lexicon**. (FILLMORE, 2006, p. 35)

This quote seems to foreshadow a distinction that Fillmore draws in his "question" about developing the Swedish Constructicon, specifically concerning preferences among the researchers in terms of approaches to the analysis of constructions in Swedish.[5] Given below, the question, with characteristic subtlety, reveals Fillmore's position on the issue.

> If you're populating the constructicon, is there any tension among the researchers between having some kind of a superficial template that has all of the positions you need for each of these things, and then you simply fill those slots? I'm thinking for example of the construction Paul Kay and I described with *What's X Doing Y*, where it's not really a question, it's sort of a complaint. So, *what's that scratch doing on my violin? What's the flag doing at half-staff? What are you doing standing outside without any shoes on?...* Is there tension between somebody who would like to say that this is a template and you could identify the slots: what-is-somebody-doing and some kind of a secondary predicate?... Or, there might be somebody who prefers to have everything be lexical. And, so in that case, the head verb is BE. And, BE has as its complement *doing*. What is something *doing*? And,

---

[5] Fillmore posed the (so-called) question to Benjamin Lyngfelt who had just finished his presentation on differences between English and Swedish constructions.

doing requires a secondary predicate (that is) *standing in the snow with no shoes on*, etc. And, that particular *do* has specified as its direct object the interrogative *what*. And, the rest of the grammar tells you that the "what" has to come at the beginning. And, the distribution of the element in the progressive has the *what* first and then a form of BE and the rest of it.[6]

The last paper included in the first part of the present issue takes up Ohara's (this volume) closing suggestion to pursue developing FrameNet-style lexica and constructicons. In their exposition of the Swedish analogue of such work, Borin, Forsberg, and Lyngfelt characterize the close encounter between the practical requirements of natural language processing (NLP) on the one hand and the theory and practice of linguistic description for both lexical and grammatical material on the other. The authors discuss several theoretical and conceptual issues that have surfaced while developing SweFN++ (BORIN ET AL., 2010) and planning the Swedish constructicon (LYNGFELT ET AL., 2012). The SweFN++ project is in the process of developing an integrated, Swedish lexical macro-resource primarily for use in language technology research and development to build NLP applications. Most of the components comprising SweFN++ are digital lexical resources that already exist; for these, the primary work of SweFN++ is ensuring the interoperability of the existing resources on as many levels as possible. Swedish FrameNet is an important new resource being created in the context of the larger SweFN++ project. The Swedish constructicon effort is the newest project currently in its early stages, which also will be added to the macro-resource. The decision to develop a Swedish constructicon, adding it to the ongoing work of SweFN++, part of which is Swedish FrameNet, demonstrates the researchers' understanding that frames and constructions belong together (v. Section 1). As more and more research groups decide to develop FrameNet-style lexica and constructicons for languages other than English, explicit attention to the question of why frames and constructions belong together will increase.[7]

3. More Ambitions for FrameNet: TIME

In a more recent communication,[8] Fillmore offered another set of recommendations, this one for the FrameNet team's future work on the concept of TIME, albeit as a suggestion for how the team might demonstrate its frame development process to prospective users of FrameNet data, including developers of NLP applications. Insofar as FrameNet only provides partial characterization of the very complex and broadly relevant concept of TIME for linguistic analysis and natural language processing, viewing Fillmore's recommendations as another ambition for the project is quite natural.

"...One way to specify a moment or a span in time is to modify a noun like *time* or *moment* with a relative clause that identifies some event." Some examples include:

(1)   The day the war ended...
(2)   The moment the baby was born...
(3)   The weeks after the tsunami....

---

[6] Fillmore pursued the distinction by explicating another construction specifically one illustrated by the following: *Were it not for the invention of the barcode reader, the lines in the supermarket checkout would be much slower than they are now*. The scrupulously exacting detail of the analyses presented for this example and the previous one belied Fillmore's recognition that he was not really asking a question (at least to MRLP).

[7] For example, work has begun on French FrameNet (https://sites.google.com/site/anrasfalda), as well as Korean FrameNet, the latter the newest addition to growing number of languages with a FrameNet lexical resource.

[8] The conversation took place on August 26, 2013, in San Francisco. Quoted material in this section comes from Fillmore's written synopsis of his part of the conversation.

Describing an event requires specifying the time of the event or, if the event is complex, the time of some part of the event. In other words, Fillmore states, "It's essential to think of times and events at the same time." Pursuing the idea of demonstrating the day-to-day work of FrameNet, including how the team conducts a free wheeling conversation about a particular concept, along with related concepts under consideration, Fillmore recommends including "sentences that express temporal notions, words to explore in the analyses, and a collection of concordance lines to get started with the corpus-based work." For example, sentences that have *time* or a time expression as the subject, show that "the verb *elapsed* more or less requires a temporal reference point—stated as a phrase or a clause, and introduced by a word like *since.* The examples (4)-(5) below are among those found in the Corpus of Contemporary American English (COCA),[9] where the subject of each sentence is a time expression.

> (4)   But there is about as much time between now and November as has **ELAPSED** since Obama held his June advantages. (Washington Post)
> (5)   By law, we can't put out a missing persons until 72 hours has **ELAPSED** since the last time she was seen." (Movie: "Chasing Sleep")

Following this line of inquiry about the use of the word *elapse*, Fillmore recommends asking, "what kind of linguistic expressions can represent the time point when the calculation begins about how much time has elapsed?" Beginning to answer that question, he adds, "A clause that represents the time the event began, a noun phrase that refers to the point in time in which the event took place," and notes the importance of "being clear about the kinds of expressions that can show up in a frame to be able to recognize paraphrases in context".

This shortened overview[10] of his recommendations for investigating TIME in Frame Semantic terms demonstrates Fillmore's approach to analyzing linguistic data in general, as well as his piercing observations about language. While none of the papers in this special issue of *Veredas* addresses the concept of TIME, they each show the benefit of adopting Fillmore's approach: gaining an understanding of the nature of language.

References

ATKINS, B. T. S. & RUNDELL, M. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press, 2008.

BORIN, L.; DANÉLLS, D.; FORSBERG, M.; GRONOSTAJ, M.; KOKKINAKIS, D. The past meets the present in Swedish FrameNet++. In: DYKSTRA, A. & SCHOONHEIM, T. (Eds.). *Proceedings of the 14th EURALEX International Congress*, Leeuwarden: Fryske Akademy, 2010, p. 269-281.

BUSSE, D. *Frame Semantik*. Berlin: de Gruyter, 2012.

FILLMORE, C. J. *The Articulation of Lexicon and Constructicon*. Paper presented at Fourth International Construction Grammar Conference, Tokyo, 2006.

---

[9] http://corpus.byu.edu/coca.
[10] What appears in this terribly short version is necessarily only a fraction of Fillmore's recommendations about analyzing TIME. This author plans to flesh out Fillmore's ideas on TIME when the recording of the conversation becomes available.

FILLMORE, C. J., R. LEE-GOLDMAN, R. & RHOMIEUX, R. The FrameNet Constructicon. In: BOAS, H. C. & SAG, I. A. (Eds.). *Sign-Based Construction Grammar*. Stanford: CSLI Publications, 2012. p.

FONTENELLE, T. *International Journal of Lexicography*, 16 (3), 2003.

LYNGFELT, B.; BORIN, L.; FORSBERG, M.; PRENTICE, J.; RYDSTEDT, R.; SKÖLDBERG, E. & TINGSELL, S. Adding a Constructicon to the Swedish resource network of Språkbanken. In: JANCSARY, J. (ed.). *Proceedings of KONVENS 2012*. Vienna: ÖGAI, 2012. p. 452-461.

RUPPENHOFER, J.; ELLSWORTH, M.; PETRUCK, M. R. L.; JOHNSON, C. R. & SCHEFFCZYK, J. *FrameNet II: Extended Theory and Practice*. Web Publication, 2010. Available at: framenet.icsi.berkeley.edu.