This article is published online at http://kdictionaries.com/kdn/kdn1304.html.

Towards Hebrew FrameNet

Miriam R. L. Petruck
miriamp@icsi.berkeley.edu

Thanks to a publicly available Hebrew language (newspaper) corpus[1], as well as other web-based resources, such as Rav-Milim[2] and Hebrew WordNet[3], the creation of **Hebrew FrameNet** (HFN) has become possible. Moreover, there are good prospects for cooperation and collaboration with the computational linguistics community in Israel (e.g. access to larger corpora for research and evaluation purposes, use of software for lemmatization and search, links to Hebrew WordNet, where appropriate).

In the context of a research project that investigates the universality of the semantic frame, initial steps have been taken towards the development of Hebrew FrameNet, an on-line lexical resource for contemporary Hebrew which will provide the semantic and syntactic combinatorial possibilites, or *valences*, for each item analyzed, through the manual annotation of example sentences in a newspaper corpus (and eventually, the automatic capture and organization of the annotation results for web-based viewing and querying). With advances in computer technology and the existence of (searchable) corpora, the work of lexicography has changed dramatically in recent years. The fine-grained semantic classification and syntagmatic information of the sort to be provided by Hebrew FrameNet will make the HFN database an invaluable resource for lexicographers and advanced language teachers/learners, as well as researchers in linguistics and natural language processing (NLP).

In accord with FrameNet[4], the first computational lexicography project of its kind (Fontenelle 2003), HFN is based on the principles of Frame Semantics (FS; Fillmore 1978, 1985, Petruck 1996), at the heart of which is the *semantic frame*, an experience-based schematization of the speaker's world against which word meaning can be understood. In Frame Semantics, a linguistic unit *evokes* a frame, whose *frame elements* (FEs), or participants and props in a scene, indicate the semantic roles that need to be filled. A Frame Semantic analysis of a lexical item relies on the identification and definition of the frame(s) in which the word participates, along with the frame-specific FEs that are recorded as triples of information about the semantic role, the phrase type and the grammatical function of the constitutent that is annotated.

To illustrate, consider the three predicates (in boldface) in the first sentence of the initial corpus used for Hebrew FrameNet: *esrot anaʃim* ***magi'im*** *mi-tailand le-israel kʃe-hem* ***nirʃamim*** *kə-mitnadvim ax le-ma'ase* ***meʃamʃim*** *sxirim zolim* (tens (of) people reach from-thailand to-israel while-they register as-volunteers but in-deed they serve laborers cheap) - 'Tens of people reach Israel from Thailand, registering as volunteers but in fact serving as cheap labor'. The verb *magi'a* - 'reach' evokes an Arriving frame, charcterizing a situation in which a Theme moves in the direction of a Goal, the latter either expressed explicitly or implied by the verb. The noun phrase *esrot anashim* fills the role of Theme, and functions as the subject of the clause; the Goal is expressed by the prepositional phrase complement *le-israel*; the example sentence also includes an optional Source expression in the prepositional phrase *mi-tailand*. *nirʃam* - 'register' evokes a Registration frame, describing a scene in which a Registrant puts an Entity on

record at an Institution as belonging to a Category or as Licensed for a specific purpose or state. The noun phrase *kʃe-hem* expresses the Registrant and functions as the subject of the clause; the noun phrase *kə-mitnadvim* fills the Category role. Finally, *meʃamʃim* evokes the Function_as frame, in which an Entity serves a Function or Purpose, the former for activities and the latter for states of affairs. Although not present in the maximal clause of the verb *meʃamʃim*, it is clear what fills the Entity role (*hem*, also indicated by the 3rd-person masculine plural ending *-im* on the verb); the object noun phrase *sxirim zolim* expresses the Purpose. Table 1, below, provides the definitions for the evoked frames and their respective instantiated frame elements.

Arriving: A Theme moves in the direction of a Goal, the latter either expressed explicitly or implied by the verb. Theme is the object that moves toward a Goal. Goal identifies any expression that indicates the final location of the Theme as a result of the motion. Source inidicates the (optionally occuring) starting point of the Theme.
Registration: A Registrant puts an Entity on record at an Institution as belonging to a Category or as Licensed for a specific purpose or state. Registrant is the person who puts the Entity on record, where the two may be co-referential. (Compare *They registered (themselves) as new immigrants* and *They registered their children as new immigrants*). Category is the group to which the Entity belongs.
Function_as: An Entity serves a Function or Purpose, the former for activities and the latter for states of affairs. Entity is the person or object that serves a Function or Purpose. Purpose indicates the state of affairs that holds for the Entity.

Table 1: Evoked Frames and Instantiated Frame Elements

Frame Element annotation for each of the three predicates is given in (1).

(1)
[esrot anaʃim Theme] **magi'im** [mi-tailand Source] [le-israel Goal]
tens (of) people reach from-thailand to-israel

[kʃe-hem Registrant] **nirʃamim** [kə-mitnadvim Category]
as/when-they register as-volunteers

ax le-ma'ase **meʃamʃim** [ovdim sxirim zolim Purpose]
but in-fact they function workers hired cheap

Tens of people arrive in Israel from Thailand, registering as volunteers, but in fact they function as cheap hired workers.

Such Frame Semantic analyses are useful for research in crosslinguistic lexicology (Subirats and Petruck 2003) and in the advanced foreign language classroom (Sato 2004). For instance, whereas the Hebrew verb *meʃameʃ* expresses the Purpose role as a direct object noun phrase (*sxirim zolim*), English *serve* expresses it as a prepostional phrase complement (*as cheap labor*). The availability of such information via the internet will facilitate studies in Hebrew linguistics as well as Hebrew language teaching/learning.

An initial goal of HFN is to produce full annotation for frame evoking elements[5] in the newspaper corpus. This serves as a means of (1) creating the infrastructure for using the FrameNet DeskTop for the analysis of Hebrew texts and (2) determining the level of linguistic description and computational representation at which the lexicon of Modern Hebrew can be characterized in terms of existing frame semantic concepts. Adapting the FrameNet DeskTop (FNDT; a suite of tools used for defining frames, FEs, and words, and annotating illustrative example sentences) for HFN will demonstrate the feasibility of using the software for a non-IndoEuropean language.[6] Investigating the linguistic expression of events and scenarios through the same or different frames will also document the different lexicalization patterns of Hebrew and English (Talmy 2000).

As with FrameNets for other languages (e.g. Spanish[7]) the HFN database will function as both a dictionary and a thesaurus. The dictionary-like features include definitions, tables summarizing the patterns of syntactic realizations of FEs that occur with a word, and sets of annotated sentences from the corpus showing the semantic information associated with each syntactic pattern. Like a thesaurus, words are linked to the semantic frames in which they participate, and frames are linked to other collections of words as well as to related frames. Once attaining sufficient coverage, HFN data will serve the needs of research in NLP for Hebrew, contributing deep semantic information for a variety of tasks, including word sense disambiguation, machine translation, information extraction, and question answering (Litkowski 2004).

Notes

1. http://mila.cs.technion.ac.il/website/english/resources/corpora/2000sentences/index.html.
2. http://www.ravmilim.co.il; see *Kernerman Dictionary News* 12, 2004.
3. http://multiwordnet.itc.it/online
4. http://www.icsi.berkeley.edu/~framenet
5. A frame evoking element is any sense of a word that brings to mind a frame.
6. Similar efforts are under way for Japanese (http://www.nak.ics.keio.ac.jp/jfn).
7. http://gemini.uab.es/SFN

References

Fillmore, C.J. 1977. Topics in Lexical Semantics. In R. Cole (ed.) *Current Issues in Linguistic Theory*. 76-138.

Fillmore, C.J. 1985. Frames and the Semantics of Understanding, *Quaderni di Semantica* 6.2: 222-254.

Fontenelle, T. (ed.) 2003. *International Journal of Lexicography* 16.3, September 2003. (Special issue devoted to FrameNet).

This article is published online at http://kdictionaries.com/kdn/kdn1304.html.

Litkowski, K.C. 2004. Senseval-3 Task: Automatic Labeling of Semantic Roles. In *Proceedings of Senseval-3: The Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, ACL: Barcelona, Spain. 9-12.

Petruck, M.R.L. 1996. Frame Semantics. In J. Verschueren, J-O. Östman, J. Blommaert, and C. Bulcaen (eds.). *Handbook of Pragmatics*. Philadelphia: John Benjamins.

Sato H. 2004. FrameNets and Language Teaching. Presentation at Crosslingual FrameNet Group Meeting. October, 2004, ICSI, Berkeley, CA.

Subirats-Rüggeberg, C. and M.R.L. Petruck. 2003. Surprise: Spanish FrameNet! Presentation at Workshop on Frame Semantics, International Congress of Linguists. July, 2003, Prague, Czech Republic.

Talmy, L. 2000. Lexicalization Patterns . In L. Talmy. *Toward a Cognitive Semantics*. Vol. 2: *Typology and Process in Concept Structuring*. Cambridge, MA: MIT Press. 21-146.