

Tracking Verbs Across Languages

Beth Levin

Department of Linguistics
Northwestern University
Evanston, IL 60208

(Invited Talk)

Current theoretical linguistic research on the lexical knowledge of native speakers of a language shows that the human lexicon is highly structured; verbs fall into classes on the basis of shared meaning, and the members of these classes have in common a variety of properties concerning the expression and interpretation of their arguments, as well as the extended meanings that they show. What is of particular importance to work in machine translation is what light linguistic research can shed on the space of cross-linguistic similarities and divergences in the organization of the verb lexicon. In this talk I will discuss some insights into these questions that have emerged from my recent work with Malka Rappaport Hovav on intransitive verbs in English and other languages. My goal will be to illustrate parallels between the syntactically-relevant semantically-defined subclasses of the intransitive verb class across languages, as well as striking, though systematic, differences in the properties that some of these subclasses exhibit across languages.

The syntactic behavior of verbs to a large extent reflects their meaning. Thus it is possible to identify semantically coherent classes of verbs whose members show shared syntactic behavior. The notion of semantic class is valuable to understanding the organization of the verb inventory of a language. The recognition of such classes can also be a powerful aid in building lexicons for NLP and MT systems. If properties of verbs follow from their class membership, then such properties do not need to be encoded in the lexical entries of the verbs themselves but can be derived from their class membership, allowing for simpler lexical entries, while permitting a more general and modular organization of the lexicon as a whole. As will be illustrated, it appears that the same classes of verbs are found across languages, often showing the same range of properties modulo certain differences between the languages. (Of course in carrying out such studies one has to be aware of purported translation equivalents that actually aren't.)

As the structure of the verb lexicon has become clearer, another dimension of complexity has become apparent: the phenomenon that Apresjan (1973) calls "regular polysemy" as it applies to verbs. Regular polysemy refers to instances of polysemy that are common to members of a set of semantically-related words. This phenomenon is best illustrated in the noun lexicon, where a noun such as *cup* can refer not simply to the container but also to the quantity of a substance that can be held in such a container (*cup of milk*). Regular polysemy is also attested in the verb lexicon; for example, one of the best known instances of verbal regular polysemy involves the use of verbs of manner of motion (e.g., *amble*, *run*, *swim*,

walk) as verbs of directed motion in some languages (*Pat limped; Pat limped into the room*). The existence of regular polysemy can also explain why the relationships between semantic class and syntactic behavior are sometimes obscured. Within a single language, some attempts at identifying verb classes have run into problems because some verbs show apparently unexpected patterns of syntactic behavior. In many instances, such verbs can be shown to belong to more than one semantic class, with the problematic conjunction of properties resolving itself once this fact is taken into account.

Recognizing and understanding regular polysemy is clearly of great importance to MT. Particularly relevant from the perspective of MT are systematic similarities and divergences across languages in terms of regular polysemy; these should be exploited in lexicon construction. This talk will present several examples of cross-linguistic similarities and divergences involving regular polysemy in the intransitive verb lexicon. The best known example of verbal regular polysemy involves the use of verbs of manner of motion as verbs of directed motion. As shown by Talmy (1975; 1985), many languages differ from English in not allowing verbs of manner of motion to have a directed motion use. English also permits verbs of sound (e.g., *clank, gurgle, screech, whistle*) to be used as verbs of directed motion (*The truck rumbled; The truck rumbled into the driveway*); this possibility appears to be open to a subset of the languages which like English allow verbs of manner of motion to be used as verbs of directed motion. Evidence will be presented in support of this multiple classification of verbs of sound. This study will also suggest that verbs of manner of motion are more like verbs of sound than they are like verbs of inherently directed motion (e.g., *arrive, come, descend, go*). In concluding this talk, I will briefly discuss what this unexpected parallel might reveal about the organization of the verb lexicon and the aspects of verb meaning that need to be included in a lexical semantic representation.

References

- Apresjan, Ju.D. (1973) "Regular Polysemy", *Linguistics* 142, 5-32.
- Talmy, L. (1975) "Semantics and Syntax of Motion", in J.P. Kimball, ed., *Syntax and Semantics* 4, Academic Press, New York, 181-238.
- Talmy, L. (1985) "Lexicalization Patterns: Semantic Structure in Lexical Forms", in T. Shopen, ed., *Language Typology and Syntactic Description 3: Grammatical Categories and the Lexicon*, Cambridge University Press, Cambridge, 57-149.