Integrating Grammatically Relevant Lexicalized Meaning into Morphological Analyzers

KATINA BONTCHEVA, NICOLAS KIMM & NATALIA MAMEROW
(Heinrich-Heine-University Düsseldorf)

The aim of this paper is to illustrate the integration of semantic information like the lexicalised property of verbs being (non-)scalar into morphological analyzers.

Quite a few commercial or open-source morphological analyzers are available. However, most of them do not contain semantic properties. In this paper, we will illustrate below how it can significantly enhance the performance of a shallow or a deep syntactic parser. Morphological analyzers typically consist of two main modules: a (lexc-)lexicon module that lists inflectional classes and describes the morphotactics of the language, and a module where the realization rules and the phonological and orthographical alternations are handled by finite-state replace rules. The lexicon module itself consists of a master lexicon and a number of text files that hold the bases of the lexemes that belong to the different inflectional classes. In the lexicon, we include lexicalized meaning that conforms with the description in Rappaport Hovav & Levin, 2010:23: "In order to distinguish **lexicalized meaning** from inferences derived from particular uses of verbs in sentences, we take lexicalized meaning to be those **components of meaning** that are entailed in all uses of (a single sense of) a verb, regardless of context" (emphasis ours). Obviously, this definition is applicable not only to verbs but to all word classes. Words lexicalize a set of attributes some of which constrain their morphosyntactic behaviour, e.g. properties of verbs that would allow the prediction of the verb's argument realization or properties of nouns that would influence the choice and use of determiners, etc. However, in this paper we will limit ourselves to the description of verbs and their lexicalized aspectually relevant properties. Since lexicalized information belongs to the lemma, it should be added directly to the base of the lexical entry. One way to do this is to append every single entry with semantic information. Obviously, this is extremely time and labour consuming. A more reasonable approach would be to identify abstract lexicalized properties that are shared across large groups of verbs and preferably split the verb lexicon of a language into disjunctive classes. This will allow the integration of semantic properties into morphological analyzers without jeopardizing the computational efficiency. One such property is (non-)scalarity (for details on scalarity cf. Rappaport Hovav & Levin, 2010:28ff.). Scalarity can be related to Vendler classes (Vendler 1957) which makes its integration into a morphological analyser even more appealing.

The account of Beth Levin and Malka Rappaport Hovav on verb classes developed over the years in a steady and consistent (Levin, 1993; Levin & Rappaport Hovav, 1991, 1995, 2005; Rappaport Hovav, 2008; Rappaport Hovav & Levin, 1998, 2001, 2005, 2010), among others. Here we will just summarize the most important ideas and implications: (i) Dynamic verbs either lexicalize scales (scalar verbs) or do not (non-scalar verbs); (ii) Non-scalar verbs lexicalize manner; (iii) Manner verbs are activities in Vendler's sense; (iv) Incremental-theme verbs do not lexicalize scale, they line up with manner verbs; (v) Scalar verbs lexicalize result; (vi) Scalar verbs lexicalize two types of scales – multi-point and two-point scales; (vii) Multi-point scalar verbs are accomplishments in Vendler's sense; (viii) Two-point scalar verbs are achievements in

Vendler's sense; (ix) The chosen aspectually relevant properties are complementary; (x) All lexical distinctions described here have grammatical consequences which are relevant to aspectual composition.

Rappaport Hovav's (2008) analysis of the original Vendler classes in terms of scalarity is summarized by Van Valin (to appear) in Figure 1.

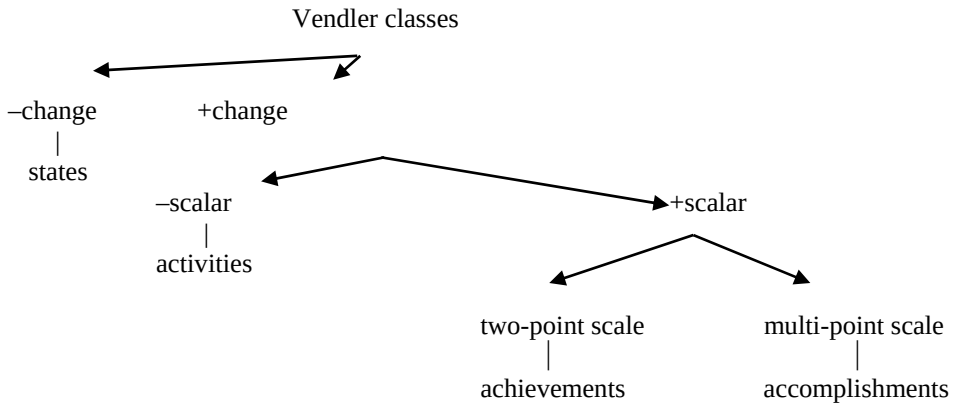

Figure 1: Vender classes and scales (following Rappaport Hovav 2008)

Vendler's classification has some disadvantages for our purpose: (i) Vendler does not classify verbs but VPs; (ii) part of the features used to differentiate between the classes are not lexicalized by the verb but can be determined at the VP level; (iii) this classification allows multiple class membership even for the same word sense. Thus *run* can be activity and accomplishment, cf. above *running/running a mile*. Levin and Rappaport Hovav's reinterpretation of Vendler's classification on the other hand has some very attractive properties: (i) The verbs fall into disjunctive classes. There is no multiple class membership (for the same word sense). (ii) The aspectual properties are lexicalized exclusively by the verb and are not computed at the VP level. (iii) The lexicalized aspectual properties constrain the syntactical behaviour of the verb. (iv) Consequently, it can be predicted what arguments and/or adjuncts a verb must/can take, and with which and how many Vendler time schemata it can be associated.

The following example shows how the integration of the lexicalized properties described above can enhance the efficiency of a shallow rule-based syntactic parser that uses the output of our morphological analyzer for English as its input. Let us slightly modify the first line of the lyrics of the song of Paul Simon "Killer wants to go to college" and analyze it with our morphological analyzer without lexicalized information. Below is the output:

1 killer	killer	+N+Nom+Sg		
2 wants	want	+N+Nom+Pl	+V+Pres+3P+Sg	
3 to	to	+Prep	+InfMark	
4 go	go	+V+Inf	+V+Pres+Non3PSg	+N+Nom+Sg
5 to	to	+Prep	+InfMark	
6 school	school	+V+Inf	+V+Pres+Non3PSg	+N+Nom+Sg
7 .	.	+Punct		

A shallow syntactic parser "sees" only the morphological tags. It will produce ambiguous output since the first word (*killer*) can be a premodifier or an NP head, the second an NP head or a finite verb, the third and the fifth a PP head or an infinitive marker and the fourth and the sixth an NP head, a non-finite or a finite verb:

	word 1	word 2	word 3	word 4	word 5	word 6
1	NPHead	MainV+F	PPHead	PPCompl	PPHead	PPCompl
2	NPHead	MainV+F	PPHead	PPCompl	Premarker	MainV-F
3	NPHead	MainV+F	Premarker	MainV-F	PPHead	PPCompl
4	[…]					

If we supply the additional information that *go* is a directed motion verb, i.e. lexicalizes a multi-point scale and is associated with the following argument realization pattern <SV|SVA> (S = subject, V = verb, A = adjunct) the shallow syntax parser will be able to produce an unambiguous analysis:

1 killer	killer	+N+Nom+Sg	NPHead
2 wants	want	+V+Pres+3P+Sg	MainV+F
3 to	to	+InfMark	Premarker
4 go	go	+V+Inf	MainV-F
5 to	to	+Prep	PPHead
6 school	school	+N+Nom+Sg	PPCompl
7 .	.	+Punct	

The example shows that even if only one of the tokens with noun/verb ambiguity has additional lexicalized information it is possible to provide unambiguous shallow syntax analysis.

This is work in progress. Currently, we extract all verbs that appear in example sentences in the relevant publications of Levin and Rappaport Hovav and add them to the respective verb classes. It is still not quite clear if we need to consider the transitivity of the verbs. For manner verbs it seems to be irrelevant since transitive verbs like *scrub* can be used with unspecified objects and intransitive verbs like *run* can be used with non-subcategorized objects. For directed motion verbs that lexicalize the path as their scale it might be beneficial to append also information for the type of path that is lexicalized and the prepositions that occur with these paths. We will also investigate the applicability and usefulness of this approach crosslinguistically, e.g., for German, Russian, Bulgarian among others.

Levin, B. (1993). English Verb Classes and Alternations: A Preliminary Investigation. Chicago: University of Chicago Press.

Levin, B. & Rappaport Hovav, M. (1991). Wiping the Slate Clean: A Lexical Semantic Exploration. In: B. Levin & S. Pinker, (eds.). Special Issue on Lexical and Conceptual Semantics. Cognition 41, 123-151.

Levin, B. & Rappaport Hovav, M. (1995). Unaccusativity: At the Syntax-Lexical Semantics Interface. Cambridge: MIT Press.

Levin, B. & Rappaport Hovav, M. (2005). Argument Realization. Cambridge: Cambridge University Press.

Rappaport Hovav, M. (2008). Lexicalized Meaning and the Internal Temporal Structure of Events. In: S. Rothstein, (ed.). Theoretical and Crosslinguistic Approaches to the Semantics of Aspect. Amsterdam: John Benjamins, 13-42.

Rappaport Hovav, M. & Levin, B. (1998). Building Verb Meanings. In: M. Butt & W. Geuder, (eds.). The Projection of Arguments: Lexical and Compositional Factors. Stanford: CSLI Publications, 97-134.

Rappaport Hovav, M. & Levin, B. (2001). An Event Structure Account of English Resultatives. Language 77, 766-797.

Rappaport Hovav, M. & Levin, B. (2005). Change of State Verbs: Implications for Theories of Argument Projection. In: N. Erteschik-Shir & T. Rapoport (eds.) The Syntax of Aspect. Oxford: Oxford University Press, 274-286.

Rappaport Hovav, M. & Levin, B. (2010). Reflections on Manner/Result Complementarity. In: M. Rappaport Hovav, E. Doron & I. Sichel (eds.). Syntax, Lexical Semantics, and Event Structure. Oxford: Oxford University Press, 21–38.

Van Valin, R. D. Jr. (to appear). Achievements, Accomplishments and Scalarity. Under review for: Proceedings of 'Scalarity in Verb-Based Constuctions'.

Vendler, Z. (1957). Verbs and Times. The Philosophical Review, Vol. 66, No. 2, 143-60.