

Integrating Aspectually Relevant Properties of Verbs into a Morphological Analyzer for English

Katina Bontcheva

Heinrich-Heine-University
Düsseldorf

bontcheva@phil.uni-
duesseldorf.de

Abstract

The integration of semantic properties into morphological analyzers can significantly enhance the performance of any tool that uses their output as input, e.g., for derivation or for syntactic parsing. In this paper will be presented my approach to the integration of aspectually relevant properties of verbs into a morphological analyzer for English.

1 Introduction

Heid, Radtke and Klosa (2012) have recently surveyed morphological analyzers and interactive online dictionaries for German and French. They have established that most of them do not utilize semantic properties. The integration of semantic properties into morphological analyzers can significantly enhance the performance of any tool that uses their output as input, e.g., for derivation or for syntactic parsing. In this paper will be presented my approach to the integration of aspectually relevant properties of verbs into a morphological analyzer for English.

In section 2 I will describe a prototypical finite-state morphological analyzer for English that doesn't utilize semantic properties. Some classifications of English verbs with respect to the aspectually relevant properties that they lexicalize will be outlined in section 3. In section 4 will be presented my approach to the integration the semantic classes in the lexicon. I will describe the modified morphological analyzer for English in section 5 and point out in section 6 the challenges that inflectionally-rich languages present to the techniques outlined in section 4.

Finally, in section 7 I will draw some conclusions and outline future work on other languages.

2 A Prototypical Finite-State Morphological Analyzer for English

English is an inflectionally-poor language which for this reason has been chosen to illustrate my approach to the integration of grammatically relevant lexicalized meaning into morphological analyzers. It has a finite number of irregular (strong) verbs. The rest of the verbs are regular and constitute a single inflectional class.

This prototypical morphological analyzer for English has parallel implementations in *xfst* (cf. Beesley and Karttunen (2003)) and *foma* (cf. Hulden (2009a) and (2009b)). It consists of a lexicon that describes the morphotactics of the language, and of phonological and orthographical alternations and realizational rules that are handled by finite-state replace rules elsewhere. The bases of the regular verbs are stored in a single text file. Here is an excerpt from the *lexc* lexicon without semantic features:

```
LEXICON Root
                                     Verb ;
...
LEXICON Verb
^VREG      VerbReg ;
...
LEXICON VerbReg
+V:0      VerbRegFlex ;
...
! This lexicon contains the morpho-
tactic rules.

LEXICON VerbRegFlex
< ["+Pres"] ["+3P"] ["+Sg"] > # ;
< ["+Pres"] ["+Non3PSg"] > # ;
< ["+Past"] > # ;
< ["+PrPart"|" +PaPart"] > # ;
< ["+Inf"] > # ;
```

3 Aspectually Relevant Properties of Verbs

The information that is provided by the prototypical analyzer described above contains lemma, W(ord)-features (morphosyntactic features that exhibit different specifications in different cells of the same inflectional paradigm) and L(exeme)-features that “specify a lexeme’s invariant **morphosyntactic** properties” (e.g., gender of nouns, cf. Stump (2001), p. 137, emphasis mine).

L-features should not be confused with lexicalized meaning. I adopt the definition in Rappaport Hovav and Levin (2010), p. 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 mine). Obviously, this definition is applicable not only to verbs but to all word classes.

However, in this paper I will limit myself to the description of lexicalized aspectually relevant properties of verbs.

3.1 Vendler’s Classification

In his famous paper “Verbs and Times” Vendler (1957) introduced his “time schemata presupposed by various verbs” (ibid.). He proposes four time schemata: **states, activities, accomplishments** and **achievements**.

It is important to point out from the beginning that although he didn’t declare explicitly that he was classifying VPs, he did imply this: “Obviously these differences **cannot be explained in terms of time alone: other factors, like the presence or absence of an object, conditions, intended state of affairs, also enter the picture.**” (ibid., p. 143, emphasis mine).

The properties that are often used to define Vendler’s classes are **dynamicity, duration** and **telicity**. States are non-dynamic, achievements are non-durative. States and activities are inherently unbounded (non-telic); accomplishments and achievements are inherently bounded. Since three features are needed to differentiate between only four classes that cannot be represented as, e.g., a right-branching tree one wonders if these are the right features to be used for the classification.

Vendler’s classification was widely accepted and is used in most current studies on aspect.

However, Vendlerian classes cannot be implemented in a *lexc* lexicon for the following reasons:

- Vendler does not classify verbs but VPs
- Part of the features used to differentiate between the classes are not lexicalized by the verb but can be determined at the VP level
- 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*.

3.2 Levin and Rappaport Hovav’s Approach to English Verb classes

Sets of semantically related verbs that share a range of linguistic properties form verb classes. There are different criteria for grouping and granularity, e.g., Levin (1993) classifies the verbs in two ways: a) according to semantic content with 48 broad classes and 192 smaller classes; b) according to their participation in argument alternations with 79 alternations. The account of Beth Levin and Malka Rappaport Hovav for verb classes developed over the years in a steady and consistent way that can be traced in the following publications: (Levin 1993; Levin and Rappaport Hovav 1991, 1995, 2005; Rappaport Hovav 2008; Rappaport Hovav and Levin 1998, 2001, 2005, 2010), among others.

Here I will just summarize the most important ideas and implications for the non-stative verbs:

- Dynamic verbs either lexicalize scales (scalar verbs) or do not (non-scalar verbs)
- Non-scalar verbs lexicalize manner
- Scalar verbs lexicalize result
- Scalar verbs lexicalize two major types of scales – multi-point scales and two-point scales
- The chosen aspectually relevant properties are complementary
- All lexical distinctions described here have grammatical consequences which are relevant to aspectual composition.

This interpretation of non-stative verbs has some very attractive properties:

- The verbs fall into disjunctive classes. There is no multiple class membership (for the same word sense).
- The aspectual properties are lexicalized exclusively by the verb and are not computed at the VP level.

- The lexicalized aspectual properties constrain the syntactical behavior of the verb.
- Manner verbs in English show a uniform argument-realization pattern: they can appear with unspecified and non-subcategorized objects.
- Result verbs are more constrained and less uniform in their argument realization patterns. Transitivity (in contrast to the manner verbs) is an issue.

4 Intersection of Semantic Classes and Inflectional Classes

The main difficulties here arise from the fact that the set of bases that belong to one inflectional class of verbs usually is not identical with the set of bases that lexicalize a particular aspectually relevant property. As a rule, it has intersections with more than one semantic class. The situation is relatively manageable in inflectionally-poor languages like English but becomes very complicated in inflectionally-rich languages.

The distribution of verbs in inflectional classes is in general complementary. There are some exceptions that will not be discussed here.

Vendler's approach to the verb classification described in 3.1 has the undesirable property that most of the verbs have multiple class membership, while the approach of Levin and Rappaport Hovav described in 3.2 has advantages which make the task easier.

Thus, for English we have the set of bases of regular verbs that is monolithic, and the same set of bases but this time divided into complementary subsets of aspectual semantic classes in the sense of Levin and Rappaport Hovav. The cross product of the number of subsets in the first set and the number of subsets in the second set equals the number of aspectual semantic classes since there is only one inflectional class of regular verbs.

5 The modified Prototypical Lexicon for English

The following modifications need to be introduced to the lexicon in order to incorporate the aspectual properties of English verbs.

The single placeholder pointing to the single file containing the bases of regular verbs must be replaced with several placeholders that point to the

files containing the complementary subsets of bases of verbs belonging to the different aspectual classes.

New continuation lexicons introducing each aspectual class must be added immediately after LEXICON Verb. Since the union of the sets of aspectual-class bases of regular verbs is identical with the set of the bases of the regular verbs, all aspectual-class lexicons have the same continuation lexicon: LEXICON VerbRegFlex. Irregular verbs get the semantic tags added to the lexical entry and suppletive verbs get them in the master lexicon.

```

Multichar_Symbols
+V +VIrrTT %<manner%>
...
LEXICON Root
    Verb ;
    VerbSuppl ;
...
LEXICON VerbSuppl
go%<resmulpo%>+V+Inf:go # ;
go%<resmulpo%>+V+Pres+3P+Sg:goes # ;
go%<resmulpo%>+V+Pres+Non3PSg:go # ;
go%<resmulpo%>+V+Past:went # ;
go%<resmulpo%>+V+PaPart:gone # ;
go%<resmulpo%>+V+PrPart:going # ;
...
LEXICON Verb
^VREGM    VerbRegManner ;
...
LEXICON VerbRegManner
+V%<manner%>:0    VerbRegFlex ;

LEXICON VerbRegFlex
...

```

Below is an excerpt from the file holding the bases of irregular verbs that build identical past-tense and perfect-participle forms by adding '-t':

```

...
{creep<manner>}:{creep} |
{feel} |
{keep} |
{sleep} |
{sweep<manner>}:{sweep} |
...

```

In order to be able to rewrite the semantic-class tags, which appear only on the lexical (upper) side of the transducer containing the lexicon, I invert the network, apply the semantic-tag rewriting rules and invert the resulting net again. The network is then composed with the realization rules and the

phonological and orthographical alternations that operate on the surface (lower) side of the transducer:

```
! Semantic-features tag-rewriting
define LEX2 [LEX1.i] ;
define LEX2 [LEX1.i] ;
define Mnr [ %< m a n n e r %> ->
             %<%+SV%>%<%+SVO%>%<%+SVOOC%> ] ;

! alternative RRG tags
!define Mnr [ %< m a n n e r %> ->
             !%<do' %(x%, %[predicate' %(x%) or
             !           %(x%, y%)%]%)> ] ;

define LEX3 [LEX2 .o. Mnr] ;
define LEX [LEX3.i] ;

! Inflectional morphology: realization
...
```

Here is the output of the analysis of 'swept' with dependency-grammar valency-pattern tags (S=subject, V=verb, O=object, OC=object complement):

```
swept
sweep<+SV><+SVO><+SVOOC>+V+Past
sweep<+SV><+SVO><+SVOOC>+V+PaPart
```

and the alternative output with Role and Reference Grammar logical structures:

```
swept
sweep<do' (x, [predicate' (x) or (x, y)])>+V+Past
sweep<do' (x, [predicate' (x) or (x, y)])>+V+PaPart
```

Valency information is necessary for syntactic parsing and has been used in Constraint Grammar shallow parsers and in dependency parsers. The advantage of this approach to already existing morphological analyzers for English is that the valency-pattern tags are added to classes of verbs rather than to individual lexical entries. The ability to provide alternative outputs for the integrated aspectually relevant semantic information is a novelty of this morphological analyzer.

6 Beyond English: the Challenges of Inflectionally-Rich Languages

We have seen a simplified example that shows the modeling and the implementation of a morphological analyzer that utilizes semantic-class tags for aspectually relevant lexical properties of English verbs.

Things become much more challenging if we want to model inflectionally-rich languages such as Bulgarian, Russian or Finnish. Bulgarian verbs, for example, can be divided (depending on the modeling) into some 15 complementary inflectional classes. This number multiplied by 4 Levin-Rappaport-Hovav classes would result in some 60 sets of verb bases that share the same inflectional class and Levin-Rappaport-Hovav class. If a finer-grained semantic classification is adopted, the number of classes will considerably increase and this will lead to a lexicon that exclusively requires manual lexicographical work.

7 Conclusion

This paper illustrates the integration of aspectually relevant properties of verbs into a morphological analyzer for English. I showed that these features can be integrated while the computational efficiency of the analyzer can still be maintained if the linguistic modelling is adequate. However, this only scratches the surface of the challenge of integrating semantic features into morphological analyzers. In the future, it is planned (together with other researchers) to extend the integration of semantic features to nouns, adjectives and adverbs. We also plan to model and implement morphological analyzers for other languages such as German, Russian, Polish and Bulgarian.

References

- Kenneth R. Beesley and Lauri Karttunen. 2003. Finite State Morphology. Palo Alto, CA: CSLI Publications
- David Dowty. 1979. Word Meaning and Montague Grammar. Dordrecht: Reidel.
- William Foley and Robert Van Valin, Jr. 1984. Functional Syntax and Universal Grammar. Cambridge: Cambridge University Press.
- Ulrich Heid, Janina Radtke and Anette Klosa. 2012. Morphology and Semantics: A Survey of Morphological Analyzers and Interactive Online Dictionaries – and Proposals for their Improvement. 15th International Morphology Meeting. Morphology and Meaning. Vienna, February 9-12, 2012
- Mans Hulden. 2009a. Finite-State Machine Construction Methods and Algorithms for Phonology and Morphology. PhD Thesis, University of Arizona.
- Mans Hulden. 2009b. Foma: a Finite-State Compiler and Library. In: Proceedings of the EACL 2009 Demonstrations Session, pp. 29-32.

- Beth Levin. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago, IL: University of Chicago Press.
- Beth Levin and Malka Rappaport Hovav. 1991. Wiping the Slate Clean: A Lexical Semantic Exploration. In: B. Levin and S. Pinker, (eds.). *Special Issue on Lexical and Conceptual Semantics*. *Cognition* 41, pp. 123-151.
- Beth Levin and Malka Rappaport Hovav. 1995. *Unaccusativity: At the Syntax-Lexical Semantics Interface*. Cambridge, MA : MIT Press.
- Beth Levin and Malka Rappaport Hovav. 2005. *Argument Realization*. Cambridge, UK: Cambridge University Press.
- Malka Rappaport Hovav. 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, pp. 13-42.
- Malka Rappaport Hovav and Beth Levin. 1998. Building Verb Meanings. In: M. Butt and W. Geuder, (eds.). *The Projection of Arguments: Lexical and Compositional Factors*. Stanford, CA: CSLI Publications, pp. 97-134.
- Malka Rappaport Hovav and Beth Levin. 2001. An Event Structure Account of English Resultatives. *Language* 77, pp. 766-797.
- Malka Rappaport Hovav and Beth Levin. 2005. Change of State Verbs: Implications for Theories of Argument Projection. In: N. Erteschik-Shir and T. Rapoport (eds.) *The Syntax of Aspect*. Oxford: Oxford University Press, pp. 274-286.
- Malka Rappaport Hovav and Beth Levin. 2010. Reflections on Manner/Result Complementarity. In: M. Rappaport Hovav, E. Doron, and I. Sichel (eds.). *Syntax, Lexical Semantics, and Event Structure*. Oxford: Oxford University Press, pp. 21–38.
- Gregory Stump. 2001. *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.
- Zeno Vendler. 1957. Verbs and Times. *The Philosophical Review*, Vol. 66, No. 2., pp. 143-60