Degrees of Stativity: The Lexical Representation of Verb Aspect

Judith L. Klavans and Martin Chodorow †
IBM T.J.Watson Research
Yorktown Heights, NY
and
†Hunter College of the City University of New York

Abstract

The automatic acquisition of lexical knowledge from large corpora has dealt primarily with coocurrence phenomena, at the expense of inherent lexical features. We present here a methodology for obtaining semantic information on verb aspect by parsing a corpus and automatically applying linguistic tests with a set of structural analysis tools. Once applied, we propose a representation for verb aspect that associates a value with weights for event types. Weights reflect typical verb use, and thus represent a measure of the resistance or ease of coercion in sentential context. The results we report here have been obtained in two ways: by extracting relevant information from the tagged Brown corpus (Francis and Kučera 1982), and by running a parser (McCord 1980, 1990) on the Reader's Digest corpus to extract more accurate information on verb usage in text.

1 Overview

Our work illustrates two points:

1. Inherent lexical properties can be determined by applying a battery of established linguistic tests to corpora. This adds to the utility of corpus analysis a dimension beyond coocurrence phenomena (e.g. mutual information, substitutability).

2. A lexical property need not be discrete but can be represented as a value to be interpreted as a tendency or probability for the lexical item to exhibit the given property in an unmarked context. Corpus-derived values are variable, i.e. DEGREE values

Linguistic tests have been automatically applied to parsed corpora to determine an initial aspectual value for stativity for a set of frequent verbs, covering over 90% of verb occurrences in a one million word corpus.

2 Event types

Aspect can be informally defined as a property which reflects the temporal organization of an event, such as duration (whether an event involves change over time), telicity (whether it has a definite endpoint or is ongoing), iterativity (whether or not it is repetitive) and so on (see Comrie 1976.) We assume three event types, following Vendler 1967, refined by many others, and recently recast from the perspective of computational lexicon building by Pustejovsky 1991, and Pustejovsky and Boguraev (to appear) [1]:

```
State(S): know, resemble, be, love
Process(P): run, walk, swim
Transition(T): give, open, build, destroy
```

A verb can enter into a construction which may change the overall phrasal or sentential event structure, as the result of event-coercion. Coercion can be defined as the process by which verbs in context appear to demonstrate some regular ambiguities, i.e. they appear to change categories. Pustejovsky argues that a verb is inherently (lexically) specified as being of a certain event type (e-type).

```
Schema: e-type(Verb) = S|P|T
        e-type(resemble) = S
        e-type(run) = P
        e-type(give) = T
```

Figure One

[1] Notice that accomplishment verbs (e.g. "paint a picture"), and achievement verbs ("reach a goal") are both considered transition verbs.

We call this the NON-DEGREE APPROACH, since there is no degree specified. Our claim is that, rather than the representation in Figure One, the event structure is a vector of values, e-type (Verb) = (S,P,T). We deal here only with the stative/non-stative distinction, since there are regular differences between states versus activities and accomplishments (Lakoff 1965). We propose a simplified representation e-type(Verb)=$V_s(x)$ with a single value for stativity, V_s, the non-stative value being merely the complement, i.e. $V_s + V_\neg s = 1$. Our position can be summarized as:

Schema: e-type(Verb) = $V_s(x)$
where $0 \leq V_s(x) \leq 1$ and
where $V_s + V_\neg s = 1$.

Figure Two

We call our position the DEGREE APPROACH since a numeric value or DEGREE is specified. We agree with the notion of assuming basic verb types, with coercions, contra the position proposed by Dowty 1979 that there be different entries for different usages.

To allow the process interpretation in sentences like:

```
(1) The child is being a fool today.
(2) She is resembling her mother
    more and more each day.
```

the phrase "more (and more)" in (2), a temporal expression, acts to "coerce" or force a non-stative event (in this case a process). [2] In fact, the verb "be", often touted as fundamentally the most stative verb in the language is frequently coerced into process and transition.

3 Computational Lexicons

The work reported here is part of an ongoing project in the Lexical Systems Group at IBM Research to extract and represent lexical knowledge in CompLex, a computational lexicon which will be able to provide information to natural language processing systems. The seeds of this project are presented in Klavans 1988, where it is argued that the building of a computational lexicon requires mapping of resources into a common central lexical data base, rather than being dictionary bound. The view is further expanded in Byrd 1989. Pustejovsky and Boguraev (to appear) argue that a theory of lexical semantics making use of a knowledge representation framework offers an expressive vocabulary for the representation of such lexical information. They outline a concrete framework consisting of four levels of lexical meaning: (1) Argument Structure; (2) Event Structure; (3) Qualia Structure; and (4) Lexical Inheritance Structure. Pustejovsky 1991 provides formal details of the event structure specification for verbs, with a formal explanation of semantic type coercion.

Specification of verb and phrasal aspect matter to NLP systems in several ways. For example, when a stative verb is used in the present tense, it involves only one occasion of the event. In contrast, a non-stative usage involves a frequentative, iterative, or repetitive interpretation. Thus:

```
John knows the answers. (single)
John runs. (repetitive)
Sue builds houses. (repetitive)
```

It is necessary to understand the interpretation of stativity for several reasons. In language generation, adverbial adjuncts which have a durational interpretation are, under most circumstances, disallowed:

```
*John knew the answers for three days.
 John ran for three hours.
 Sue built houses for three decades.
```

For text analysis, if a verb which is usually stative is used with a non-stative adverb, such as "deliberately", it can be inferred that the event is non-stative, and not the reverse. If the event is non-stative, then it can be inferred that it could be a repetitive or iterative event. This can effect not only the semantic interpretation of the text itself, but also translation and the choice of adverb. [3]

[2] Our position might be viewed in terms of assigning probabilities to the arc transitions in a system such as that proposed by Moens and Steedman 1988, in which coercions are recursive but carefully constrained along several key parameters defining possible transitions within a finite-state transition network.

[3] Many of these issues are discussed in the CL Special Issue on Tense and Aspect (June, 1988) in articles by Hinrichs, Moens and Steedman, Nakhimovsky, Passoneau, and Webber. For example, Passoneau demonstrates how, without an accurate specification of the aspectual tendencies of the verb coupled with the effect of temporal and aspectual adjuncts, messages, which tend to be in the present tense, are not correctly understood nor generated in the PUNDIT system. For instance, "the pressure is low" must be interpreted as stative, whereas "the pump operates" must be interpreted as a process. In machine translation, for example, the verb *parecerse* meaning 'to resemble one another' is even less stative in Spanish than in English. Thus, sentence (2) above should be translated into Spanish

Aspect is complex. Stativity is not a simple feature such as $animate(+/-)$. To observe this, it suffices to look in any comprehensive grammar, for example, Quirk et al. 1972, in which lexical verbs are divided into the classes "dynamic" and "stative", with the caveat that "it would be more accurate to speak of 'dynamic' and "stative' uses of verbs" (p. 94-95). Dowty 1979 observes that the issue of interpretation and aspect involves " the thorny problems of polysemy and homophony"(p. 62). Since some verbs are "more stative" than others, meaning that the most common unmarked use is as a marker of sentential or event stativity, the lexicon must embody this lexical fact. Our proposal provides the capability in the lexicon of representing that variability, combined with automatic means of inducing variability values.

4 Procedure

Some standard tests for the stative/non-stative distinction are given in Dowty 1979, such as the progressive, imperative, complement of force/persuade, intentional adverbs, and pseudo-cleft, as in: [4]

```
Progressive: non-statives
* John is knowing the answer.
  John is running.
  John is building a house.
Complement of force/persuade:
* Gail forced John to know the answer.
  Gail persuaded Amy to run.
  John forced Bill to build a house.
Adverbs, e.g. deliberately, carefully:
* Gail deliberately knew the answer.
  Evelyn ran carefully.
  Sue carefully built a house.
```

Even though tests for stativity involve interactions between semantic and syntactic facets of a sentence, we have chosen three tests for stativity, the most robust being the tendency for a verb to occur in the progressive. Unlike other tests, the progressive itself is a statement of duration and process. We hypothesized that degree of stativity, i.e. a value for stativity V_s, could be inferred by determining the ratio of total frequency of occurrence of a verb in a progressive usage, past or present, over its frequency as a verb in the same corpus:

$$V_s = \frac{F(Prog(X))}{F(Verb(X))}$$

$$\texttt{where} \quad 0 \leq V_s \leq 1 \quad .$$

A value closer to 0 indicates that a verb prefers stativity. This basic value can be modified by other tests, such the force/persuade test, and the deliberately/carefully test.

We are aware that this is an oversimplification of the property of stativity. However, our goal is to search for the most robust and pragmatically possible tests to start with. Our technique has given results which concur with our intuitions, although there are limitations discussed below. [5] In order to do this, we needed a text tagged for part of speech. Otherwise, instances such as "hearing aid", "a knowing look" would be taken as instances of progressive verbs.

5 Results

Tagged Kučera and Francis Data

The Brown corpus[6] provided a convenient starting point, since words are tagged for part of speech. However, a closer look at the tag set itself reveals a weakness which could bias our results. The tag VBG is used to tag "verb, present participle, and gerund." Thus, there is no distinction in the label for the different usages of the "-ing" form, although some "-ing" forms are labelled NN for noun or JJ for adjective. Despite this problem, we chose to use the Brown data, knowing that the numbers might be distorted. We started with a list of the 100 most frequent words labelled as verbs (i.e. the 100 most frequent verbs, which account for over 90verbs which have been discussed in the literature on stativity (such as "resemble", "matter", "intend"). Figure Three lists some results, ordered by degree of stativity.

`e-type(try)=` $V_s(.3326)$
`e-type(work)=` $V_s(.3064)$
`e-type(sit)=` $V_s(.2929)$
`e-type(run)=` $V_s(.2853)$
`e-type(play)=` $V_s(.2552)$

in the simple present, not the progressive, as in English. `Cada día ella se parece más a su madre.`
For activity verbs, the progressive in English translates to the progressive in Spanish.

[4] Activites and accomplishments are subject to other distinguishing tests which are not the subject of this paper.

[5] Pustejovsky, personal communication, has pointed out some problem cases where the progressive falsely indicates that root is non-stative, such as lie/sit verbs (*The book is lying on the shelf, The cup is sitting on the counter*), and mental attitude verbs (*John is thinking that he should go home now, *John is knowing that he should go home now, Mary is suspecting that John will propose tonight.*) Such problems can be resolved by fine-tuning tests.

[6] We thank Slava Katz for pointing us to this resource.

```
e-type(move)= Vs(.2326)
e-type(go)= Vs(.2304)
e-type(follow)= Vs(.1234)
e-type(give)= Vs(.0783)
e-type(become)= Vs(.0718)
e-type(hear)= Vs(.0669)
e-type(feel)= Vs(.0637)
e-type(appear)= Vs(.0481)
e-type(know)= Vs(.0332)
e-type(want)= Vs(.0253)
e-type(need)= Vs(.0197)
e-type(be)= Vs(.0173)
e-type(seem)= Vs(.0120)
```

Figure Three

As can be seen, the ranking roughly reflects intuitions about stativity, so, for example, *seem* is more stative than *hear*, which is in turn more stative than *run*.

Parsing with English Slot Grammar

The second more refined method utilizes a parser to analyze text, and to record verb usages. For this purpose, we used the English Slot Grammar (McCord 1980, 1990) a broad-coverage parser written in PROLOG. [7] To obtain counts of verb usages from the representations produced by ESG, we used a tool for querying trees (QT), built by the second author, also in PROLOG. The corpus is the Reader's Digest (RD) corpus, consisting of just over one million words. We took the same list of the 115 most frequent and most frequently discussed verbs that was used for obtaining values from the Brown corpus. We extracted all sentences under 30 words containing the inflectional variants of these verbs from the RD corpus. We then ran the parser on this subcorpus, ran QT, and obtained values for the different verb usages. Unlike the Brown data, distinctions are made between the gerundive and participial usages. Figure Four gives results for some verbs, listed in the same order as in Figure Three, with n indicating the number of tokens:

```
e-type(try)= Vs(.2167)   (n = 286)
e-type(work)= Vs(.1311)   (n = 244)
e-type(sit)= Vs(.1506)   (n = 146)
e-type(run)= Vs(.1565)   (n = 230)
e-type(play)= Vs(.2315)   (n = 95)
e-type(move)= Vs(.0798)   (n = 213)
e-type(go)= Vs(.2901)   (n = 1071)
e-type(follow)= Vs(.0375)   (n = 133)
e-type(give)= Vs(.0209)   (n = 430)
e-type(become)= Vs(.0507)   (n = 493)
e-type(hear)= Vs(.0000)   (n = 242)
e-type(feel)= Vs(.0317)   (n = 315)
e-type(appear)= Vs(.0272)   (n = 147)
e-type(know)= Vs(.0000)   (n = 630)
e-type(want)= Vs(.0000)   (n = 466)
e-type(need)= Vs(.0000)   (n = 258)
e-type(be)= Vs(.0124)   (n = 12482)
e-type(seem)= Vs(.0000)   (n = 260)
```

Figure Four

[7] An exception to the quality is imperatives, where there were some errors in the parsing; they were removed from our calculations.

Additional Syntactic Tests

The progressive test is only one of several tests, and in and of itself is certainly inadequate. Several tests must be run, and then event values must be computed for each linguistic test. Two parameters are involved: the strength of each test as an indicator of e-type, and the sparsity of data.

We have preliminary results on two additional tests: the force/persuade test and the deliberately/carefully test. Synonyms and taxonyms were collected for each (ad)verb, data were extracted from the corpus and parsed. For example, the following shows how a sentence with "force" was analyzed. However, more data is needed, from a larger corpus, for the results to be significant. The same applies for the adverb test.

```
Difficulties forced him to abandon ...
verb(force) inf_comp_verb(abandon)
```

Figure Five - Verb "Force"

Results of running and computing the weights of different tests on larger corpora will be reported in future publications.

6 Discussion

As expected, the results from each corpus differ considerably; we believe this is due primarily to surface tagging vs. full parsing. The results from the second method using ESG do not carry the noise from the ambiguous VBG tag from the Brown corpus. However, there are two important points to be made: (1) One million words is simply not enough. More data need to be (and will be) run to get a more complete and accurate count. These are to be viewed as preliminary data, usable but not complete. (2) The value $V_s(.0000)$

cannot be considered categorical. Verbs are generally adaptable in context.[8] It is a known fact that values either for words with low frequencies or words in low frequency constructions must be computed on very large corpora (Liberman 1989.)

The current limitations of this approach must be clearly stated. First of all, this method conflates the polysemous usages of certain verbs and, in English, of the verb-particle construction. It could be argued that with enough corpus data, this would become unimportant, but we believe this position not correct. What is required is a fuller analysis of adjuncts in order to know if a verb has been coerced. For example, it could be the case that a verb which is S in the unmarked case (i.e. in a neutral context) tends to appear as a T verb frequently, since that verb might not occur frequently in a null context at all. As another example, consider the case of a typical stative verb "know". With the object "answer", "know" becomes typically inchoative, e.g. "know the answer by tomorrow" meaning "become knowledgeable of the answer", or it could be used in the transition sense, e.g. "he will know the answer by tomorrow" meaning "he does not know now and will know then." Thus, it could be underlying semantic structure, and not surface syntactic behavior, that determines coercion possibilities. [9] In conclusion, The DEGREE APPROACH captures the fact that verbs have degrees of e-type, i.e. that some verbs are more pliable than others. Thus, rather than the non-degree values in Figure One, we argue for entries like:

`e-type(resemble) =` $V_s(.0740)$
`e-type(go) =` $V_s(.2304)$
`e-type(seem)=` $V_s(.0120)$

A corpus-based method can be used to automatically derive values for e-type, i.e. under a certain cut-off, the verb is stative, but alterable in context. More importantly, it gives a degree of likelihood that given any context, the verb will be used statively or non-statively.

[8] This is a fact which any semanticist who is trying to argue a firm point can attest to.

[9] Also, there appear to be some clashes on the resulting values and intuitions, thus leading to the suggestion that either our intuitions are not correct or that the method is unreliable. We have not, in fact, addressed the issue of underlying semantic representation (e.g. in terms of primitives) in this paper. It has been suggested that syntactic tests for aspect might be flawed, and that the only way to distinguish aspectual classes is via the semantic consequences of a stative vs. nonstative proposition. If correct, the approach of extracting values based on syntactic tests will fail by definition, regardless of whether the values are assigned manually or automatically.

7 References

1. Byrd, Roy. 1989. "Discovering Relationships among Word Senses", *Dictionaries in the Electronic Age: Proceedings of the Fifth Annual Conference of the University of Waterloo Centre for the New Oxford English Dictionary*

2. Comrie, Bernard 1976. *Aspect.* Cambridge University Press: Cambridge, English.

3. Dowty, David. 1979. *Word Meaning and Montague Grammar.* Dordrecht: Holland.

4. Francis, W. Nelson and Henry Kučera. 1982. *Frequency Analysis of Enlglish Usage: Lexicon and Grammar.* Houghton Mifflin: Boston.

5. Klavans, Judith L. (1988) "Building a Computational Lexicon using Machine Readable Dictionaries", *Proceedings of the Third International Congress of the European Association for Lexicography.* Budapest, Hungary.

6. Lakoff, George. 1965. "On the Nature of Syntactic Irregularity", PhD Dissertation, Department of Linguistics. Indiana University: Bloomington, Indiana.

7. Liberman, Mark. 1989. "How Many Words Do People Know?" Invited Talk delivered at the 27th Annual Meeting of the Association for Computational Linguistics. Vancouver, B.C., Canada.

8. McCord, M. C. 1980. "Slot Grammars" *Computational Linguistics*, Vol 6:31-43.

9. McCord, M. C. 1990. "SLOT GRAMMAR: A System for Simpler Construction of Practical Natural Language Grammars," in R. Studer (Ed.), *International Symposium on Natural Language and Logic*, Lecture Notes in Computer Science, Springer Verlag.

10. Moens, Marc and Marc Steedman. 1988 "Temporal Ontology and Temporal Reference" *Computational Linguistics.* Vol. 14:2:15-28.

11. Pustejovsky, James. 1991. "The Syntax of Event Structure" *Cognition* Vol. 41:103:47-82.

12. Pustejovsky, James and Bran Boguraev. to appear. "Lexical Knowledge Representation and Natural Language Processing" *IBM Journal of Research and Development.*

13. Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, Jan Svartvik. 1972. *A Comprehensive Grammar of the English Language*, Longman: London.

14. Tenny, Carol Lee. 1987. "Grammaticalizing Aspect and Affectedness", PhD Dissertation, Department of Linguistics. M.I.T., Cambridge: Massachusetts.

15. Vendler, Zeno. 1967. *Linguistics in Philosophy*, Cornell University Press: Ithaca, New York.