Inheriting Verb Alternations*

Adam Kilgarriff
Longman Dictionaries
Burnt Mill
Harlow, Essex CM20 2JE
England

April, 1993

Abstract

The paper shows how the verbal lexicon can be formalised in a way that captures and exploits generalisations about the alternation behaviour of verb classes. An alternation is a pattern in which a number of words share the same relationship between a pair of senses. The alternations captured are ones where the different senses specify different relationships between syntactic complements and semantic arguments, as between *bake* in "John is baking the cake" and "The cake is baking". The formal language used is DATR. The lexical entries it builds are as specified in HPSG. The complex alternation behaviour shared between families of verbs is elegantly represented in a way that makes generalisations explicit, avoids redundancy, and offers practical benefits to computational lexicographers.

1 Introduction

The paper shows how the verbal lexicon can be formalised in a way that captures and exploits generalisations about the alternation behaviour of verb classes. An alternation is a pattern in which a number of words share the same relationship between a pair of senses. The kinds of alternations to be captured are ones where the different senses specify different relationships between syntactic complements and semantic arguments, as in the relation between *bake* in "John is baking the cake" and "John is baking", or between *melt* in "the chocolate melted" and "Mary melted the chocolate".[1] Given that compactness and non-redundancy are a desideratum of theoretical descriptions, the different usage-types for *bake* and *wipe* should not require us to introduce different primitives into the lexicon. Moreover, as the alternations are shared with other verbs, they should be described at some general node in a hierarchically organised lexicon, and inherited.

DATR is a formal language in which the such relationships and generalisations can be simply stated.

*I would like to thank Gerald Gazdar and Roger Evans for their many valuable comments, and SERC for the grant under which the work was undertaken.

[1]The morphosyntactic distinctions between, for example, *bake* and *is baking* are not addressed here. Extensive DATR treatments of morphology are provided in various papers in [Evans and Gazdar, 1990].

Much has been written about verb alternations and their syntactic corollaries. Here we do not add to the evidence or construct new theory, but simply formalise other people's accounts: those of [Atkins *et al.*, 1986] and [Levin and Rappoport Hovav, 1991]. The first investigates the range of alternations between transitive and intransitive forms of verbs. The second, titled *Wiping the Slate Clean*, explores the relations between meaning and subcategorisation possibilities for 'wipe' verbs, 'clean' verbs, and related groupings. The language used is DATR, a default inheritance formalism designed for lexical representation. We follow Levin and Rappoport Hovav in taking a distinct subcategorisation frame as defining a distinct word sense, and also in working with commonsense verb classes such as 'cooking verbs', since classes such as this serve to predict the alternations a verb participates in with some accuracy.

An important constraint is that the lexical entries are of a kind specified by a grammar formalism, so can be used for parsing and semantic interpretation. The formalism chosen in this paper is HPSG [Pollard and Sag, 1987].

Below we present detailed formal accounts for alternations involving cooking verbs and physical-process verbs. After motivating the DATR treatment and considering related work, we describe how verb entries appear in HPSG, then represent alternations as mappings between HPSG lexical entries, then introduce the main constructs of DATR and define a translation from HPSG notation to DATR. Finally we build a DATR inheritance network which represents the alternate verb forms by inference, without the lexicographer having to explicitly say anything about them.

The analysis presented in this paper is a part of a larger lexicon fragment which describes a further five alternations relating seven verb classes and formalises much of the structure described in both articles. The complete fragment, illustrated in Fig. 1. is presented in full in [Kilgarriff, 1992].

1.1 Why DATR?

As 'lexicalism' — the doctrine that the bulk of the information about the behaviour of words should be located in the lexicon — has become popular in computational and theoretical linguistics, so formalisms for expressing lexical information have been developed. The syntax, semantics and morphology of most words is shared with that of many others, so the first desideratum for any such formalism is to provide a mechanism for stating information just once, in such a way that it is defined for large numbers of words. Inheritance networks serve this purpose. If words are arranged into a taxonomy or some other form of network, then a fact which applies to a class of words can be stated at a nonterminal node in the network and inherited by the words to which it applies. Work in knowledge representation has addressed questions of different kinds of network, and the kinds of machinery needed to retrieve inherited information, in detail (see, e.g., [Brachman and Levesque, 1985]).

The next requirement is that exceptions and subregularities can be expressed. It must be possible to describe concisely the situation where a word or class of words are members of some superclass, and share the regular characteristics of the superclass in most respects, but have different values for some feature or cluster of features. Several lexical representation formalisms addressing these desiderata have been proposed, e.g.