THE STRUCTURE OF THE MERRIAM-WEBSTER POCKET DICTIONARY

Robert Alfred Amsler

Department of Computer Sciences
University of Texas at Austin
Austin, Texas 78712-1188

TR-164 December 1980

Research supported in part by the National Science Foundation under grant MCS77-01315 and by the Computer Science Department and Linguistics Research Center of the University of Texas at Austin.

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to the chairman of my supervisory committee, Professor Robert F. Simmons, whose direction, aid, and encouragement have been invaluable, and to the members of the committee, Dr. Ron Wyllys, Dr. John Loehlin, and Dr. Brian Stross. I would also like to thank Dr. Nicholas Hopkins and Dr. Mary Sanches whose ethnosemantics courses made significant contributions to the initial dissertation topic choice.

Thanks should also go to Dr. Winfred P. Lehmann, Director of the Linguistics Research Center, Jonathan Slocum, Michael K. Smith, Dr. John S. White, Gretchen Hazard and Dr. Helen-Jo Hewitt for their contributions to The Dictionary Project.

I also wish to thank John Olney, Sally Sedelow and the National Science Foundation for making it possible to explore in greater detail the structure of the lexicon than would have been possible without their initial efforts and funding support.

Finally, I wish to thank the graduate students and faculty whom I have been associated with during these many years at The University of Texas, as well as those I have met via the ARPANET or repeated visits to old alma maters who have offered encouragement and counsel over the many years this research has taken.

Robert A. Amsler

The University of Texas
Austin, Texas
November, 1980

STATEMENT OF PURPOSE

This dissertation has as its purpose the exploration and discussion of the structure of a machine-readable copy of an ordinary pocket dictionary with particular attention to the utility of the information contained therein for future application in computational linguistics, ethnosemantics, and information science. This structure was first explored by hand using two concordance-like printouts of the Merriam-Webster Pocket Dictionary's contents prepared from magnetic tapes produced by John Olney at System Development Corporation [Olney 1967,1968].

Initially the goal of the analyses was the determination of whether useful semantic information could be derived from dictionary definitions. Once it was determined by several hand analyses that such data did provide a new source of information about the lexicon, the larger goal of assembling a complete taxonomy of the words in the dictionary was conceived and undertaken.

A hand analysis of the rich semantic information contained in the dictionary for verbs defined in terms of "move" is presented in chapter 2. This componential analysis of a set of definitions revealed the potential of dictionaries for use as the basis of numerous additional studies of high-frequency verbs based upon their usage in defining more specific verbs.

Chapter 3 presents the results of a hand analysis of the taxonomy of "vehicle" terms developed from a large sample of definitions based upon the word "vehicle" and its descendants. This study revealed that indeed the dictionary did contain large, coherent, and computationally useful information in a taxonomic-like organization that could be revealed by connecting together definitions on the basis of their defining terms.

Chapter 4 deals with the steps involved in designing, loading, and selectively accessing large databases containing all the definitions of nouns, verbs, and adjectives contained in the dictionary. Statistical information on the frequencies and nature of the part of speech data contained in the dictionary is given in chapter 4 and appendix 3. A "word sense meaning" representation for uniquely specifying dictionary senses and the problems of semantic ambiguity in lexical usage are also illustrated.

In chapter 5, statistics on the frequency of dictionary defining vocabulary and measures of semantic ambiguity are given. Section 5.2 contains an extensive discussion of the steps taken to perform semantic disambiguation on dictionary definition terms along with further statistics on the frequencies of such disambiguated words in the dictionary. Appendix 4 gives part of a disambiguation protocol session transcribed from a tape-recording and demonstrating the nature of the human disambiguation task carried out on the dictionary. Finally, chapter 5 concludes with a discussion of the methods used to computationally assemble and enumerate the complete taxonomic structure of the dictionary's noun and verb definitions.

Chapter 6 discusses some of the findings about the nature of the dictionary's taxonomies of nouns and verbs. It discusses the manners in which dictionary taxonomies ultimately terminate in primitive root concepts, relationships to other taxonomies, case argument relations to verbs, and in partitives and collectives.

Finally, chapter 7 concludes with a discussion of a possible means of automating the analysis procedures of chapters 4 and 5 to perform fully automatic parsing and disambiguation of dictionary entries. This discussion also demonstrates the application of the proposed disambiguation technique to natural language processing and computational linguistics in general.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

PREFACE

This dissertation has combined techniques and goals from three disciplines: ethnosemantics, information science, and natural language processing/ computational linguistics in artificial intelligence.

Ethnosemantics has developed numerous empirical linguistic techniques for extracting, analyzing, and presenting semantic information about the basic concepts of a culture. I chose to apply some of these techniques to the study of the lexical knowledge of an advanced civilization as represented in a small sample of some 45,000 English dictionary definitions.

Information science has developed extensive classifications of knowledge in the form of library classification systems. It has a well-developed body of knowledge dealing with computational manipulation of lexical data, database design and use, and efforts to span the knowledge of a whole civilization in one systematic method.

Natural language processing/computational linguistics in artificial intelligence are integrally involved in the effort to build computer programs with human-like intelligence. To do this they will require a large amount of "artificial memory" to provide such programs with the necessary background information which humans use to perform complex tasks such as lexical disambiguation. This dissertation is motivated by the desire to provide this "artificial memory" for intelligent computer programs by deriving it from existing published dictionaries.

There are many fields which are related to the dictionary, but which the author cannot claim to have adequately treated in this dissertation. These include linguistics, philosophy, psycholinguistics, socio-linguistics, lexicography, dictionary history, dictionary data base theory, classification theory, etc. I explicitly mention this because I believe the dictionary is capable of being analyzed within these fields to provide further insights about our lexical knowledge and the way it is structured in dictionaries by lexicographers.

While one dictionary cannot suffice to serve as the definitive source of information about English lexical semantic structure; I do believe that one dictionary represents far more than just the opinions of the small group of lexicographers who wrote it. It represents a snap-shot view of the mental lexicon which we all share and which permits us to communicate intelligibly to each other.

I believe every intelligent entity that wishes to communicate in English will have to have a shared lexical taxonomy of concepts. This dissertation is a first step toward providing such a taxonomy.

Robert A. Amsler
November 27, 1980

CHAPTER I INTRODUCTION

There would seem to be a considerable amount of human understanding involved with knowledge of taxonomic relationships -- knowledge which is in some sense immutable and common to all speakers of the language and thus capable of codification for computational use. One convenient source of this knowledge is the set of definitions in an ordinary dictionary.

1.1 Dictionaries as a Source of Taxonomic Knowledge

Definitions are ideally structured for taxonomic organization. Each definition is written to employ a "genus" term and "differentia" which distinguish the genus term of one definition from that of otherwise identical definitions. This is to say, each definition contains a key word or expression which tells what the concept being defined "is a type of". This relationship has been conveniently referred to as an ISA relationship, "x ISA y" meaning that the concept x is a more specialized instance of the concept y, e.g. "cat ISA animal".

Dictionaries then expand this ISA relationship by augmenting it with differentia, such as "with paws" and "kept as a pet" to form a full definition such as,

cat - an animal with paws kept as a pet.

By finding all the definitions in which "animal" is used as the genus term one can assemble a taxonomy containing one ancestor and many descendants - e.g. figure 1-1.

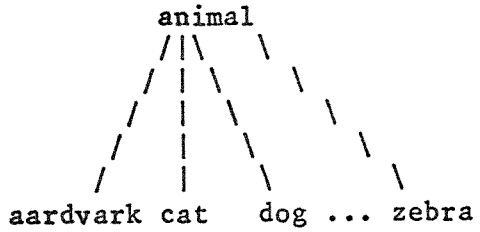

Figure 1-1 A One-Level Taxonomy

Likewise looking up the definition of "animal" we might find it defined as a "living thing capable of voluntary motion". Together with the other members of the "living thing" taxon we would find "plant" and perhaps "microbe" allowing creation of a two level taxonomy such as is shown in figure 1-2.

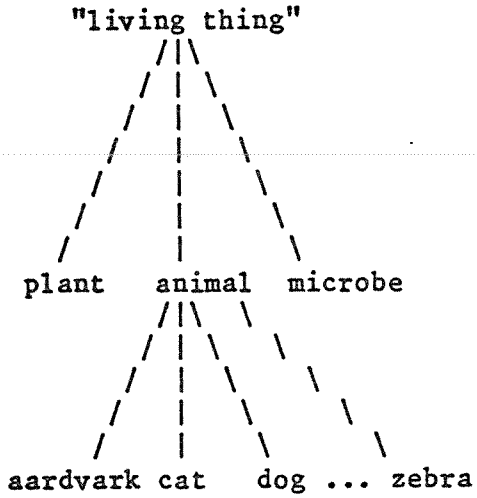

Figure 1-2 A Two-Level Taxonomy

1.2 Recent Progress in Natural Language Processing Relevant to the Dictionary

A significant current trend in cognitive science and data base management has been the heightening of interest in ISA hierarchies (i.e., "is a" chains of semantic triples of the form "CAT ISA MAMMAL", "MAMMAL ISA ANIMAL", etc.) [Fahlman 1977; Hayes 1977a, 1977b; Browse 1978; Martin 1978; Mylopoulos et al. 1978; Schneider 1978]. Another relevant development lies in the potential for a new computer architecture as described by Fahlman [1977], based upon very large taxonomic data structures, and the likelihood that such hardware will be available at Carnegie-Mellon University for research use over the ARPANET [Fahlman 1978].

Although natural language processing systems have long utilized some form of taxonomic information there have only recently been investigations of the potential utility of very large taxonomic structures and the processing techniques needed to handle them. E.g., in artificial intelligence, large taxonomies are important data base components of computer "consultants" such as (MYCIN) [Shortliffe 1976] and (PROSPECTOR) [Duda et al. 1977]. Scott Fahlman [1977] has demonstrated the potential viability of specific hardware in processing very large taxonomies, and Philip Hayes [1975, 1977a, 1977b] has addressed some of the processing algorithms which can be applied to a very large taxonomic lexicon to perform complex tasks such as automatic disambiguation.

Fahlman points out that existing problem-solving systems based upon deductive search are extremely slow even with the use of complex search-optimizing algorithms. This is of concern not only because of the necessity of writing new search-optimizing algorithms for each new

knowledge domain, but also because this type of search is trivially easy for humans. Thus something is wrong with our existing hardware which results in efforts to avoid using very large taxonomies as an integral component of such processing systems.

Fahlman [1978] has further indicated that it should be possible within a few years to develop hardware capable of processing very large taxonomies, on the order of 40,000 nodes in size, in which,

> the search for most implicit properties and facts ... will take only a few machine-cycles, and [in which] the time required is essentially constant, regardless of how large the knowledge base might become.
>
> [Fahlman 1977, p. 11]

The knowledge base of which Fahlman speaks is precisely the type which the dictionary contains as an ISA hierarchy. Thus one may view the further development of dictionary taxonomies as leading toward a data base which may be tested in such hardware.

1.3 Extracting Taxonomies from Machine-Readable Dictionaries

The typical dictionary contains tens of thousands of definitions which form the basis of an enormous taxonomic system representing part of our knowledge of the meanings of the words of the English lexicon. In terms of the needs of artificial intelligence and computational linguistics detailed explicit knowledge of this taxonomic system may be required to carry out human-like tasks with language.

If the extraction of such taxonomies had to be accomplished entirely by hand it could take decades of effort. Fortunately a number of dictionaries, complete with definition texts, exist in machine-readable form. Two such dictionaries are the Merriam-Webster Pocket Dictionary (MPD) and the much larger Merriam-Webster Seventh Collegiate Dictionary (W7), both of which were rendered machine- readable by John Olney et al. [1967]. For the purposes of this study, the 4.5 megabyte MPD was used.

In addition to preparing a machine-readable dictionary, Olney prepared a complete inverted index to the definition texts of the dictionary ("the concordance index") which listed all defined vocabulary ("main entries") under the words which occurred in their definitions. An extract from this concordance index for the occurrences of "move" is shown in figure 1-3.

18 56 SEPT 17,1968 PAGE 4654

	ENTRY NAME	TYPE	POS	MAIN VBILL UNOTE	SENSE NUMBR	MAIN OR R-ON	SEQ NO.
MOVE	ADVANCE	1	VB	M	1	M	A04
	AGILE		AJ	M	0	M	A03
	AGITATE		VB	M	1	M	A02
	ANGLE	2	VB	M	0	M	A04
	ANIMATE	2	VB	M	3	M	A05
	APPROACH		VB	M	1	M	A02
	ASCEND		VB	M	1	M	A02
	BANG	1	VB	M	2	M	A07
	BARGE	2	VB	M	2	M	A02
	BISHOP		N	M	3	M	A06
	BLOW	1	VB	M	1	M	A02
	. . . .						
	ZOOM		VB	M	1	M	A02

259 OCCURRENCE(S) * 3* N =17 VB=235 AJ=6 AV=1

Figure 1-3 Partial entry from the Olney Concordance Index listing definition main-entries in which "MOVE" occurs

This concordance index of some 10,000 pages provides a means of finding all definitions which contain a given genus term. It does not, however, distinguish the part of speech of an occurrence, nor whether an occurring word is in fact a genus term or merely used in another position in a definition. Thus, if one were looking for "animal"'s usages in other definitions one would be directed to definitions such as "a cage used to confine an animal" and "a man (as opposed to an animal) who does such a task". Thus not every instance of "animal" which occurs in the dictionary is a taxonomic usage.

Using the concordance index one can look up each and every occurrence to determine how it is being used. Chapter 2 gives an example of how this data is used in exploring the breadth or horizontal dimension of meaning in the dictionary for the verb "move".

To explore the depth or vertical dimension of taxonomics in the dictionary the concordance index begins to become a less and less tractable tool. This led in 1975 to the development of a specialized "taxonomic" concordance index which attempted to automatically eliminate as many taxonomically irrelevant entries from the original index as possible. Using the taxonomic concordance index one could in a few weeks

build taxonomies such as that developed in chapter 3 for the "vehicles".

To progress beyond the taxonomic concordance index required extensive human effort. With the assistance of the National Science Foundation (NSF Project MCS77-013515, Development of a Computational Methodology for Deriving Natural Language Semantic Structures via Analysis of Machine-Readable Dictionaries), the development of the taxonomic description of the MPD was undertaken in 1976-1978. Using human informants and procedures derived from cognitive anthropology (ethnosemantics) the senses of the "kernel" terms of definitions were disambiguated and a taxonomic tree of the lexicon for nouns and verbs was computationally grown. The procedures for doing this are given in chapters 4 and 5, and the resultant structure of the dictionary is discussed in chapter 6.

Finally, after this development of the taxonomic structure of a representative subset of the English lexicon, chapter 7 is devoted to some possible future applications of this data and conclusions about the structure of the dictionary.

CHAPTER II THE VERBS OF MOTION: A CASE STUDY

This chapter attempts to show how meaning can be analyzed on the basis of the definitions provided in an ordinary pocket dictionary. While the source of the original data may be thought to be so inadequate as to preclude the attainment of valid results, the techniques to be described are such as to compensate for deficiencies in their original material. Whereas a single definition or even a small set of definitions may fail to cover the complete meaning of the words they define, access to a complete set of taxonomically related definitions may be used to compose a nearly complete portrait of the meaning of their root concept.

Specifically, one may build an accurate description of the numerous arguments which the verbs of motion may have by studying not isolated definitions of individual verbs, but by dealing with all dictionary definitions which use a single verb of motion as their defining kernel. This chapter will serve as an illustrative example of this technique which may be applied to numerous other verb classes (verbs of communication, thinking, creating, destroying, acquiring, etc). While other authors have dealt with motion verbs (e.g. [Miller 1972; Levelt, Schreuder and Hoenkkamp 1976]), the analysis presented here is believed to be the only one based directly on dictionary entries rather than informant's responses or linguistic intuition.

2.1 The Senses of the Verb "Move"

The verb "move" has 8 major senses in the Merriam-Webster Pocket Dictionary. These senses are listed in table 2-1.

Table 2-1 Senses of "move" in the MPD

move 1a - to go or cause to go from one point to another
move 1b - ADVANCE
move 1c - DEPART

move 2 - to change one's residence

move 3a - to change or cause to change place, position, or posture
move 3b - SHIFT

move 4 - to show marked activity

move 5a - to take or cause to take action
move 5b - PROMPT

move 6 - to make a formal request, application, or appeal

move 7 - to stir the emotions

move 8 - EVACUATE

These definitions do not however reflect the case-argument differences between the senses of "move". On the basis of their case argument role differences there are only six senses which need be distinguished. These are the given in figure 2-1.

```
move 1' (examples - travel,march)
        AGENT:
        THEME: <Phys-obj>
        SOURCE (from): Location1
        GOAL (to): Location2
        LOCUS: <medium>

move 2' (examples - agitate,fidget)
        AGENT:
        THEME: <Phys-obj part> or <Phys-obj>
        LOCUS: Location1
        SOURCE: Position1 or Posture1
        GOAL: Position2 or Posture2

move 3' (examples - migrate)
        AGENT: <animate>
        THEME: "center of activity"
        SOURCE (from): Location1
        GOAL (to): Location2

move 4' (examples - "metaphorical" uses of move)
        THEME: <intangible>
        SOURCE (from): State1
        GOAL (to): State2

move 5' (examples - touch,persuade,stir-2)
        THEME: <sentient>
        INSTRUMENT: <event>
        SOURCE (from): State1
        GOAL (to): State2

move 6' (examples - "move for the court to recess")
        AGENT: <human>
        THEME: <event>
        SOURCE: State1
        GOAL (to): State2
```

Figure 2-1 Case Argument Patterns of the Senses of "move"

The patterns of figure 2-1 require additional explanation. If we construct a decision tree for the arguments we can see how the senses above could be eliminated during disambiguation of a particular "move" usage.

1. Has a physical object changed location ?

(yes) 1.1 Has a physical object changed its habitual location ?

(yes) 1.1.1 MOVE3'
"the birds moved south for the winter"
"We just moved to town"
"the company moved its headquarters out of the city"

(no) 1.1.2 MOVE1'
"The cars moved down the road"
"The chess master moved (his chess piece)"
"Have your bowels moved today"

(no) 2 Has a part of a physical object changed location ?

(yes) 2.1 MOVE2'
"The trees moved gently in the breeze"
"He moved restlessly in his sleep"
"She pressed a button and the printing press began moving"
"The boat moved slowly from side to side at the dock"

PHYSICAL CHANGES (THEME is <phys-obj> or <phys-obj-part>;
SOURCE & GOAL are locations)

MENTAL CHANGES (SOURCE & GOAL are states)

(no) 3 Has the rate at which something was happening changed ?

(yes) 3.1 MOVE4'
"The plot moves quickly here"
"the melody moved gracefully from note to note"
"for a while there was nothing to do, but suddenly things really began to move"

(no) 4 Has some action which is part of a plan or procedure been proposed or performed ?

(yes) 4.1 MOVE6'
"moved for a recess"
"revolutionaries must move carefully"

(no) 5 Has someone's emotional state changed ?

(yes) 5.1 MOVE5'
"moved to tears"

Figure 2-2 Decision Tree for the Verb "Move"

2.2 Move and its Descendants

The most logical initial choice for a motion verb would seem to be the verb "move", being itself the word from which the nominalized form "motion" is derived. The only instrument required for this investigation is in fact a concordance to an authoritative dictionary, such as one of the Merriam-Webster dictionaries. Such a concordance is available for the Merriam-Webster Pocket Dictionary (MPD).

As we saw in chapter 1, the mere occurrence of a word within another word's definition does not imply its utility for describing the features of that word. "Move" occurs 17 times within the definitions of nouns (e.g. bishop, caster, drive, electromotive force, gambit, ...), 6 times within the definitions of adjectives (e.g. agile, eloquent, hypersonic, inert, ...) and even once within the definition of an adverb (clockwise). Thus we lose 24 occurrences of "move" to usage in other than kernel roles in non-verb definitions. Of course there are other applications of the concordance which might critically use these occurrences, e.g. to study case-arguments by examining the nouns which use "move" in their definitions, or to study the role of nominal verbs, "move" being itself a noun (as in "chess moves"), but for this task's needs only verb occurrences of "move" are required.

Of the 235 occurrences of "move" within verbs, all are within the body of the definition, but this need not be the case. A word may occur within a usage example supplied to illustrate the definition being given, as in the occurrence of the verb form "ran" in the definition of "behind" as <ran behind the house>. Such usage examples are also a good subject for investigation, but they do not serve our current needs. We seek only occurrences of "move" in the defining text of another verb's definition.

2.3 Kernels of Definitions

As well as not being used within the body of the definition a word may be used in some capacity other than as the "kernel" of a definition. Definitions of nouns and verbs in typical modern dictionaries follow a rigid stylistic form. They consist of a single sentence almost always composed of two parts, a <genus> or kernel term and a <species> or descriptor portion. This allusion to biological taxonomics is the unavowed theory behind dictionary definitions. The dictionary bears a remarkable resemblance to a tree-structured description of meaning when viewed in this manner. Later I shall discuss the serious deviations from this strict taxonomic form of description and what they portend about semantics, but for now the reader may assume dictionaries are built as true taxonomic data structures with the kernels of definitions indicating

the more-general ISA-type nodes of more-specific words being defined.

The few deviations from this defining structure are definitions by enumeration of examples, as for:

furniture-1b - moveable articles (as chairs, tables, or beds) for a room.

green-1b - the color of fresh growing grass or of the emerald.

Finally, even if a verb's definition does use "move" as its kernel, this "move" must be in the sense of "move1" as described in table 2-1 and figures 2-1 and 2-2. For the verbs listed in the concordance index under "move", table 2-2 lists the 171 definitions in which "move" managed to pass all these tests. Appendix 2 contains the full definitions of these verbs for your convenience.

Table 2-2 171 MPD Verbs Defined in Terms of "MOVE1"

AGITATE 1.1	FREEWHEEL 1	REVOLVE 12	TILT 1
APPROACH 1.1	FUNNEL 2	RIDE 2	TOSS 4
ASCEND 1.1	GLIDE 1	RING 2.2	TRAVEL 2
BARGE 2.2	GO 1.1A	RISE 7	TRAVEL 5
BLOW 1.1	GRAVITATE 1	ROCK 1	TREAD 2
BLUNDER 1.1	GRAVITATE 2	ROLL 1	TREMBLE 2
BOB 1.1	GRIND 6	ROLL 2	TRIP 1
BOLT 1.1	HASTEN 2	ROLL 3	TROOP 1
BULLDOZE 1.1	HIKE 1	RUN 15	TUG 3
BUMP 1.1	HITCH 1	RUN 7	TURN 1.1A
BUSTLE 1.1	HOP 1	RUSH 1	TWITCH 1
CARRY 1.1	HOVER 1	SCOUR 1	TWITCH 2
CIRCLE 2.2	HURL 1	SCREW 3	UNSETTLE 1
CIRCULATE 1.1	HURRY 3	SCUD 1	UP 3.3A
COAST 2.2	HURTLE 1	SHAKE 1	VERGE 3.2
CRAWL 1.1	INCH 2.1B	SHUFFLE 3	WADE 2
DANCE 2	JERK 2	SHUTTLE 1	WAGGLE 1
DANDLE 1	JIGGLE 1	SIDLE 1	WALK 1
DASH 7	JOLT 1	SKIP 1	WALTZ 2
DELAY 3	KEDGE 1	SKULK 1	WANDER 1
DIVERGE 1	LASH 1	SLIDE 1	WARP 3
DODGE 1	LABOR 2	SLINK 1	WASH 6
DRAG 2	LUMBER 1	SMACK 1	WAVE 4
DRAW 1.3A	LURK 1	SMASH 2	WHIP 10
DROP 9	MARCH 1	SNEAK 1	WHIP 1
EDGE 2	MIGRATE 1	SPIN 6	WHIRL 1
FALL 1.12	MILL 2	SPIRAL 1	WHIRL 4
FALTER 1	NOD 2	SPRING 1	WHIR 1
FAN 2	NOSE 2	SPRING 3	WHISH 1
FIDDLE 2	NOSE 6	SQUELCH 2	WHISK 1
FIDGET 2.1	PADDLE 2.1	STEAM 3	WHISK 2
FLAP 3	PADDLE 3.1	SURGE 2	WIGGLE 1
FLING 1	PASS 1.3	SWASH 1	WIND 4.2
FLIRT 1	PLAY 2.2	SWEEP 4	WOBBLE 1
FLOAT 2	PLAY 2.6	SWEEP 6	WORK 3.19A
FLOUNCE 1	POUND 4	SWERVE 1	WORM 1
FLUCTUATE 1	PROGRESS 1	SWING 1	WREATHE 1.2B
FLUTTER 2	PUTTER 1	SWING 5	WRENCH 1
FLUTTER 4	RATTLE 2	TEAR5 5	WRIGGLE 1C
FLY 1	RECEDE 1	TEETER 1	WRITHE 1
FLY 2	RECIPROCATE 1	THRASH 3	ZIP 1.1
FLY 6	REMOVE 1.1	THROW 1.7A	ZOOM 1
FORGE 1	REMOVE 1.2	THUD 1	

2.4 Motion Verb Descriptors

The remainder of a definition excluding the kernel may be termed the descriptors. This phrase like its biological counterpart, denotes how this species of word differs from the genus class in which it belongs. These descriptors are strongly organized. There are perhaps two dozen different types of species descriptors for any given verb. Just as a biological life form might be described by specific valued descriptors for size, habitat, life expectancy, etc., so are verbs (and other words - e.g. nouns) described by giving values for many attributes these words may possess, such as the familiar case argument values shared by all verbs.

Following Simmons [1973] we have used case argument names such as:

1. AGENT (Human subject)
2. THEME (Direct object)
3. SOURCE (Prior state, condition, location, etc. of the theme)
4. GOAL (Effective state, condition, location, etc. of the theme after or while the verb in question operates)
5. INSTRUMENT (Associated agent of the verb, often the only tangible effector, but nevertheless conceptually a tool of some other agent)
6. LOCUS (The spacial descriptor of where the verb performs its action)

We may readily assign some of the verb of motion descriptors to these case argument categories, but this system fails to account for all the descriptor categories which may be specified for a set of verbs such as those dealing with movement. I have developed an expanded set of motion attributes for the verbs of motion. These categories are enumerated in figure 2-3.

AGENT
MOTIVE
EMOTIVE - accompanying mental/physical state

LOCUS - (specialized to mean PATH-of-motion)
MEDIUM
DISTANCE - feature: <+distant><-distant>
SOURCE - from
GOAL - to

THEME - object moved
VELOCITY - feature: <+rapid><-rapid>
ACCELERATION - complex-feature: <+acceleration+>,etc.
FORCE - feature: <+force><-force>
RESISTANCE - feature: <+resistance><-resistance>
STEADINESS - feature: <+steady><-steady>
CONTINUITY - feature: <+continuity><-continuity>
INSTRUMENT
ORIENTATION

ACCOMPANYING ACTION
MINIMUM NUMBER OF MOVING OBJECTS REQUIRED
MINIMUM NUMBER OF ACCESSORY
(NON-MOVING) OBJECTS REQUIRED

Figure 2-3 Descriptors of the Verbs of Motion

This extended set of verb argument roles is illustrated in the following sentence pattern:

> [1] <agent> moves [2] <theme> [3] <orientation> over [4] <distance> using [5] <instrument> along [6] <path> with [7] <steadiness> and [8] <continuity> at [9] <velocity> from [10] <source> to [11] <goal> with [12] <acceleration> through [13] <medium> because of [14] <motive> using [15] <force> against [16] <resistance> displaying [17] <emotive>.

Normally only a few of these case arguments would be used in a sentence, such as, "[1] The boy moved [2] the rock [6, 11] downhill from [10] atop the mesa.", but for a demonstration one can create an (admittedly awkward) sentence using all the arguments, such as:

> [1] The mountaineers moved [2] their tired bodies [3] sideways [4] twenty feet along [6] the rock shelf using [5] ropes and pitons, with [8] continuous [7] steady [9] slow steps from [10] the windswept north slope of the mountain to [11] the sheltered west slope with [12] ever increasing speed over [13] the rock-hard surface [14] to escape the imminent storm, using [15] great effort against [16] the opposing wind and with [17] immense courage.

Before discussing the nature of the specific verb of motion descriptors it is noteworthy that two classes of such descriptors may be delineated. These are the classic "features" most notably referred to by J. J. Katz [1964]. Features are descriptors which have binary values such as <+animate> and <-animate>. As such they are usually not referred to as values of the attribute "animate", but rather just noted as being present or absent for a given word. Features may also refer to scales of values, such as age, height, weight, etc. which may be abridged to a binary choice <+old>,<-old>; <+tall>,<-tall>; <+heavy>,<-heavy>; etc.

Ordinary descriptors are not of this limited a range of values. The case argument descriptors are a good example. Their range of values even when specialized to movement verbs is quite broad. Thus, instruments involved in motion verbs are frequently vehicles such as cars, buses, trains, etc. and the source and goal of a motion verb are locations. The locus of a movement verb may be termed the path or course of the movement. Note that this is not the same as the medium of a motion verb, which concerns what the theme moved through, on, or within while following a

given path or course. This also differs from the orientation of the object(s) moving. Path may be visualized as the trace left by a moving object through space, orientation as the relationship of the object's "front" to this traced path.

This specialization of case argument names should not surprise many, for case arguments are an attempt to deal with the semantic attributes of a verb on something approaching a syntactic level. As such, they have been made as broad as possible and are in fact linked very closely with prepositions and the positional information in a sentence. It seems questionable whether such a case-based approach, which attempts to stuff meanings into a limited number of case roles and to delineate these roles solely on the basis of their position in the sentence or their introductory preposition, can succeed. Certainly it seems clear that not all verbs have the same subset of case argument roles available. The motion verbs, for example, definitely have a "medium" as distinct from the more general "locus".

Locus - (including Path and Course)

The descriptor whose value is most often specified for a motion verb is that of locus, or more specifically path. Initially I was somewhat apprehensive that there were actually two locus descriptors for motion verbs, one dealing with the area in which the motion occurred, the other with the path or course the moving object followed. For example, "to dandle" is defined as, "to move up and down in one's arms or on one's knee in affectionate play (as with a baby)." Thus it might seem that the path of the motion was up and down, and the locus was in one's arms or on one's knee, but this is not true. The phrases "in one's arms" or "on one's knee" refer to a secondary action, i.e. to the full phrase "while holding in one's arms or on one's knee". This and similar problems involving the use of ambiguous prepositions such as "about" in the definitions of "thrash", "swash", "play", "flutter", and "dance" (i.e. all defined using "move about ...") seem best interpreted as referencing the path of the motion rather than the spacial area in which it occurs.

If we now consider the values associated with this descriptor in the actual dictionary definitions of the MPD "move" verbs we can arrange these in a hierarchy as follows:

I. on a course,in a circuit - go
  A. in a particular direction - verge, tend
    1. aside from a straight line or course, sideways - swerve, sidle
    2. in a downward direction,downhill,downward - fall, drop
    3. upward - rise,up,ascend
    4. in relation to another part - work
      a. around an axis or center, in a circle or similar curve, around, as if by completing a revolution - turn,whirl,roll,circle
        i. spirally - screw,ring
          /a/ in spirals
          /b/ in a spiral course - spiral
        ii. in an arc - swing
        iii. <circularly>
          /a/ in circles - wreathe
          /b/ in a circle,in a ring - circulate,mill,ring
            /i/ in orbit,in an orbit - revolve

  B. in different directions - diverge
    1. up and down - nod,fluctuate,bob,dance,dandle
    2. rocking
      a. to and fro - wiggle
        i. backward and forward, back and forth - waggle,reciprocate,hover,shuttle,rock
      b. from side to side - wobble
    3. over and over - roll

II. without a fixed course,irregular - shake,agitate,wander,flirt
  A. about - dance,play,flutter,swash,thrash

Figure 2-4 Hierarchy of Locus Descriptor Values for Verbs of Motion

Velocity and Acceleration

Velocity and acceleration are closely related motion descriptors. Acceleration is the rate of change of velocity with respect to time, i.e. if one changes velocity one has accelerated to do so. Thus starting to move and slowing to a stop are motions produced by accelerating (or decelerating).

The English language does not seem to distinguish well between velocity and acceleration. To move quickly may mean to begin moving with a rapid acceleration or to be moving at a constant rapid velocity. One word, "sudden", does seem to be associated with acceleration and this was the only term taken as conclusive evidence for this descriptor being present.

Both velocity and acceleration are polar descriptors (i.e. features) and are only specified when below normal or above normal values are required. Thus, velocity is normal unless it is "slow", "gradual", "dragging", etc. or "rapid", "quick", "fast", etc. and acceleration is 0, i.e. motion is with constant velocity unless it is "sudden". There is an additional value required for acceleration, namely whether the effect of the acceleration is to increase or decrease the velocity of the moving object, i.e. whether it is true or positive acceleration or negative acceleration (deceleration). These distinctions can be denoted by using a signed value before the feature to denote its "suddenness", + being "sudden" and - being "gradual", and a signed value after the feature as well to denote the effect upon the object's velocity, + indicating an increase in velocity and - a decrease.

Thus sudden deceleration would be marked as <+acceleration-> and sudden acceleration would be <+acceleration+>, with gradual acceleration being <-acceleration+> and gradual deceleration being <-acceleration->. This doubly specified feature is in keeping with the nature of acceleration as a second level descriptor of a motion parameter. Velocity is of course a simple feature being either <+rapid> or <-rapid> if specified.

To illustrate these motion features consider the sentences of figure 2-5.

Example Usage	Descriptor
he slowed suddenly	<+acceleration->
he slowed gradually	<-acceleration->
he moved slowly	<-rapid>
he moved suddenly	<+acceleration+>
he moved gradually	<-rapid> or <-acceleration+>
he slowly accelerated	<-acceleration+>
he moved rapidly	<+rapid> or <+acceleration+>

Figure 2-5 English Sentences illustrating Velocity and Acceleration Features for Verbs of Motion

If we now enumerate the English verbs of motion which use a velocity or acceleration descriptor in their definition we find the groups of table 2-3.

Table 2-3 Verbs of Motion Marked for Velocity and Acceleration taken from MPD definitions using "Move"

1(a). <+rapid>

Descriptor value	Verb(s)
rapid	agitate
rapidly	scour,shuttle,spin,swing
brisk	bustle-1.1a,play-6
briskly	step,whip-10a,whisk-2
quick	flutter-2,hop,grind,trip,wiggle
quickly	dance,hasten,spring-4,spring-1a,whip,whisk-1
speed	dash,sweep,whirl,zip
speedily	scud
with haste	tear,hurry
hastily	fling,rush
swiftly	fly-6
very swiftly	streak
nimbly	whip-10a,whisk-1

1(b). <-rapid>

Descriptor value	Verb(s)
slow	drag
slowly	crawl,inch-2.1b,delay,hose-6,worm
gradually	draw,edge,forge
with a dragging gait	shuffle

1(c). Unmarked but noted (superfluous marking as normal)

Descriptor value	Verb(s)
at a natural unhurried gait	walk

2. <+acceleration+>

Descriptor value	Verb(s)
sudden	jolt,twitch-1a
suddenly	bolt,dodge,dash,spring-1a
abrupt	jerk
abruptly	swerve

Source and Goal

Source and goal for a motion verb are spacial locations. They are not typically specified in definitions seemingly because they are typically specified in many sentences using motion verbs. This is a general definitional trait, i.e. to leave blank an argument position which in usage is nearly always filled in. By analogy one could consider a definition to be a business form with certain information printed on it beforehand because some values are invariant whereas the most important information, the information whose presence is necessary to complete the form, is simply marked with a predesignated blank line and a descriptor to await a value which will be filled in when the form is used.

It should be noted that often in definitions a case argument position will be prespecified as requiring a value if the verb is to be used. The classic distinction between "transitive" and "intransitive" verbs is one example of this type of prespecification for the theme descriptor.

Predesignated case arguments may also be noted through the use of a broad generic value being specified instead of a specific descriptor. Thus, for source and goal of motion verbs the generic terms place, point, country, locality, and position are used to indicate that these case argument descriptors will be present when the verb is used and in all likelihood assume some more specific value. This "descriptor binding" process is vital to integrating definitional material into the understanding of actual utterances and figures prominently in the task of questioning a speaker to fill-in missing information during a conversation. Thus one knows what questions are appropriate to ask by recalling which generic descriptor categories are specified in the definitional form of a verb and noting whether these descriptors have been "bound" at a sufficiently detailed level or are capable of further specification. For example consider the following hypothetical statement and question pairs for motion verbs having prespecified source and/or goal descriptors.

(a) remove-.1a - to move from one place to another

(b) (REMOVE1A (SOURCE: ONE PLACE)
(GOAL: ANOTHER PLACE))

(c) Speaker1: "I removed the old trunk from the attic."

(d) ((REMOVE (TIME: PAST)
(AGENT: I)(THEME: (THE OLD TRUNK))
(SOURCE: (ONE PLACE) . (THE ATTIC))
(GOAL: (ANOTHER PLACE) . ?????))

(e) Speaker2: "Where did you put it?,"

Initially the dictionary definition (a) was translated into a case argument/descriptor symbolism (b) noting that two arguments SOURCE and GOAL were prespecified as ONE PLACE and ANOTHER PLACE. After the utterance (c) by Speaker1, Speaker2 possessed the augmented definitional structure given in (d). Speaker2 thus has added values for AGENT, THEME and SOURCE and noted with question marks (?????) that no value was given by Speaker1 for the GOAL argument present in Speaker2's definitional form of "remove". This then prompted Speaker2 to question Speaker1 about this missing or "unbound" variable in Speaker2's data structure.

The significance of this exchange is that: (1) A direct translation of a dictionary entry into a case argument data structure is demonstrated, and (2) the application of this data structure to guiding dialogue is shown. This approach explains why Speaker2 did not question the LOCUS (path-of-motion) or INSTRUMENT arguments of the removal event, i.e. these arguments are not specified as being in need of further binding in the definition of "remove". It also explains why Speaker2 chose to ask a where-type question instead of a more general GOAL query such as, "What has become of it?"; i.e. Speaker2 was stating the most precise formulation of Speaker2's knowledge involving ANOTHER PLACE.

Theme

Motion verb definitions do not specify their themes any more than they do their SOURCE or GOAL arguments. Those few themes which are specified are all physical objects, though at least one is not a solid physical object (i.e. air). These are listed in table 2-4.

Table 2-4 Verbs of Motion Pre-Marked for THEME

VERB	THEME
barge	oneself
throw	a lever
fan	air
fiddle	hands,fingers
kedge,warp	ship
smack	lips

Motive

Motive was taken as being indicated either by its absence, as in "aimlessly", "idly", "without a fixed goal", or due to the use of the defining phrase, "so as to ...".

Causation is an important problem-solving concept in artificial intelligence and as such its correspondences in language are important because they provide the basis for imputing motives or implementing goals as stated in English expressions. Thus if we know that the definition of verb "x" involves a motive we may write rules allowing inductive and deductive inference programs to use this knowledge to obtain a result produced by verb "y", where verb "x" causes verb "y".

In addition to providing causality links for problem-solvers, the "so as to ..." causal-relation is used to relate various homographs of a given word to each other. "To thud" is "to move or strike so as to make a thud." Far from being uselessly circular such a definition tells us of the integral relationship between the noun and verb forms of the spelling "thud".

A "thud" is "a dull sound, a THUMP or a BLOW." The previous verb definition tells us how such a "thud" was produced, i.e. something moved (or struck) something else.

An inductive (or deductive) problem solver may utilize these two definitions to form plans in which:

GOAL: (a) It is plausible to infer that if it wishes to make a sound it can move something or cause something to strike something else.

SOURCE: (b) It is plausible to infer that if it has heard a dull sound that this was produced by something striking something else or moving.

EFFECT: (c) It is plausible to infer that if it causes a robot or something to move or strike something else that this will produce a dull sound (hence it may have to avoid this form of action if one of its operational constraints is silence).

Accompanying physical/mental state

Motion, like any other human activity, possesses a set of verbs which are marked upon a "human" emotive descriptor in addition to whatever purely physical descriptors (velocity, acceleration, etc.) they possess. These verbs cannot be fully described without specifying some state-of-mind for their participant or a metaphorical interpretation ascribed to their theme or agent.

Since an emotive component to language can best be described using the semantic differential [Osgood, Suci, and Tannenbaum 1957], I have choosen this framework to detail the characteristics of these connotatively tinged verbs of motion. Using the three primary dimensions of connotative space, best represented by the polar adjective pairs GOOD-BAD, STRONG-WEAK, and ACTIVE-PASSIVE, figure 2-6 presents a spacial chart representing these three axes in a two-dimensional graph. This consequently shows a separation of the verbs into six major clusters.

i. GOOD-STRONG-ACTIVE verbs: waltz

ii. GOOD-WEAK-ACTIVE verbs: dandle

iii. BAD-STRONG-ACTIVE verbs:
(a) tear, thrash, wrench, fling
(b) barge, blunder, lumber, march

iv. BAD-STRONG-PASSIVE verbs: lumber, pound, march

v. BAD-WEAK-ACTIVE verbs: flutter-4, flirt, bolt, rush, fidget, fiddle, toss, freewheel, bustle

vi. BAD-WEAK-PASSIVE verbs:
(a) worm, slink, lurk, skulk, sneak
(b) drag

Figure 2-6 Connotative Classification of Verbs of Motion with Emotive Descriptor Values

It should be emphasized that this is a rating not of the verbs themselves, but of their descriptors. Only those verbs which were marked in their definitions by emotive descriptors were so rated. Since the MPD is weakest in the presentation of reliable connotative meaning, a

deficiency only partially compensated for in the larger Seventh Collegiate and Third International dictionaries through the use of explanatory synonym paragraphs, a considerable argument may be made for deferring the acceptance of these isolated connotative values until a full rating of all the motion verbs has been completed. Certainly words such as "shuffle", "wiggle", "forge", and "crawl" have as strong a connotative meaning as the ones noted in the previous list as possessing connotative descriptors. It appears at this point that connotation will have to be studied with other techniques [Amsler 1969] than dictionary analysis.

Accompanying actions

Among the most interesting aspects of the semantics of natural languages is the decomposition of complex actions into simpler actions. Quite distinct from the hierarchical organization of verbs which is based upon the genus vs. species type of distinction, this decomposition process identifies sub-processes which are either consecutively or concurrently taking place whenever the verb in question is appropriately used. These accompanying actions are a microscopic view of the macroscopic event their verb describes. In order for a verb to qualify as a sub-part of another verb it must describe an action which is in and of itself not equivalent to the whole verb's action, yet it must as well be necessary to the performance of the global verb's action.

Distinguishing sub-parts of verbs from more specialized verbs is more difficult than for nouns. With nouns it is not likely that one will confuse a "door" which is a part of a "house" with a "cottage", which is a type of "house"; but for verbs these distinctions are not typically based on spacial properties of the word meanings, nor even strictly determined by temporal properties. For example, suppose we were to notice a pelican in the sky and utter the statement, "I saw something moving in the sky". The more specific verb, "fly", could be used to replace the phrase, "moving in the sky", resulting in "I saw something flying." This would be a genus/species distinction.

If on the other hand we were to state that we saw "something moving by alternately flapping its wings, soaring, and gliding", we would be further detailing the event (as we did in "flying"), but this time by dividing the event into sub-parts, i.e. "flapping", "soaring", and "gliding". The subtle point here is that "flapping", "soaring", and "gliding" are themselves a species of movement, but their combined usage in the phrase, "alternately flapping its wings, soaring and gliding," clearly indicates that the pelican performed each of these actions as a part of the total moving event we observed. In this case these actions were consecutively and cyclically linked, but they need not have been so, i.e. the pelican might also have been "searching for food" while he was "moving", i.e. a

concurrent action.

The only other distinction to be made here is that the above illustration was based upon a "usage example" of the words "flap", "soar", and "glide" and not upon a definition. Usage example sentences clearly can contain more complex sub-parts than definitions, i.e. there is no word whose meaning is "to move through the air by alternately flapping one's wings, soaring, and gliding in search of food." This particular event has not been granted definitional status in the language. We might define the verb "to pelican" with the above phrase, but only if we did so would or should we regard this particular conglomerate of sub-verbs as having more than passing interest in the construction of a lexical world model.

Verbs possessing accompanying actions in their surface definitions do not necessarily do so in their underlying semantic representations. In some cases surface form will contain redundant verbs whose presence adds nothing to the description of the event. Such is the case with,

remove 2a - to move by lifting or taking off or away

The presence of redundancy is readily seen by considering whether it is possible to "lift off, lift away, take off, or take away" something without moving it. Since certainly "lift" requires motion, and the move component specified no further information than is already specified by "lift" (It could have done so if it specified a path-of-motion as in, "to move forward by lifting, ..."), the use of move is redundant and "remove 2a" could just as easily been defined as,

remove 2a - to lift or take off or away

Once redundancy has been checked for there still remains a question of primacy of the syntactic kernel verb. "Carry" is a good example of a definition in which this primacy does not hold up.

carry 1 - to move while supporting

The critical question here is whether one could define "carry" as "to support while moving." Since this would seem to be an equivalent semantic representation we are forced to concede that either verb could have been the kernel of this definition. This definitely suggests that a strictly taxonomic interpretation of definitions does not always work. Furthermore, any lexical language processing system's representations must allow recognition and generation of either of these surface forms from its deep semantic data structure.

Later this will be seen to be but one form of a far deeper semantic unification principle underlying dictionary definitions. For now the reader must only bear in mind that underlying the meaning of a word bound to a given part-of-speech lies the meaning of a morphological form which is independent of syntactic part-of-speech representation.

Force and Resistance

Force and resistance are closely related but not identical descriptors. The lack of resistance may be accompanied by either great force (as in the case of someone trying to fight their way out of a paper bag) or very little force. Both descriptors are features and may be represented as <+force>, or <-force>; <+resistance>, or <-resistance>.

Table 2-5 Verbs of Motion Marked for Friction and Resistance

1(a). <+force>

Descriptor value	Verb(s)
forcefully	whip
forcibly	blow
force	smash,sweep,tear,whirl,edge
vigorously	hurl,lash
vigor	zip
great effort	labor

1(b). <-force>

Descriptor value	Verb(s)
light	trip

2(a). <+resistance>

Descriptor value	Verb(s)
difficulty, with labor	wade,grind
friction	grind

2(b). <-resistance>

Descriptor value	Verb(s)
freely	freewheel
smoothly	glide,slide
effortlessly	glide,coast
as if on wheels	run
easily	waltz

Steadiness and Continuity

Steadiness and continuity of motion are most difficult to separate because some descriptors clearly imply values for both simultaneously. They are however different. The steadiness of a moving object refers to its spacial position while in motion. To move forward on a given path does not preclude the presence of some "microscopic" perturbation to this motion which marks it as less than flawlessly steady. This type of perturbation may originate from the object itself vibrating or twisting or turning while it moves or it may be a product of the object's interaction with a surface it contacts, as in the case of a car traveling down a bumpy road. Steadiness is a feature and may be denoted by <+steadiness> or <-steadiness>.

Continuity is a temporal phenomenon. An object may be moving with a steady motion, but this motion need not be continuous. The object may move, stop, and move again in an endless series of motions each of which is steady. A good example would be the motion of a mountain climber carrying a bottle of nitroglycerin up a mountainside. While his upward climb may be interrupted many times as as he pauses to insert a piton, whenever he moves he does so very very steadily, avoiding any sudden jars or jerks. The interaction of steadiness and continuity is most apparent in such ambiguous phrases as "irregular" motion. Irregular may indeed also refer to the path of motion itself.

Continuity may be represented as a feature with <+continuity> or <-continuity> representing the two marked values. Motion which is unmarked for continuity is presumably <+continuity> though.

Table 2-6 Verbs of Motion Marked for Steadiness and Continuity

1(a) & 2(a). <+steadiness> and <+continuity>

Descriptor value	Verb(s)
steadily	draw,forge
steady	stream

1(b). <-steadiness>

Descriptor value	Verb(s)
unsteadily	blunder,falter,teeter
turbulently	toss

2(b). <-continuity>

Descriptor value	Verb(s)
repeatedly	bob
-	crawl,fluctuate,jerk,jiggle,twitch

Instrument

There are several types of instruments involved in a motion verb. First, and most commonly, there is the instrument which is doing the moving and which might be thought of as a generalized sense of "vehicle". This includes the usual meaning of cars, planes, ships, etc., but has as its essential part merely the concept of being an instrument-actant, rather than a passive means of travel. Second, there are the non-actant instruments. This too includes some "vehicles", but these do nothing to cause motion, they merely provide a means for carrying passengers, e.g. surf board, raft, sled, etc. Finally, there are the bystander instruments, those which would never be confused with "vehicles" because they are only incidental to the movement. The rope and pitons of a mountain climbing, the poles used in skiing, the face mask or snorkel used in a skin diving. Few definitions actually specify an instrument of any of these types. Motion verbs are far more concerned with the method of motion than the instrument, leaving instruments to be filled in by the

speaker since pre-specifying an instrument severely restricts a motion verb, as can be seen by the few applications the following motion verbs have (figure 2-7).

kedge 1 - to move a ship by hauling on a line attached to a small anchor dropped at the distance and in the direction desired

warp 3 - to move (a ship) by hauling on a line attached to some fixed object (as a buoy, anchor, or dock)

Figure 2-7 Motion Verb Definitions having Pre-Specified Instruments

Medium

Medium is one of the better examples of a case argument specialized to a verb class. Motion verbs rather uniformly allow specification of a medium on, in, or through which the motion occurs. Once again, as with the other standard case argument roles, a pre-defined medium is not common, though I think a number of motion verbs have rather strong "typical" mediums or achieve designation of a medium through alternate means. For example, "walk" does not mention a medium per se, but it does specify "on foot". Since "feet" are an instrument for motion on a surface, most often the ground, this seems a tempting indication of such a specification. However, some insects have "feet" and "walk" on the surface of the water or other non-specific "surfaces", so the lexicographers may have been cleverer than we suspect in avoiding specification of a too restrictive medium. "Kedge" and "warp" are move verbs specified for a "ship", it is tempting to say that the medium "water" is present, though I guess one could "kedge" a spaceship through a vacuum. Table 2-7 gives those motion verbs whose media were pre-specified.

Table 2-7 Verbs of Motion Marked for Medium

Medium	Verbs
air	fly-1,fly-2,fan-2
current	drop-9 (e.g. "we dropped down the harbor and were soon steering south")
fluid	float-2
ground	crawl-1.1a
soft mud	squelch
water	ride-2 (e.g. "ride at anchor"),paddle

Orientation

Orientation refers to the specialization of the motion verb because of the shape or alignment of the moving object(s). Ordinarily orientation would be unmarked and the motion could take place in any direction without further specification of the alignment of the moving objects to the path. Occasionally, a type of movement is defined not just in terms of what the motion itself consists of, but the position or orientation of the theme or agent during the motion, e.g.

sidle - to move sideways or side foremost

Thus in the verb "sidle" the moving entity must be oriented such that the direction of motion is at a right-angle to "front". This seems to require the entity be orientable (i.e. have a "front"). Note too that orientation is distinct from path. One can sidle in a circle, for instance.

Minimum Number of Objects (Moving and Background)

Finally, the move verbs often require certain minimum numbers of objects to be involved for their actions to take place. By default these numbers would be one (moving) and zero (background). Table 2-8 gives some exceptions:

Table 2-8 Motion Verbs with Minimum Requirements on Numbers of Moving and Background Objects

No. Objects (Moving, BackGround)	Verbs	Specified Values
(2,)	carry-1.1a	("that which is supported")
(1, 1) or (2,)	dodge-1	("the object dodged")
(2,)	diverge-1a	("the common point diverged from")
(2,)	bulldoze	("that which gets bulldozed")
(2,)	remove	("that which gets lifted, taken off, or away")
(, 1)	work-3.19a	("the other part")
(, 1)	pass-1.3	("the thing passed")
(, 1)	dandle	("one's arms, one's knee")
(, 1)	coast-2.2	("bicycle, sled")
(, 1)	crawl-1.1a	("the body, the ground")
(, 1)	approach-1.1a	("thing moved nearer to")
(, 1)	throw	("lever", "clutch or switch")
(, 2)	warp	("line; buoy, anchor, or dock")
(, 2)	kedge	("line and anchor")

2.5 Technique I - Analysis of the Case Argument Pattern of a High Level Verb

The procedure used to analyze "move" in this chapter is summarized in figure 2-8.

Steps:

(1) Secure from a concordance of the dictionary all definitions which have as their kernel the verb to be studied.

(2) Determine the SENSE of this verb on the basis of the full definitions provided by the Pocket, Seventh Collegiate, and Third International dictionaries.

(3) Eliminate from the definitions selected in step 1 all verbs whose sense is other than the one being studied, forming a validated set of definitions.

(4) Segregate the descriptors which occur in the validated definitions into sets using componential analysis techniques.

(5) Analyze each descriptor set to determine the attribute being described.

(6) Rewrite the definitions in terms of the attributes and values of the descriptors, checking to see that these descriptors are independent.

(7) Produce the final verb and descriptor pattern

Figure 2-8 Analysis of the Case Argument Pattern of a High Level Verb

CHAPTER III TAXONOMIC DEVELOPMENT

The analysis of "move" carried out in chapter 2 only required examining one page of data from the Olney concordance index, that which contained the entry for "move". However, the original intent of this project had been to explore the extended taxonomic descendants of nodes in the dictionary and to develop a taxonomy of a representative portion of the English language. A second procedure (figure 3-1) was formulated to use a concordance index to develop the taxonomic structure of a semantic domain.

Steps:

(1) Choose a semantic domain to be developed, selecting a "best" representative verb (or noun) or set of verbs (or nouns) from the domain.

(2) Using the taxonomic concordance secure all the definitions using these "best" representative(s) as their kernel(s).

(3) Validate the senses of these usages and form an EXPAND-list of verbs (or nouns).

(4) For each element of the EXPAND-list repeat steps 1-3 until no element on the EXPAND-list of any verb (or noun) has any USAGES in the dictionary not already on the EXPAND-list.

(5) Form a data base of the disambiguated verbs (or nouns) and their associated EXPAND-list descendants.

(6) Computationally connect the data base and print out its structure as a tree, detecting and enumerating any LOOPS which occur.

(7) Resolve any loops occuring in step 6 by merging nodes of the tree to form synonym senses or by reevaluating the disambiguation phase of step 5 for those elements.

(8) Display the final structure of the several hundred or thousand nodes of the connected tree.

Figure 3-1 Technique II: Tree Development

3.1 Case Studies in Tree Development

The tree development procedure relied upon many of the same steps as the verb argument development procedure (section 2.5). Words had to pass a series of tests for acceptable taxonomic relatedness including a disambiguation test in which they were validated as descendants of the correct sense of the word being studied at each level. However, the tree development procedure made no attempt to componentially analyze the descriptors of a set of definitions. The "move" verbs were studied using the tree development procedure of figure 3.1.

The taxonomic structure of "move" turned out to be quite shallow. Most verbs defined using "move" were either terminal or terminated within one further level. "Move" thus seemed a poor choice for understanding how the dictionary represented deep taxonomies of words.

It seemed that taxonomic development would surely be well illustrated for some nouns. A noun which had been suggested by "move", namely "vehicle" was finally selected as a candidate for taxonomic development.

The "vehicle" taxonomy turned out to be surprisingly rich and cleanly delineated. It was rich in terms of the number of levels and non-terminals. It was cleanly delineated because there were few ambiguous nouns in the tree and because it was very nearly a true taxonomy [Kay 1971]. Starting at essentially one node, "vehicle", a tree of great breath and considerable depth was uncovered (figure 3-2).

While I could have shuffled these data into an appendix I felt it sufficiently worthwhile to include in the body of the text at this point. The reason is that this is a small tree. It covers six pages of data, but represents no more than a snapshot of one scene in the dictionary landscape. It is important to get a feel for just how big a taxonomic development task was obviously needed to survey the whole landscape of dictionary taxonomics.

(VEHICLE
(AMBULANCE)
(AUTOMOBILE
(BLOODMOBILE - an automobile equipped for collecting blood from doners)
(BROUGHAM - a coupe automobile ; esp. one electrically driven)
(COACH - a closed 2-door automobile
(CHARABANC - <Brit.> - a sight-seeing motor coach))
(COMPACT - a small automobile)
(CONVERTIBLE - an automobile with a top that may be lowered or removed)
(COUPE - a 2-door automobile with an enclosed body and separate luggage compartment
(BROUGHAM - a coupe automobile ; esp. one electrically driven)
(EIGHT - an 8-cylinder automobile)
(FLIVVER - a small cheap usu. old automobile)
(HACKNEY, HACK - an automobile kept for hire)
(HARDTOP - an automobile resembling a convertible but having a rigid top)
(HOT ROD - an automobile rebuilt or modified esp. for high speeds)
(JALOPY - a dilapidated automobile)
(LANDAU - an enclosed automobile with a top whose rear quarter can be opened or folded down)
(PHAETON - an open automobile with two cross seats)
(ROADSTER - an open automobile with one cross seat)
(SEDAN - an enclosed automobile usually with front and back seats
(BROUGHAM - a sedan having no roof over the driver's seat)
(LIMOUSINE - a large luxurious often chauffeur-driven sedan)
(SALOON - <Brit.> SEDAN))
(SIX - a 6-cylinder automobile)
(STATION WAGON, WAGON - an automobile having an interior longer than a sedan's, one or more folding or removable seats to facilitate trucking, and no separate luggage compartment)
(TAXICAB, TAXI, CAB - a chauffeur-driven automobile for hire that usually carries a device for automatic registering of the fare due)
(BICYCLE, BIKE, CYCLE, WHEEL
(TANDEM - a bicycle for two persons sitting one behind the other))
(BUCKBOARD - a 4-wheeled vehicle with a springy platform carrying the seat)

(BUS, COACH, OMNIBUS
(JITNEY - a small bus that carries passengers over a regular route according to a flexible schedule)
(SHUTTLE BUS - a bus traveling back and forth over a short route)
(TROLLEY BUS - a bus powered by electric power from two overhead wires))
(CARRIAGE - a wheeled vehicle
(BAROUCHE - a 4-wheeled carriage with a high driver's seat in front and a folding top)
(BROUGHAM - a light closed horse-drawn carriage with the driver outside in front)
(BUGGY - a light carriage)
(CAB - a light closed horse-drawn carriage)
(CABRIOLET - a light 2-wheeled one-horse carriage)
(CHAISE - a two-wheeled carriage with a folding top)
(COACH - a closed 2-door 4-wheeled carriage with an elevated outside front seat for the driver
(FLY - a horse-drawn public coach)
(STAGECOACH - a coach that runs regularly between stations))
(COUPE - a closed carriage for two persons inside with an outside seat for the driver in front
(CABRIOLET - a convertible coupe))
(DOGCART - a light one-horse carriage with two seats back to back)
(EQUIPAGE - a horse-drawn carriage usu. with its attendant servants)
(FLY - <Chiefly Brit.> - a light covered carriage or cab)
(GIG - a light 2-wheeled carriage)
(HACKNEY - a carriage (or automobile) kept for hire)
(LANDAU - a 4-wheeled carriage with a top
divided into two sections that can be lowered, thrown back, or removed)
(HANSOM - a 2-wheeled covered carriage with the driver's seat elevated at the rear)
(PERAMBULATOR - <Chiefly Brit.> a baby carriage)
(POST CHAISE - a 4-wheeled closed carriage for rapid travel)
(RIG - a carriage with its horse or horses)
(SURREY - a 4-wheeled 2-seated horse drawn carriage)
(TANDEM - a 2-seated carriage with horses hitched tandem)
(TRAP - a light 2-wheeled or 4-wheeled one horse carriage on springs)
(TROLLEY, TROLLY - a wheeled carriage running on
an overhead rail or track (as on a parcel railway in a store))
(VICTORIA - a low 4-wheeled carriage with a
folding top and a raised seat in front for the driver))

(CART, TUMBRIL, TUMBREL
(BARROW - a cart with boxlike body and two shafts for pushing)
(CHAISE - a light carriage or pleasure cart)
(DRAY - a strong low cart for carrying heavy loads)
(OXCART - a cart drawn by oxen)
(PUSHCART - a cart or barrow pushed by hand)
(WAGON - a child's 4-wheeled cart))
(CAR - a vehicle moved on wheels
(BOXCAR - a roofed freight car usu. with sliding doors)
(CABLE CAR - a car moved; along rails by an endless cable operated by stationary engine; or along an overhead cable)
(CABOOSE - a car usu. at the rear of a freight train for the use of the train crew and railroad workmen)
(COACH - a railroad passenger car esp. for day travel)
(CRUISER - a police car equipped with radio to maintain communication with headquarters)
(DINER - a railroad dining car)
(FLATCAR - a railroad freight car without permanent raised sides, ends, or covering)
(GONDOLA - a railroad car with no top designed for bulky materials)
(HANDCAR - a small 4-wheeled railroad car propelled by a hand-operated mechanism or by a small motor)
(PULLMAN - a railroad passenger car with specially comfortable furnishings; esp. : one with berths)
(RATTLETRAP - an old car which is rickety and full of rattles)
(SLEEPER - a railroad car with berths for sleeping)
(TRAM - a boxlike car running on a railway (tramway) in a mine or a logging camp)
(TROLLEY - a passenger car that runs on tracks and gets its electric power through a trolley)
(VAN - a closed railroad freight or baggage car)
(WAGON-LIT - a railroad sleeping car))
(CHARIOT)
(EXPRESS)
(FLOAT)
(FOUR-IN-HAND)
(HALF-TRACK)
(HEARSE)
(JINRIKISHA)
(LOCOMOTIVE, ENGINE)
(MOTORCAR, AUTOMOBILE see AUTOMOBILE)
(MOTORCYCLE, CYCLE)
(MOTORTRUCK, LORRY
(PATROL WAGON - an enclosed police wagon or motortruck for carrying prisoners)
(TRACTOR - a motortruck with short chassis for hauling a trailer)
(VAN - an usu. enclosed wagon or motortruck for moving goods or animals))

(MOTOR VEHICLE)
(PEDICAB)
(PHAETON)
(TRACTOR)
(TRAILER
    (CAISSON - an ammunition chest mounted on two wheels and joined as a trailer to form an ammunition wagon)
    (TRUCK-TRAILER - a combination of a trailer and the truck that draws it))
(TRANSPORT)
(TRICYCLE
    (VELOCIPEDE - a child's tricycle))
(TROLLEY)
(TROLLEY BUS)
(TRUCK - a vehicle (as a small flat-topped car on small wheels, a 2-wheeled heavy wagon or automobile) designed for carrying heavy articles
    (BOOKMOBILE - a truck that serves as a traveling library)
    (DOLLY - a small wheeled truck used in moving heavy loads)
    (MOTORTRUCK - an automobile truck for transporting freight)
    (PICKUP - a light truck with open body and low sides)
    (TRUCK-TRAILER - a combination of a trailer and the truck that draws it)))
(TUMBRIL, TUMBREL)
(SHUTTLE)
(SLED
    (BOBSLED - a short sled usu. used as one of a joined pair; a compound sled formed of two bobsleds and a coupling)
    (TOBOGGAN - a long flat-bottomed light sled made of thin boards curved up at one end))
(SLEDGE)
(SLEIGH
    (CUTTER - a light sleigh))
(SPACESHIP)
(STREETCAR, TRAM)
(SUBWAY, METRO)
(SULKY)
(UNICYCLE)
(VELOCIPEDE)
(WAGON
    (BANDWAGON - a wagon carrying musicians in a parade)
    (CAISSON - an ammunition chest mounted on two wheels and joined as a trailer to form an ammunition wagon)
    (CART - a two-wheeled wagon)
    (CHUCK WAGON - a wagon equipped with a stove and provisions for cooking)
    (CONESTOGA - a broad-wheeled covered wagon usu. drawn by six horses and formerly used esp. for transporting freight across the prairies)
    (FLY - a horse-drawn public coach or delivery wagon)

(LORRY - a large low horse-drawn wagon without sides)
(PATROL WAGON - an enclosed police wagon or motortruck for carrying prisoners)
(PRAIRIE SCHOONER - a covered wagon used by pioneers in cross-country travel)
(RUNABOUT - a light wagon, automobile or motorboat)
(VAN - a usu. enclosed wagon or motortruck for moving goods or animals)
(WAGONETTE - a light wagon with two facing seats along the sides behind a cross seat in front)
(WAIN - a usu. large heavy farm wagon))
(WHEEL see BARROW)
(WHEELBARROW)
(WRECKER)

Figure 3-2 "Vehicle" Taxonomy from Merriam-Webster's New Pocket Dictionary

CHAPTER IV DICTIONARY DATABASES

The two preceding studies were entirely based upon words presented with indications of their status and parts of speech. The problems all of these context-less treatments led to was becoming apparent. The meaning of a word does not exist apart from its context. The meaning of the text of a definition is more than the sum of the words in the definition. The meaning lies within the human mind and is in effect unlocked by the reading and interpretation of the information in the textual definition. To study meaning one must study words in context.

Each of the prior experiments with the Pocket Dictionary had required several months of effort. Most of this time was spent in transcribing dictionary definitions because of the lack of appropriate computational facilities. The basic research tools used for those analyses were the listings of the Olney and taxonomic concordance indexes to the dictionary, consisting respectively of some 10,000 and 2,000 pages of computer output. Data found in these listings still had to be transferred into computer-readable form through interactive CRT keyboarding and editing so that it could then be sorted and annotated with sense meanings and definitions for further on-line use.

The operations of keyboarding dictionary definitions, editing these definitions to introduce semantic analysis information, and sorting the resultant data, were of course the basic component capabilities of a good data base management system. Use of such a system would have saved one or two months of effort on each domain studied and provided computer assembly of the required full texts of definitions on the basis of their constituent words. Subsequent sorting and processing of these definitions would also have occurred more rapidly if human on-line interaction could have been reduced to the minimum required for cognitive tasks such as disambiguation -- rather than typing, editing, and writing utility programs to perform minor transformations on the data.

Finally, difficulties specifically related to working on separate experimentally selected subsections of the dictionary began to appear and implied a serious integration problem should these separate sections ever need merging into larger analyses. A study of two or three isolated semantic sub-domains in the dictionary does not involve very much redundancy in terms of re-evaluating interconnected nodes in semantic networks; but as the number of sub-domains analyzed increases, their overlap is an inevitable problem. The task of retaining all of the information derived from prior analysis of a node in one sub-domain, such that it will be available to analyses involving other domains, makes imperative the unification of the source data base and the annotated data

base containing the prior analyses.

All of these reasons seemed to imply a data base would have to be created containing the entire text of the dictionary. It would have to provide for extraction of word usages in context as well as the definitions by main entries. It should also provide for the storage of additional information such as would be necessary to distinguish the separate senses of a usage as well as its correct part of speech when these were obtained. The only practical way this could be done was to abandon the concordance indexes and return to the original MPD dictionary tapes.

4.1 Preparations for Creating Dictionary Databases

In order to fit the dictionary into a database system, it was necessary to segment the full text of the dictionary into separate collections of data on the basis of the parts of speech of the main entries. This was due to the size limitations of the computing facilities available, but additionally served to make future retrievals simpler without harming the taxonomic connectivitity of the data. Taxonomies are limited to relationships between words having the same part of speech, hence databases containing all the definitions of only one part of speech nevertheless are complete taxonomic data sets.

A by-product of dividing the lexicon into separate databases by parts of speech was a tabulation of the frequencies of dictionary main entries by their parts of speech and complete lists of main entries belonging to any one part of speech.

4.1.1 Dictionary Part-of-Speech Information

There are ten standard grammatical "Parts Of Speech" (POS) in the MPD, namely: nouns (N), verbs (VB), adjectives (AJ), adverbs (AV), prepositions (PP), pronouns (PN), conjunctions (CJ), interjections (IJ), and the definite and indefinite articles (DA,IA). Extensive examples of these are given in appendix 3. Additionally there are five classes of suffixes: noun suffixes (NS) such as "-ade", "-age", "-ry", and "-ship"; verb suffixes (VS) such as "-es" and "-s"; adjective suffixes (JS) such as "-ally", "-al", "-ment", and "-ways"; adverb suffixes (AS) such as "-ly" and "-ward"; and non-specific suffixes (SF) such as "-est"; There is also one class of prefixes (PF) including "anti-", "sub-", "un-" etc. and one class of "combining forms" (CF) containing "multi-" and "self-". One entry is classified as a transitive verb (VT), i.e. "vail" and another as an imperative verb, i.e. "vide", but the former is probably a minor error left-over when the Seventh Collegiate dictionary was pruned down to make the MPD. A couple of hundred lexical main entries have no specified part of speech, mostly being noun phrases such as "master of ceremonies" or

"chief petty officer".

When discussing the frequency of occurrence of parts of speech a problem occurs. A word is often referred to as "having a part of speech", but in fact words can "have" several parts of speech. What then does one refer to as the frequency of the parts of speech in a dictionary? If one selects "pure" parts of speech, i.e. words which only occur as main entries of one part of speech -- then one has lost many occurrences. If one does not make this distinction, and only requires each word to occur as an element of this part of speech, then one is counting many words more than once. To treat this matter adequately, I will deal with each method of counting "part of speech" forms separately. In both cases the same data will be used, though in one case the data will be allowed to be sampled more than once, whereas in the other unique parts of speech such as "noun-verbs" will be referred to.

In ranked order there 3 parts of speech which collectively account for 98% of all the main entries in the MPD: nouns, verbs and adjectives. The low frequency of adverbs in the MPD (less than 2%) demonstrates that they are not considered in need of separate definitions. The MPD contains only 54 adverbs ending in -LY and only 495 others not ending in -LY (demonstrating also that adverbs formed with the -LY suffix are not generally considered to require a separate definition from that of the the noun, verb or adjective main entry from which they derive).

No matter what approach to the frequency of parts of speech is taken, nouns are always the most common part of speech. If we admit only "pure" nouns, the statistics indicate 12,466 nouns vs. 3805 adjectives and 2779 verbs; If we allow noun-verbs, noun-adjectives, adjective-verbs, and noun-verb-adjectives and combinations of these with the other "functional" parts of speech, then there are 15,166 nouns, 5029 verbs, and 4677 adjectives. Thus in either case the number of nouns is roughly 3 to 4 times the number of adjectives or verbs.

After the N,VB,AJ trilogy follow adverbs (AV) as the 4th most frequent part of speech of dictionary main entries. There are 332 "pure" adverbs and this number only rises to 549 with the addition of mixed parts of speech, giving adverbs roughly 1/10 the frequency of adjectives or verbs and 1/30 the frequency of nouns.

Following adverbs come the "functional" parts of speech, i.e. words which are normally treated as not bearing the specific "content" of a text, but only serving to provide grammatical structure to the other content-bearing words. Thus we have 51 "pure" prepositions, 45 "pure" pronouns, 28 "pure" conjunctions, and 19 "pure" interjections and that's it. If we allow multiple parts of speech then we have 120 prepositions, 82 pronouns, 55 conjunctions, 22 interjections, and the indefinite and definite articles "a,an" and "the".

The remaining "parts of speech" listed in the dictionary are actually prefixes and suffixes, being individualized to 36 Noun-Suffixes, 26

Adjective-Suffixes, 10 Verb-suffixes, 4 Adverb-Suffixes, 6 prefixes, 1 "suffix", and an unused category, "combining form". (An erroneous use of the "transitive verb" (VT) POS occurs 3 times; and a VP POS code for "verb phrase" is also used twice). These statistics are listed in table 4-1 and the full listings of the adverbs, prepositions, pronouns, conjunctions, interjections, suffixes and prefixes is given in appendix 3.

Table 4-1 Part of Speech Statistics for the Merriam-Webster New Pocket Dictionary (Non-disjoint).

Rank	Mixed+Pure POS	Frequency	%	
1	N	15,166	66%	**********
2	VB	5,029	22%	*****
3	AJ	4,677	20%	*****
4	AV	549	2%	*
5	PP	120	-	.
6	PN	82	-	.
7	CJ	55	-	.
8	NS	36	-	
9	JS	26	-	
10	IJ	22	-	
11	VS	10	-	
12	PF	6	-	
13	AS	4	-	
14	VT	3	-	
15	DA	2	-	
16	VP	2	-	
17	SF	1	-	
18	CF	0	-	
		22,817	100%	

4.1.2 Semantic Ambiguity

While the elimination of syntactic ambiguity can be facilitated by segregating the definitions into separate data sets on the basis of their parts of speech, there remains the problem of distinguishing the semantic senses in which words sharing the same part of speech are being used.

To understand what this entails, consider the entry (figure 4-1) for "table" taken from the taxonomic concordance index.

```
TABLE                    64 OCCURRENCE(S) N =56 VB=4 AJ=2 PP=2
N  TREE OF   55 MEMBERS
    173  1.18A02 FORM          -0    . 0A08 LAZY SUSAN
      4  1. 5A02 BOARD         -0    . 0A08 WAITRESS
      4  1. 4A02 BOARD          0    . 2A09 DRAWER
      0   . 2A02 BENCH          0    . 0A09 DINNERWARE
      0   . 1A02 DESK          -0    . 0A09 LAPBOARD
     -0   . 0A02 CARREL         0    . 0A09 BILLIARDS
      0   . 1A02 TIMETABLE      2  1. 2A10 CHART
     -0   . 0A02 WORKTABLE      1  1. 2A10 CUSHION
      0   . 2A02 ALTAR          0  2. 2A10 POOL
     -0   . 0A03 ESCRITOIRE    -0    . 0A10 CRUET
     -0   . 0A04 CHABLIS        0    . 0A10 SIDEBOARD
     -0   . 0A04 NUTPICK        0    . 0A11 FOOTMAN
     -0   . 0A04 CLARET        -0    . 0A11 FINGER BOWL
      0  3. 0A04 ROSE           0    . 0A11 CASTER
     -0   . 0A04 BEAUJOLAIS    -0    . 1A11 VALANCE
     -0   . 0A05 CHIANTI       64  1. 4A12 TABLE
     -0   . 0A05 SAUTERNE      -0    . 0A12 SWEET CORN
     -0   . 2A05 CHRONOLOGY    -0    . 0A13 EPERGNE
     -0   . 0A05 BURGUNDY      -0    . 0A13 NAPKIN
      1   . 1B06 WAITER         0    . 0A13 TABLEWARE
      0   . 2A06 BOOTH         -0    . 0A14 MARQUETRY
      0   . 2B06 CRYSTAL       -0    . 0A14 TUREEN
     -0   . 0A06 NAPERY        -0    . 0A15 CROUPIER
      0   . 0A07 TABLECLOTH     1  1. 3A16 LEAF
     -0   . 0A07 ORTOLAN       -0    . 0A18 SPIRITUALISM
     -0   . 0A08 CHAFING DISH   0    . 0A19 PEWTER
      0   . 0B08 CENTERPIECE    4  1. 1A21 SILVER
```

Figure 4-1 "Table" noun entry from Taxonomic Concordance Index

For example, the occurrences of "table" include two distinct semantic fields, field 1 includes words such as "timetable", "chronology", and "form" (as in "racing form"), while field 2 contained the "furniture"

items ("desk", "carrel", "escritoire", "lapboard", etc.). Additionally a syntagmatic use of "table" as in "table wine" brought in many words such as "chablis", "claret", "rose", "beaujolais", "chianti", "sauterne", and "burgundy". No matter what organization of these entries was presented, it would still be the case that the text of their definitions would have to be consulted to semantically disambiguate them.

The solution to this was to design a database representation distinguishing each of the senses of a word and to use this notation throughout the database. A separation of these two forms led to labeling the conventional representation of a word as its SPELLING, and the disambiguated form as a WORD SENSE MEANING or just SENSE of the word.

4.1.3 Composition of MPD entries

Each definition in the Pocket Dictionary can be uniquely identified by a "sense number" comprising some combination of its homograph integer (if any), a "decimal" point, the definition integer, and a suffixed letter if subsense divisions are present.

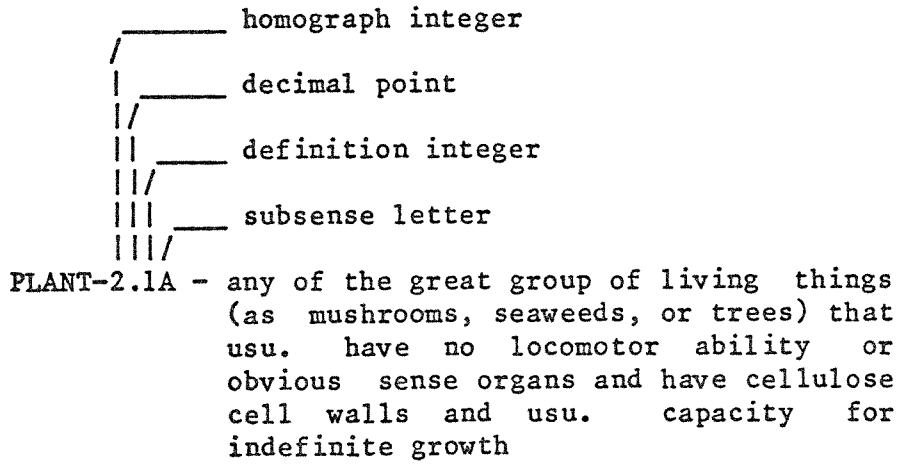

Figure 4-2 Representation of a Pocket Dictionary Word Sense

Homographs are defined as main entries with exactly the same written form as another that appears either before or after it. Occasionally these are words with the same part of speech but different etymologies, but usually they are the same word manifest as different parts of speech, e.g., "act" as a noun and as a verb. They are marked in the dictionary by super-scripted integers preceding such main entries which have more than one identical occurrence.

Following the pronunciation and part of speech indication of all main

entries which have more than one sense, full-sized boldface integers precede each major definition passage.

Subsenses are indicated in the dictionary by the use of colons (always in boldface type) preceding definition passages. As a convenient means of labeling such subsenses the letter A was assigned to the definition following the first colon, B the second, C the third, etc. More than two subsenses are infrequent, more than three are very rare. Whether or not a main entry has definition integers, colons are always present, and the computational representation was always given at least an A subsense letter.

In the absence of a homograph integer the decimal point begins the identification sequence. In the absence of any definition integer a zero is used. (The convention of using zero when there are no sense numbers allows rapid identification of main entries whose definitions are unambiguous according to the Merriam-Webster lexicographers.) In the absence of a subsense, the letter A is appended.

Thus the occurrence of .0A for a definition identification indicates that there are no homographic forms of the word or word phrase being defined and that it has only one major sense in the lexicon. A 1.1A identification would imply the existence of at least one homograph, and at least a second major sense; i.e., both 2.1A (or 2.0A) and 1.2A are also present.

Definitions in the Pocket Dictionary are ordered by their usage historically, and consequently homograph numbers occur without fixed association with part of speech. Where two parts of speech are based upon the same root form the homographs appear alphabetically by the part of speech with which they are associated, e.g., adjective before noun, noun before verb, etc.

A definition of a noun, verb, or adjective is either a textual definition, a synonymous cross-reference, or a usage note.

Usage notes are rare in definitions and restricted to word senses of a highly functional (rather than content) role. Such usage notes occur in the definitions of the verbs BE-.4a, .5a, .6a, and .7a; COULD-.0b, HAVE-1.15a, DO-.23a, and of other functional parts of speech such as the indefinite article A-2.0b, preposition IN-1.5a, and 1.6a; conjunction OR-.0a, pronoun ONE-2.3a, interjection LO-.0a, and noun plural suffix -ES-1.1a; e.g.,

> to (prep) 1.18a - used for marking the following verb as an infinitive and often used by itself at the end of a clause in place of an infinitive suggested by the preceding context.

Definitions by synonymous cross-reference serve both as the listing of a synonym, and as a one-word definition. These are described in the Pocket Dictionary as not being definitions, but "an indication that a definition at [their] boldface equivalent can be substituted at the place where the small capitals are used." The difficulty this raises is that the boldface equivalent may itself be defined by another -- often reciprocating — synonymous cross-reference, resulting in a loop. Sometimes, as in the definitions of figure 6-4, the circularity is an integral part of the process of defining a primitive semantic concept; but in cases such as,

admit-.2a - PERMIT, ALLOW

allow-.3a - ADMIT, CONCEDE,

and

account-2.1a - CONSIDER < I account him lucky >

consider-.1a - THINK, PONDER
consider-.2a - HEED, REGARD
consider-.3a - JUDGE, BELIEVE

in which no text appears, there would seem to be an error.

The texts which define nouns are noun phrases in syntactic structure. Textual definitions of verbs begin with an infinitive verb, usually preceded by "to", and consist of an "infinitive verb phrase".

Textual definitions are usually given in terms of a more general word and associated distinguishing properties which specify the restrictions on the more general word and make its use appropriate in a definition of the main entry involved.

For nouns and verbs it is generally possible to identify a single word or set of conjoined words which are the syntactic "kernel" of the definition. Such a word or words may be loosely described as being in an "ISA" ("is a") relationship with the main entry subsense in whose definition they appear. For example, the word "plant" in the definition of "tree-1.1a",

tree-1.1a - a woody perennial plant usu. with a single main stem and a head of branches and leaves at the top

or the conjoined words "bed" and "shelter" in the definition of the noun "nest-1.1a",

nest-1.1a - the bed or shelter prepared by a bird for its eggs and young

or the word "make" in the definition of the verb "beautify-.0a",

beautify-.0a - to make more beautiful

or the conjoined words "move" and "pass" in the definition of the verb "fly-1.1a",

fly-1.1a - to move in or pass through the air with wings

In most cases, the kernel of a definition itself occurs as a main entry in the dictionary. Thus, the property of closure exists for the ISA relation between main entries and their definition kernels. Main-entry/kernel pairs are created by determining what homograph number, sense, and subsense are to be attributed to the kernel. Upon the discovery of this information for each definition's kernel words, the resulting paired defining and defined words can be used as the basis for constructing a taxonomy of the vocabulary of the entire dictionary. Since definitions can contain "or" between their kernel terms the taxonomy can have more than one upward path, a property described as being "tangled" or "multi-rooted".

Given both a means of uniquely representing the word senses of spelling forms and separation of the dictionary into data sets containing only definitions from one part of speech, it was possible to create databases from dictionary definitions from which one could selectively extract all the noun or verb word senses whose definitions contained a given spelling.

There were many by-products from the loading of the dictionary definitions into databases. These included frequency data on the number of occurrences of each spelling form used in the dictionary and frequency data on the number of different senses each spelling form of a given part of speech had as a main entry (a measure of the "ambiguity" of a word). Some of these will be discussed in Chapter 5. The more important result was that for the first time a full-scale random-access DBMS containing the dictionary definitions had been constructed and was available for retrieval of definitions satisfying simple Boolean restrictions over combinations of occurring words.