

## Part-of-Speech Implications of Affixes

by Lois L. Earl,\* Lockheed Missiles and Space Company, Palo Alto, California

*This paper describes a systematic investigation of the extent to which the part of speech of words can be identified from their prefixes and suffixes. The results indicate that it is possible to determine, with 95 per cent accuracy, the inclusive part of speech of an affixed word from a consideration of its prefixes, suffixes, and length. By "inclusive" parts of speech we mean a string that will include all of the parts of speech assigned by both dictionaries considered but that may include one or two extraneous parts of speech. The extra parts of speech will differ according to the class of words, as adjectives may have an extra part-of-speech "noun" or "adverb," while nouns may have an extra part-of-speech "verb." The part-of-speech implications of seventy-two prefixes and of eighty-seven suffixes are given.*

In a highly inflected language, the structure of a word is indicative of its syntactic role. A relationship between form and part of speech might also be expected in English, a language not highly inflected but closely related to more inflected languages. Such a relationship was noted by J. Dolby and H. Resnikoff,<sup>1</sup> who show that a high percentage of a set of words called "elementary words" (roughly equivalent to the set of one-syllable words) can be used as nouns, adjectives, or verbs, while a high percentage of the remaining multi-syllable words can be used only as nouns or adjectives. If this relation can be regarded as a general rule, and if subrules can be developed to cover the considerable number of exceptions to the general rule, it will be possible to identify part of speech by algorithm. Intuitively, it would be expected that prefixes and suffixes are key structural elements; this expectation is reinforced by the structure of the European languages whose beginnings and endings indicate the grammatical properties of words.

A logical step in an effort to classify words from their structure is to examine the relationship between the affixes of words and their part-of-speech possibilities as listed in a dictionary. The part-of-speech information from *The Shorter Oxford Dictionary*<sup>2</sup> and from the *Merriam Webster New International Dictionary*<sup>3</sup> was recorded on magnetic tape. A computer was used to correlate the affixes of words with their part-of-speech possibilities. A total of 73,582 words was recorded, but, of course, not all of these words contain affixes.

The first problem encountered is that of selecting a list of affixes. Two sets of affixes have been selected, the first being the operationally defined affixes derived from dictionaries solely on graphemic evidence<sup>4,5</sup> and the

second being all "beginnings or endings" listed in *A Dictionary of Modern English Usage*<sup>6</sup> which were not already on the first list. Both lists are given in Table 1.

TABLE 1  
AFFIXES SELECTED FOR CORRELATION  
A. AFFIXES SET I

Prefixes			Suffixes		
a	deni	mi	a	ful	less
ab	dis	mis	able	ia	let
ac	e	ob	age	ial	lock
ad	ec	out	ah	ic	man
al	em	photo	al	ie	ment
an	en	pro	an	ier	ness
ana	epi	re	ant	ile	o
ap	ex	sa	ar	in	ock
apo	hy	sac	ard	ine	on
at	hyper	sub	at	ion	or
auto	i	sun	ate	is	ot
be	im	tele	ee	ish	ow
ca	in	trans	el	ism	tation
circum	incon	un	en	ist	ue
com	inex	uncon	ent	ite	um
con	inter	vi	eon	ium	ure
cy	irre	vul	er	ive	us
de	lu		et	land	ward
demo	ma		ey	ler	y

B. AFFIXES SET II

Prefixes			Suffixes		
aero	deci	semi	ae	genic	lily
air	demi	super	al	iana	logy
bi	for	vice	as	ible	phile
brain	fore	yester	cy	ise	th
by	hecto		eer	ist	ty
bye	homo		ern	ity	ular
centi	non		est	ix	valent
co	para		ette	ize	ways
deca	self		ex	latry	worthy

\* This work was supported in part by the U. S. Navy (Office of Naval Research); the computer time was supported by the Independent Research Program of Lockheed Missiles and Space Company. The author wishes to thank Dan L. Smith, who wrote the computer program referred to in this paper, and J. L. Dolby and H. L. Resnikoff, who have acted as consultants to Lockheed on the ONR contract.

The inflectional suffixes *ed* and *ing* and the adverbial *ly* were not considered in this study because they have well-recognized implications. It is believed that the number of words ending in *ed*, *ing*, or *ly* whose parts of speech differ from the expected is small enough so that such words can be listed as exceptions.

The second problem encountered is that of determining when an affixing unit is acting as an affix in a given word, as *re* is a prefix in *react* but not in *read*. This problem is complicated by an uncertainty as to what the words "prefix" and "suffix" signify. It is difficult to determine from the definitions currently in use to what unit an affix is expected to attach (word, stem, or syllable), to what extent the function of an affix is semantic, and to what extent the affix should indicate phonetic syllabic boundaries (as *pre* indicates syllabic boundaries in *prefix* but not in *preface*). Since we hope to use affixes in determining part of speech from form alone, we will use a formal definition. For purposes of this study, an affix will be recognized as an affix under only two formal and reproducible conditions. First, the unit to which any affix attaches must contain one or more vowel strings. Second, the unit to which any prefix attaches must begin with an admissible initial consonant string, and the unit to which any suffix attaches must end with an admissible terminal consonant string. The admissible initial and terminal strings, whose derivation is given by Dolby and Resnikoff,<sup>1</sup> are listed in Table 2. It is possible to refine these rules to produce a closer correspondence with any given definition, but these criteria seem adequate for our purposes.

To correlate the affixes in Table 1 with parts of speech, a computer program was written to examine all double-standard words with two or more vowel strings. (To avoid the complication of considering archaic or little-used words, only words having a standard meaning in both dictionaries were used.) It sorted out all words that had an affix, that is, a beginning or ending that matched a member of the affix list and met the established criteria. Each of these words had a part-of-speech string given for it, that is, the list of parts of speech possible for that word. The parts of speech recorded on tape are as follows: noun [N], adjective [AJ], verb [V], adverb [AV], preposition [PR], conjunction [CJ], pronoun [PN], interjection [IJ], past verb [PV]. The category *other* [OT] was used whenever the dictionary gave some part of speech other than the nine listed; OT comprises mainly participles and collective nouns.) Since the dictionaries do not always agree, the string is taken as the parts of speech that are associated with standard meanings of the word in either dictionary. The program associated the part-of-speech string of a given word with that word's prefix or suffix. Up to nine different strings could be associated with an affix. For each affix, a count of the number of words with that affix was made for each encountered part-of-speech string, with the counts divided

TABLE 2  
INITIAL AND TERMINAL STRINGS

A. ADMISSIBLE INITIAL CONSONANT STRINGS OF CVC WORDS

B	N	BL	GL	SH	TR	SCH
C	P	BR	GN	SK	TW	SCR
D	Q	CH	GR	SL	WH	SHR
F	R	CL	KN	SM	WR	SPH
G	S	CR	KR	SN		SPL
H	T	DR	PH	SP		SPR
J	V	DW	PL	SQ		STR
K	W	FL	PR	ST		THR
L	Z	FR	RH	SW		THW
M		GH	SC	TH		

B. ADMISSIBLE FINAL CONSONANT STRINGS OF CVC WORDS NOT ENDING WITH E

B	BB	LT	RF	WD	GHT
C	CH	MB	RK	WK	LCH
D	CK	MM	RL	WL	LPH
F	CT	MN	RM	WN	LTH
G	DD	MP	RN	XT	MPH
H	FF	ND	RP	ZZ	MPT
K	FT	NG	RR		NCH
L	GG	NK	RT		NTH
M	GH	NN	SH		NTZ
N	GN	NT	SK		RCH
P	LD	NX	SM		RSH
R	LF	PH	SP		RST
T	LK	PT	SS		RTH
W	LL	RB	ST		SCH
X	LM	RC	TH		TCH
Z	LP	RD	TT		

according to the number of syllables in the words. The following example will help to clarify.

The result for the prefix *inter* is shown in Table 3. A 1 indicates presence in the dictionary of the part of speech identified by the abbreviation at the head of the column. Thus, the first line of Table 3 indicates that the first part-of-speech string encountered in the words prefixed with *inter* was noun and verb and that there were twenty-three total words with this part-of-speech string, one of them a two-vowel-string word and twenty-two of them three-vowel-string words. The next line shows that there were three total words with the string noun, adjective, and verb, one of them a two-vowel-string word and two of them three-vowel-string words. Thus the nine lines indicate the first nine part-of-speech strings encountered. When a tenth string was found, the program terminated the examination of this affix and printed a notation to that effect. Note that the column headed "Total" shows the distribution according to part of speech of all words prefixed with *inter* and that the columns headed "N vs" show the distribution according to part of speech of words with N vowel strings. The distribution according to vowel strings was obtained because it had been noted that there was a general tendency for the percentage of noun-adjective

TABLE 3  
EXAMPLE OF AFFIX STATISTICS OUTPUT BY THE COMPUTER PROGRAM FOR THE PREFIX "INTER"\*

N	NO. OF OCCURRENCES VS. VOWEL STRINGS†															
	AJ	V	AV	PR	CJ	PN	IJ	PA	OT	2 VS	3 VS	4 VS	5 VS	6 VS	7 VS	Total
1	0	1	0	0	0	0	0	0	0	1	22	0	0	0	0	23
1	1	1	0	0	0	0	0	0	0	1	2	0	0	0	0	3
1	0	0	0	0	0	0	0	0	0	1	7	27	22	4	3	64
0	0	1	0	0	0	0	0	0	0	0	25	10	6	3	0	44
0	1	0	0	0	0	0	0	0	0	0	1	23	15	7	0	46
0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1
1	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1
1	1	0	0	0	0	0	0	0	0	0	2	7	4	3	0	16
0	1	0	0	0	0	0	0	0	1	0	1	7	4	1	0	13

\* More than nine unique part-of-speech patterns were encountered. N, noun; AJ, adjective; V, verb; AV, adverb; PR, preposition; CJ, conjunction; PN, pronoun; IJ, interjection; PA, past; OT, other; VS, vowel string.

† VS columns show the part-of-speech distribution for two-, three-, four-, five-, six-, and seven-vowel string words. The "total" column shows the part-of-speech distribution in the total sample.

words to increase with the number of syllables.

Study of the part-of-speech distributions of the words with affixes in Set I (Table 4) shows that the words with a given affix have an average of eight or more part-of-speech combinations associated with them, and, in general, there is wide distribution of the words among the different part-of-speech strings. In fact, the results indicate that it will be impossible to assign a 100 per cent unique part-of-speech string to a word on the basis of its affixes. What should be possible is to establish an algorithm which will be 95 per cent correct in assigning an "inclusive" part-of-speech string, by which we mean a string that will include all of the dictionary-assigned parts of speech but that may include some extraneous parts of speech.

Since, as already noted, the majority of multisyllable words can be used only as nouns or adjectives, this will be the point of departure in deriving a part-of-speech algorithm. All words that do not behave as nouns, or adjectives, or nouns and adjectives only are to be considered exceptional, to be listed or to be identified as exceptional by examination of their affixes. The algorithm will be constructed to identify the exceptions and leave the rest to be given the basic assignment of noun-adjective for multisyllable words or noun-adjective-verb for one-syllable words.

Because they are manageably few, all adverbs not ending in *ly* and all prepositions, conjunctions, interjections, and irregular past-tense verbs can be removed and put in a special exception list. This leaves combinations of noun, adjective, verb, and "other" to deal with, where "other" comprises participial forms and collective nouns. Regular forms of participles can be recognized by the inflectional endings *ing* or *ed*, and, irregular forms of participles and collective nouns are few enough so that they can be added to the exception list. (So also can all words that end in *ing* or *ed* but are not

participial forms.) Seven possible part-of-speech combinations remain:

- |                               |         |
|-------------------------------|---------|
| (1) Noun                      | N       |
| (2) Adjective                 | AJ      |
| (3) Noun and adjective        | N-AJ    |
| (4) Verb                      | VB      |
| (5) Noun and verb             | N-VB    |
| (6) Adjective and verb        | AJ-VB   |
| (7) Noun, adjective, and verb | N-AJ-VB |

Since most nouns can be used as adjectives, and since the AJ-VB combination is uncommon except for participles, which are already taken care of, the seven combinations can be reduced to four by merging (3) with (1), and (5) and (6) and (7), to give:

- |                                      |       |
|--------------------------------------|-------|
| (1) Noun and adjective               | NA    |
| (2) Adjective                        | AJ    |
| (3) Verb                             | VB    |
| (4) Verb and (noun and/or adjective) | NAVVB |

To put it another way, there are two large classes of multisyllable words, NA and NAVB, which must be distinguished. In addition, the class AJ must be distinguished from the NA and the class VB from the NAVB. Whenever these distinctions cannot be made with 95 per cent accuracy, assignments will be made to the inclusive set.

The construction of the algorithm thus becomes quite simple, a matter of studying the distribution of the part-of-speech strings for each affix, ignoring any part of speech other than noun, adjective, or verb. In accordance with the 95 per cent criterion, an affix for which 95 per cent of the words with that affix have a single part of speech, either AJ or VB, will be classified as "adjectival" or "verbal," respectively, and the algorithm will simply assign words containing such an affix to the AJ or the VB class instead of to the basic NA

class. Affixes for which 95 per cent of the words are nouns and/or adjectives, but not verbs, may be con-

TABLE 4  
PART-OF-SPEECH (POS) IMPLICATIONS OF AFFIXES  
A. AFFIXES SET I

Prefixes	POS Code	Suffixes	POS Code
a	NAVB	a	Neutral
ab*	NAVB	able	Neutral
ac	NAVB	age*	NAVB
ad*	NAVB	ah	Neutral
al*	NAVB	al	Neutral
an*	NAVB	an	Neutral
ana	NAVB	ant*	NAVB
ap	NAVB	ar*	NAVB
apo	Neutral	ard	NAVB
at*	NAVB	at*	NAVB
auto	Neutral	ate	NAVB
be*	NAVB	ee*	NAVB
ca	NAVB	el*	NAVB
circum	NAVB	en	NAVB
com*	NAVB	ent	NAVB
con	NAVB	eon*	NAVB
cy	Neutral	er	NAVB
de	NAVB	et*	NAVB
demo	NAVB	ey*	NAVB
deni	Neutral	ful	Neutral
dis	NAVB	ia	Neutral
e	NAVB	ial	Neutral
ec	Neutral	ic	Neutral
em	NAVB	ie*	NAVB
en	NAVB	ier*	NAVB
epi	NAVB	ile*	NAVB
ex	NAVB	in*	NAVB
hy	NAVB	ine	NAVB
hyper	Neutral	ion	NAVB
i*	NAVB	is*	NAVB
im	NAVB	ish*	NAVB
in	NAVB	ism	Neutral
incon	Neutral	ist	Neutral
inex	Neutral	ite*	NAVB
inter	NAVB	ium	Neutral
irre	Neutral	ive	Neutral
lu	Neutral	land	Neutral
ma	NAVB	ler	Neutral
mi	NAVB	less	AJ
mis	NAVB	let*	NAVB
ob	NAVB	lock*	NAVB
out*	NAVB	man	Neutral
photo*	NAVB	ment	NAVB
pro	NAVB	ness	Neutral
re	NAVB	o	NAVB
sa*	NAVB	ock*	NAVB
sac	Neutral	on	NAVB
sub	NAVB	or*	NAVB
sun	Neutral	ot*	NAVB
tele*	NAVB	ow*	NAVB
trans	NAVB	tation	Neutral
un	NAVB	ue	NAVB
uncon	Neutral	um*	NAVB
vi*	NAVB	ure	NAVB
vul	Neutral	us	Neutral
		ward	Neutral
		y	NAVB

B. AFFIXES SET II

Prefixes	POS Code	Suffixes	POS Code
aero	Neutral	ae	
air	Neutral	al	Neutral
bi*	NAVB	as	Neutral
brain		cy	Neutral
by	Neutral	eer	NAVB
bye		ern	NAVB
centi	Neutral	est	NAVB
co*	NAVB	ette*	NAVB
deca	NAVB	ex	NAVB
deci	NAVB	genic	AJ
demi	NAVB	iana	
for*	NAVB	ible	Neutral
fore	NAVB	ise	NAVB
hecto		ist	Neutral
homo	Neutral	ity	Neutral
non	Neutral	ix	NAVB
para	NAVB	ize	VB
self	Neutral	latry	Neutral
semi	Neutral	lily	Neutral
super	NAVB	logy	Neutral
vice		phile	
yester		th	NAVB
		ty	Neutral
		ular	Neutral
		valent	
		ways	
		worthy	Neutral

C. AFFIXES SET III

Suffixes	POS Code
ous	AJ
sial	AJ
sion*	NAVB
sive	Neutral
tial	Neutral
tion	Neutral
tious	AJ
tive	Neutral

sidered as "neutral," since words containing them behave as nouns and/or adjectives in accordance with the general rule. An affix, however, for which 5 per cent of the words (and more than five words) have a verb usage will be classified "noun-verbal," and words containing such an affix will be assigned to the NAVB class. As already indicated, all words that do not contain an affix and that are not in an exception list are classified as NA if multisyllable and NAVB if one syllable.

It must be realized that a good many ambiguities will be introduced by this algorithm. For example, for words prefixed with *inter*, 71 of the 211 words in our data set have a verbal usage, with further breakdown as follows:

Noun and verb	23			
Noun, adjective, and verb	3	or	NAV B	27
Adjective and verb	1			
Verb	44		VB	44

Accordingly, words beginning with *inter* will be assigned to the NAVB class, obtaining the correct inclusive part of speech for 71 words at the cost of introducing the extraneous part-of-speech VB to the 140 well-behaved NA words. The situation is worse in the ambiguity between the AJ and the NA classes. For example, although about 8 per cent of words ending in the suffix *ful* are adjectives, 34 out of the total 169 have a noun usage, so rather than take a 20 per cent error of omission, *ful* is regarded as a neutral suffix, and an extra part of speech has been introduced in 80 per cent of the words. By stretching a point, the suffix *less* can be considered adjectival, since it is 94 per cent adjectival, but many other adjective-tending affixes encountered cannot (*ic*, 54 per cent; *able* 79 per cent; *ish*, 70 per cent; *ial*, 61 per cent; *us*, 87 per cent; *mis*, 61 per cent).

A part-of-speech implication of either NAVB, VB, AJ, or neutral (i.e., NA) has been determined for all of the affixes. These implications are listed in Table 4. When there were fewer than five words with a given affix, no assignment was made. The implications of the operational affixes and of the *Dictionary of Modern English Usage*<sup>6</sup> affixes break down statistically as follows:

	Operational	English Usage
Neutral ...	33	20
NAV B ....	77	17
AJ .....	1	1
VB .....	0	1

In Table 4, some of the affixes have asterisk superscripts. These are affixes with an NAVB implication, which in words of four or more syllables may be regarded as neutral, since in the dictionary there were fewer than three four- to eight-vowel-string words with these affixes that possessed verbal usages. NAVB affixes that are neutral for five- to eight-vowel-string words were not considered because there are only about 1,250 of these, while there are about 11,250 four- to eight-vowel-string words.

There are some words, of course, that have both prefix(es) and suffix(es). As the part-of-speech tabulations for suffixes were independent of prefixes, and vice versa, there was a possibility of a particularly influential and common affix introducing an extra part of speech into the part-of-speech counts of other affixes. For example, suppose that all the words with the prefix *trans* were always nouns except those that end in verbal suffixes, such as *er* or *ate*, as in *transfer* and *translate*. Then *trans* would have been assigned the implication NAVB when it should have been neutral. To test this possibility, the Set I prefix counts were repeated

with all words having non-neutral suffixes omitted from the data set. However, the part-of-speech implication of all prefixes remained the same. Since none of the part-of-speech implications of the prefixes changed, it was decided that it was unnecessary to test suffixes on a set from which prefixed words had been removed.

Prefixes were chosen for the test because the suffixes seem to have a stronger influence than prefixes in multi-affixed words, as, for example, the neutral *ism* wins over the NAVB *ex* in *exorcism* and the verbal *ize* wins over the neutral *vul* in *vulcanize*. Suffixes would thus cause much more of a problem in the prefix counts than prefixes in the suffix counts. The one easily noted exception to the rule of suffix ascendancy is for such words as *automation* and *vulcanization*, in which the neutral *auto* and *vul* seem to be ascendent over the NAVB *ion*. However, a consideration of other words in which both prefix and suffix are NAVB, as in *demolition*, *construction*, *accession*, etc., indicates that there is a group of important suffixes beginning with *t* or *s* that failed to show up in the operational definition of affixes. To test this hypothesis, these possible suffixes were subjected to the part-of-speech tests for affixes with the following results:

Suffix	POS Implication
tion	Neutral
sion*	NAV B
tial	Neutral
sial	AJ
tive	Neutral
sive	Neutral
tious	AJ

Examination of the suffix *tious* led to examination of the weak suffix possibility *ous*, which, like *tious*, turned out to have strongly adjectival implications. Undoubtedly, these suffixes do exist and have strong part-of-speech connotations. For the sake of completeness, they have been added to Table 4 as Set III.

Whether or not the use of the part-of-speech implications reported in this paper will be adequate to produce 95 per cent accurate part of speech by algorithmic assignment remains to be seen. They are, of course, guaranteed to produce 95 per cent inclusive accuracy on words with listed affixes. It is not yet known how many non-affixed words there are or how well they fit the general rules. Before comprehensive testing can take place, it may be necessary to develop more definitive rules for determining when an affix is acting as an affix in a given word.

Received February 4, 1966

## References

1. Dolby, J., and Resnikoff, H., "On the Structure of Written English Words," *Language*, Vol. 40, No. 2 (April-June, 1964).

2. *The Shorter Oxford English Dictionary on Historical Principles*. 3d ed., revised with addenda. Oxford: Clarendon Press, 1959.
3. *Webster's Third New International Dictionary of the English Language*. Springfield, Mass.: G. C. Merriam Co., 1961.
4. Resnikoff, H., and Dolby, J., "The Nature of Affixing in Written English," *Mechanical Translation*, Vol. 8, Nos. 3, 4 (June and October, 1965).
5. Earl, L. L., "Structural Definition of Affixes in Multi-syllable Words," this issue.
6. Fowler, H. W., *A Dictionary of Modern English Usage*. Revised and edited by Sir Ernest Gowers. 2d ed. New York: Oxford University Press, 1965.