# The Nature off Affixing in Written English, Part II

by H. L. Resnikoff and J. L. Dolby, The Institute for Advanced Study,
Princeton, New Jersey, and R & D Consultants Company, Los Altos, California

*This is a continuation of the authors' paper of the same title which appeared in Volume 8 of this journal. The present part extends the authors' definitions of prefix and suffix (in written English) to corpora of three-vowel-string words, and implements them on a corpus K consisting of 19,329 graphemically distinct three-vowel-string words from the* Shorter Oxford Dictionary. *The notion of a parasitic affix is introduced, and the parasitic suffixes for K are determined.*

This paper is a continuation of reference 1 (which will be called Part I throughout). In that paper[1] a systematic procedure for finding English affixes was briefly described, and the results of applying the procedure to the CVCVC words in the *Shorter Oxford English Dictionary* were given.

Here we will present several refinements of the procedure used in Part I and apply the technique to the study of affixes in the three-vowel-string words, that is, CVCVCVC words.

There are some novelties which arise. Among these, the most important is certainly the occurrence of suffixes which primarily occur attached to other suffixes. Evidently these could not be found from an investigation of the two-vowel-string words, and so they did not make their appearance in Part I. Another new feature is the occurrence of two-vowel-string affixes, which cannot occur in two-vowel-string words for obvious reasons.

Except where otherwise noted, the terminology and definitions are those used in Part I.

The reader should note the recently published work of Monroe,[2] which forms an interesting complement to our investigations,

## Notational Refinements

Before coming to the proper subject of this paper we would like to make corrections to Part I and to introduce some minor refinements of notation.

The weak suffix -Y should be added to Table III of Part I. The classes Cls(NCH/Y) and Cls(FF/Y), among others, testify to the existence of this affix. Also, in the penultimate paragraph, *read* -IST *for* -ous.

We turn now to the notational refinements. From Volume 2 of *The English Word Speculum*[3] it can be seen that the letter Q in initial position is always followed by the letter U with only one occurrence of the sequence QY. Since there are fewer than four exceptions to the statement that Q is always followed by U in initial position, this will be taken as a universal property of English words. Using the terminology of Part I, the sequence QU is the only admissible initial sequence beginning with Q.

Similarly, from Volume 3 of *The English Word Speculum,* we find that the only words that end with the letter Q are SEQ and ESQ. Again there are fewer than four words ending with Q, and so it is clear that Q alone does not occur admissibly in either initial or final position in English words.

A somewhat more tedious examination of *Speculum 3* (this mode of reference to particular volumes of reference 3 will be used hereafter) shows that Q is always followed by U with the exceptions noted above. For this reason, the letter sequence QU can be treated as a single unit in the words in which it occurs. Such a letter sequence which functions as a distinct unit in all contexts will be called a "generalized letter," and all generalized letters are classified as consonants. Throughout this paper we will assume that the sequence QU is a generalized letter and hence a consonant. With this assumption it is worth noting that the string QUE is an admissible final-consonant string, occurring in words like MASQUE.

Because only the admissible final-consonant strings not ending with E were used to determine the affixes in Part I, the addition of the admissible final-consonant string QUE does not influence the results of that paper. However, the generalized letter QU should replace the letter Q in the first section of Table I of Part I.

The fact that QU is assumed to be a generalized letter will have an effect on the syllabic decomposition of certain words constructed like QUADRILATERAL, where the first vowel string must now be interpreted as the single vowel A, since QU is a consonant.

From Table 3 of a previous study,[4] we see that X is the only consonant that is not an admissible initial consonant, and Table 6 of the same paper shows that J and QU are not admissible final consonants. Hence, in the terminology of Part I, there is a mandatory decomposition point as indicated in each of the following sequences:

$$V_1X \text{ —— } V_2, V_1 \text{ —— } JV_2, \text{ and } V_1 \text{ —— } QUV_2,$$

where $V_1$ and $V_2$ are arbitrary vowel strings. In order to simplify both the notation and presentation, we will make the convention that these letter sequences be interpreted as standing for the sequences

$$V_1X\phi V_2, \quad V_1\phi JV_2, \quad \text{and} \quad V_1\phi QUV_2,$$

where $\phi$ denotes the blank consonant. In this way the definitions of prefixes and suffixes given in Part I become applicable to words containing the letters X, J, and QU without any alteration.

The procedure just described was tacitly followed in Part I; Table I there showed that the mandatory decomposition points given above exist for these letters. The only consequence drawn from these assumptions in Part I was that EX- is a strong prefix in the two-vowel-string corpus that was examined there. This conclusion is not altered by our present conventions.

## Modified Definitions

The affix definitions given in Part I referred specifically to a two-vowel-string corpus. Here we will consider a three-vowel-string corpus, and so the definitions must be modified accordingly.

Let K be a fixed corpus of three-vowel-string words, and let the words belonging to K be given in the form

$$C_1V_1C_2V_2C_3V_3C_4.$$

*Definition P1.* Let $P = C_1V_1C_2'$ (resp. $P = C_1V_1C_2V_2C_3'$) be a fixed initial-letter string. P is called a *strong prefix* (with respect to K) if there exist two distinct classes of words from K, $Cls(P/C_I'')$ and $Cls(P/C_{II}'')$, each of which contains more than three words, such that $C_2'C_I''$ and $C2'C_{II}''$ (resp. $C_3'C_I''$ and $C'C_{II}''$) are mandatory decomposition points of the second consonant string $C_2$ (resp. the third consonant string $C_3$).

This definition parallels that given in Part I, but makes it possible to consider two-vowel-string prefixes. The corresponding definition of a *strong suffix* is this:

*Definition S1.* Let $S = C_3''V_3C_4$ (resp. $S = C_2''V_2C_3V_3C_4$) be a fixed final-letter string. S is called a *strong suffix* (with respect to K) if there exist two distinct classes of words from K, $Cls(C_I'/S)$ and $Cls(C_{II}'/S)$, each of which contains more than three words, such that $C_I'C_3''$ and $C_{II}'C_3''$ (resp. $C_I'C_2'$ and $C_{II}'C_2''$) are mandatory decomposition points of the third consonant string $C_3$ (resp. the second consonant string $C_2$).

In an analogous fashion, the definitions of *weak prefix* and *weak suffix* given in Part I are generalized to apply to a three-vowel-string corpus.

*Definition P2.* Let $P = C_1V_1$ (resp. $P = C_1V_1C_2V_2$) be a fixed initial-letter string. P is called a *weak prefix* (with respect to K) if there exist two distinct classes of words from K, $Cls(P/C_I)$ and $Cls(P/C_{II})$, each of which contains more than three words, such that $C_I$ and $C_{II}$ are admissible initial-consonant strings. Here $C_I$ and $C_{II}$ are the entire second (resp. third) consonant strings of words from K.

*Definition S2.* Let $S = V_3C_4$ (resp. $S = V_2C_3V_3C_4$) be a fixed final-letter string. S is called a *weak suffix* (with respect to K) if there exist two distinct classes of words from K, $Cls(C_I/S)$ and $Cls(C_{II}/S)$, each of which contains more than three words, such that $C_I$ and $C_{II}$ are admissible final-consonant strings. Here $C_I$ and $C_{II}$ are the entire third (resp. second) consonant strings of words from K.

It will turn out to be necessary to consider a still weaker definition of affixes, but this must wait until the consequences of the four definitions presented above have been examined.

The admissible initial- and final-consonant strings of English words play a critical role in the application of all four of the definitions, because the notion of a mandatory decomposition point, as defined in Part I, is rooted in explicit knowledge of the admissible consonant strings. This information, taken from reference 4, and presented in Table I of Part I, will be used repeatedly in the application of the definitions given in later sections of this paper.

One other matter must be decided before the definitions can be applied. It may happen, for instance, that the sequence P' is a prefix and the longer sequence P" = P'X is also a prefix, where x is a non-blank letter string. It is intuitively unsatisfactory to permit a word belonging to an admissible class $Cls(P''/Y'')$ to appear in one of the defining classes $Cls(P'/Y')$. Therefore, we make the convention that words appearing in an admissible class for an affix A are to be excluded from membership in all classes for affixes contained in A. Thus a word belongs to the admissible class of the longest affix it contains.

As a concrete illustration, consider the suffixes -LY and -Y. Since -LL is a popular admissible final-consonant string, there are many three-vowel-string words ending with -LLY. If -Y is under examination, we would be tempted to consider $Cls(LL/Y)$ to show that -Y is a suffix. Since -LY is a suffix, it is not clear that the decomposition LL-Y is appropriate; perhaps L-LY is correct in certain circumstances. Application of the convention requires that the decomposition L-LY be considered; according to the definition, only classes with mandatory decomposition points can be considered to determine the strong suffixes. Since -LL is an admissible final-consonant string, L-LY is not a mandatory decomposition point, and so $Cls(L/LY)$ cannot be considered as a defining class for -LY either. Hence the effect of the convention is to delete from the corpus the words of the form -LLY which may involve more than one distinct suffix.

As a second illustration, consider the suffixes -ICAL and -AL. The convention requires that the words in the admissible classes defining -ICAL not be used in the classes defining -AL. For the corpus described in the next section, this means that words ending with -PTICAL

# TABLE 1

## Prefix Data for the Corpus k

| Prefix | Entry | N | Prefix | Entry | N | Prefix | Entry | N |
|---|---|---|---|---|---|---|---|---|
| A- | (A/BL) | 8 | | (DIS/J) | 4 | OC- | (OC/C) | 12 |
| | (A/BR) | 12 | | (DIS/R) | 12 | OUT- | (OUT/L) | 4 |
| | (A/CR) | 15 | | (DIS/V) | 6 | OVER- | (OVER/H) | 8 |
| | (A/DR) | 4 | | | | | (OVER/W) | 11 |
| | (A/GR) | 9 | E- | (E/CL) | 7 | | | |
| | (A/QU) | 12 | | (E/GR) | 4 | OX- | (OX/φ) | 19 |
| | (A/SC) | 10 | | (E/QU) | 21 | | | |
| | (A/SP) | 12 | | | | PA- | (PA/TR) | 11 |
| | (A/SPH) | 4 | EN- | (EN/C) | 37 | | | |
| | (A/TR) | 7 | | (EN/F) | 13 | PE- | (PE/TR) | 12 |
| | | | | (EN/H) | 10 | | | |
| AB- | (AB/D) | 4 | | (EN/L) | 5 | PRO- | (PRO/CR) | 4 |
| | | | | (EN/R) | 6 | | (PRO/GR) | 4 |
| AC- | (AC/C) | 34 | | (EN/V) | 8 | | (PRO/J) | 5 |
| | (AC/QU) | 5 | | | | | (PRO/SP) | 5 |
| | | | EX- | (EX/φ) | 68 | | (PRO/TR) | 8 |
| AD- | (AD/H) | 4 | | (EX/C) | 31 | | | |
| | (AD/J) | 12 | | (EX/CR) | 8 | PROX- | (PROX/φ) | 5 |
| | (AD/M) | 14 | | (EX/H) | 7 | | | |
| | (AD/V) | 18 | | (EX/P) | 34 | PU- | (PU/BL) | 5 |
| | | | | (EX/PL) | 9 | | (PU/TR) | 6 |
| AL- | (AL/C) | 7 | | (EX/PR) | 5 | | | |
| | (AL/H) | 4 | | (EX/S) | 8 | QUA- | (QUA/DR) | 14 |
| | | | | | | | | |
| AN- | (AN/C) | 14 | FIX- | (FIX/φ) | 4 | QUIN- | (QUIN/QU) | 5 |
| | (AN/H) | 6 | | | | RE- | (RE/CL) | 4 |
| | | | HEX- | (HEX/φ) | 9 | | (RE/CR) | 5 |
| ARC- | (ARC/T) | 4 | | | | | (RE/FL) | 6 |
| | | | HUS- | (HUS/B) | 7 | | (RE/FR) | 12 |
| AR- | (AR/QU) | 4 | | | | | (RE/PL) | 10 |
| | | | HY- | (HY/DR) | 32 | | (RE/PR) | 11 |
| AT- | (AT/L) | 4 | | | | | (RE/QU) | 6 |
| | | | IN- | (IN/C) | 105 | | (RE/TR) | 23 |
| AX- | (AX/φ) | 5 | | (IN/F) | 46 | | | |
| | | | | (IN/FL) | 12 | SAX- | (SAX/φ) | 6 |
| BE- | (BE/SL) | 4 | | (IN/FLEX) | 4 | | | |
| | | | | (IN/FR) | 10 | SE- | (SE/CR) | 7 |
| BIO- | (BIO/GR) | 4 | | (IN/H) | 14 | | (SE/QU) | 9 |
| | | | | (IN/J) | 7 | | | |
| CIRCUM- | (CIRCUM/FL) | 5 | | (IN/L) | 6 | SUB- | (SUB/C) | 4 |
| | | | | (IN/QU) | 7 | | (SUB/D) | 5 |
| COAD- | (COAD/J) | 4 | | (IN/V) | 48 | | (SUB/J) | 6 |
| | | | | | | | (SUB/M) | 11 |
| COM- | (COM/F) | 4 | IRON- | (IRON/W) | 5 | | (SUB/T) | 9 |
| | | | | | | | (SUB/TR) | 4 |
| CON- | (CON/C) | 54 | LAND- | (LAND/L) | 4 | | | |
| | (CON/F) | 23 | | | | SUC- | (SUC/C) | 27 |
| | (CON/J) | 10 | LAX- | (LAX/φ) | 4 | | | |
| | (CON/V) | 35 | | | | TAX- | (TAX/φ) | 6 |
| | | | LI- | (LI/QU) | 6 | | | |
| CY- | (CY/CL) | 12 | | | | TECH- | (TECH/N) | 4 |
| | | | LUX- | (LUX/φ) | 6 | | | |
| DE- | (DE/CL) | 6 | | | | TE- | (TE/TR) | 20 |
| | (DE/CR) | 8 | MA- | (MA/J) | 7 | | | |
| | (DE/FL) | 12 | | | | TRANS- | (TRANS/F) | 13 |
| | (DE/J) | 4 | MAX- | (MAX/φ) | 5 | | | |
| | (DE/PL) | 4 | | | | UN- | (UN/B) | 13 |
| | (DE/PR) | 15 | MI- | (MI/CR) | 10 | | (UN/C) | 15 |
| | (DE/SC) | 6 | | | | | (UN/F) | 4 |
| | (DE/TR) | 8 | MIS- | (MIS/B) | 8 | | (UN/H) | 9 |
| | | | | (MIS/D) | 7 | | (UN/R) | 10 |
| DI- | (DI/PL) | 9 | | (MIS/F) | 5 | | (UN/W) | 7 |
| | | | | (MIS/R) | 6 | | | |
| DIS- | (DIS/B) | 9 | | | | UNDER- | (UNDER/W) | 6 |
| | (DIS/CL) | 4 | MULTI- | (MULTI/PL) | 4 | | | |
| | (DIS/CR) | 6 | | | | U- | (U/PR) | 5 |
| | (DIS/D) | 4 | OB- | (OB/J) | 6 | | | |
| | (DIS/F) | 7 | | (OB/T) | 4 | VI- | (VI/BR) | 6 |
| | (DIS/G) | 6 | | (OB/V) | 5 | | (VI/TR) | 11 |
| | (DIS/GR) | 4 | | | | | | |

## TABLE 2

### SUFFIX DATA FOR THE CORPUS K

**Column 1**

- **-A**
  - (CC/A) 4
  - (ND/A) 7
  - (NG/A) 7
  - (LG/A) 6
  - (LL/A) 62
  - (NN/A) 7
  - (RR/A) 5
  - (SS/A) 4
  - (NT/A) 5
  - (TT/A) 17
- **-MA**
  - (G/MA) 5
- **-IC**
  - (RG/IC) 4
  - (LL/IC) 9
  - (CT/IC) 23
  - (NT/IC) 13
  - (PT/IC) 15
- **-OID**
  - (LL/OID) 4
- **-ANCE**
  - (ND/ANCE) 6
  - (RD/ANCE) 4
  - (CT/ANCE) 4
  - (TT/ANCE) 4
- **-ENCE**
  - (ND/ENCE) 6
  - (LG/ENCE) 4
  - (RG/ENCE) 4
  - (RR/ENCE) 6
- **-IDE**
  - (NT/IDE) 4
  - (X/IDE) 5
- **-EE**
  - (NT/EE) 6
- **-AGE**
  - (LL/AGE) 4
  - (NT/AGE) 6
- **-LIKE**
  - (N/LIKE) 5
- **-ALE**
  - (NG/ALE) 6
- **-ABLE**
  - (CT/ABLE) 5
  - (NT/ABLE) 6
  - (LV/ABLE) 4
- **-IBLE**
  - (NG/IBLE) 4
  - (NS/IBLE) 10
  - (RS/IBLE) 5
  - (SS/IBLE) 14
  - (CT/IBLE) 7
  - (PT/IBLE) 5
  - (RT/IBLE) 6
- **-ILE**
  - (CT/ILE) 13
  - (NT/ILE) 4
- **-OME**
  - (RS/OME) 6
- **-INE**
  - (ND/INE) 6
  - (RD/INE) 7
  - (LL/INE) 8
  - (RM/INE) 7
  - (NT/INE) 26
  - (RT/INE) 9
- **-UNE**
  - (RT/UNE) 7
- **-ORE**
  - (RM/ORE) 5

**Column 2**

- **-URE**
  - (NS/URE) 4
  - (SS/URE) 5
  - (CT/URE) 12
  - (NT/URE) 7
  - (RT/URE) 5
  - (XT/URE) 7
- **-ISE**
  - (NCH/ISE) 6
  - (RT/ISE) 4
- **-OSE**
  - (LL/OSE) 4
  - (MP/OSE) 4
  - (RP/OSE) 4
- **-ATE**
  - (RC/ATE) 4
  - (ND/ATE) 5
  - (LL/ATE) 23
  - (MN/ATE) 5
  - (RN/ATE) 12
  - (RR/ATE) 4
  - (NS/ATE) 5
  - (SS/ATE) 5
  - (CT/ATE) 8
  - (NT/ATE) 9
- **-ULATE**
  - (NG/ULATE) 5
- **-ITE**
  - (ND/ITE) 6
  - (NG/ITE) 5
  - (LL/ITE) 11
  - (NT/ITE) 7
  - (RT/ITE) 7
- **-IVE**
  - (LS/IVE) 8
  - (NS/IVE) 19
  - (RS/IVE) 15
  - (SS/IVE) 32
  - (CT/IVE) 62
  - (NT/IVE) 14
  - (PT/IVE) 20
  - (MPT/IVE) 8
  - (RT/IVE) 7
  - (X/IVE) 4
- **-IZE**
  - (ND/IZE) 4
  - (RD/IZE) 6
  - (RN/IZE) 5
- **-ING**
  - (RL/ING) 12
- **-ISH**
  - (ND/ISH) 4
  - (RN/ISH) 5
- **-ICAL**
  - (PT/ICAL) 4
  - (RT/ICAL) 4
- **-EAL**
  - (NG/EAL) 5
- **-AL**
  - (RM/AL) 7
  - (RN/AL) 17
  - (RS/AL) 8
  - (SS/AL) 6
  - (NT/AL) 16
  - (TT/AL) 6
- **-IAL**
  - (NN/IAL) 10
  - (NT/IAL) 16
  - (X/IAL) 5
- **-EL**
  - (NN/EL) 4

**Column 3**

- **-FUL**
  - (T/FUL) 7
  - (ST/FUL) 6
- **-YL**
  - (X/YL) 5
- **-ISM**
  - (RCH/ISM) 4
  - (NT/ISM) 4
- **-UM**
  - (ND/UM) 10
  - (LL/UM) 11
  - (NT/UM) 8
- **-IUM**
  - (NN/IUM) 6
  - (NG/IUM) 4
- **-AN**
  - (LL/AN) 5
  - (LM/AN) 4
  - (RM/AN) 14
- **-IAN**
  - (LL/IAN) 5
  - (SS/IAN) 5
  - (NT/IAN) 4
- **-EN**
  - (GHT/EN) 4
- **-AIN**
  - (RT/AIN) 8
- **-IN**
  - (RK/IN) 7
  - (LL/IN) 6
  - (SS/IN) 4
  - (RT/IN) 4
- **-CON**
  - (L/CON) 4
- **-ON**
  - (ND/ON) 4
  - (LL/ON) 5
  - (MM/ON) 5
- **-ION**
  - (LL/ION) 19
  - (LS/ION) 12
  - (NS/ION) 25
  - (RS/ION) 28
  - (SS/ION) 63
  - (CT/ION) 111
  - (NT/ION) 15
  - (PT/ION) 39
  - (MPT/ION) 10
  - (RPT/ION) 4
  - (RT/ION) 15
  - (X/ION) 11
- **-ATION**
  - (ND/ATION) 4
  - (LL/ATION) 4
  - (RN/ATION) 4
  - (SS/ATION) 4
  - (CT/ATION) 5
  - (NT/ATION) 5
  - (PT/ATION) 4
  - (RT/ATION) 4
  - (X/ATION) 4
- **-O**
  - (CC/O) 6
  - (ND/O) 8
  - (RD/O) 4
  - (NG/O) 6
  - (RG/O) 4
  - (LL/O) 6
  - (NT/O) 10
  - (RT/O) 7
  - (TT/O) 22

| Affix | Class | N | Affix | Class | N | Affix | Class | N |
|---|---|---|---|---|---|---|---|---|
| -AR | (LL/AR) | 4 | -OUS | (ND/OUS) | 5 | | (RT/MENT) | 8 |
| -ER | (BB/ER) | 5 | | (LL/OUS) | 8 | | (ST/MENT) | 7 |
| | (MB/ER) | 16 | | (RM/OUS) | 5 | -OST | (RM/OST) | 12 |
| | (LD/ER) | 15 | -IOUS | (ND/IOUS) | 4 | -OW | (LL/OW) | 10 |
| | (ND/ER) | 67 | | (NT/IOUS) | 7 | -Y | (RD/Y) | 5 |
| | (RD/ER) | 8 | | (PT/IOUS) | 4 | | (RG/Y) | 8 |
| | (NG/ER) | 28 | | (X/IOUS) | 4 | | (RCH/Y) | 22 |
| | (CK/ER) | 6 | -EROUS | (ND/EROUS) | 4 | | (RTH/Y) | 8 |
| | (LL/ER) | 23 | -UOUS | (CT/UOUS) | 6 | | (NN/Y) | 13 |
| | (NN/ER) | 4 | | | | | (PS/Y) | 7 |
| | (RN/ER) | 5 | -ET | (CK/ET) | 4 | | (SS/Y) | 4 |
| | (MP/ER) | 6 | | (LL/ET) | 5 | | (X/Y) | 5 |
| | (RR/ER) | 4 | -ANT | (ND/ANT) | 12 | -CY | (N/CY) | 69 |
| | (NS/ER) | 5 | | (RD/ANT) | 4 | | | |
| | (SS/ER) | 5 | | (LL/ANT) | 9 | -IFY | (LL/IFY) | 4 |
| | (CT/ER) | 4 | | (RR/ANT) | 4 | -LY | (ND/LY) | 4 |
| | (GHT/ER) | 5 | | (SS/ANT) | 6 | | (RD/LY) | 6 |
| | (LT/ER) | 5 | | (CT/ANT) | 6 | | (N/LY) | 19 |
| | (NT/ER) | 28 | | (NT/ANT) | 5 | | (NT/LY) | 9 |
| | (PT/ER) | 6 | | | | | | |
| | (RT/ER) | 13 | -ENT | (MB/ENT) | 5 | -ARY | (MM/ARY) | 4 |
| | (TT/ER) | 18 | | (ND/ENT) | 9 | | (CT/ARY) | 4 |
| | (LV/ER) | 5 | | (LG/ENT) | 5 | | | |
| -CER | (N/CER) | 6 | | (NG/ENT) | 8 | -ERY | (BB/ERY) | 9 |
| -OR | (LL/OR) | 4 | | (RG/ENT) | 8 | | (ND/ERY) | 4 |
| | (NS/OR) | 4 | | (LL/ENT) | 7 | | (GG/ERY) | 13 |
| | (SS/OR) | 15 | | (RR/ENT) | 8 | | (TCH/ERY) | 4 |
| | (CT/OR) | 45 | | (NT/ENT) | 4 | | (CK/ERY) | 7 |
| | (NT/OR) | 11 | | (LV/ENT) | 4 | | (LL/ERY) | 5 |
| | (PT/OR) | 5 | | | | | (MM/ERY) | 5 |
| -ATOR | (CT/ATOR) | 4 | -IENT | (NT/IENT) | 4 | | (NN/ERY) | 8 |
| -IS | (PS/IS) | 9 | -MENT | (ND/MENT) | 5 | | (PP/ERY) | 4 |
| -ESS | (RL/ESS) | 24 | | (CH/MENT) | 6 | | (TT/ERY) | 13 |
| | (NN/ESS) | 7 | | (NCH/MENT) | 4 | -RY | (L/RY) | 7 |
| | (RN/ESS) | 9 | | (SH/MENT) | 13 | | (N/RY) | 12 |
| -US | (NTH/US) | 6 | | (GN/MENT) | 6 | -ORY | (CT/ORY) | 4 |
| | (LL/US) | 10 | | (RN/MENT) | 6 | -ETY | (CK/ETY) | 4 |
| | (SS/US) | 6 | | (T/MENT) | 6 | | | |
| | | | | (NT/MENT) | 7 | | | |

and -RTICAL are not included in classes of the form Cls(C/AL).

## The Corpus

The definitions presented in the previous section make it apparent that the set of affixes (that is, prefixes and suffixes) that they determine depend implicitly on the corpus K. In general, a small corpus will not provide all of the affixes that can be obtained from a larger corpus, so that it is desirable to implement the definitions on as large a corpus as is practical. On the other hand, there is no a priori assurance that the set of affixes becomes stable once the corpus includes some certain fixed subcorpus. That is, it might be the case that continually increasing the size of the corpus continually increases the size of the affix set. This is a difficult problem, for which a direct answer is not likely to be obtainable. There are certain indirect ways of investigating whether the affix set tends to become stable

for sufficiently large corpora, but these are all rather elaborate and require an extensive analysis which cannot be attempted here. Nonetheless, the importance of this problem should not be overlooked.

We have chosen to implement the affix definitions on the corpus K of three-vowel-string words given in *Speculum 2*. Note that the collection of three-vowel-string words in *Speculum 3* coincides with this corpus. The corpus can also be described as the collection of all three-vowel-string boldface left justified words from the *Shorter Oxford English Dictionary* which have the property that their parts of speech (as indicated by either the *Shorter Oxford* or the *Merriam-Webster New International Dictionary,* 3d edition) are included in the categories "noun," "adjective," "verb," "adverb."

The primary reason for choosing K in this way is that this corpus is displayed in the *Speculum* in a manner convenient for the implementation of the affix definitions. Its size is another attraction: it consists of 19,329

graphemically distinct words and thus is reasonably large but still permits detailed human examination. It may be helpful to remark that the total number of three-vowel-string words in the *Shorter Oxford English Dictionary* is 20,762, so that the corpus K contains about 93 per cent of all of the three-vowel-string words in this medium-size dictionary.

## Results

The results of applying the definitions given above to the corpus K are assembled in Tables 1 and 2, devoted to prefix data and suffix data, respectively. In each of these tables the letter string under examination is listed, and those admissible classes containing the given letter string are shown together with the number of words they contain. Since only admissible classes are tabulated, the corresponding numbers are all greater than 3.

For convenience, the class Cls(X/Y) has been written in the abbreviated form (X/Y) in the tables.

In accordance with the procedures described by the definitions and augmented by our conventions, the strong and weak affixes with respect to K are precisely those letter strings that correspond to at least two classes in Tables 1 and 2.

Examining Table 1, we see that of the sixty-three initial-letter strings represented, twenty-two are prefixes; from Table 2, of the seventy-six letter strings, forty-seven are suffixes. Thus the procedures used in constructing these tables produce a relatively high proportion of affixes compared to the total number of letter strings corresponding to admissible classes.

The set of affixes that compose Table 3 is somewhat different from the set of affixes found in Part I from the two-vowel-string corpus. There are fifteen prefixes that appear in both Part I and Table 3 of Part II, but Part I lists the six prefixes

BE-, CY-, I-, OUT-, SUN-, TRANS-,

that do not appear in Table 3, while the seven prefixes

AN-, OB-, OVER-, PRO-, PU-, SE-, VI-,

are in Table 3 but not in Part I. Of these latter, OVER- is a two-vowel-string prefix and so could not have appeared in Part I.

There are twenty-six suffixes that are common to Part I and Table 3 of Part II. The following twenty-five suffixes are in Part I but not in Part II:

-ED, -LAND, -ARD, -WARD, -EE, -IE,
-ING, -LING, -AH, -OCK, -LOCK, -EL,
-MAN, -EN, -EON, -IER, -LER, -LESS,
-IS, -NESS, -AT, -LET, -OT, -OW, -EY,

and twenty-one suffixes are in Table 3 of Part II but not in Part I:

-ANCE, -ENCE, -IDE, -ABLE, -IBLE,
-ISE, -OSE, -ATE, -IZE, -ICAL, -IAL,
-ISM, -IUM, -IAN, -ATION, -ESS, -OUS,
-IOUS, -ARY, -ERY, -RY.

Of these, -ICAL, -ATION, -ARY, and -ERY are two-vowel-string suffixes, and so could not have appeared in Part I.

### Difficulty of Vowel-String Decomposition

Our procedures have been based on the recognition of inadmissible consonant strings in English words. The essential hypothesis regarding strong affixes is that an inadmissible consonant string implies the existence of either a compounding unit or an affix whose point of attachment in the word lies in the inadmissible consonant string.

We will now consider what happens if this idea is modified to admit the consideration of inadmissible vowel strings, and the corresponding hypothesis. Figure 5 of reference 4 graphically shows that the only admissible multiletter English vowel strings are

AI, AU, AY, EA, EE, EI,
IE, OA, OI, OO, OU;

all others are inadmissible. Using the obvious modifications of the definitions above, and applying them to the corpus K, certain new classes are joined to the collection of admissible classes in Tables 1 and 2.

Only suffix classes will be treated in detail. All of the suffix classes obtained from K by means of an inadmissible vowel-string decomposition are listed in Table 4. These lead to only four new suffixes, namely,

-ALIZE, -AR, -ATOR, -ALIST.

### TABLE 3

#### PREFIXES AND SUFFIXES WITH RESPECT TO K

**PREFIXES**

| | | |
|---|---|---|
| A- | E- | PU- |
| AC- | EN- | RE- |
| AD- | EX- | SE- |
| AL- | IN- | SUB- |
| AN- | MIS- | UN- |
| CON- | OB- | VI- |
| DE- | OVER- | |
| DIS- | PRO- | |

**SUFFIXES**

| | | | | |
|---|---|---|---|---|
| -A | -URE | -IAL | -ATION | -ENT |
| -IC | -ISE | -FUL | -O | -MENT |
| -ANCE | -OSE | -ISM | -ER | -Y |
| -ENCE | -ATE | -UM | -OR | -LY |
| -IDE | -ITE | -IUM | -ESS | -ARY |
| -ACE | -IVE | -AN | -US | -ERY |
| -ABLE | -IZE | -IAN | -OUS | -RY |
| -IBLE | -ISH | -IN | -IOUS | |
| -ILE | -ICAL | -ON | -ET | |
| -INE | -AL | -ION | -ANT | |

TABLE 4

## VOWEL-STRING-DECOMPOSITION DATA FOR SUFFIXES IN K

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| -ENA | (A/ENA) | 7 | -ALISM | (U/ALISM) | 8 | -US | (E/US) | 10 |
| -AC | (I/AC) | 12 | -UM | (E/UM) | 17 | | (I/US) | 12 |
| -OTIC | (I/OTIC) | 4 | | (I/UM) | 183 | -OLUS | (E/OLUS) | 4 |
| -AD | (I/AD) | 7 | -EAN | (A/EAN) | 6 | -OUS | (E/OUS) | 104 |
| -ANCE | (I/ANCE) | 9 | -AN | (I/AN) | 360 | | (I/OUS) | 166 |
| -ABLE | (I/ABLE) | 40 | -ON | (E/ON) | 19 | | (U/OUS) | 36 |
| | (U/ABLE) | 5 | | (I/ON) | 396 | -ACEOUS | (I/ACEOUS) | 9 |
| -ATE | (I/ATE) | 96 | -ATION | (I/ATION) | 21 | -ITOUS | (U/ITOUS) | 5 |
| | (U/ATE) | 27 | | (U/ATION) | 12 | -ET | (U/ET) | 7 |
| -OLITE | (I/OLITE) | 4 | -O | (I/O) | 32 | -ANT | (I/ANT) | 13 |
| -ATIVE | (I/ATIVE) | 6 | -OSO | (I/OSO) | 5 | | (U/ANT) | 5 |
| -ALIZE | (I/ALIZE) | 7 | -AR | (I/AR) | 4 | -ENT | (U/ENT) | 15 |
| | (U/ALIZE) | 5 | -ER | (U/ER) | 5 | -OT | (I/OT) | 5 |
| -ONIZE | (I/ONIZE) | 5 | -ONER | (I/ONER) | 6 | -AST | (I/AST) | 7 |
| -ARCH | (I/ARCH) | 4 | -OR | (I/OR) | 8 | -ALIST | (I/ALIST) | 5 |
| -UCH | (E/UCH) | 4 | -ATOR | (I/ATOR) | 6 | | (U/ALIST) | 6 |
| -AL | (I/AL) | 142 | | (U/ATOR) | 4 | -ONIST | (I/ONIST) | 7 |
| | (U/AL) | 28 | -ASIS | (I/ASIS) | 9 | -IST | (U/IST) | 4 |
| -ONAL | (I/ONAL) | 13 | -ONLESS | (I/ONLESS) | 6 | -ARY | (I/ARY) | 23 |
| | | | | | | | (U/ARY) | 19 |

Comparing this with the number of suffixes previously obtained from K, that is, forty-seven suffixes, indicates that the vowel decomposition is a relatively unproductive way to search for affixes. In fact, of the four suffixes listed above, both -ALIZE and -ATOR can be decomposed into sequences of suffixes already obtained. We have -AL-IZE and -AT-OR. The suffix -AR is new, but -ALIST appears to the intuition to be the sequence -AL-IST; unfortunately, none of the techniques that have been described thus far has managed to produce the sequence -IST as a suffix. This must be considered a defect of the methods described, but it is clearly as much of a defect for the vowel-decomposition technique as for the earlier described consonant-decomposition method. In a later section we will introduce still another procedure which will produce -IST in a natural way. Noting that -AR appears in the suffix tables in Part I will permit us to interpret each of the four suffixes given above either as a suffix from Part I or a sequence of suffixes produced by either the consonant-decomposition method or by the still to be described technique. Hence we can conclude that nothing is gained by the introduction of the vowel-string-decomposition procedure discussed in this section, and so henceforth this method will not be used. There is a more serious reason for restricting the affix-defining procedures to consonant strings. Table 4 lists the forty-four distinct letter strings for which there are admissible suffix classes with vowel-string-decomposition points. Of these letter strings, fully twenty are two-vowel-string sequences. The corresponding data for Table 2 are seventy-six letter strings of which ten are two-vowel-string sequences. This shows that the inadmissible vowel-string decomposition is relatively much more sensitive to two-vowel-string affixes (or to sequences of one-vowel-string affixes) than to one-vowel-string affixes. This is reflected in the fact that three of the four new affixes derived from vowel-string decompositions are two-vowel-string affixes. The combination of insensitivity to one-vowel-string affixes and low rate of production of affixes makes it probable that the mechanism involved in vowel-string decompositions is different from that for consonant-string decompositions, and so it seems most wise to try to keep these two notions well separated, at least until they are better understood.

**Parasitic Affixes**

There are two popular vowel-beginning letter sequences which intuition would undoubtedly call suffixes, but which did not appear as weak suffixes in Part I. They are

TABLE 5

### TABLE 5

PARASITIC SUFFIX DATA FOR THE CORPUS K

**Column 1**

| Suffix | Pair | Count |
|---|---|---|
| -A | (IC/A) | 17 |
| | (ING/A) | 5 |
| | (AL/A) | 12 |
| | (EL/A) | 5 |
| | (AN/A) | 18 |
| | (EN/A) | 17 |
| | (IN/A) | 26 |
| | (ON/A) | 8 |
| | (AR/A) | 17 |
| | (ER/A) | 14 |
| | (OR/A) | 13 |
| | (US/A) | 4 |
| | (AT/A) | 9 |
| | (ET/A) | 4 |
| | (OT/A) | 4 |
| -IA | (AL/IA) | 5 |
| | (EL/IA) | 6 |
| | (AN/IA) | 8 |
| | (EN/IA) | 6 |
| | (IN/IA) | 6 |
| | (ON/IA) | 12 |
| | (AR/IA) | 9 |
| | (ER/IA) | 6 |
| | (OR/IA) | 5 |
| -IC | (AL/IC) | 11 |
| | (EL/IC) | 6 |
| | (AN/IC) | 20 |
| | (EN/IC) | 10 |
| | (IN/IC) | 14 |
| | (ON/IC) | 54 |
| | (ION/IC) | 6 |
| | (AR/IC) | 9 |
| | (ER/IC) | 17 |
| | (OR/IC) | 10 |
| | (AT/IC) | 58 |
| | (ET/IC) | 47 |
| | (ANT/IC) | 6 |
| | (OT/IC) | 36 |
| -ED | (ER/ED) | 7 |
| | (AT/ED) | 10 |
| -OID | (ER/OID) | 7 |
| | (AT/OID) | 5 |
| -ANCE | (EN/ANCE) | 10 |
| | (ON/ANCE) | 4 |
| | (ER/ANCE) | 12 |
| -ENCE | (IN/ENCE) | 7 |
| | (ER/ENCE) | 12 |
| -ADE | (ON/ADE) | 10 |
| -IDE | (IC/IDE) | 18 |
| -ODE | (AT/ODE) | 4 |
| -UDE | (ET/UDE) | 5 |
| -EE | (AR/EE) | 7 |
| | (ENT/EE) | 4 |
| -AGE | (AN/AGE) | 5 |
| | (EN/AGE) | 5 |
| | (IN/AGE) | 4 |
| | (ON/AGE) | 10 |

**Column 2**

| Suffix | Pair | Count |
|---|---|---|
| | (AR/AGE) | 4 |
| | (ER/AGE) | 15 |
| | (OR/AGE) | 4 |
| | (US/AGE) | 4 |
| | (OT/AGE) | 4 |
| -IE | (ER/IE) | 14 |
| -ALE | (ING/ALE) | 6 |
| -ABLE | (IC/ABLE) | 14 |
| | (EN/ABLE) | 7 |
| | (IN/ABLE) | 9 |
| | (ON/ABLE) | 6 |
| | (AR/ABLE) | 6 |
| | (ER/ABLE) | 8 |
| | (US/ABLE) | 4 |
| | (AT/ABLE) | 5 |
| -IBLE | (IS/IBLE) | 5 |
| | (US/IBLE) | 4 |
| -ILE | (AT/ILE) | 8 |
| -OLE | (AR/OLE) | 6 |
| -ULE | (IC/ULE) | 15 |
| -OME | (OT/OME) | 8 |
| -INE | (IC/INE) | 9 |
| | (ARD/INE) | 5 |
| | (AL/INE) | 7 |
| | (EL/INE) | 13 |
| | (UM/INE) | 5 |
| | (AN/INE) | 5 |
| | (IN/INE) | 5 |
| | (ON/INE) | 4 |
| | (AR/INE) | 14 |
| | (ER/INE) | 18 |
| | (OR/INE) | 6 |
| | (AT/INE) | 7 |
| | (ANT/INE) | 11 |
| | (ENT/INE) | 11 |
| | (OT/INE) | 9 |
| -URE | (AT/URE) | 34 |
| | (ENT/URE) | 7 |
| -OSE | (IC/OSE) | 6 |
| | (ER/OSE) | 5 |
| -ATE | (IC/ATE) | 47 |
| | (AL/ATE) | 4 |
| | (EL/ATE) | 8 |
| | (UM/ATE) | 4 |
| | (AN/ATE) | 10 |
| | (EN/ATE) | 7 |
| | (IN/ATE) | 53 |
| | (ON/ATE) | 20 |
| | (AR/ATE) | 4 |
| | (ER/ATE) | 34 |
| | (OR/ATE) | 26 |
| | (ET/ATE) | 4 |
| | (ENT/ATE) | 9 |
| -ITE | (AL/ITE) | 11 |
| | (EL/ITE) | 14 |
| | (AN/ITE) | 16 |

**Column 3**

| Suffix | Pair | Count |
|---|---|---|
| | (EN/ITE) | 15 |
| | (IN/ITE) | 18 |
| | (ON/ITE) | 36 |
| | (AR/ITE) | 6 |
| | (ER/ITE) | 37 |
| | (OR/ITE) | 7 |
| | (IS/ITE) | 6 |
| | (AT/ITE) | 10 |
| | (ET/ITE) | 4 |
| | (ANT/ITE) | 5 |
| -ETTE | (IN/ETTE) | 5 |
| | (ON/ETTE) | 7 |
| -UE | (EN/UE) | 4 |
| | (IN/UE) | 4 |
| -IVE | (IS/IVE) | 5 |
| | (US/IVE) | 23 |
| | (AT/IVE) | 61 |
| | (ET/IVE) | 12 |
| | (ENT/IVE) | 13 |
| | (OT/IVE) | 4 |
| -IZE | (IC/IZE) | 13 |
| | (ARD/IZE) | 6 |
| | (AL/IZE) | 46 |
| | (EL/IZE) | 6 |
| | (AN/IZE) | 16 |
| | (EN/IZE) | 5 |
| | (IN/IZE) | 8 |
| | (ON/IZE) | 28 |
| | (ION/IZE) | 5 |
| | (AR/IZE) | 13 |
| | (ER/IZE) | 13 |
| | (OR/IZE) | 22 |
| | (AT/IZE) | 11 |
| | (ET/IZE) | 7 |
| | (OT/IZE) | 6 |
| -ING | (ET/ING) | 6 |
| -EUCH | (AT/EUCH) | 4 |
| -OUGH | (OR/OUGH) | 5 |
| -ISH | (AN/ISH) | 4 |
| | (EN/ISH) | 5 |
| | (ON/ISH) | 7 |
| | (ER/ISH) | 10 |
| -I | (AL/I) | 6 |
| | (AR/I) | 12 |
| -ICK | (ER/ICK) | 4 |
| -AL | (IC/AL) | 72 |
| | (AN/AL) | 4 |
| | (EN/AL) | 5 |
| | (IN/AL) | 34 |
| | (ON/AL) | 10 |
| | (ION/AL) | 13 |
| | (ER/AL) | 23 |
| | (OR/AL) | 22 |
| | (IS/AL) | 9 |
| | (US/AL) | 7 |
| | (AT/AL) | 4 |
| | (ET/AL) | 7 |
| | (ENT/AL) | 10 |

**-EAL**
(ER/EAL) 4
(OR/EAL) 5

**-IAL**
(ED/IAL) 4
(ON/IAL) 4
(AR/IAL) 6
(ER/IAL) 8
(OR/IAL) 26
(ENT/IAL) 14

**-EL**
(IC/EL) 4
(ON/EL) 5
(ER/EL) 11

**-IL**
(IC/IL) 4

**-OL**
(IN/OL) 4
(ER/OL) 7

**-IM**
(AT/IM) 4

**-OM**
(ED/OM) 7

**-ISM**
(IC/ISM) 19
(AL/ISM) 31
(AN/ISM) 20
(EN/ISM) 9
(IN/ISM) 13
(ON/ISM) 18
(AR/ISM) 7
(ER/ISM) 14
(OR/ISM) 8
(AT/ISM) 12
(ET/ISM) 6
(OT/ISM) 11

**-UM**
(IC/UM) 5
(AN/UM) 9
(IN/UM) 4
(AR/UM) 4
(AT/UM) 5
(ENT/UM) 7

**-IUM**
(AN/IUM) 5
(EN/IUM) 7
(IN/IUM) 7
(ON/IUM) 11
(AR/IUM) 17
(ER/IUM) 4
(OR/IUM) 8
(AT/IUM) 5

**-YM**
(ON/YM) 11

**-AN**
(IC/AN) 11
(ER/AN) 7

**-IAN**
(IC/IAN) 18
(AL/IAN) 10
(EL/IAN) 9
(AN/IAN) 13
(EN/IAN) 10
(IN/IAN) 10
(ON/IAN) 49
(AR/IAN) 31
(ER/IAN) 24
(OR/IAN) 7
(US/IAN) 5
(AT/IAN) 6

**-EN**
(UM/EN) 6

**-IN**
(IC/IN) 9
(AL/IN) 5
(EL/IN) 11
(AN/IN) 5
(ON/IN) 6
(AR/IN) 10
(ER/IN) 7
(OR/IN) 5
(AT/IN) 7
(ET/IN) 9

**-ON**
(IC/ON) 9
(ED/ON) 4
(EL/ON) 4
(ER/ON) 10
(IS/ON) 21
(ET/ON) 9

**-ION**
(EL/ION) 6
(ER/ION) 5
(IS/ION) 19
(US/ION) 31
(AT/ION) 223
(ET/ION) 12
(ENT/ION) 14
(OT/ION) 9

**-ORN**
(IC/ORN) 7

**-O**
(IC/O) 10
(ED/O) 9
(AL/O) 6
(AN/O) 7
(IN/O) 12
(AR/O) 4
(ER/O) 10
(AT/O) 11
(ET/O) 7
(ENT/O) 5

**-ARP**
(IC/ARP) 4

**-ER**
(IC/ER) 4
(AL/ER) 4
(EN/ER) 19
(IN/ER) 13
(ON/ER) 14
(ION/ER) 6
(ER/ER) 40
(IS/ER) 4
(AT/ER) 16
(ET/ER) 18
(ANT/ER) 8
(ENT/ER) 10
(OT/ER) 4
(OW/ER) 15

**-EER**
(IN/EER) 5
(ON/EER) 4
(ET/EER) 9

**-IOR**
(ER/IOR) 7

**-OR**
(IS/OR) 8
(US/OR) 4
(AT/OR) 51
(ENT/OR) 8

**-IS**
(AL/IS) 4
(AR/IS) 4

**-ESS**
(ER/ESS) 6
(OR/ESS) 6

**-US**
(EL/US) 5
(AN/US) 5
(EN/US) 5
(ER/US) 7
(OR/US) 6

**-OUS**
(IC/OUS) 5
(AL/OUS) 6
(EL/OUS) 4
(AN/OUS) 7
(EN/OUS) 7
(IN/OUS) 17
(ON/OUS) 12
(ER/OUS) 38
(OR/OUS) 32
(ET/OUS) 4

**-EOUS**
(IC/EOUS) 6
(AN/EOUS) 6
(IN/EOUS) 8

**-IOUS**
(IC/IOUS) 9
(ON/IOUS) 4
(AR/IOUS) 13
(OR/IOUS) 8
(ENT/IOUS) 7

**-ET**
(EL/ET) 27
(AN/ET) 7
(EN/ET) 5
(IN/ET) 14
(ON/ET) 9
(AR/ET) 9
(ER/ET) 13

**-IT**
(IC/IT) 7
(ED/IT) 4
(ER/IT) 6

**-ANT**
(IC/ANT) 16
(EN/ANT) 7
(IN/ANT) 12
(ON/ANT) 5
(ER/ANT) 8
(OR/ANT) 9
(AT/ANT) 4

**-ENT**
(ED/ENT) 4
(ING/ENT) 5
(AL/ENT) 7
(AN/ENT) 5
(IN/ENT) 7
(ON/ENT) 7
(AR/ENT) 4
(ER/ENT) 12

**-IENT**
(IC/IENT) 8
(ED/IENT) 4
(ENT/IENT) 4

**-OT**
(IC/OT) 4

**-EST**
(ER/EST) 4
(OR/EST) 4

**-IST**
(IC/IST) 7
(AL/IST) 28
(AN/IST) 14
(EN/IST) 6
(IN/IST) 9
(ON/IST) 15
(ION/IST) 7
(AR/IST) 8
(ER/IST) 4
(OR/IST) 14
(AT/IST) 8
(ET/IST) 5

**-Y**
(ED/Y) 4

| | | | | | | |
|---|---|---|---|---|---|---|
| (AN/Y) | 12 | (AR/Y) | 135 | | (ET/Y) | 24 |
| (EN/Y) | 8 | (ER/Y) | 198 | | (OW/Y) | 4 |
| (IN/Y) | 8 | (OR/Y) | 31 | | | |
| (ON/Y) | 35 | (IS/Y) | 4 | -IFY | (AT/IFY) | 5 |

-ISM and -IST.

One can say that these sequences are not generally attached to one-vowel-string sequences to form two-vowel-string words. The data in Table 2 show that -ISM appears as a suffix for the three-vowel-string corpus K, but that -IST still does not turn out to be a suffix with respect to K. It can be concluded that while -ISM can be generally attached as a suffix to two-vowel-string sequences to form three-vowel-string words, this is not true of -IST. However, it turns out that there are twelve admissible classes of the form Cls(X/IST) where X denotes a consonant-ending suffix with respect to the two-vowel-string corpus investigated in Part I. The classes are

| | | | |
|---|---|---|---|
| Cls(IC/IST) | 7 | Cls(ON/IST) | 15 |
| Cls(AL/IST) | 28 | Cls(AR/IST) | 8 |
| Cls(AN/IST) | 14 | Cls(ER/IST) | 4 |
| Cls(EN/IST) | 6 | Cls(OR/IST) | 14 |
| Cls(IN/IST) | 9 | Cls(AT/IST) | 8 |
| Cls(ION/IST) | 7 | Cls(ET/IST) | 5. |

In each case the suffix ends with a single consonant which is both an admissible initial and an admissible final consonant, and so these classes make no contribution to the set of affixes produced by the definitions above.

Suffixes can be thought of as forming a natural generalization of the notion of admissible final-consonant strings which are not also admissible initial-consonant strings, unless, of course, the suffix is simultaneously a prefix (for example, A, AL, AN, etc.). If it is agreed that a prefix-suffix ambiguity occurring internally in a word cannot be a prefix (resp. suffix) unless it is preceded (resp. followed) by another prefix (resp. suffix), then the procedures used to define the weak affixes can be extended in a natural way to produce intuitively reasonable suffixes like -IST. In particular, affixes produced by such a procedure are generally found attached to other affixes. Hence they will be called *parasitic affixes.* Furthermore, parasitic affixes with respect to a three-vowel-string corpus cannot have more than one vowel string. For otherwise words of the corpus defining the parasitic affixes would consist entirely of affixes, which does not occur admissibly in English.

Another restriction occurring in the following definitions will be explained after they are stated.

*Definition P3.* Let $P = C_1 V_1$ be a fixed-letter sequence in initial position. P is a *parasitic prefix* (with respect to K) if there exist two distinct classes of words from K, Cls(P/P') and Cls(P/P''), each of which contains more than three words, such that P' and P" are prefixes with respect to the two-vowel-string corpus investigated in Part I.

*Definition S3.* Let $S = V_3 C_4$ be a fixed-letter sequence in final position, S is a *parasitic suffix* (with respect to K) if there exist two distinct classes of words from K, Cls(S'/S) and Cls(S"/S), each of which contains more than three words, such that S' and S" are suffixes with respect to the two-vowel-string corpus investigated in Part I.

Note that the definitions require that a parasitic prefix (resp. parasitic suffix) end (resp. begin) with a vowel. For otherwise we should expect to have found the affix using the consonant-decomposition-point method outlined above.

The English language forms the majority of its word inventory by attachment of successive prefixes and suffixes to short admissible forms. Although there are many words that contain sequences of prefixes, it is far more common to observe several suffixes in sequence in long words. In this sense, the investigation of parasitic suffixes assumes somewhat greater importance than the corresponding investigation of parasitic prefixes.

Table 5 gives the parasitic suffix data consisting of admissible classes for the corpus K. There are seventy-seven letter sequences represented. Of these, fifty-three are parasitic suffixes. The following twelve are new, that is, they do not appear in Part I or in Table 3 of this part.

-IA, -OID, -ETTE, -I, -EAL, -OL,
-EER, -EOUS, -IT, -IENT, -EST, -IST.

Note in particular that -IST is a parasitic suffix. The present study has shown that -IST is not obtained as a suffix with respect to the two-vowel-string corpus (of Part I), and that it does not precede suffixes in the corpus K. This latter fact can be deduced from the data in Table 5. But it would be erroneous to infer that -IST can only occur in final position, for examination of the four-vowel-string corpus in *Speculum 3* shows, for instance, that -IST precedes -IC. This simply means that in general -IST is not attached to one-vowel-string letter sequences to form English words.

The typical size of classes in Table 5 seems to be about the same as for the classes in Table 2. But the suffix -Y corresponds (in Table 5) to the classes QS(AR/Y) and Cls(ER/Y) with 135 and 198 members, respectively. These extremely populous classes contain the sequence -RY, which is a suffix with respect to K, but not with respect to the two-vowel-string corpus of Part I. It is likely that instances of -A-RY and -E-RY are

mixed in with those of -AR-Y and -ER-Y in Table 5. This does not matter for the questions that have been studied thus far, since -Y is shown to be a parasitic suffix by the existence of nine other admissible classes where such ambiguities do not arise.

*Received February 10, 1966*

## References

1. Resnikoff, H. L., and Dolby, J. L. "The Nature of Affixing in Written English," *Mechanical Translation,* Vol. 8 (1965), pp. 84-89.
2. Monroe, G. K. "Phonemic Transcription of Graphic Post-base Affixes in English: A Computer Problem," Dissertation, Brown University, 1965.
3. Dolby, J. L., and Resnikoff, H. L. *The English Word Speculum,* Vol. 2: *The Forward Word List;* Vol. 3: *The Reverse Word List.* Sunnyvale, Calif.: Lockheed Missiles & Space Co., 1964.
4. ---------- . "On the Structure of Written English Words," *Language,* Vol. 40 (1964), pp. 167-196.