COMPUTER-PRODUCED REPRESENTATION OF DIALECTAL VARIATION: INITIAL FRICATIVES IN SOUTHERN BRITISH ENGLISH

W. Nelson Francis, Jan Svartvik, and Gerald M. Rubin
Department of Linguistics, Brown University

It has become apparent to some dialectologists that dialectology, particularly in its interpretive phase, is a branch of linguistics particularly adapted to the use of computers. The dialectologist typically deals with large bodies of data, usually in the form of single words and short phrases, and he is interested in sorting and comparing individual items on many bases: phonological, morphological, lexical, and geographical. The major obstacle that has prevented widespread use of computers in dialect study is the fact that the data for most of the great dialect surveys have been collected, recorded, and in most cases edited prior to the computer age.[1] Thus the problem of preparing large bodies of data, much of it in narrow phonetic transcription, for computer use has been formidable. One of the aims of the present paper is to show that results can be obtained relatively easily by computerized sorting and mapping that would take endless hours by traditional methods, and hopefully to encourage others to invest time and money in preparing data for the computer rather than in time-consuming hand sorting and map-making.

Accordingly we sought a problem that would be complex enough to reveal the advantages of computerized dialectology while at the same

ıe involving a body of data small enough to be quickly prepared. ıce two of the three authors are specialists in English (the third a computer specialist), we naturally turned to the published lumes of the *Survey of English Dialects*,[2] which embody carefully ıtrolled data, collected with professional skill, and presented in ıvenient tabular form in meticulously edited and printed volumes. ı since one of the two areas covered by the volumes in print at the ıe the study was undertaken (May 1969) was the south of England, ; problem of the voicing of initial fricatives in the southwest ;urally suggested itself. This problem had the further advantage, ; our purposes, of dealing with consonants (simpler than vowels in ;t varieties of English) in initial position, hence easily sorted ı examined. The selection of this problem has proved to be a ;py one.

The area covered by Volume 4 of *SED* comprises the ten southern- ;t counties of England, which, with their key numbers in the *Survey*, ; 31 Somersetshire, 32 Wiltshire, 33 Berkshire, 34 Surrey, 35 Kent, Cornwall, 37 Devonshire, 38 Dorsetshire, 39 Hampshire, and 40 ;sex. A total of 75 localities in this area were reported on by ; *Survey*; in what follows these will be identified by a four-digit ıber, the first two digits indicating the county and the remaining ; the locality. Thus 3906 stands for Burley, the sixth locality ;ted in Hampshire, according to the numbered list on p. 31 of the ;roduction to the *Survey*. The data selected for examination

included all those words beginning with graphic *f-*, *s-*, or *th-* followed by a vowel or voiced consonant which were starred in the *SED* questionnaire.[3] To this list we later added a few non-starred words which showed universal distribution and were otherwise of interest. The final list contained 68 words, of which 27 are *f*-words, 22 *s*-words, 16 *th*-words and 3 *sh*-words (i.e. words beginning with /š/ in standard English). We took only the first recorded form from each locality; this is presumably a citation form, produced by an informant in response to a question, and recorded in narrow IPA transcription. The 59 cases where no response was given were coded XXX in our computer code.

The corpus thus comprised 68 x 75 or 5100 items, including the 59 blanks. Our computer expert then produced 68 decks of punch-cards, one for each word, each deck containing 75 cards, one for each locality. These were numbered at the left for locality and on the right for the reference number of the item in the *SED* questionnaire.[4] A coding system was devised which preserved all significant features of the phonetic transcription while passing over apparently irrelevant fine points (see Appendix A), and the words were transcribed in this code directly onto the cards for the guidance of the key-puncher, who then punched the coded words in a fixed place on the cards. Subsequently the standard spelling was inserted by the computer to the left of the coded phonetic spelling. This whole process took about a dozen hours of the investigators' time (not counting the

latively simple programming involved) and about the same amount of
e key-puncher's time. The result was a body of data consisting of
)0 entries of the following sort:

)1 FINGER F19G)R. 6 7 7

:s is to be interpreted as indicating that at locality 3101 (Weston
Somerset) the word *finger*, which appears as item VI.7.7 of the
) questionnaire, is pronounced [fɪngəɹ·] (or perhaps more accu-
:ely /fɪngər·/ in the quasi-phonemic transcription used).

The nature of the problem with which we are dealing may be most
ıply introduced by an excerpt from the full treatment given to the
.cing of initial fricatives in Middle English by Horn and Lehnert
'54, Vol. II, §437):

> In gewissen Mundarten sind in alter Zeit im Wortanlaut die starken und stimmlosen Reibelaute *f*, *s* und *þ* vor Vokalen und schwachen, stimmhaften Konsonanten schwach und stimmhaft geworden: *f-* > *v-*, *s-* > *z-*, *þ-* > *đ-* . . . Die neuenglische Hochsprache hat einige Wörter aus den Mundarten aufgenommen. . .
>
> Die starken und stimmlosen Reibelaute im Wortanlaut sind in Kent und im südöstlichen Mittelland schwach und stimmhaft geworden. In me. Handschriften wird *v* für anlautendes *f* vor vor Vokalen geschrieben, *z* für anlautendes *s* vor Vokalen, . . . Wir dürfen annehmen, daß auch stimmloses *th-* stimmhaft geworden ist, da heutige Mundarten für *th-* ebenso wie für *f-* und *s-* stimmhafte Laute haben. . .
>
> Aus der Tatsache, daß französische Lehnwörter im Me. *f-* und *s-* beibehalten . . . geht hervor, daß die stimmlosen Reibelaute in den englischen Wörtern schon vor der Aufnahme der französischen Lehnwörter stimmhaft geworden waren. . .

> Der Lautwandel hat sich im Laufe der Zeit von Kent aus über die südlichen und angrenzenden östlichen Grafschaften ausgedehnt, und zwar hat er heute in diesen Gebieten anlautendes *þ* ebenso ergriffen wie *f* und *s*. Vereinzelt ist auch anlautendes *š* zu *ž* geworden. . . In den heutigen Mundarten von Kent, Surrey und Sussex sind die anlautenden stimmhaften Laute unter dem Einfluss des Hochenglischen durch stimmlose ersetzt worden, während das östliche Herefordshire, Teile von Gloucestershire, das westliche Berkshire, und besonders Wiltshire, Dorsetshire, Somersetshire und Devonshire stimmhafte Reibelaute im Anlaut aufweisen.

Horn-Lehnert go on to point out that by analogy and under the influence of Standard English, initial *f-* and *s-* in French loanwords have become voiced. This is borne out in our corpus by the words *flour, farmer,* and *flowers,* which have initial *v-* in more than half the localities included. As our subsequent discussion and maps will show, Horn-Lehnert should have included in the initial voiced fricative area the western half of Hampshire and all but the southwest tip of Cornwall. Since the *SED* records for Herefordshire and Gloucestershire have not yet been published, we have not been able to include these counties in our survey.

The traditional assumptions, then, are that *f-*, *s-*, and *þ-* became voiced (except before voiceless consonants) in initial position at an early date—certainly before the 13th century—in the southeast; that this affected all native words; that this change subsequently spread into the old West Saxon area of the southwest; that

er this spread voiceless initial fricatives were restored in the theast; and that analogy, dialect borrowing, and the influence of ndard English worked variously to blur the exceptionless charac- of the sound change, to produce voicing of initial *f-* and *s-* French loanwords and of initial *š-* in native words like *shilling* well as French words like *sure*, and otherwise to create a mixed uation in the whole southern area.[5] Our project was to see what ht the records of the *SED* can throw on this situation by exploit- them in some of the many ways made possible by computer techno- y.

The first step was to sort the data in as many ways as we felt ld be productive. Accordingly our computer expert produced four ts of the 5100 items, sorted as follows:

List I: sorted first by keyword (the standard English graphic word identifying the item) alphabetically; then by locality. This list presents the data in the same kind of order in which it appears in the *SED* records and allows easy inspection of all versions of each word in one list.

List II: sorted first by locality and then by keyword alphabetically. This list brings together in one place all the records from each locality and thus permits comparison of the amount of voicing recorded from various localities.

List III: sorted first by locality and then by citation (the recorded local form) alphanumerically. This list even more

graphically reveals the amount of initial voicing; it also makes it easy to surmise from inspection whether or not incidence of initial voicing might be influenced by the following vowel or consonant.

List IV: sorted first on the second and following characters of the citation form, in linguistically significant order (i.e. by vowels and consonants in articulatory order), then by the first character. This list greatly facilitates looking into the question of the possible influence of following sounds on initial voicing.

These lists, though interesting in themselves, were considered primarily as intermediate diagnostic procedures, to be used to guide us in future sorting, counting, and eventually mapping. Even the most cursory inspection of them revealed what we had suspected from our first look at the data: that there is tremendous variation within the relevant area both from word to word and from locality to locality. It certainly seemed, at least so far as this feature in this region is concerned, that the maxim attributed to Gilliéron, "Every word has its own history," is true.

Accordingly we asked our computer expert for various more sophisticated sortings and counts, and for two kinds of maps, to be produced by the CalComp plotter (see Appendix B). These were the following:

1. A table for each of the four sets of words, listing all words

he set and counting the number of each occurring initial conson-
the words to be ranked in descending order of the number of
ial voiced consonants recorded. These are reproduced here as
es 1 - 4.

2. A list of all possible vowels, diphthongs, and second conson-
in each set of words, counted and tabulated in terms of each
ible initial consonant. A portion of this list is reproduced
with as Table 5.

3. An individual map for each of the 68 words, showing the
ial consonant recorded for each locality. Eight of these are
oduced here as Maps 6-13.

4. Four proportional maps, one for each set of words, indica-
at each locality the voiceless:voiced ratios; thus for the
ords the legend 3/13 occurring at 3905 indicates that at that
lity only three of the 16 *th*-words begin with voiceless conson-
. Map 1 shows the proportional map for the *s*-words as it came
the plotter, while maps derived from these proportional maps to
al the varying distributions more clearly are included as Maps
5.

Tables 1 - 4 support our suspicion that each word has its own
ie distribution with regard to the initial consonant. Thus Table
ows that among the *f*-words the proportion of voiceless to voiced
es from 20:54 in FELLIES to 52:22 in FOAL. Even in those cases
e the proportions are the same, recourse to List I reveals that

TABLE 1, F-Words

Word	Ref.	[v]	[ð]	[f̬]	[f]	[θ]	Vcls. Total	Vcd. Total	Missing
1. FELLIES	1.9.9	49	5	ø	19	1	20	54	1
2. FURROW	2.3.1	53	ø	ø	21	ø	21	53	1
3. FOOT	6.10.1	46	ø	1	28	ø	28	47	ø
4. FRIDAY-1	7.4.4	46	ø	ø	26	3	29	46	ø
5. FERN	4.10.13	45	ø	ø	30	ø	30	45	ø
6. FINGER	6.7.7	45	ø	ø	30	ø	30	45	ø
7. FIRE	5.3.1	45	ø	ø	30	ø	30	45	ø
8. FRIDAY-2	7.4.7	44	ø	ø	28	ø	28	44	3
9. FIVE	7.5.6	42	ø	ø	33	ø	33	42	ø
10. FLEAS	4.8.4	42	ø	ø	33	ø	33	42	ø
11. FROGS	4.9.6	42	ø	ø	33	ø	33	42	ø
12. FIND	9.3.2	41	ø	ø	34	ø	34	41	ø
13. FOX	4.5.11	41	ø	ø	34	ø	34	41	ø
14. FURTHER	9.2.1	41	ø	ø	34	ø	34	41	ø
15. FLIES	4.8.5	40	ø	ø	34	ø	34	40	1
16. FLOUR	5.6.1	40	ø	ø	35	ø	35	40	ø
17. FIRST	7.2.1	39	ø	ø	36	ø	36	39	ø
18. FLOOR	5.2.7	38	ø	1	36	ø	36	39	ø
19. FARMER	8.4.7	38	ø	ø	37	ø	37	38	ø
20. FLITCH	3.12.3	38	ø	ø	26	ø	26	38	11
21. FLOWERS	8.5.13	36	ø	ø	39	ø	39	36	ø
22. FORTY	7.1.14	35	ø	ø	40	ø	40	35	ø
23. FATHER	8.1.1.	31	ø	ø	44	ø	44	31	ø
24. FIGHT	3.13.6	31	ø	ø	43	ø	43	31	1
25. FORD	4.1.3	22	ø	1	41	ø	41	23	11
26. FOAL	3.4.1	22	ø	ø	52	ø	52	22	1
27. FORKS	1.7.9	22	ø	ø	40	ø	40	22	13

TABLE 2, S-Words

8B

rd	Ref.	[z]	[ž]	[s]	[š]	Vcls. Total	Vcd. Total	Missing
TURDAY	7.4.5	46	ø	29	ø	29	46	ø
E	6.3.2	46	ø	29	ø	29	46	ø
X	7.1.5	42	1	32	ø	32	43	ø
W-N	3.8.6	43	ø	31	ø	31	43	1
CK	3.7.1	42	ø	32	ø	32	42	1
VEN	7.1.6	41	ø	34	ø	34	41	ø
DDLE	1.5.6	40	ø	33	ø	33	40	2
UTH	7.6.25	40	ø	35	ø	35	40	ø
GHT	8.2.9	39	ø	36	ø	36	39	ø
LVER	7.7.7	39	ø	35	ø	35	39	1
W	5.10.3	37	ø	38	ø	38	37	ø
OT	5.4.6	37	ø	38	ø	38	37	ø
EAT	6.13.5	33	1	41	ø	41	34	ø
COND	7.2.3	31	ø	44	ø	44	31	ø
ET	5.7.6	30	ø	44	1	45	30	ø
EAR	8.8.9	23	ø	52	ø	52	23	ø
XTON	8.5.4	22	ø	53	ø	53	22	ø
OW	7.6.13	15	1	59	ø	59	16	ø
OKE	5.1.4	14	1	60	ø	60	15	ø
OUT	3.9.1	4	ø	70	ø	70	4	1
EDGE	1.9.1	3	ø	68	ø	68	3	4
ITCH	5.10.4	1	ø	74	ø	74	1	ø

TABLE 3, TH-Words

	Word	Ref.	[ð]	[d]	[v]	[θ]	[s]	Vcls. Total	Vcd. Total	Missing
1.	THATCH-N	2.7.6	31	8	6	29	ø	29	45	1
2.	THATCH-V	2.7.5	32	7	6	29	ø	29	45	1
3.	THUMB	6.7.6	45	ø	ø	30	ø	30	45	ø
4.	THISTLE	2.2.2	19	22	3	28	2	30	44	ø
5.	THIRSTY	6.13.10	43	ø	ø	31	ø	31	43	1
6.	THIRTEEN	7.1.11	43	ø	ø	32	ø	32	43	ø
7.	THRESH	2.8.1	1	42	ø	31	ø	31	43	1
8.	THURSDAY	7.4.3	43	ø	ø	32	ø	32	43	ø
9.	THREE	7.1.3	ø	42	ø	33	ø	33	42	ø
10.	THIMBLE	5.10.9	41	ø	ø	34	ø	34	41	ø
11.	THIRD	7.2.4	41	ø	ø	34	ø	34	41	ø
12.	THIRTY	7.1.13	41	ø	ø	34	ø	34	41	ø
13.	THOUSAND	7.1.16	40	ø	ø	35	ø	35	40	ø
14.	THREAD	7.10.2	ø	39	ø	36	ø	36	39	ø
15.	THUNDER	7.6.21	39	ø	ø	36	ø	36	39	ø
16.	THIGH	6.9.3	33	ø	ø	41	ø	41	33	1

TABLE 4, SH-Words

	Word	Ref.	[ʒ]	[ʒ̥]	[z]	[ʃ]	Vcls. Total	Vcd. Total	Missing
1.	SUGAR	5.8.10	20	1	1	52	52	22	1
2.	SHILLING	7.7.5	18	ø	ø	57	57	18	ø
3.	SURE	9.7.12	13	ø	ø	62	62	13	ø

TABLE 5

FOR A WE HAVE:
47 > FROM T WORDS
8 D FROM T WORDS
62 F FROM F WORDS
15 S FROM S WORDS
82 V FROM F WORDS
6 V FROM T WORDS
80 Z FROM S WORDS
7 8 FROM T WORDS

FOR A. WE HAVE:
2 S FROM S WORDS

FOR A% WE HAVE:
2 $ FROM S WORDS
6 > FROM T WORDS
4 S FROM S WORDS
1 V FROM F WORDS
13 Z FROM S WORDS
2 8 FROM T WORDS

FOR A: WE HAVE:
2 > FROM T WORDS
49 F FROM F WORDS
31 V FROM F WORDS
4 Z FROM S WORDS

FOR A1 WE HAVE:
7 > FROM T WORDS
17 D FROM T WORDS
36 F FROM F WORDS
10 S FROM S WORDS
19 V FROM F WORDS
6 Z FROM S WORDS
10 8 FROM T WORDS

the geographical distribution is not identical. Thus items 5 - 7 on the table all have a voiceless-voiced ratio of 30:45, but the voiceless localities, at least in the area to the west of the major isogloss to be described below, are not the same. FERN (see Map 13) is voiceless at 3102, 3604, 3902, and 3907 and voiced at 3405 and 4003; FINGER is voiceless at 3405, 3902, 3907, and 4003 and voiced at 3102 and 3604; FIRE is voiceless at 3102, 3405, 3604, and 4003, and voiced at 3902 and 3907. In tabular form:

	3102	3405	3604	3902	3907	4003
FERN	F	V	F	F	F	V
FINGER	V	F	V	F	F	F
FIRE	F	F	F	V	V	F

The same kind of discrepancy is shown by items 9 - 11, with 33:42 ratio, and items 12 - 14, with 34:41.

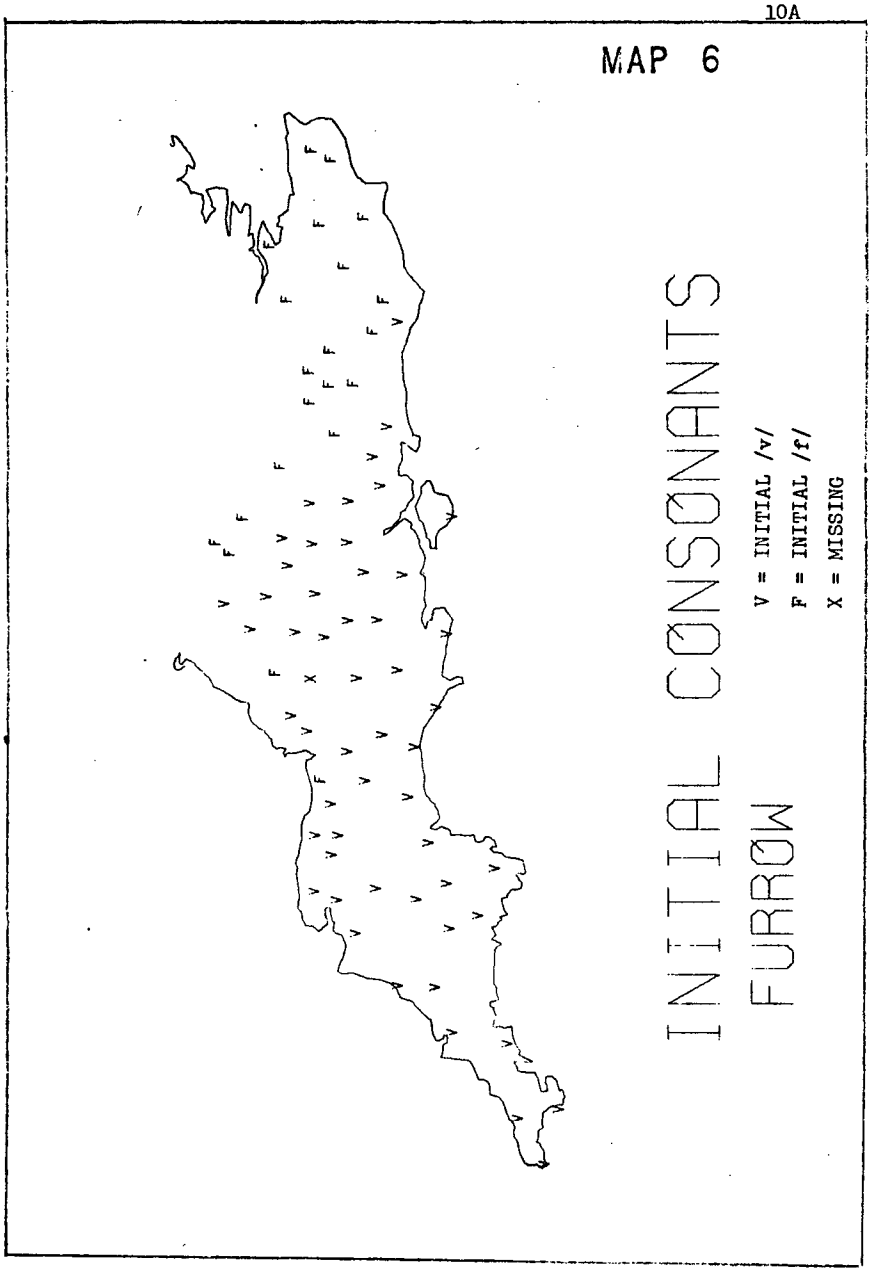
Table 1 also shows that neither etymology nor following consonant seems to affect the distribution markedly. The three French words, FLOUR, FARMER, and FLOWERS, appear in the middle of the list, with ratios of 35:40, 37:38, and 39:36 respectively. The words with initial *fl-* are also in the middle, ranging from FLEAS 33:42 to FLOWERS 39:36.

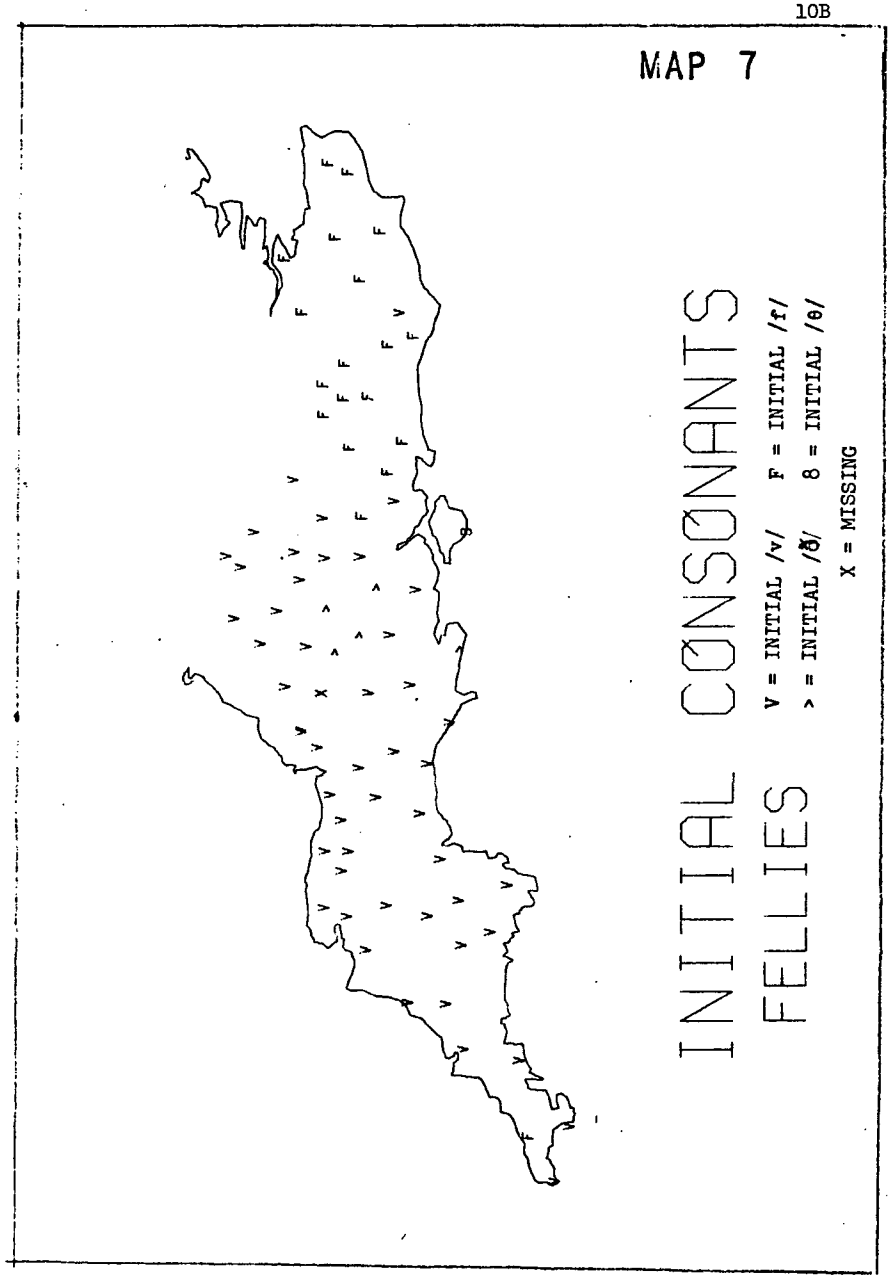
The three last words on the list, with ratios 41:23, 52:22, and 40:22, all show some peculiarity in the recordings indicating that for many of the informants they were unfamiliar, learned, or bookish words, hence more likely to have standard English pronunciation. Thus

FORD, the expected response to the question "Sometimes there is no bridge (over a rivulet). What do you call that shallow place where you can walk across?" (*SED* IV.1.3), was not known, not found, or not recorded in seven localities, replaced by a local term (*splash, sluice*) in four others, and given as a "suggested word" (i.e. one pronounced first by the field worker) in nine others. Though FOAL is recorded from all but one locality in answer to III.4.1, it is revealed by the answers to another question (III.4.6) that in many localities the preferred word is *colt*. The records also show that FORKS (the agricultural kind), not recorded from 13 localities, yields to other preferred local words (*picks, prongs, spuds*) in many others. Likewise it has a mixed etymology, being derivable from Latin *furca* either through OE *forca* or Anglo-Norman *fourque*.

At the top of the list two words, FELLIES and FURROW, show unusual distribution in that voiced forms extend well into the usually voiceless areas of Berkshire, Sussex, and the tip of Cornwall (see Maps 6 and 7). FELLIES also shows six instances of the substitution of the dental fricative for the labial, which otherwise occurs only in FRIDAY, and which is reversed in six occurrences of /v/ in THATCH. If these five words are set aside, the range of voiceless-voiced ratios, from 28:47 to 43:32, much more closely resembles that of the *th*-words (Table 3), which range from 29:45 to 41:33.

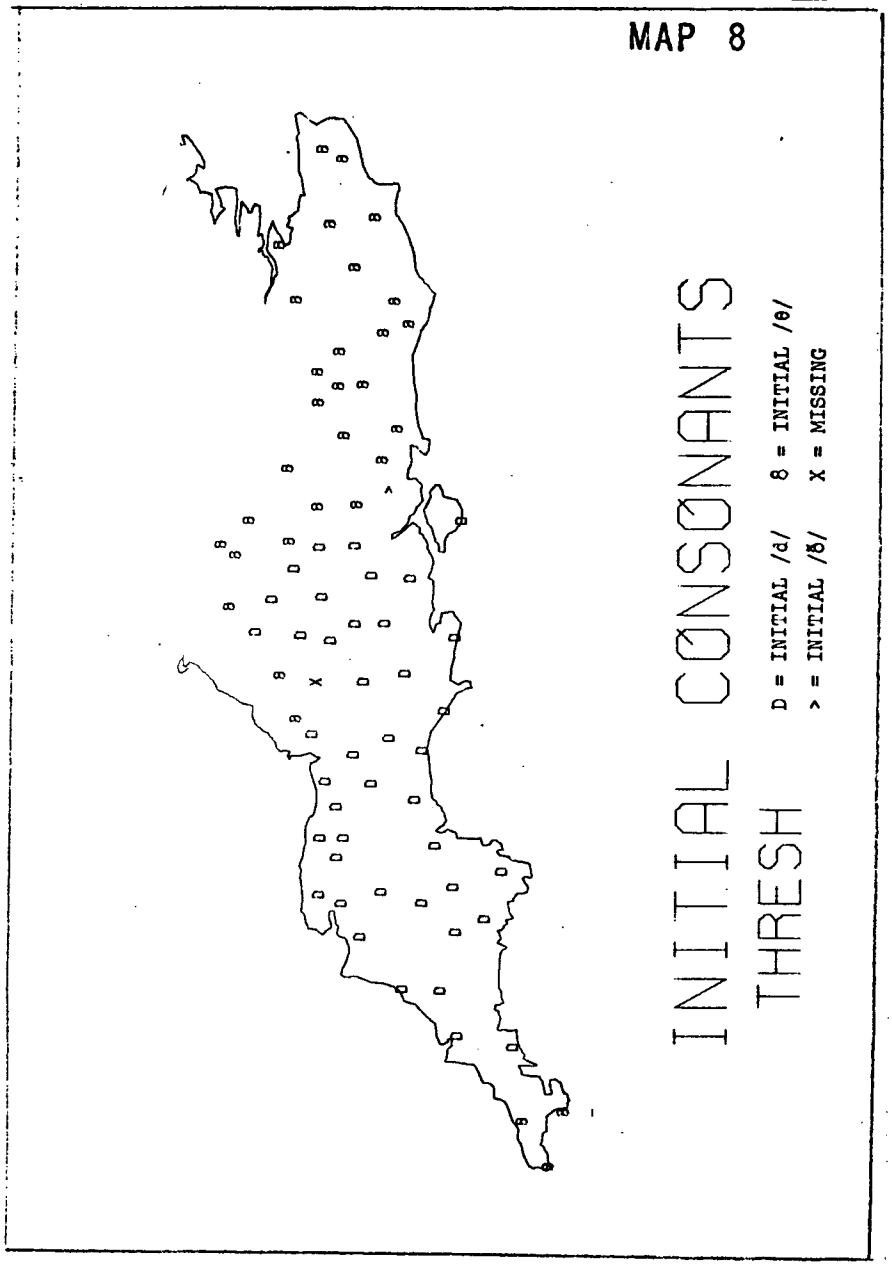
Observations in similar detail can also be made on the basis of Tables 2 - 4, but exigencies of space suggest abridgement.[6] The

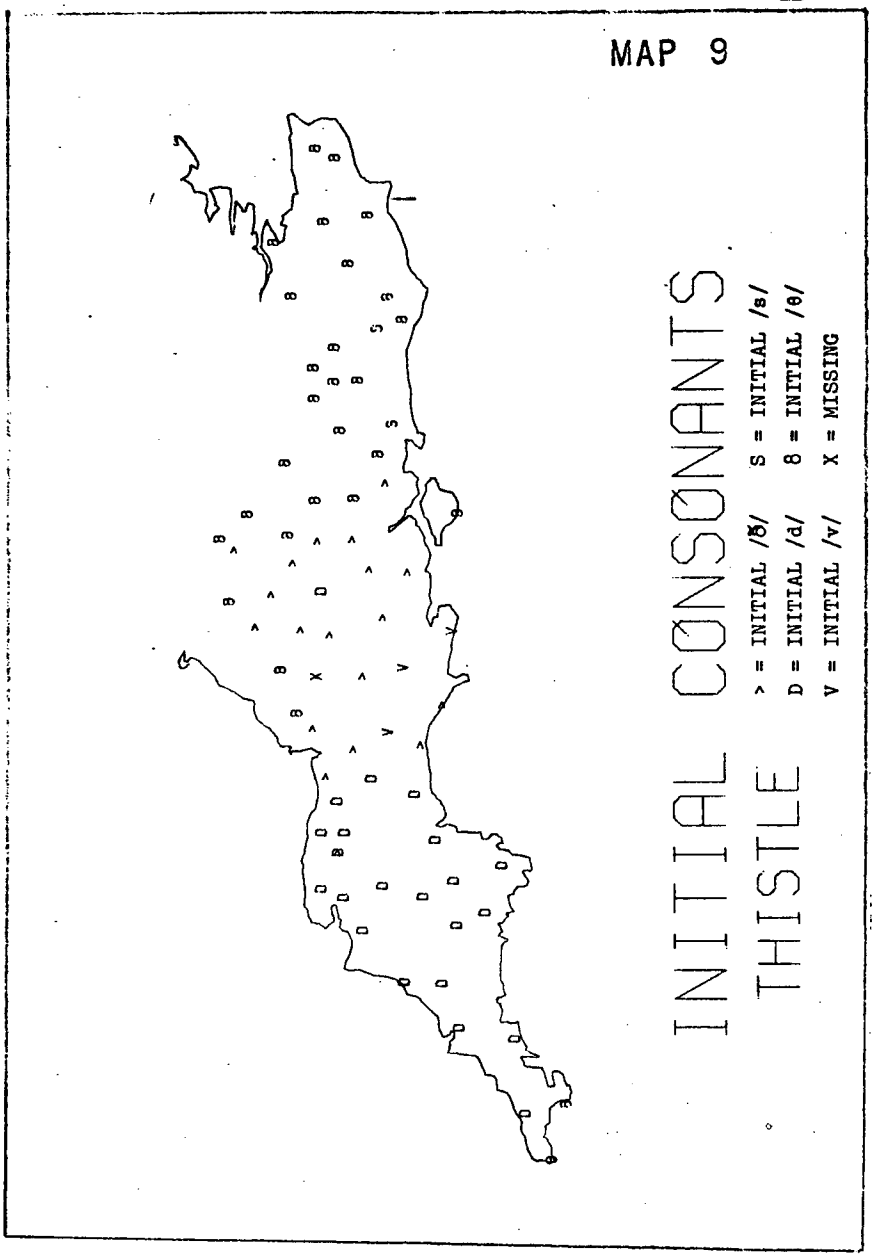

MAP 6
INITIAL CONSONANTS
FURROW
V = INITIAL /v/
F = INITIAL /f/
X = MISSING

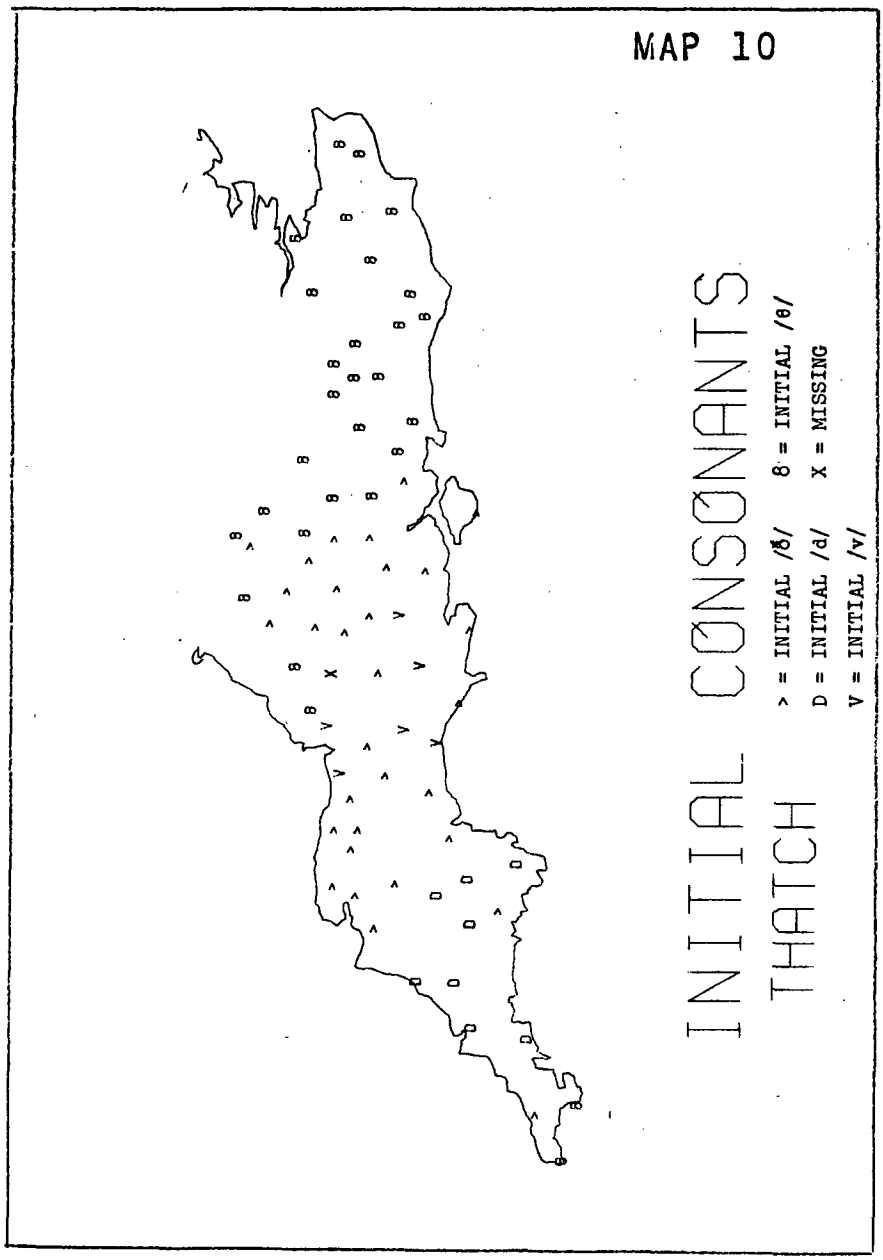

MAP 7
INITIAL CONSONANTS
FELLIES
V = INITIAL /v/
F = INITIAL /f/
> = INITIAL /ð/
8 = INITIAL /θ/
X = MISSING

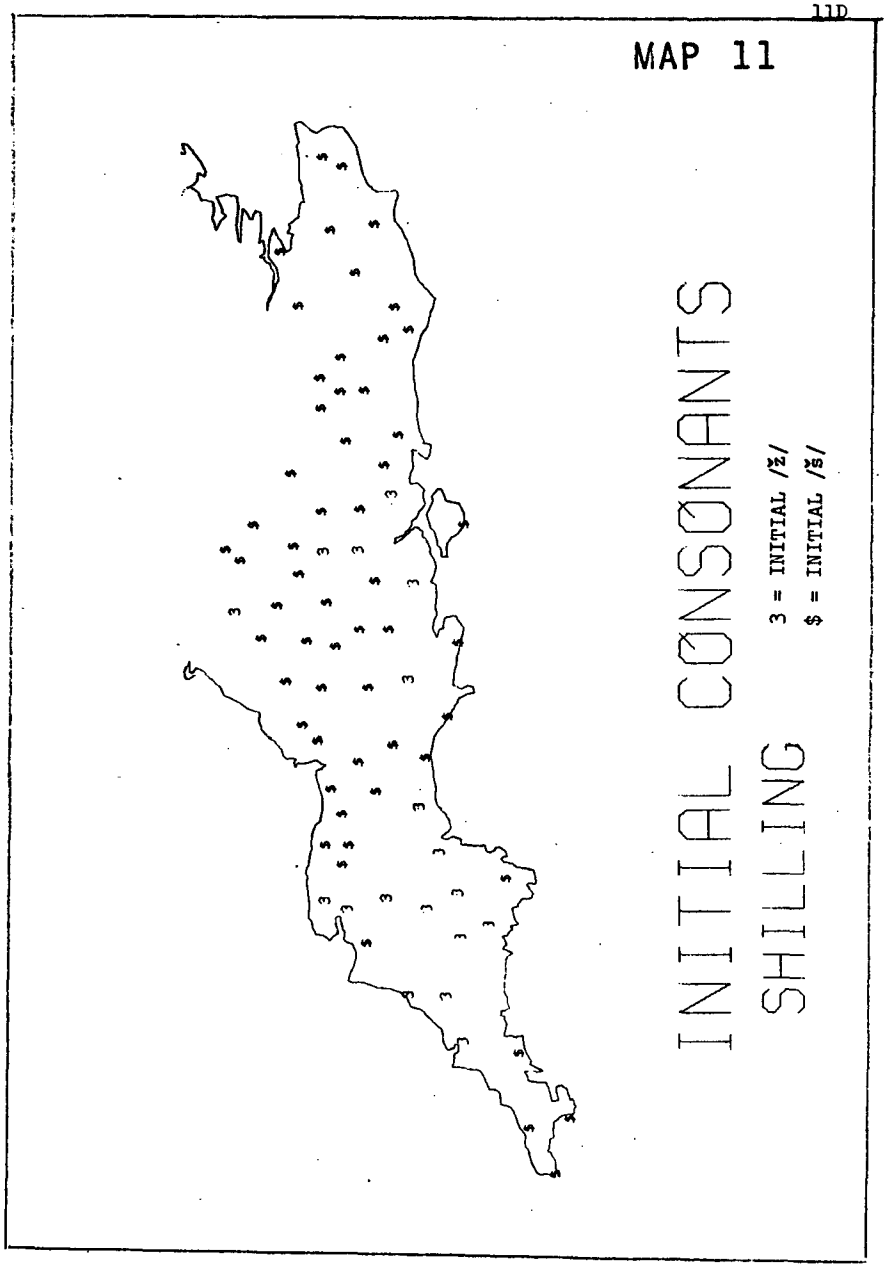
marked drop-off in voiced forms in the last five words of Table 2 implies that, in contrast to initial *f-*, the voicing of initial *s-* is much reduced by a following nasal or /l/ (this will be discussed further below). Voicing is not to be expected before a voiceless consonant; we included STITCH because it shows one freakish occurrence of initial [z] at 3901. In this set the French words, SECOND, SUET, and SEXTON, do appear well down the list, with ratios of 44:31, 45:30, and 53:22 respectively. If the words of French origin and those with a consonant after the initial *s-* are set aside, the remaining range of ratios, from 29:46 to 38:37, comes much closer to that of the other sets.

The *th*-words in Table 3 are, of course, all native words, since neither Norman nor Central French has had initial dental fricatives at any time when borrowing into English could occur. Likewise the only consonant that occurs at all frequently after initial /θ/ or /ð/ is /r/. Hence the range of ratios of the whole set is narrower than those of the *f-* and *s*-words. Noteworthy is the fact that throughout the voiced area, the initial cluster /θr/ appears as /dr/, presumably as a result of a later change of /ð/ to /d/ in this environment.[7] (See Map 8 for THRESH). This same change has occurred in THISTLE in western Somerset, Devon, and Cornwall (see Map 9), and in THATCH (Map 10) in south Devon and eastern Cornwall. These two words also show initial /v/ sporadically in Somerset and Dorset, presumably a substitution the opposite of that occurring in FELLIES in five localities.


MAP 8
INITIAL CONSONANTS
THRESH
D = INITIAL /d/
8 = INITIAL /θ/
> = INITIAL /ð/
X = MISSING


MAP 9
INITIAL CONSONANTS
THISTLE
> = INITIAL /ð/
D = INITIAL /d/
V = INITIAL /v/
S = INITIAL /s/
8 = INITIAL /θ/
X = MISSING


MAP 10
INITIAL CONSONANTS
THATCH
> = INITIAL /ð/
D = INITIAL /d/
V = INITIAL /v/
8 = INITIAL /θ/
X = MISSING


MAP 11
INITIAL CONSONANTS
SHILLING
3 = INITIAL /ʒ/
$ = INITIAL /ʃ/

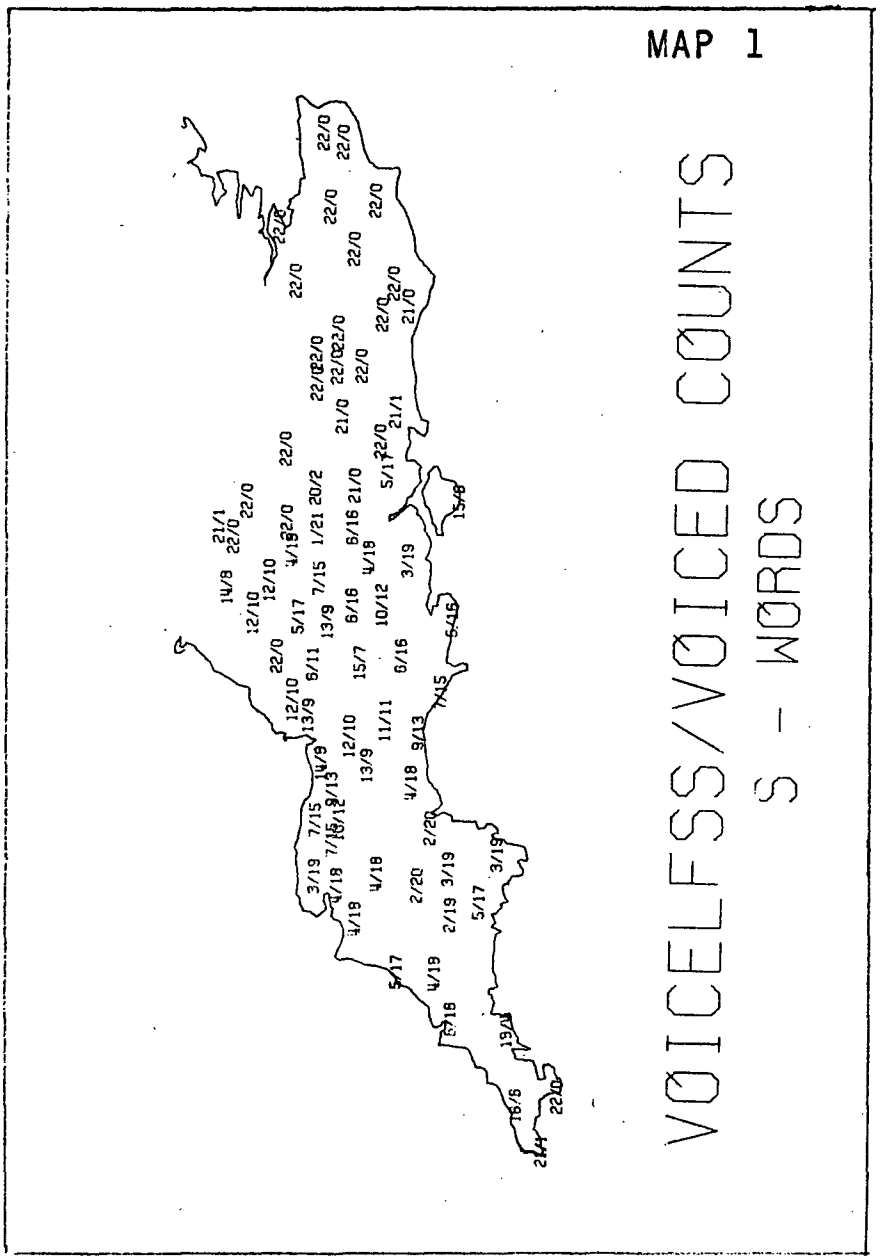
One locality, 3805 (Kingston, Dorset), even has /v/ in THISTLE and /ð/ in FELLIES. Why only two of the 16 *th*-words should show this kind of substitution remains unexplained. There is certainly nothing in their etymology to account for it.

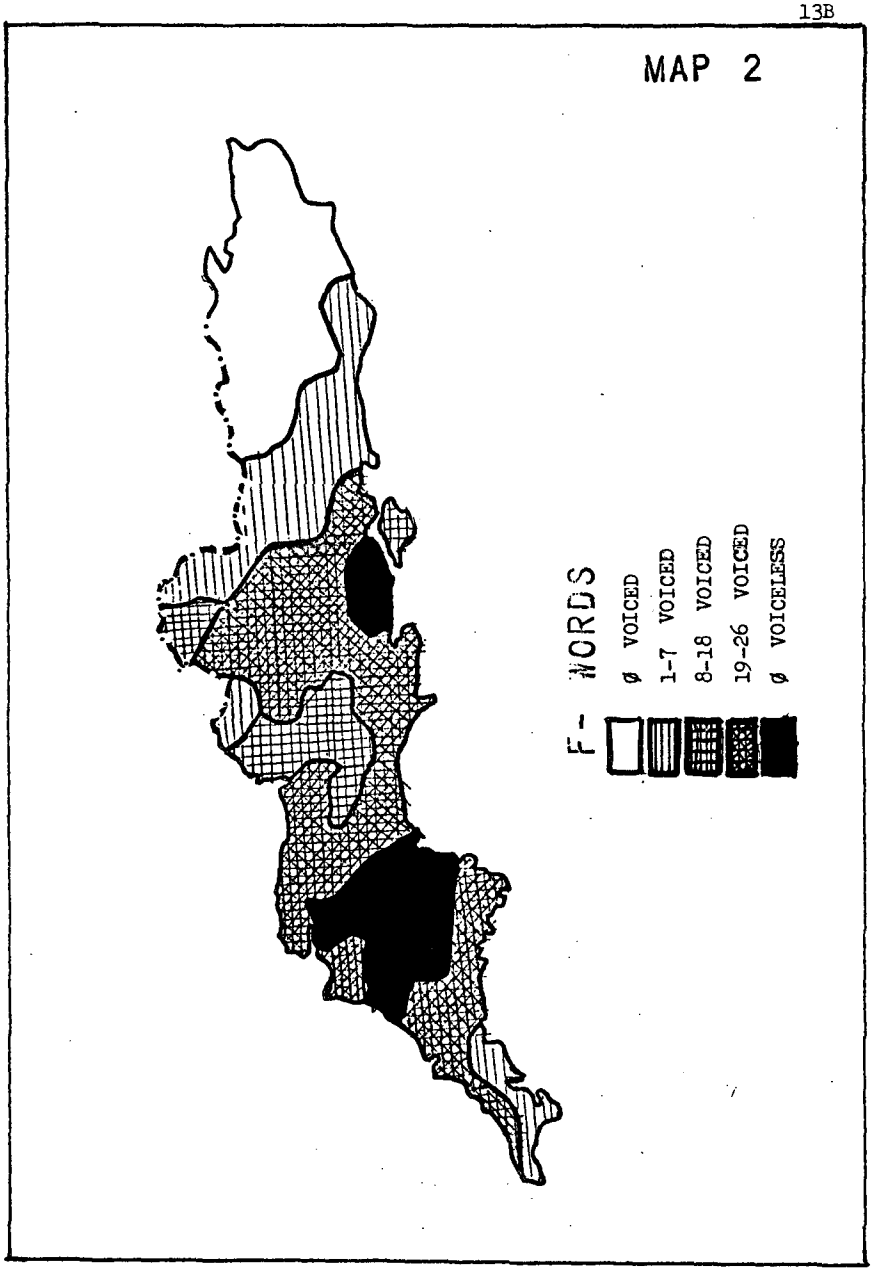
Table 4 lists the only three *sh*-words that show more than a few scattered instances of initial /ž/. The other starred words of this sort, with the number of occurrences of initial voicing for each, are *sheaf* 9, *sheep* 1, *shelf* 3, *shoe* 2, and *shovel* 3. Even the three words listed have a ratio characteristic of the low end of the other lists: 52:22 for SUGAR (a French loanword), 57:18 for the native word SHILLING, and 62:13 for another French word, SURE. Presumably the two French words had something like initial [sj-] at the time of borrowing, while SHILLING and the other native words had /sk-/ until later OE. Instances of voicing in these words is to be attributed wholly to analogy, presumably with the *s*-words. As might be expected, the heaviest concentration of voiced forms is in the "hard-core" voicing areas in Devonshire and western Hampshire (see Map 5), but there are scattered instances in Somerset, Dorset, and Wilts. (see Map 11 for SHILLING).

Just as Tables 1 - 4 reveal the wide variety in voiceless-voiced ratio for individual words, so Maps 2 - 5 show the same variety for individual localities. These maps have been simplified from the original computer-produced proportional maps described above, which were in the form shown in Map 1. Taken together, they demonstrate

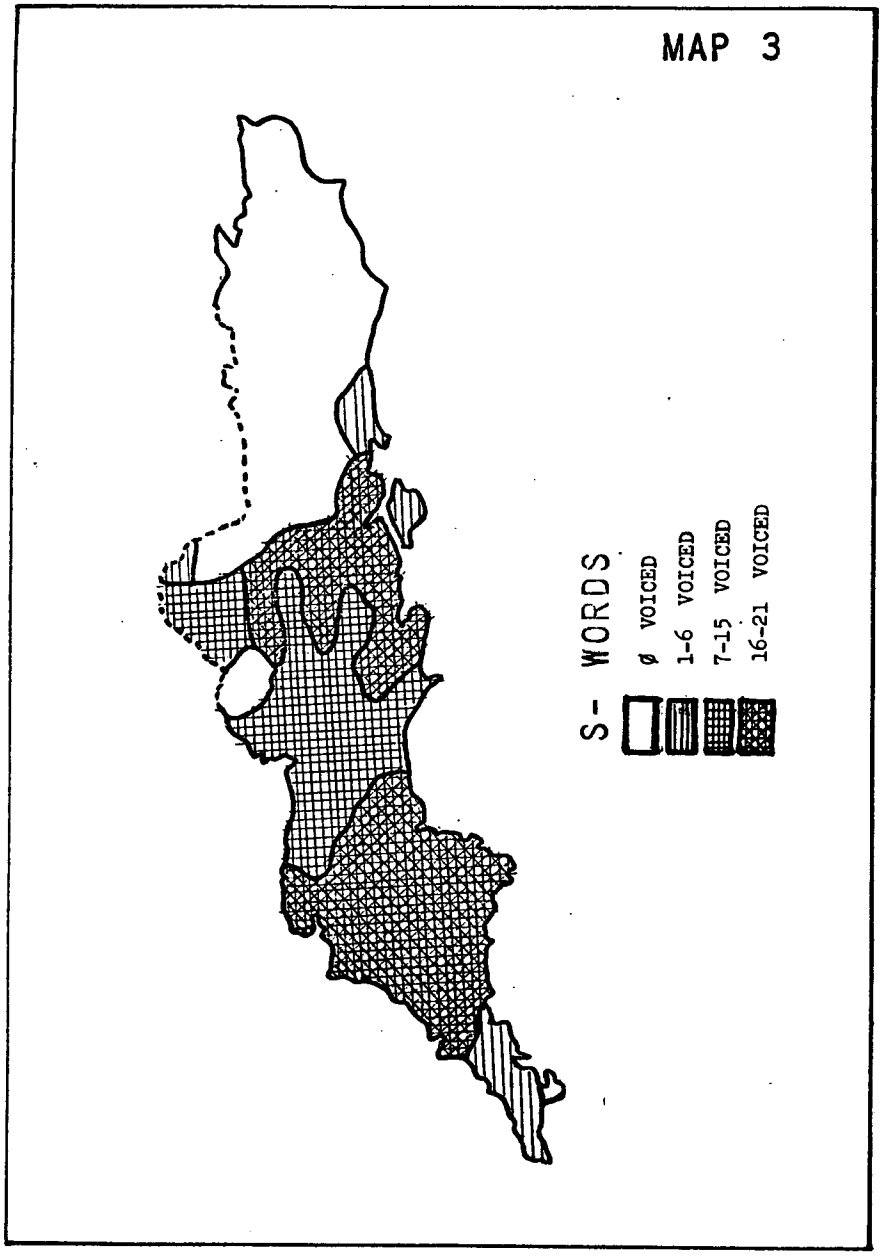
several interesting points. All four of them show a strongly marked isogloss on the eastern boundary of the voicing area. This line starts on the coast east of Portsmouth, trends in a northwesterly direction across the middle of Hampshire, follows the Wiltshire - Berkshire line for a short distance, and then turns west across northern Wiltshire toward the Cotswolds. The plotting of its subsequent course must await the publication of the records from Gloucestershire and Hereford. It marks a sharp division quite different from the grading noticeable within the voicing area. At one point about ten miles north of Winchester it passes between two localities (3901 and 3902) showing respectively total voicing and total voicelessness in the *th*-words. A little to the north of this it separates two localities—Burbage, Wilts., and Inkpen, Berks.—which are only 8 miles apart and yet show voiceless voiced ratios of Ø:16 and 15:1 respectively in the *th*-words. It is equally well marked for the other sets, though in the case of the *f*-words the inexplicable eastern extension of voicing in two words—FELLIES AND FURROW—creates the appearance of a transitional area (Map 2).

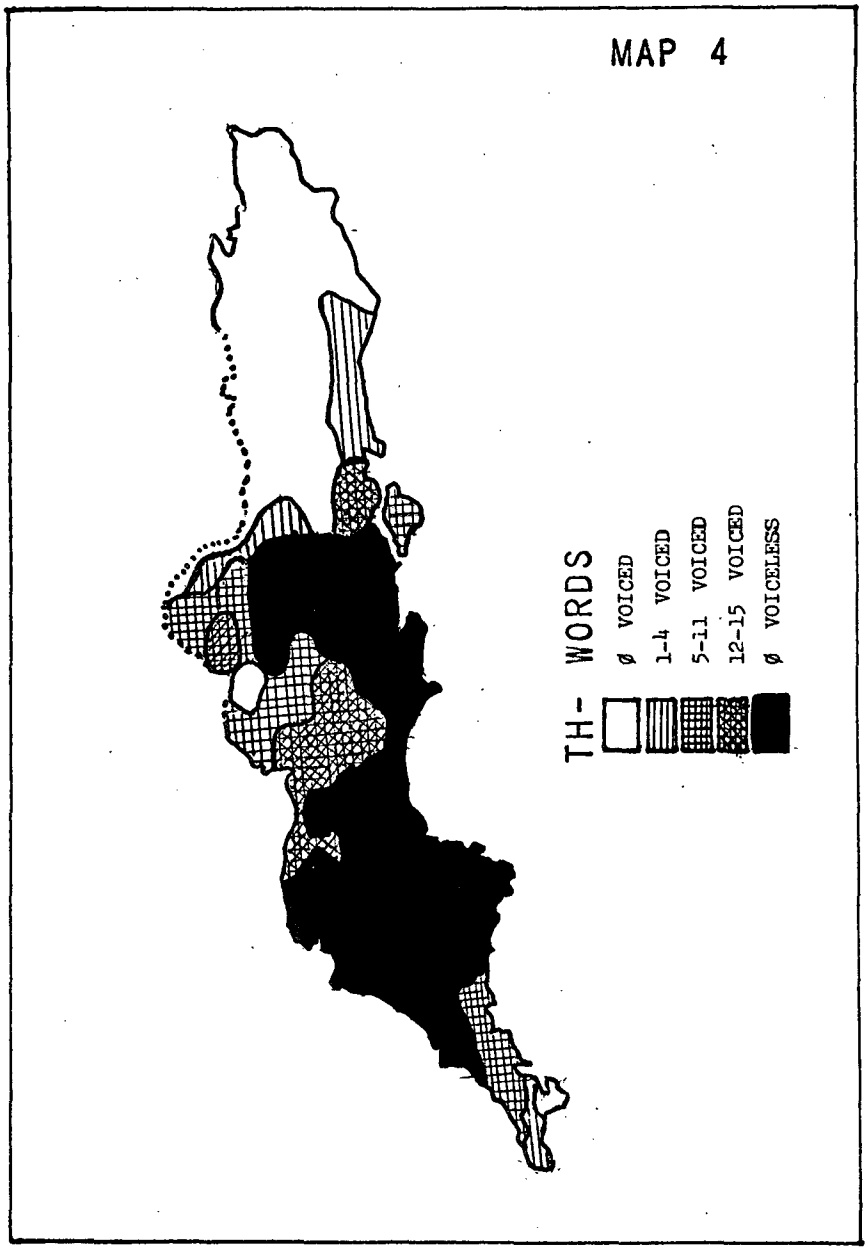
The maps also show that the voicing did not extend to the southwest tip of Cornwall except in a few words. Presumably the English brought into this formerly Celtic area was more strongly influenced by standard English. The extreme case is represented by 3607, which has voicing in only two of the 68 words: FELLIES and FURROW, which, as we have seen above also extend beyond the voicing area on the east.


MAP 1
VOICELFSS/VOICED COUNTS
S - WORDS

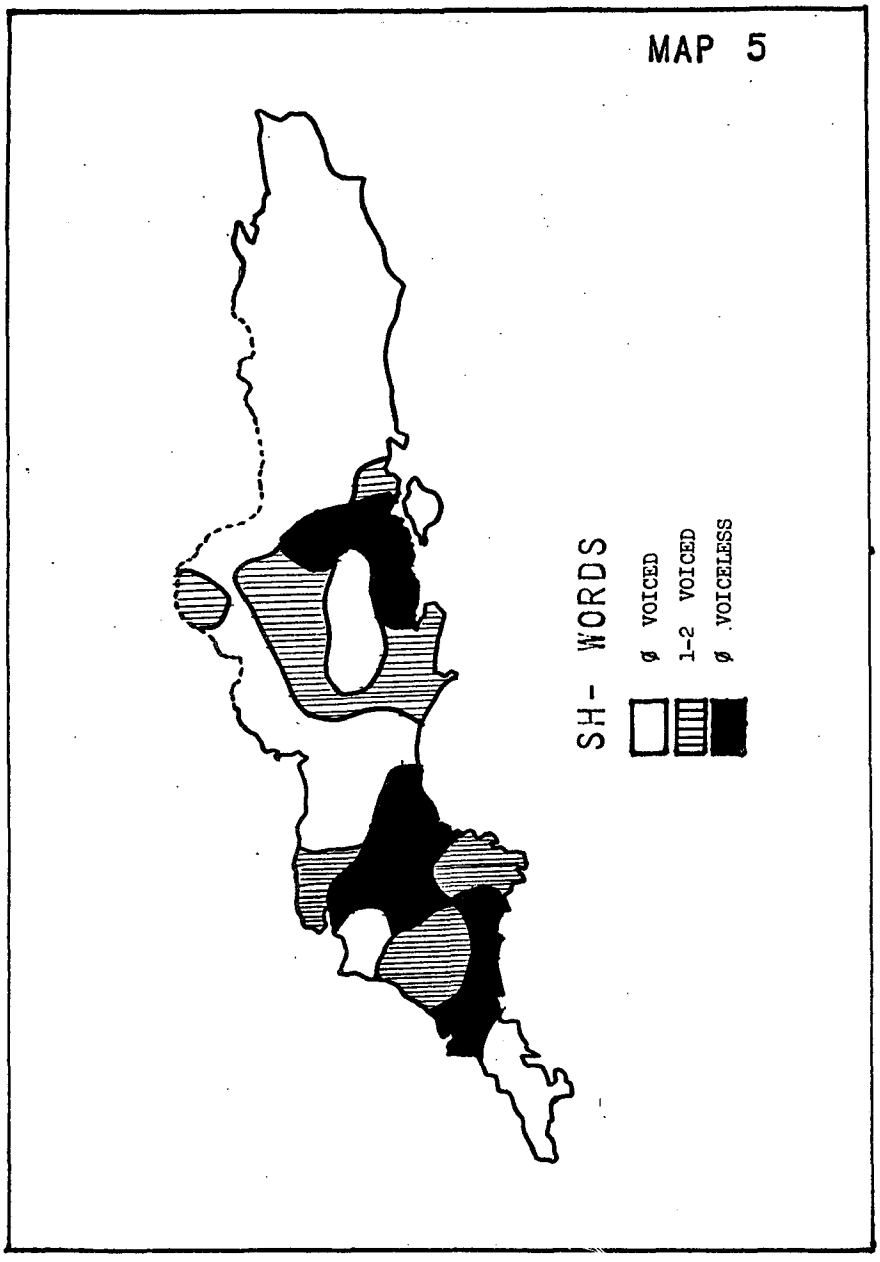

MAP 2
F- #ORDS
ø VOICED
1-7 VOICED
8-18 VOICED
19-26 VOICED
ø VOICELESS

13C




MAP 4
TH- WORDS
ø VOICED
1-4 VOICED
5-11 VOICED
12-15 VOICED
ø VOICELESS

13E



Maps 2 - 5 also reveal rather similar patterns of voicing in different parts of the region. Just west of the main eastern isogloss is an area where voicing is virtually total for all sets of words. The area differs in size from one set to another, being larger for the *f-* and *th*-words than for the *s-* and *sh*-words, but its heart is western Hampshire, Dorset, and part of southern Wiltshire. It is separated from the other area of almost total voicing in Devonshire by a mixed area in Somerset and Wiltshire where the proportion of voicing among the words examined ranges from 25% to 75%. Finally there are fringe areas in the east and in southern Cornwall where only a few of the words—those high on the lists in Tables 1 - 4—show voicing. In the light of the traditional historical view of a feature spreading west from Kent, as expressed in the quotation from Horn-Lehnert above, this distribution is a bit puzzling. Full exploration of its implications must be postponed to a later study.

One possible explanation of the variation in the incidence of voicing from one word to another is to be sought in the influence of the sounds immediately following the initial fricative. Chart I displays some of the characteristics of the four sets with regard to to following phonemes. The *sh*-words have far less initial voicing than the other sets. No more than 22 localities have initial voicing for any single word in this set. On the other hand, the *s*-words demonstrate the widest range: from voicing in 46 localities to a long tail of words with no initial voicing anywhere. The most stable is

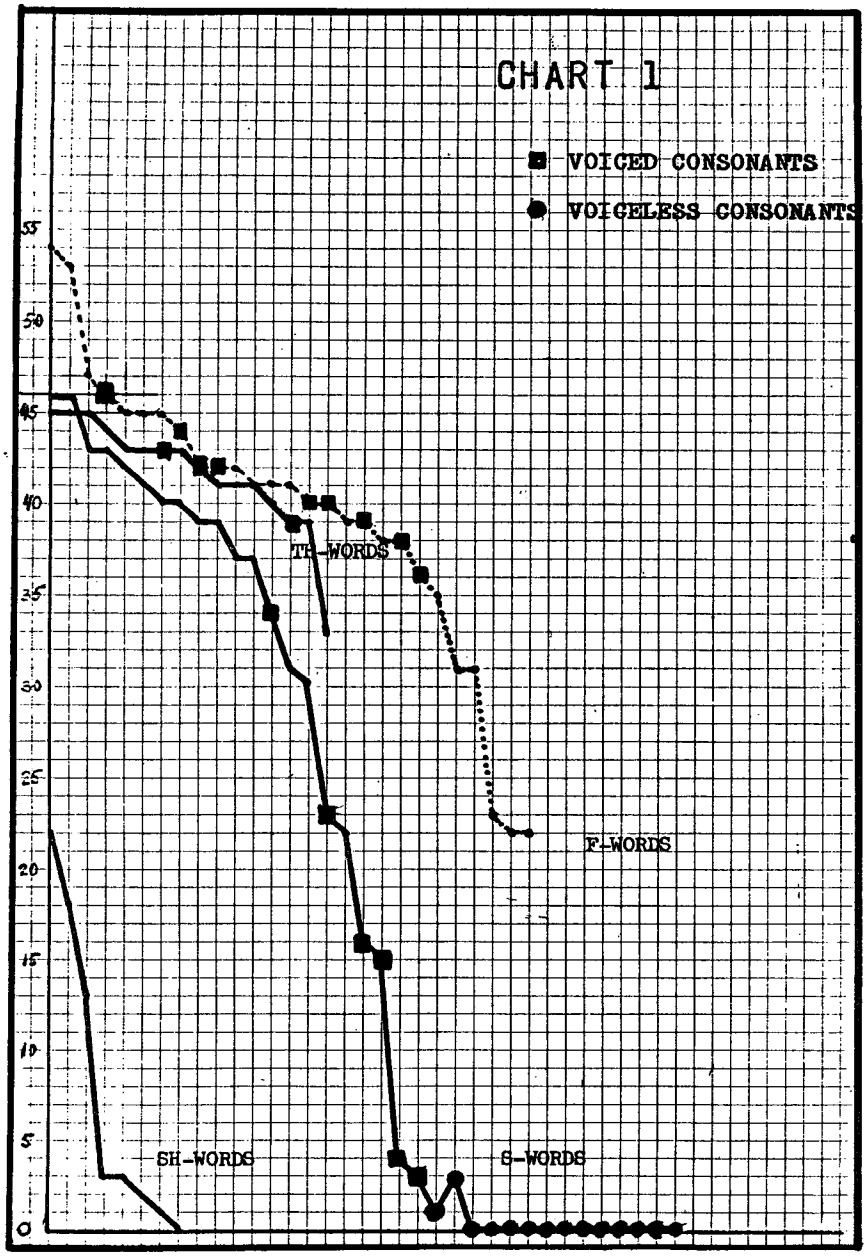


CHART 1
VOICED CONSONANTS
VOICELESS CONSONANTS
55
50
45
40
35
30
25
20
15
10
5
0
TH-WORDS
F-WORDS
SH-WORDS
S-WORDS

the *th-* set, which has between 45 and 33 localities with voicing. This means that for all words but one in this set, over half the localities have a voiced initial fricative. This is the highest proportion for any set.

The sound following the initial fricative has been indicated on Chart I in terms of three categories:

+ vowel (unmarked)

+ voiced consonant (marked with square)

+ voiceless consonant (marked with dot)

There is a striking correlation between voiceless initial fricative and a following voiceless consonant in the *s*-words (no other set has a voiceless consonant in second position). With the exception of *spring* (72:3) and STITCH (74:1), the association is absolute for

⟨s⟩ + { /k/ *school, scratch, sky, squirrel*
{ /p/ *spade, speak, spokes*
{ /t/ *stars, steal, stile, stool, straw*

Only STITCH has been included in the computation.

Voiced consonants in second position of *s*-words also associate with initial /s/ rather than /z/:

⟨s⟩ + { /w/ SWEAT (41:34), SWEAR (52:23)
{ /n/ SNOW (59:16), SNOUT (70:4)
{ /m/ SMOKE (60:15)
{ /l/ SLEDGE (68:3)

In the case of the other sets, however, there is no clear connection:

⟨th⟩ + /r/ THRESH (31:43), THREE (33:42), THREAD (36:39)

⟨f⟩ + /r/ FRIDAY-1 (29:46), FRIDAY-2 (28:44) FROGS (33:42).

⟨f⟩ + /l/ FLEAS (33:42), FLIES (34:40), FLOUR (35:40), FLOOR (36:39), FLITCH (26:38), FLOWERS (39:36)

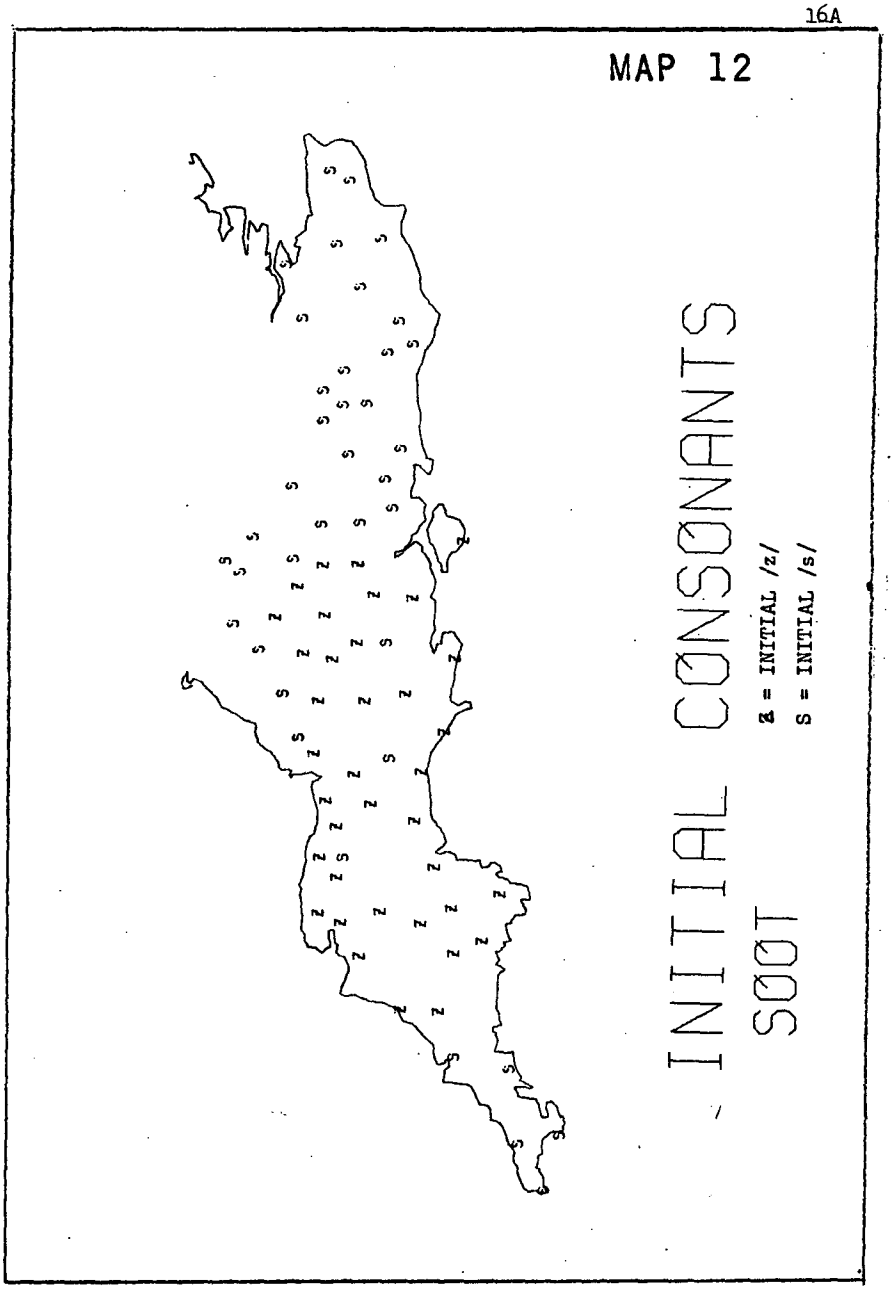
Vowels present a more complex situation. Table 6 shows that for *f*-, *s*-, and *th*-words, high and low front vowels /y, i; æ/ and low central vowels /ɑ, a/ associate with initial voicing. For the three sets taken together the proportion of voiced initial fricatives occurring with these vowels following is between 60 and 63 per cent.

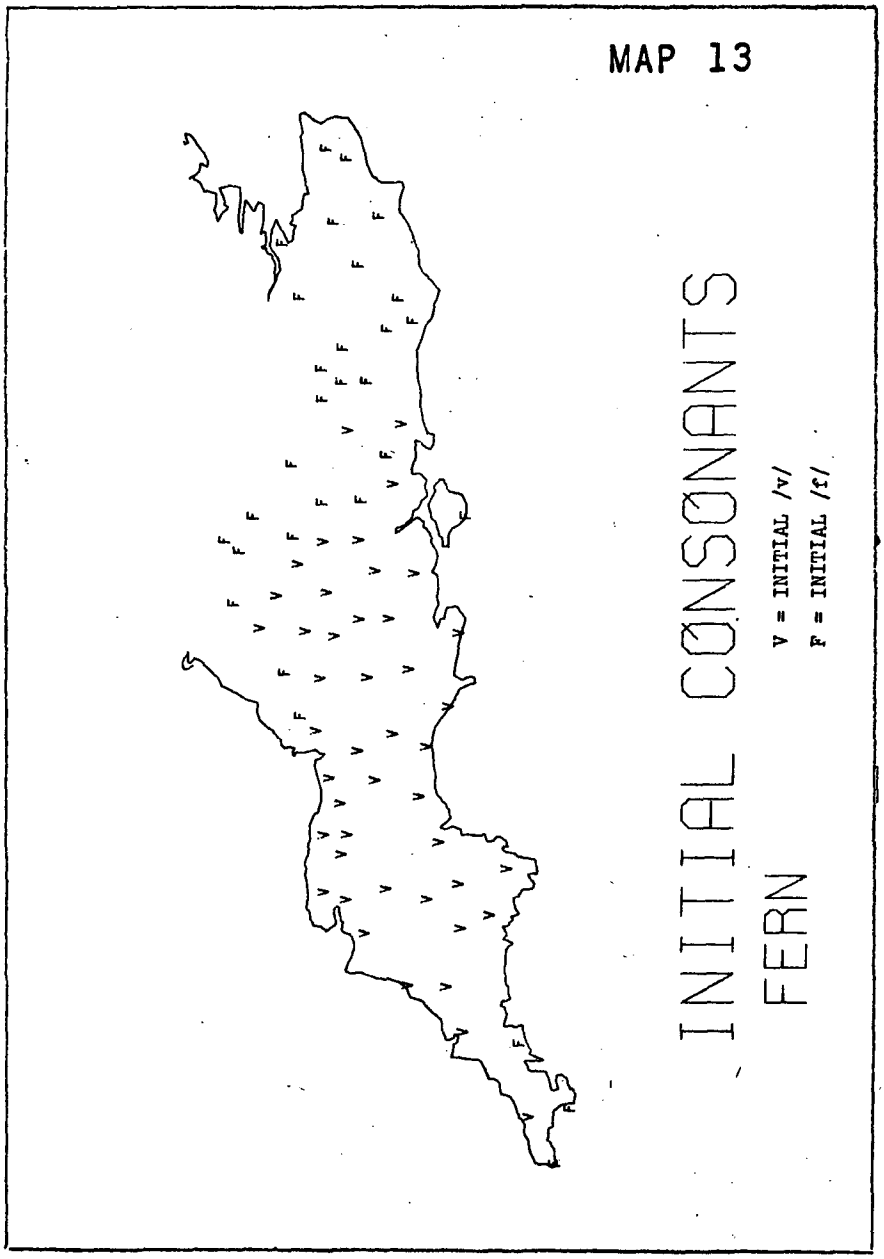
TABLE 6

PROPORTION OF VOICELESS TO VOICED INITIAL FRICATIVES IN RELATION TO THE FOLLOWING VOWEL IN F-, S-, TH-WORDS

	Front	Central	Back
High	62% 189/308	6/0	50% 109/108
Mid	36% 248/140	55% 435/540	54% 43/50
Low	60% 119/176	63% 197/329	40% 252/167

There is negative association between voicing and mid front /e, ɛ, œ/ as well as low back vowels /ɒ, ɔ/. A study of the values for the individual sets reveals that initial voicing is particularly associated with high and low front vowels in *f*-words (72% and 78% respectively) and with low central vowels in *s*- and *th*-words (77% and 83%).


MAP 12
INITIAL CONSONANTS
SOOT
Z = INITIAL /z/
S = INITIAL /s/


MAP 13
INITIAL CONSONANTS
FERN
V = INITIAL /v/
F = INITIAL /f/

In *f*- and *th*-words low back vowels have a particularly low association with initial voicing (40% and 30%, respectively). Vowel length appears to have no appreciable correlation with the voicing or unvoicing of the initial fricative.

Rounded high front vowels /Y, y/ showed a more marked association with initial voicing than unrounded high front vowels /I, i/. The ratios for all *f*- and *s*-words with rounded front vowels are: FOOT Ø:16; SOOT 1:15, SUCK Ø:7, SUET 1:15. If we make a table for all the second vowels in SOOT (Map 12), we get:

	Front	Central	Back
High	1:15	Ø:1	32:19
Mid		1:2	
Low			

There is, of course, a possibility here of pseudo-correlation, caused for instance by the possibility that those localities which have the fronted /Y/ for standard English /U/ in words like SOOT happen to be located in the area of strongest voicing. Investigation of this possibility must await further study.

The tendency for high front vowels generally to co-occur with voiced initial fricatives is, however, obvious in words like FERN (Map 13) where no rounding is involved:

	Front	Central	Back
High	Ø:16		
Mid	Ø:2	30:21	
Low		Ø:6	

Finally, to illustrate the difference in correlation with low central and low back vowels, we may cite the figures for FORTY. The voiceless:voiced ratio with a low central vowel [a, a:, ɑ, ɑ:] is 13:17, while with a low back [ɒ, ɔ] it is 27:18.

* * *

It is clear that even more study, of individual words and individual localities, is needed before all the complications of this one dialect feature can be unraveled. We should, for example, take into account the second and third responses for many of the words, many of which were taken from incidental conversation and hence are inclined to be more natural. Even casual inspection of the data indicates that they show a much higher incidence of initial voicing than do the citation forms. But we hope that this paper has shown that, given adequate and convenient data, the computer can be of inestimable aid to the dialectologist.

NOTES

[1]An exception is the Dictionary of American Regional English (DARE), being prepared at the University of Wisconsin under the direction of Frederic G. Cassidy, which is employing some sophisticated computer techniques.

[2]See under Orton and Dieth in Bibliography. This work will henceforth be referred to as *SED* or the *Survey*.

[3]The starred words are those which were included primarily for their phonological importance. Fieldworkers were instructed to obtain them at all costs, even if they had to suggest the word and ask the informant to pronounce it. In most cases words were chosen that have universal distribution in the dialects, but occasionally a word thought to be common turned out to be unfamiliar or even unknown, as in the case of FORKS, FORD, and FLITCH in our corpus.

[4]The questionnaire is divided into nine books, each of which is subdivided into sections containing several questions. An item is thus identified by a three-part number, e.g. VIII.4.6, indicating question 6 in section 4 of book VIII. We changed the Roman numerals to Arabic in the interest of simpler coding.

[5]For an alternative theory, holding that initial fricatives were already voiced in the language of the Jutes and Frisians who settled Kent, see Bennett 1955 in Bibliography.

[6]We hope to explore the linguistic implications of this project more fully in a later article.

[7]One instance of /ðr-/ in THRESH is reported from 3905 Hambledon, Hants., which is just within the eastern border of the voicing area.

Bennett, W. H., "The Southern English Development of Germanic Initial [f s þ]," *Language* 31:367-71 (1955).

Elworthy, F. T., *The Dialect of West Somerset.* English Dialect Society, Series D, vol. 20, 1875.

Hewett, Sarah, *The Peasant Speech of Devon.* London, 1892

Horn, W., and M. Lehnert, *Laut und Leben: Englische Lautgeschichte der neueren Zeit,* vol. II. Berlin, 1954.

Jacobson, Ulf, *Phonological Dialect Constituents in the Vocabulary of Standard English.* Lund Studies in English 31, Lund, 1962.

Jennings, J. K., *The Dialect of the West of England, Particularly Somersetshire. . .* 2nd ed., London, 1869.

Jordan, R., rev. H. C. Matthes, *Handbuch der Mittelenglischen Grammatik,* vol. I, §159. Heidelberg, 1934.

Kjederqvist, J., *The Dialect of Pewsey (Wiltshire).* Transactions of the Philological Society, London, 1903-04.

Kruisinga, E., *A Grammar of the Dialect of West Somerset.* Bonner Beiträge zur Anglistik 18, 1905.

Kurath, H., "The Loss of Long Consonants and the Rise of Voiced Fricatives in Middle English," *Language* 32:435-45 (1956).

Matthews, W., "South Western Dialect in the Early Modern Period." *Neophilologus,* 24:193-209, 1939.

Moore, Samuel, rev. by A. H. Marckwardt, *Historical Outlines of English Sounds and Inflections.* Ann Arbor, 1951. p. 115.

Mossé, F., tr. by J. A. Walker, *A Handbook of Middle English.* Baltimore, 1952. §44, pp. 39f, also Fig. 6, p. 38.

Orton, Harold, and Eugen Dieth, *Survey of English Dialects. Introduction,* by Harold Orton (Leeds, 1962); *The Basic Material,* vol. 4, *The Southern Counties,* ed. H. Orton and M. F. Wakelin (Leeds, 1967).

Widén, B., *Studies on the Dorset Dialect.* Lund Studies in English 16, Lund, 1949.

Wilson, Sir James, *The Dialect of the New Forest in Hampshire.* Publications of the Philological Society 4, Oxford, 1913.

Wright, Joseph, *The English Dialect Grammar.* Oxford, 1905, repr. 1968. §§278, 310, 313, 320, 322.

----------, *An Elementary Middle English Grammar,* 2nd ed. Oxford, 1928. §236.

APPENDIX A, CODING SYSTEM

The following system was used in coding the data for the computer.

VOWELS

I = [i ɨ ᵻ ï]
1 = [ɪ ɪ̞]
Y = [y Y]
E = [e]
(= [ɛ ɛ̞ ɛ̃]
& = [œ]
@ = [æ ɐ̞ ɐ̟]
A = [a a̲ ä ɑ ɑ̞ ä]
) = [ə ə̞ ɐ]
7 = [ʌ ʌ̟ ʌ̃ ʌ̈]
% = [ɒ ɒ̞ ɒ̞̈ ɒ̈]
U = [u u̞]
5 = [ɔ̈]
4 = [ü]
O = [o o̞ o̞ õ]
# = [ɒ ɒ̟ ɒ̈ ɔ o̞ o̞ ɔ̈]

DIACRITICS

: = ː
. = ˑ

CONSONANTS

8 = [θ]
T = [t t̬]
> = [ð̞]
D = [d]
6 = [d̪]
B = [b]
V = [v]
Z = [z z̞]
< = [z̪]
F = [f]
2 = [f̬]
$ = [ʃ] (i.e. [š])
W = [w]
L = [l ɫ]
R = [r ɾ] also "r-coloring"
N = [n n̟ n̠]
9 = [ŋ]
M = [m]
J = [j]
3 = [ʒ] (i.e. [ž])
; = [d̻ʒ̻]
? = [ʔ]
XXX = n.a., n.k., n.r., etc.

APPENDIX B, PREPARING THE MAPS

by Gerald M. Rubin

When it was first decided to produce dialect maps by computer, several methods were discussed. Output could be intricately placed on a printed page, and then a map outline could be superimposed on that page by hand. In this way we could achieve our basic goal, that of having the computer do the tedious task of sorting and tallying the linguistic data to be displayed, while leaving a minimum of work to the researcher. But this idea was not pliable enough to let us represent a map as it actually is. The squareness of the format and the constant distance between characters on the printed page made it impossible to reproduce any map with sufficient accuracy.

A second method discussed was to output the entire map and data on a visual display unit such as a CRT scope. Here we could draw the map, but our printing format was again too strict. This method also is expensive, since it requires the use of an on-line scope.

We finally decided upon an off-line plotter. The one we used was a CalComp #563 Digital Plotter. This machine takes a computer tape which has been produced by an on-line computer program and draws the date in the tape onto a roll of paper. In order to visualize how the plotter works, imagine a set of coordinate axes with a y-axis about 12" long and an infinitely long x-axis. This grid is the piece of paper to be plotted on. The instructions to the plotter are simple. They boil down to two: lower or raise the pen point (so it will or will not write as it moves) and move the pen in a straight line to location (x,y) on the grid. In this manner anything can be drawn, from straight lines to circles, letters, and numbers (curves are actually made up of very short straight line segments).

The routines used to produce CalComp tapes are FORTRAN subroutines. It was therefore necessary to write a FORTRAN program whose input would be (1) instructions for drawing the outline of a map and plotting localities within it, and (2) linguistic data in a specially processed format. The output would be the tape which directs the plotter.

Two methods were tried out for producing the map outline. In the first, a transparent grid was placed over a map and coordinates of "bends" in the outline were recorded. Thus a map would be produced by moving in a straight line from one bend to another. The map produced by this method gave only a rough approximation of the original because we could not find a grid small enough to represent accurately all the bends and curves in the outline of the map. The second method was much more successful. A digitizer was used to measure and record the coordinates of bends in the outline and the positions of the localities within the area covered. The digitizer (used by courtesy of the American Mathematical Society, Providence, R.I.) consists of a large table and a "bomb-sight" connected to a paper-tape punch. The crosshairs of the bombsight are moved along the outline of the map, and all coordinates are accurately measured and punched onto paper tape. This paper tape then becomes the direct input to the computer. Special codes were used to instruct the plotter to lift the pen (to move away and draw an island, for example) or to indicate that the next coordinates would be the positions of localities on the map. When the program was completed, the instructions for drawing the outline of England and marking the coordinates of the 311 localities of the *Survey of English Dialects* required about 4000 x and y measurements. [For the present paper only the bottom quarter of the map, with 75 localities, was used.]

The raw linguistic data was keypunched onto cards using the phonetic coding discussed above (Appendix A). Also on the cards were the numbers of the county and locality and the keywords. For example, the card for SUGAR, county 31, locality 4, was

3104 SUGAR ;%G)R: 5. 8.10

Since the corpus of data included 68 words for each of the 75 localities in Southern England, the card input consisted of 5100 cards. These records could be sorted by phonetic word (citation), by locality, by keyword, or by any combination of these.

A PL/I program was written to examine the records and to output the linguistic results which were to be drawn onto a map. This output was then read in by the FORTRAN plotting program. The input to the plotting program was of a standard format:
MAP NUMBER; COUNTY; LOCALITY; NUMBER OF SYMBOLS; LINE NUMBER; MESSAGE
For example, the card

```
3    33     5     7     3-F/1-V
```

meant "on map #3 draw the seven characters 3-F/1-V at locality 5 in county 33." "Line number" was used to produce the legend at the bottom of each map. The legend was drawn whenever the county and locality numbers were Ø. For example, the instruction

```
2    Ø     Ø    18     1     INITIAL CONSONANTS
```

meant "on map #2 write the above message at the position for line 1 of the legend."

The input to the plotting program came in sorted by map number. Whenever a new map number was read in (meaning that all the data for the previous map had been plotted), the program would issue instructions to move over on the plotting paper and draw a new map outline. Data would then be drawn on this new map until a different one was requested. Thus one run of each of the PL/I and FORTRAN programs could produce any number of related or even unrelated maps.

It should be apparent that the great virtues of this method of producing both working and finished dialectal maps are speed and accuracy. Anyone who has produced even rough working maps by stamping or drawing symbols on an outline map knows how time-consuming and tedious this process is, and how subject it is to error. The computer never wearies and never (we trust!) makes an error. As a result, the dialectologist can experiment with all kinds of working maps and select those which are interesting or significant. This should add a new dimension to the study of linguistic geography.