Joy Jenkyns

<sup>28</sup> Lincoln Archaeological Trust, Sixth Annual Report 1977-78 (Lincoln, 1978), 11; J. Campbell, ed., The Anglo-Saxons (Oxford, 1982), 54, plate 55; B. Gilmour, 'The Anglo-Saxon church at St Paul-in-the-Bail, Lincoln', Medieval Archaeology XXIII (1979), 214-18, and fig. 2.

<sup>29</sup> For a recent survey of such artefacts, see R. Bruce-Mitford, 'Ireland and the hanging-bowls – a review', in M. Ryan, ed., *Ireland and Insular Art A.D.* 

500-1200 (Royal Irish Academy, Dublin, 1987), 30-9.

<sup>30</sup> Sixth Annual Report, as in n.28, above.
<sup>31</sup> The Development of OE y and &o in South-Eastern Middle English, Lund Studies in English XLII (Lund, 1972), 60-4, 122, and map on p. 123. I am grateful to Professor M. L. Samuels for referring me to this work. The sole example of OE byden given by Ek (p. 44) is Bensted Green, PNEssex, 267. For Bedlar's Green and Bedwell, Ek follows PNEssex (p. 35) and PNHerts. (p. 138) and adduces OE \*byde, thus ignoring EPN, I, 72, s.v. byden.

<sup>32</sup> As above, n.12. <sup>33</sup> For ME bede-hous, see MED, s.v.  $b\bar{e}d(e \text{ n., 2b.}$  Cf. OE bed-hus, EPN, I, 24

and the derived Welsh bettws, ibid. 32.

<sup>34</sup> Both the spellings in question are from local documents rather from those written at Westminster. For analogical ME spellings in e in the SW. of England as a reflex of OE y, see A. McIntosh, M. L. Samuels, M. Benskin, eds, A Linguistic Atlas of Late Mediaeval English, 4 vols (Aberdeen, 1987), I, maps 399 (DID), 414 (FIRST), 972 (BURY, BURIED), and 1040 (KIND, MIND); II, map 125 (4 and 5 FIRST: ferst(e)). I am grateful to Mr Victor Watts for referring me to these maps.

35 Cf. McIntosh et alii, Linguistic Atlas, I, 3-28.

This paper 1 describes the setting up of a computer-searchable database of Old English boundary clauses, based on the Toronto Dictionary of Old English Corpus. 2 It goes on to outline some preliminary investigations into the application of computing techniques to this material. Emphasis will be on methodology rather than on specific results.

The boundary descriptions contained within Anglo-Saxon charters are of interest within a number of different disciplines, engaging the attention not only of the linguist and onomast but also of the archaeologist, historical geographer, agrarian historian, historical botanist, local historian and so on. Many of the questions addressed to the material from these differing viewpoints require the bounds to be seen in relation to each other, whether it be to examine the general distribution of one or more place-name elements geographically over the whole country, or to focus on their diachronic relationships. However, each separate set of bounds (essentially *text*) has its own unique set of references (essentially *data*), and they are no more directly comparable as regards provenance and date than they are with respect to reliability. The computer, with its ability to manipulate data combined with its capacity for text analysis, seems, therefore, to be particularly applicable here.

The aim of this project, at first merely exploratory, gradually crystallized into the following: the establishment of a database of bounds which could not only be readily augmented and corrected, but could also be sorted, searched and analysed by a variety of programs. The essential point is that it was not set up in order to answer predefined questions, but rather to provide the material for studies within as wide a range of *potential* applications as possible.

The first step was the identification and extraction from the Toronto Corpus of all the citations containing Old English boundaries. These were assembled into a separate 'file' and extraneous matter was deleted. Each individual perambulation was then referenced with respect to Sawyer number,<sup>3</sup> purported date of grant, date of manuscript, name of grantor, estate name and county, and printed source. The assembled data was then converted into a format for processing by the information retrieval program 'Famulus77' and sorted into order of Sawyer number. In cases where several sets of bounds exist within one charter each set was individually itemized and referenced. The perambulations themselves were defined within their contexts by placing 'tagging characters' at beginning and end,

allowing subsequent programs to focus either on these alone or to include surrounding text.

The material could at this point be checked against Sawyer<sup>3</sup> to determine whether or not all items cited in this work as containing English bounds were included. Those which were missing were inserted, and erroneous references were corrected (see below, Appendix I). The sources used by Toronto for the charter material are almost exclusively printed editions, and variant manuscript versions are omitted. A description of the Toronto texts has been published by Healey and Venezky.<sup>4</sup>

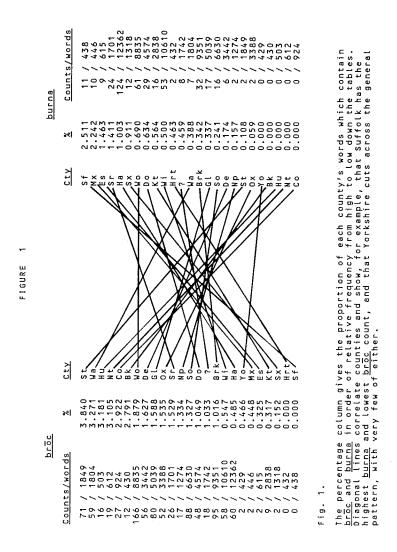
The referenced Bounds Corpus, numbering 880 perambulations, was finally re-sorted by Famulus into alphabetical order by county and estate. In this form it starts with 'Bedfordshire: Aspley Guise' and runs through to 'Yorkshire: Sherburn-in-Elmet', with 27 unidentified sets of bounds at the end. When the material is arranged in this way, similar versions of the same estate boundary, occurring in different charters, are thrown together and can be compared. Any relational work must clearly take account of such repetition, although only close topographical studies, county by county, can assess the degree of duplication of place-name elements in interrelated sets.

It was also possible at this stage to assess the degree of accuracy of the charter section of the Toronto Corpus by checking the material for the county of Hampshire against the manuscripts. In the 112 sets of bounds (12,362 words of boundary) for this county, there were found to be only seventeen instances of words miscopied by Toronto. Most of the emendations to the texts consisted of the correction of faults already existing within the printed versions (see below, Appendix II).

The material, once assembled and referenced, was edited into various formats for text and data analysis. This was finished by late Autumn 1987, and the writer then had until the end of March 1988 to experiment with its application. It is this somewhat cursory exploration which formed the basis of the Swansea paper 1 and constitutes the rest of this article.

An attempt was made to assess the relative usefulness of various computing techniques. The elements chosen for this purpose were the watercourse generics broc and burna, in the light of Margaret Gelling's analyses and of Ann Cole's preliminary regional investigations, both primarily concerned with the elements as they occur in major place-names.<sup>5</sup>

The Corpus had by this time been formatted for an adaptation of an interactive Search Program which ran on the ICL 2988 at Oxford and was



subsequently rendered inoperable by the demise of that machine at the end of December 1987.<sup>6</sup> Applying this, printouts were obtained of all the examples of the character-strings '%broc%' and '%brok%', and of '%burn%', '%brun%', '%born%' and '%bourn%', where % signifies any or no characters. In the present form of the Corpus the computer can only search for *strings*, not *headwords*, and unless specifically alerted will of course miss such aberrant forms as 'hroc' and 'burm'. Individual instances located by this method have to be checked; in the case of broc, for example, of which there were 1058 contenders, some broken items had to be discarded, together with one or two suspected badgers (brocen, brocc). Using this material, and counting examples county by county, a pilot study was undertaken, which I outline at this stage more as an indication of potential than as an examination of the two elements *per se*.<sup>7</sup>

SEARCHTEXT located all the examples of broc and burna as defined above and printed them out, within their immediatetextual context, in county/estate order. Two different methods were then applied to make comparisons with respect to the *relative* proportion of each county's material containing the two words. Firstly, the total number of mentions of broc/burna for each county was compared with the total number of words of boundary for that county, and secondly, the total number of discrete sets of bounds which contained at least one broc/burna was compared with the total number of bounds, county by county.

Figure 1 shows the results of the first method. Overall counts were 1033 brocs and 455 burnas. No distinction was made between their occurrence as unqualified nouns and as qualified place-name or quasi-place-name elements, nor was any attempt made at this stage to eliminate repetition within the boundaries. Only counties with a minimum of 400 words of boundary were included.

The percentage *figures* themselves in Figure 1 are fairly meaningless; the number of 'elements' has been compared with the number of 'words' and clearly many words themselves consist of several elements. The same method has been applied equally to broc and to burna, however, and it is the general *order* of the counties which is significant. The lines in Figure 1 correlate counties and show the essentially complementary distribution of the two terms. Figure 2, although 'busy', shows the same data in a way designed to give a clearer idea of *relative* distributions, and to throw into relief idiosyncrasies in the patterns. Here the counties are set not simply in descending order, as in Figure 1, but are also placed in their percentage positions on the scale.

The complementary nature of the broc/burna relationships is better displayed in such a graph as in Figure 2; the group of high-frequency broc counties correlates with the low-frequency burna counties, there is a levelling across the

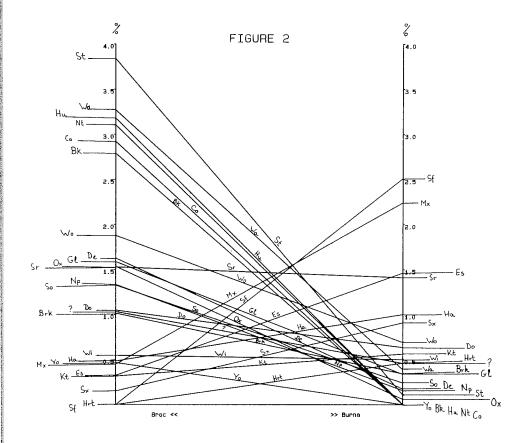


Fig. 2 SIMPLEPLOT Graph  $^8$  showing the percentage of words per county containing broc (from Sf and Hrt with 0% to St with 3.840%) up the left axis, and burna (from Co, Nt, Hu, Bk and Yo with 0% to Sf with 2.511%) up the right axis.

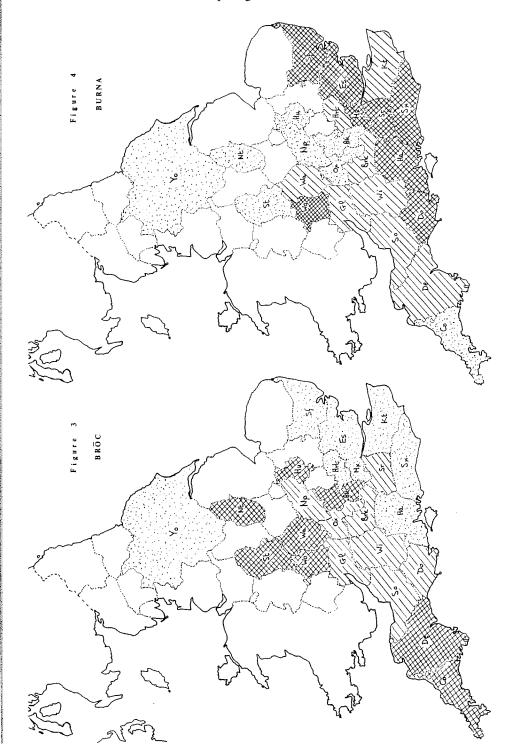
counties in the middle, and the low broc counties generally correspond with the highest of the burna counties.

Figures 3 and 4 are maps which attempt to illustrate this distribution. Of the 24 counties, the first eight with the highest counts (Group 1) are cross-hatched, the next eight (Group 2) are hatched, and the eight counties with the lowest counts (Group 3) are stippled.

Alterations to the broc map using the second method of analysis, which eliminates most of the duplication but narrows retrieval to no more than one count per boundary, would have moved Northamptonshire into Group 1, and just nudged Devon into Group 2. These two counties would have also just exchanged position in the burna map, as would Wiltshire and Sussex. Neither of these methods is altogether satisfactory; particularly sensitive are counties with a very small amount of material, such as Suffolk and Middlesex which lead the burna table. Nevertheless, it could be seen that the second approach did not disturb the overall picture established by the first, even though absolute positions on the scale varied.

In the study described above, place-names within Latin bounds were not included, nor was any account taken of dates (either of grants or of manuscripts) or of the authenticity of individual texts. It is in these respects that further work can use the facilities of the computer to refine analysis, and the programs be re-run in ways which take account of these factors. Instances of broc and burna in such phrases as on bone broc need to be distinguished from cases where these elements are qualified in some way. In this subsequent exercise, interpretation of the maps and tables with regard to Cornwall is modified. Here it can be observed that, although the word broc occurs relatively frequently in this county, there is only one possible candidate for place-name status – on wone broc in Landrake. Adding an ascender to a form using the character 'wynn' for 'w' (to emend to bone) strongly suggests that there are in fact no examples of broc used as a qualified place-name element in the Old English boundary material for this county.

This is a rough and ready, but rapidly generated, way of looking at relationships between semantically related pairs of words. It was encouraging to see that the first method of analysis suggested distributions for burna as a lexical item very similar to that outlined from major place-names by Margaret Gelling in 1984,<sup>5</sup> and interesting to observe that the broc data would seem to give a complement to this pattern. The method will be applied to other elements whose general toponymic distribution is already known and, if the results continue to be promising, to those where this is as yet largely unexplored.



The experiment using SEARCHTEXT gave county/estate distribution, and it will be obvious that producing output which orders words in this way will greatly facilitate the assembling and location of place-name elements and their subsequent plotting at county or national level. There are also clear implications for detailed regional topographical studies. However, for the purposes of this exercise, attention turned away from geographical distribution to the facilities available for looking at the *nature* of brocs and burnas through a study of their qualifying terms. The Corpus was thus reformatted for the application of a second text-analysis tool, the 'Oxford Concordance Program' (hereafter OCP). 10

By means of a combination of editing and applying OCP, lists of words containing broc and burna were generated in quite different sequences to that illustrated above. Instead of operating on the *references* to order the material, OCP now sorted by criteria within the *texts* themselves. Firstly, output was produced in alphabetical order of the word to the left of the keyword if the latter was immediately preceded by a space, but otherwise in alphabetical order of the word containing the 'string'. This ensured that compounds in forms with and without spaces were grouped together (e.g. hunig burna with hunigburna). Figure 5 shows an extract from the broc printout where, in addition, the strings 'broc' and 'brok' were declared as equal for sorting purposes in the OCP command file.

Ordering the data in this way allows us to address such questions as the following:

- 1) Are the descriptive words applied to burna of the same general semantic range as those for broc?
- 2) Are there any particular adjectives or other elements which occur quite frequently with one but rarely or never qualify the other?
  - 3) Is one more likely to be preceded by a personal name than the other?
- 4) Is their occurrence as the *first* element of a compound of similar relative frequency?
- 5) In counties where broc and burna are both infrequent, what other watercourse terms are used? What is the distribution of terms like wella, &well, &welm, bace, flode, lacu, lad, sic, ripig etc? In what sorts of compounds and with what range of qualifiers do these occur?
- 6) Are such differences as can be observed with respect to these and other elements synchronically, diachronically or semantically related?

It is not within the scope of this paper, nor was it within the scope of the pilot study, to extrapolate from the results of this investigation; a few observations must suffice.

The most immediately obvious point to emerge was the difference in the rates

'n
GURE

CLY         ESTAGE         SANYOR         MS.           SO         BATCOMBE         462         940         14           WO         PENSAX etc.         1595          12           WO         PENSAX etc.         1595          12           OX         CUDDESDON         587         956         10           OX         CUDDESDON         587         956         10           SO         KINGSBURY         1570          13           SO         KINGSBURY         1165         x674         13           SO         PENNARD MINSTER         578         955         14           SO         PENNARD MINSTER         253         950         19           SO         PENNARD MINSTER         253         955         10	Lext	of bare hulle on folanbrok	Of soln hama broc into suð fordes broc, of suð fordes broc of suð fordes broc to ðan lange sice	On fost broc. Of fost broce on pone hlið weg	On fost broc. Of fost broce on bone hlið weg.	adoun to Foulenbroke Alang Foulenbroke out on a Moure	Fram pe Hechche endlonge pes frithesbroke to pere hore epeldure	Of pare hegoean nyper innan bone fulan broc of dam fulan broce and lang pass atar sceatas on pona fulan broc on and lang fulan brocaso on ge rihtæ to procbyrga on pona fulan broc and lang fulan brocaso And lang hnut cumbes east. On fulan broc. Of fulan broce Of shiren pole on fulan broc. Of fulan broce bronch burb bar fulan broc. Of fulan broce bronch burb bar fulan broc. Of fulan broke bronch burb bar fulan broc. Of fulan broke bronch burb bar fulan broc. Of pam broce and lang streames best hit fehå on trentan par fulan broc scyet on Irentan bar swa dun on done fulan broc	on bonæ fulan broc and lang fulan brocæs utt on bytnan burnan	of Jone fulan brocc Rlang brocms	innan bone fulan broc of ðam fulan broce wip westan randes æsc bæt thanon rihte on Fulan broce. And swa and lang broces gæ rihtæ to broc briggæ on bonæ fulan broce On fulan broc. Of fulan broce on þa miclan dic	Fram þam steorte andlang þæs fulan broces oð bliðan Andlang weges to fulan broces heafdon, andlang broces	on bone fule broc, on lang broces Andlang Bæle in Fulebroc. Of Fulebroce in Sihtfyrð	
Estate   Document	MS. date	14	12	10	10	ជ្	13		12	12	1240	10		
Estate   Double   D	Date	076		926	926		¥674	903 826 939 958 955 8674 956	939	626	903 950 826 955	944	994 1042	
BATCOMBE PENSAX etc. CUDDESDON CUDDESDON CUDDESDON KINGSBURY KINGSBURY KINGSBURY KINGSBURY KINGSBURY EGHAM RISBOROUGH DROXFORD PROXFORD PROXFORD PROXFORD PROXFORD PROXFORD PROXFORD PROXFORD PROXFORD DROXFORD DROXFORD PROXFORD PR	Sawyer no.	797	1595 1595	587	587	1570 1570	1165		955	977	367 553 276 563	* * *	1380 1395	
K	Estate	BATCOMBE	PENSAX etc. PENSAX etc.	CUDDESDON	CUDDESDON	KINGSBURY KINGSBURY	ЕСНАМ	RISBOROUGH DROXFORD DROXFORD PROXFORD PENNARD MINSTER CHERTSEY & THORPE MERSTHAM DARLASTON THYFORD etc.	DROXFORD	DROXFORD	RISBOROUGH PUCKLECHURCH DROXFORD PENNARD MINSTER	BADBY cyngbrycge		5
	Cty	So	99	ŏ	ŏ	So So	S	Synookaak Kananorite	æ	no ==	8 E G E S	N d	S t	

of occurrence of broc and burna as unqualified nouns – simply 'the brook' and 'the bourn'. This is illustrated by an examination of the words as they appear after the preposition 'along'. For this the output defined above and illustrated in Figure 5 was used, supplemented by that generated by a listing of the word andlang with all its variant spellings grouped together. This latter listing was produced in alphabetical order of the word to the *right* of the preposition (see Figure 6).11

The following observations could then be made:

a) There were 103 examples of burna preceded by the preposition andlang, compared with 245 of broc in this position. 39 (approx. 38%) of these burna examples were andlang plus inflected burna with no intervening word (andlang burnan etc.); 3 (3%) were andlang plus definite article plus burna (ondlong bære burnan etc.); 56 (54%) were andlang plus burna within a 'place-name', commonly in compound form but also qualified by an inflected adjective or a personal name (andlang greot burnan, & lang limburnan, andland [sic] bradan burnan and so on). The remaining five instances were andlang plus burna as the first element of a compound (in each case burnstow).

b) In marked contrast to this, of the 245 brocs 178 (approx. 73%) were of the andlang broces type; 19 (8%) were andlang plus definite article plus broc; only 43 (18%) were andlang plus broc with a qualifier. There were two examples of andlang plus broc as the first element of a compound and three of the preposition plus definite article plus qualified broc-name.

This information modifies our reading of the tables and the map distributions and would seem to have clear implications with regard to the relative dating of these terms. Analysis, however, needs to be further broken down by location, date and textual background. In this way a series of maps and overlays could be produced to reflect these more subtle analyses. If this proved successful, the technique could be applied to other elements where, unlike the case for broc and burna, such marked differences in use might not have been predicted.

When attention turned to the qualifying terms themselves questions 1 and 2 above could be addressed. After a broken-down barrow and an unaspirated hawk had been rejected and re-alphabetized respectively, 12 both lists could be seen to start with the element alor 'alder', with several mentions each over several counties. Other elements qualify both words: myln 'mill', a significant word for dating purposes, occurs in three counties (Somerset, Dorset and Wiltshire) with burna, and in three (Dorset, Hampshire and Berkshire) with broc. The table in Figure 7, however, gives the distribution of those qualifying elements which appear to show a distinct bias towards one or other of the watercourse generics broc and burna. This table gives the number of separately named estate

ţ	Estate tate	Sawyer no.		Text		
	-	į				
ı,		679	andlang	proces		
<b>.</b>	HILTON NR WOLVERHAMPTON etc.	1380	on lang	broces	in Brunesford ŏer hit er of eode	
	REDMARLEY D'ABITOT	1338	0nd long	broces	in Ledene.	
æ	WORMLEIGHTON	588	andlang	broces	in to cranmere	
æ	ARLEY	898	ondlong	broces	in ba burnan	
יס	CHALGROVE	396	anlang	broces	in banne oðerne ford	
0	HARTLEBURY; WHITELINGE IN	1320	and lang	broces	innan ceolmes gemære	
×	EYNSHAM	911	Andlang	broces	innon Bladene.	
0	STOKE; SOUTH	769	andlang	broces	innon cameler	
<u>ب</u> د	OLNEY	834	Andlang	broces	inon Use. Andlang Use	
æ	OLDBERROW	42	Onlong	broces	midstreame on ba lytle riðig.	
_	HAWKESBURY	786	andlang	broces		
e	ASHFORD IN BURLESCOMBE etc.	653	andlang	broces	on anne stanihtne ford	
_	PUCKLECHURCH	553	and lang	broces	on Baddanham	
<b></b>	timbinctun NTH OF THE TYRL BROOK	141	andlang	broces	on blacan mere	
æ	WORMLEIGHTON	1574	andlang	broces	on blacan mere	
0	TWYFORD etc.	1599	onlong	proces	on burhwyllen	
	PUCKLECHURCH	553	and lang	broces	on Bydincel	
¥	WOOTTON	858	andlang	broces	on cealdan wyl	
0	WOLVERLEY; COOKLEY IN	726	andlang	broces	on cenunga ford	
	PATNEY	715	and lang	broces	on cnihta bryc	
0	LITTLE WITLEY	1372	Andlang	proces	on doferic	
0	TWYFORD etc.	1599	Onlong	broces	on domnipol.	
	NECNHAR	226	andlang	broces	on fearnhylles sic	
¥	GINGE; WEST	673	andlang	broces	on garing broc	
4	DARLASTON	602	and lang	broces	on gean stream on fulan ford	
¥	WOOTTON & SUNNINGWELL	290	and (ang	broces	on gorgraues bar hit ar on feng	
_	CONINGTON	649	& Lang	broces	on gyruwan fen	
	madanlieg	801	ondlong	broces	on hedenan mos	
×	CUDDESDON	587	Andlong	proces	on herpab ford	
63	ARLEY	898	Ondlong	broces	on hwitan lege	

NUMBER ER OF ER OF

orted <u>andlang</u> output, with variant spellin ting purposes. Extract from '<u>andlang</u> broc boundaries which contain the particular qualifier. For instance, a combination of the adjective hol 'hollow' + broc, although mentioned altogether 39 times, in fact occurs in 28 discrete perambulations in fourteen counties. This removes all duplication except where two different boundaries name the same physical feature. The figure for mearc is the only one which is significantly altered when this duplication is removed, as the seven instances appear to refer to only five features. The other boundaries are scattered throughout their counties, and a study of the topographical work published for them confirms that there would appear to be relatively little overlap here.

FIGURE 7

	br	ōc	burna
	County	<u>Estate</u>	$\frac{\text{County}}{0}$ $\frac{\text{Estate}}{0}$ 'boundary'
mearc	4	0.5	
(ge)mære	+ 7	25	
hreod	2	2	4 4 'reed'
hol	14	28	1 1 'hollow'
$f\overline{u}l$	10	17	0 0 'foul'
brād	0	0	4 4 'broad'
smæl	5	6	0 0 'narrow'
$\overline{\text{woh}}$	0	0	6 'twisted'
winter	0	0	7 14 'winter'

+There can be some confusion of elements here, and a few ambiguous contenders were rejected from the broc table. Those included all appear to be OE (ge)mære 'a boundary'.

We can see from the table that the words for boundary (mearc and (ge)mære) together qualify broc in a total of 32 different perambulations, whereas they qualify burna in only two. The adjectives hol and ful together occur with broc a total of 45 times but there is only one occurrence (hol) with burna. Adjectives do occur proportionately less frequently with burna than with broc; the adjective brad, however, describes burna in four separate estates in four counties, but does not appear to qualify broc in this material; woh would appear to share this predilection.

OCP was then programmed to order alphabetically the words immediately to the right of the headwords broc and burna. The resulting output, when supplemented by a strictly alphabetical listing of words commencing with the key elements, enables their occurrence as first elements of compounds to be studied. Are they frequent or rare in this position? What is the particular range of words

occurring as the second element in compounds with burna as opposed to those with broc? Choosing more commonly occurring first elements, such as mearc or wudu 'wood', or selecting an adjective, would give a better indication of the potential of such right-sorted OCP runs. Both broc and burna are rare in this position. Burna only appears in the compound burnstow, although this occurs in eight separately named estates in three different counties (?seven distinct features). One of the particular advantages of a concordancing facility such as OCP is the rapidity with which a line of enquiry can be pursued. A list containing all the other words compounded with stow, with contexts and references, can readily be generated and printed out. Figure 8 shows just the headword list, with frequency count, for the various forms containing the element. The full list, too long to reproduce here, shows further that the many hegestowes are all in Worcestershire and include much duplication, whereas there are three separately occurring cwealmstowes (in Hampshire, Huntingdonshire and Wiltshire). Confusion between 'p' and 'wynn' in many manuscripts, illustrated by the spel/swel forms, has to be taken into account and explains the inclusion of 'stop' in the search criteria.

### FIGURE 8

Belstowa	1	hegestowe	10
Belstowe	1	hegstowe	4
burestowe	1	holding stowe	1
burnestowe	2	mære stowe	2
burn stowæ	5	pleginstowe	1
burnstowe	10	spelstowe	1
burstowe	1	${\sf swelstowe}$	1
cot stowa	1	þære stowe	1
cotstowe	2	ðere stowe	2
cotstowum	1	halige stoue	1
cwealmstowe	4	þe stourewe	1
Finstowe	1	wodestou	1
halgan stowe	1	hæg stope	3
halige stowe	1	hæge stope	1

Frequency list for the words containing the strings 'stow', 'stou' and 'stop', excluding the stream-name 'stour'.

The results of the broc/burna study suggest that the material and the programs are well able to provide the basic information to approach a wide variety of questions, of which the following provide a sample:

a) What other elements commonly occur in conjunction with the words for boundary, (ge)mære and mearc? Is there a noticeable distinction between these

terms? What is their distribution? (If, for example, a compound such as mearcbroc only occurred in late charters or in early charters with dubious credentials, or if it had instead a particular regional bias or indeed a connection with a particular scriptorium, such patterns should be discernible.)

- b) Which elements are frequently qualified by a personal name, and which, though perhaps just as common, are rarely or never so qualified?
- c) What prepositions are most commonly used with the particular elements under investigation?
- d) Do some elements more than others tend to be preceded by the definite article when following, say, the preposition andlang? How would the addition of information with regard to date and provenance supplement this picture?
- e) With respect to an element such as dīc 'ditch, dyke', roughly what proportion of the total count is described by the adjective (e)ald 'old', and what other features are commonly qualified by this adjective?
- f) What distinctions can be seen between the occurrence of  $d\bar{t}c$  as a masculine as opposed to a feminine noun, and between the use of inflected and uninflected forms?
- g) What sorts of features do the boundaries most commonly go along? How often does the boundary cross, as opposed to follow, linear features of various types? What distinctions are there in this respect between the 'way-names' stræt, weg, pæb, herepæb and others?

A further feature of OCP is its ability to generate word lists, word frequency lists, and indexes. At present, for the Bounds Corpus, this is simply an alphabetical listing, with frequency count, of every separately occurring discrete string of characters – i.e. any sequence of characters preceded and followed by a space or a line end. It is only when the Bounds Corpus has been glossed with headwords that something approaching an Index of place-name elements, personal names and so on can be generated. The string-lists do, however, show the range of spellings and inflections for each element and will provide the basis for glossing programs. 13 Figure 9 shows such a list, and Figure 10 gives the forms for the place-name element byrgels, selecting Sawyer number as the reference; other references such as County or Date could be substituted here.

#### FIGURE 9

æfes	1	ælfnoðes	1
æfesce	1	ælfredes	6
æffan	1	ælfric	3
æfisc	4	ælfrices	4
æfne	1	ælfriðe	1
æfre	6	Ælfrucge	1
æfsan	2	ælfsige	1
æft	90	ælfsiges	3
æftær	2	ælfstan	1
æfte	1	ælfstanes	4
æfter	136	ælfðryðæ	1
æfter%	2	ælfþryðe	4
ægan	2	ælfweardes	1
ægces	1	ælfwenne	1
æge	1	ælfwerdes	1
ægelwardes	2	ælfwiges	3
æglardes	1	ælfwine	1
ægles	1	ælfwines	4
æglesuullan	1	ælfwoldes	2
ægscce	1	ælgares	1
ægþeres	1	ælla	1
ækera	2	ællan	2
ækeran	3	ællen	2
ækeras	1	ællenstubbe	1
ækergeat	1	ælles	2
ælbroc	2	ælmarches	1
ældan	4	ælmundes	2
ældæn	1	ælphwines	1
ælde	1	ælr	2
ældred	1	ælrenan	4
ælesbeorge	1	ælsyges	1
ælferes	1	ælwilme	1
ælfflede	3	æmbrihtes	1
ælfgares	2	æmenan	1
ælfgyðe	1	æmices	1
ælfheages	2	$a_n$	1
ælfheres	1	ænd	1
ælflæde	3	ændæ	4
Ælflede	5	ænde	8
ælfnodes	1	ændlang	1

Word list (<u>æfes</u> to <u>ændlang</u>) with frequency count.

[Đ, þ, ổ and þ are declared as equal in the command file for this run, as are æ and Æ. It is an idiosyncrasy of OCP that the form which appears in the list is simply the first encountered in the file. So, for example, the four instances of <u>ælfþryðe</u> could well include spellings <u>ælfðryþe</u>, <u>Ælfðryðe</u> etc. % = a reconstructed form.]

146

### FIGURE 10

Form	Sawyer no.
beorgelese	714
beorgelse	714
berielese	744
birgilsond	541
birigelsan	738
birigelsum	276
burcels	802
burgilsan	368
byrcelse	802
byregels	104
byrgeles	496, 673
byrgelese	673
byrgels	367, 413, 495, 586, 590,
	605, 690, 756, 820, 820,
	1566, 1568, 1568
byrgelsan	43, 43, 379, 413
byrgelsas	449, 487, 558, 977
byrgelsæ	585
byrgelse	366, 438, 495, 496, 585,
	585, 586, 647, 647, 685,
	685, 690, 1566
byrgelson	977
byrgelsum	414, 496, 582
byrgilis	366
byrgils	784
byrgilse	784
byrgylsas	517
byrgylse	800
byrieles	645
byriels	651, 1248, 1588, 1599
byrielse	651, 1248, 1599, 1599
byrigæls	427
byrigels	104, 317, 381, 503, 575
byrigelsan	695, 695, 699, 699, 1013
byrigelsas	414, 523, 608
byrigelsæ	317, 427, 503
byrigelse	179, 377, 800
byrigelsum	446
byrygels	874
heggenberles	568
50	

OCP Index extracted from a printout which isolated forms \*b\*r\*l\*s\*, where \* represents any or no characters. Most of these forms represent OE byrgels 'a burial place', and show the range of spellings in the Corpus for this element.

After this exploratory OCP work had been carried out, the whole of the Old English boundary material was converted for storage as a Relational Database, using the program INGRES. This is a powerful analytical tool which can not only re-generate the material in a form upon which OCP can be run, but is also able to further process the data itself. The texts and references can now be continually updated and corrected as spellings, checked against manuscript forms, are revised, and as references are modified. Superior versions can be substituted: Toronto's text for the estates of Cowley, Cutslow, and Whitehill in Tackley, Oxon. (Sawyer 909), following Kemble, 14 would appear to be a less satisfactory version than that found in the St Frideswide's Cartulary and will in due course be replaced. This is one of several such examples. New material can be added; the Latin boundaries represent an obvious lacuna in the Corpus, but some post-Conquest charters might merit inclusion. Much information can also be inserted from the extensive work being carried out on charters at a detailed regional level elsewhere. As the material becomes more reliable, so the range of applications will grow.

It is important to stress that certain investigations (phonological, dialectal, orthographical, etc.) can only legitimately be carried out on the manuscript-checked material. Studies of the broc/burna type, however, can justifiably be based on the whole Corpus. These will necessarily be preliminary statements to be followed up by more detailed work. The pilot OCP executions outlined above were carried out at the most basic level of analysis, making no distinction between each boundary text with respect to date, manuscript provenance, reliability, etc. These references, which were described at the beginning of this paper as essential to any *full* investigation, are now carried in the Ingres database, and OCP work will henceforth take account of them. Only when this is done will we be able to address the last of those questions listed with respect to broc and burna (point 6 above, page 138).

The work described here has grown from exploratory beginnings into a large project. It can be seen that both the manipulation and analysis of a mass of texts are made difficult when each one not only possesses a *unique* combination of attributes covering a broad geographical and chronological spread, but also contains a kaleidoscopic juxtaposition of place-name elements, adjectives, personal names and so forth. The computer can rapidly search and organize this material and, although rarely providing instant answers, can generate the data for a wide range of investigations. These observations apply equally to the analysis of major place-names.

The availability of the Dictionary of Old English Corpus has stimulated many fields of Anglo-Saxon scholarship over the last few years and it is appropriate to

end this paper by acknowledging my indebtedness to the Toronto Project. Their work not only provided the raw material upon which the study was based, but also acted as a stimulus for the examination of research facilities only relatively recently available. Such computing resources must surely have a role to play in the future of name-studies.

#### ST EDMUND HALL, OXFORD

## APPENDIX I

The following are emendations to Sawyer<sup>3</sup> noted during the course of the work described above. Unless otherwise specified, all references are to the section between the Sawyer no. and the details of MS(S). No attempt has been made to add material dating from after 1968.

p. 8 Add Finberg 1958 H.P.R. Finberg, "Three Anglo-Saxon
Boundaries', Trans. Shrops. Arch. Soc. 56 (1958), pp. 28-33.
p. 56 <u>Under B.M., Lansdowne 269 for 452 read 980</u>
Under B.M., Lansdowne 247 for 452 read 980
p. 80 no. 35 After ? Redham in Cliffe insert , Kent
p. 85 no. 59 After Latin add with bounds
p. 05 no. 99 Por Letin with English hounds read Letin with hounds
p. 91 no. 80 For Latin with English bounds read Latin with bounds
p. 92 no. 84 After Latin with English add /Latin
p. 97 no. 101 After Latin add with bounds
p. 108 no. 146 After Latin add with English
p. 114 no. 168 After in English) add, all in Kent
Under Comments: for Ward 1934 read Ward 1934/1
no. 169 After Swarling add, Kent
<u>Under Comments</u> : for Ward 1934 read Ward 1934/1
no. 170 Ditto
p. 116 no. 177 Ditto
no. 178 Ditto
p. 135 no. 254 For Cearn, Somerset read Charmouth, Dorset
p. 138 no. 268 Under MS.: For Claud. B xi read Claud. B vi
Under Comments: Add Crawford 1922, pp. 68-9, with map
facing p. 81.
p. 148 no. 311 After Latin with bounds add in Latin and English
p. 154 no. 338 After Latin add with bounds
p. 155 no. 342 For Cheselborne read Cheselbourne
p. 155 no. 342 For Cheselborne read Cheselbourne p. 157 no. 353 For Latin with bounds read Latin with English bounds
p. 159 no. 359 Under MS: For s.xi read s.xii
no. 360 <u>Under MSS</u> : <u>After 3. B.M.</u> , Harley 1761, fo. 48 (s.xvi
add; no bounds [For s.xvi read s.xiv (A.R.R.)]
p. 163 no. 374 <u>Under MSS</u> : <u>After 1. B.M., Harley 1761, fo. 47v</u> (s.xvi
add; no bounds [For s.xvi read s.xiv (A.R.R.)]
no. 377 <u>Under <i>MSS</i></u> : <u>After 2</u> . B.M., Add. 15350, fo. 57rv (s.xii
add; no bounds
p. 165 no. 383 <u>Under Comments</u> : <u>Add</u> Grundy, <i>Hants.</i> , 1926, pp. 133-5;
Crawford 1922, p. 76
Clawfold 1722, p. 70

p. 166 no. 387 Under Comments: For Finberg 1954 read Finberg 1953 no. 388 Ditto p. 170 no. 407 After Latin add with bounds p. 174 no. 427 For of Enford, Hants. read of Enford, Wilts. p. 176 no. 433 Under Comments: For Finberg 1954 read Finberg 1953 p. 178 no. 442 For Lyme Regis, Dorset read ?Uplyme, Devon p. 180 no. 452 Delete bounds Under MSS: Delete versions 1 and 2 [these belong with no. 980] p. 187 no. 485 For Cheselborne read Cheselbourne p. 195 no. 526 For Latin with English bounds read Latin with bounds in Old English, Middle English and Latin p. 197 no. 536 Delete bounds p. 221 no. 653 <u>Under Comments</u>: For PN Devon, i read PN Devon, ii p. 228 no. 689 <u>Under MSS</u>: <u>After 2. B.M., Cotton Claud. C ix, fo. 119 (s.xii)</u> add: no bounds no. 690 Under MSS: After 3. B.M., Cotton Claud. C ix, fos 119v-20 (s.xii add; no bounds p. 245 no. 771 For and read land p. 256 no. 804 For Latin with English bounds read Latin with bounds in Latin and English Under MSS: After 4. Winchester, DC. 'St Swithun's Cartulary', pt 3, fo. 136v, no. 546 (s.xiv add; no bounds Under Printed: For ex MS. 4 read ex MS. 5 p. 259 no. 818 For Tichbourne read Tichborne p. 263 no. 840 <u>Under MSS</u>: After 2. B.M., Add. 15350, fos 68v-69 (s.xii add; no bounds p. 264 no. 845 Under MSS: After 1. B.M., Harley 1761. fo. 33rv (s.xvi add; no bounds p. 267 no. 862 For Latin with English bounds read Latin with bounds p. 268 no. 870 After Wilton add, Wilts. p. 269 no. 874 Under MSS: Delete entry 2 [This is a different text in Middle English] p. 270 no. 879 After Latin add with English bounds p. 275 no. 901 <u>Under Comments</u>: For Glones read Gloues. p. 277 no. 907 <u>Under MSS</u>: After 5. B.M., Add. 5811, fo. 2 (s.xviii add; no bounds Under Comments: For PNEssex, p.532 read PNEssex, p.530 p. 281 no. 922 After Latin add with English bounds p. 287 no. 955 For Cheselborne read Cheselbourne p. 293 no. 980 <u>Under MSS</u>: Add B.M., Lansdowne 269, fo. 98rv (s.xvii) B.M., Lansdowne 447, fo. 25rv (s.xvii) p. 297 no. 993 <u>Under MSS</u>: <u>After 2. B.M., Cotton Claud. C ix. fo. 130 (s.xii)</u> add: no bounds p. 302 no. 1015 After Latin add with English bounds p. 347 no. 1180 After Latin add with English p. 371 no. 1280 After Latin with English add bounds p. 376 no. 1301 <u>Under Printed</u>: <u>After 144-45 add ex MS. 1 and p. 352 ex MS. 2 no. 1303 <u>Under MSS</u>: <u>Delete entry 3 [This is a different and merits a different and different and different and different and different and differe</u></u>

no. 1305 After English add with bounds

p. 391 no. 1380 After Upper Arley insert, Worcs.

p. 405 no. 1443 After English add with bounds p. 433 no. 1552 For English read Latin

p. 378 no. 1311 Under Comments: After p. 299, insert cf. B 636 where

p. 377 no. 1311 Delete bounds

p.	437	no.	1577	Ditto
-		no.	1578	Ditto
p.	438	no.	1585	Ditto
D.	506			Under Froxfield, Hants.: Delete 1263
Ι.				Insert Froxfield, Wilts., 1263
D.	510			Under Highclere, Hants.: For 381 read 383
Γ.				Under Himbleton, Worcs.: Add 174
p.	517			Under Mersea, Essex: Add 1015
Ď.	522			For Rainham, Essex read Rainham, Kent
Ď.	526			After Southampton, add Hants.,
'n.	530			For Tichborne, Hunts. read Tichborne, Hants.
p.	535			Under Winchester: Add 871 and 904.

#### APPENDIX II

It might be useful to comment here on the reliability of the charter section of the Toronto Old English Corpus, in so far as it can be assessed from the texts relating to Hampshire; the relevant material for this county forms a total of 12,362 words.

One short boundary (Sawyer 1558) was rejected as being altogether too inaccurate a representation of its original. This is a case where the application of ultra-violet light to the manuscript revealed many mistakes in the printed edition

cited in Sawyer. A more accurate version was substituted.

This apart, there were 214 cases where the Toronto form did not reflect that found in the manuscript in a way which was deemed to be significant. Only seventeen of these were due to mistranscription by Toronto; the rest perpetuated differences already existing within their sources. The 214 cases fell into two main groups, the great majority belonging to the first. These are errors which would not affect the sort of work illustrated in this paper. Many are instances of editing by Birch and Kemble etc., often representing their 'correction' of the text, whether deliberate or otherwise. 14 Typical examples are and and for MS and land, endlang, anlang etc.; noro for MS nord; six for syx; bonne for bonnæ; ærest for ærert; bonon ut for bonot ut; bæt for bet; fold for fald; weard setle for weard secle; litlan for littlan; onfeng for orfeng and ærest for arest. Also included under this heading are the omission of banon, bæt, etc., and faulty treatment of contractions such as bon for bon, where the general sense of the original is not significantly altered. Many of the forms in this group would be upheld in a normalized edited version; indeed, computer-readability of the texts would often be improved as regards the correct interpretation of items of Old English vocabulary. Such complacency would be entirely inappropriate, however, with respect to detailed linguistic investigations, whether phonological or orthographical. For such studies, working from manuscript-checked versions of the Corpus would be essential.

The second category of misrepresentation comprises instances where running a program on the uncorrected Corpus would be likely to lead to faulty identification of the word or its grammatical form, or to failure to locate it altogether. Representative examples are and for MS andlang; be sgares for bes gares; bocmere for bocmeres; mere for mere; mores for meres; brennungrafe for brenninggrafe; wone for bone; be for be; be for be; herde for hende; af for of; ap for yp; heore for heort and herbabes for herpabes. These are much more serious faults, of which there were 44 examples in the Hampshire perambulations. Curiously, although the level of accuracy of Toronto's transcription was seen to be very high, six of these more significant errors came from among the seventeen Toronto mistakes: tweora for MS treowa; wyrteles for wryteles; boh for poh; wæne for bæne; witwara for citwara and a transposition of two phrases within a clause.

A minor category of emendation to the Toronto Corpus consisted of differing

preferences in cases where the letters 'wynn' and 'p' are indistinguishable in the manuscript – for example, <u>pitleage</u> preferred to Birch's and Toronto's <u>witleage</u>.

Punctuation, capitalization and word division are as idiosyncratic in the Corpus as they are in the printed editions and, indeed, in the manuscripts themselves.

### **ACKNOWLEDGEMENTS AND GENERAL COMMENTS**

This work was carried out at Oxford University Computing Service, and I acknowledge with gratitude all those who gave such generous help and advice.

FAMULUS77 proved an unwieldy tool for sorting and updating. INGRES (see above), which was not initially available at OUCS, would have been better for

these purposes.

Commands which compare two texts and list all the differences between them (the DIF command on VAX systems running VMS for example), were very useful for the rigorous checking which had to be carried out regularly, although after major reformatting there were usually no short cuts to careful proof-reading at all

stages.

The editing programs used for the repeated manipulation of text were ECCE (Edinburgh Compatible Context Editor) and, more rarely, the editor EDT. ECCE is a versatile repetitive editor which was able to effect all the format changes needed, with the exception of the sequential numbering necessary for the conversion to INGRES. I am grateful to Catherine Griffin at OUCS who kindly wrote a SPITBOL program to do this. I am also indebted to Lou Burnard who modified his interactive program SEARCHTEXT and thus enabled me to apply it to my material, and to David Miles who wrote the SIMPLEPLOT program which produced Fig 2 (see below, n. 8).

#### NOTES

<sup>1</sup> This is a revised version of a paper given on 27th March 1988 at the XXth Annual Study Conference organized by the Council for Name Studies, held at the University College of Wales, Swansea. The county abbreviations used are those of the English Place-Name Society. Underlined elements or name-forms that appear in the present article are quoted from the Bounds Corpus.

<sup>2</sup> This material is described in the Introduction to the 'Toronto Corpus – Version 2, January 1988' as 'a complete record of all surviving Old English texts, excluding variant manuscripts ... prepared at the University of Toronto's Centre for Medieval Studies as a part of the Dictionary of Old English project'. It is generously made available by the Editors, via the Oxford Text Archive.

<sup>3</sup> P. H. Sawyer, Anglo-Saxon Charters: An Annotated List and Bibliography

(London, 1968).

<sup>4</sup> A. di Paolo Healey and R. L. Venezky, A Microfiche Concordance of Old English: The List of Texts and Index of Editions (Toronto, 1980).

<sup>5</sup> M. Gelling, *Place-Names in the Landscape* (London, 1984), pp. 14-20. Ann Cole is currently examining the stream-terms in relation to their geological setting along the lines of her 'Topography, hydrology and place-names in the chalklands of southern England: \*funta, &wiell, and &wielm', ante IX (1985), pp. 3-17.

<sup>6</sup> SEARCHTEXT – software developed at Oxford University for rapid searching of specific texts in the Oxford Text Archive when these were held on the ICL mainframe. See further L. Burnard, 'CAFS: A new solution to an old problem', Literary and Linguistic Computing II (1987), pp. 7-12.

<sup>7</sup> I should like to acknowledge the help and encouragement of an Oxford External Studies class during this early stage of the investigation. I am particularly grateful to Michael Pont who wrote the program to produce tables such as that illustrated in Figure 1. This operates on the output from the search programs, making the necessary calculations to set absolute figures against the amount of available data, thus providing the percentages and county rankings.

<sup>8</sup> SIMPLEPLOT – a graphics package developed by Bradford University Software

Services.

9 Future applications of this technique will make use of a Graphics facility to shade the counties in slowly decreasing intensity of stippling according to their order and percentage figures in the tables. This will eliminate the need to classify counties into what are somewhat arbitrary groupings, and will give a better visual

representation of geographical and/or topographical patterns.

Oxford Concordance Program – 'a machine independent text analysis program to produce word lists, indexes and concordances'. See S. Hockey and J. Martin, 'The Oxford Concordance Program Version 2', Literary and Linguistic Computing, II (1987), pp. 125-31. The geographically ordered output generated by

SEARCHTEXT can also be readily produced by this program.

11 OCP has a facility which enables specific sequences of characters to be declared as equal for the purposes of ordering output. However, it will not include within this definition strings containing a space. In order to get round this problem, forms such as and lang, & long etc. need to be temporarily joined prior to running the OCP program. The output from which Figure 6 was extracted was produced by applying the following two ECCE commands to the text-file in order to produce a temporary file:(f/ lan/s/lan/)\* and (f/ lon/s/lon/)\*, where [f] means 'find', [s] 'substitute', [/] is the string-delimiter and [\*] means 'repeat throughout the file'. The dummy character [1] (occurring nowhere else in the file), joins all the words starting with the strings [lan] and [lon] to the previous word. ['] was then declared as 'padding' in the OCP command file, which means that it was effectively ignored. The search criteria were defined as 'headword = @n\*1@n\* &1@n\*', where [@] represents one character and [\*] represents any or no

characters. This causes the program to treat the strings [@n\*1@n\*] and [&1@n\*] as equal for sorting, but not for printing. The original space could subsequently be returned to the output file by the ECCE command (f/'/s//)\*, and the temporary file be deleted. Other forms picked up by these criteria, but not illustrated in the extract, include 'endelang', '&langes', 'and lancg', and 'endelangweies'. The full printout contains 3192 citations, approximately 65 of which did not represent phrases containing and and had to be removed. When the material has been glossed such long-winded routines can be dispensed

with, as OCP will be able to operate directly on forms tagged by the headword. <sup>12</sup> abrocenan beorg and afoc broc.

13 The string-lists also serve to check whether any forms will have slippped the net of searches such as that outlined in note 11, above. In this latter case the list showed that there are no words beginning with the letters 'læn' in the Corpus, so that, although 'ændlæng' forms occur, there was no necessity to include 'læn' in the joining edit carried out by ECCE.

14 W. de G. Birch, Cartularium Saxonicum, 3 vols. & index (London, 1885-99), and J. M. Kemble, Codex Diplomaticus Ævi Saxonici, 6 vols. (London, 1839-48).

#### SOCIETAS ET FRATERNITAS

A Report on a Research-Project based at the Universities of Freiburg and Münster

#### JAN GERCHOW

Over the past quarter-century, German research into the so-called 'commemorative documents' surviving from the early Middle Ages has been concentrated in a research-team, directed by Karl Schmid and by Joachim Wollasch at the universities, respectively, of Freiburg and of Münster. The works of both scholars and their teams are of great moment to anthroponymists, especially those concerned with the 'one-name period' extending to c.1100, because commemorative documents such as confraternity books and obituaries (or necrologies) primarily contain personal names, often lacking even such qualifying additions as titles, or indications of relationship to other persons mentioned, or of origin, or of date. The sheer bulk of the material and the often daunting chaos of its arrangement in the unique manuscripts have hindered generations of historians and anthroponymists from fruitful exploitation of it. To give some figures: the largest known confraternity book, that from Reichenau Abbey on Lake Constance, begun by c.824 with material partly dating back to 762, contains nearly 40,000 personal-name items; many of its 164 folios are filled with entries made, over three or even more centuries, by up to 300 different scribes. Another example: the obituaries from the Cluniacensis ecclesia survive only in nine manuscripts mostly dating from the late tenth to the thirteenth centuries (in one case, the fifteenth) and containing in all about 96,000 name-items, mainly referring to Cluniac monks of the eleventh and twelfth centuries; until 1982, only excerpts from these manuscripts had been published, and their inter-relationships were unknown.

The way that continental historians and anthroponymists have learnt to handle these huge name-corpora concerns their English colleagues also, not least because Anglo-Saxon names form part of the Germanic tradition. The new appraoch is, besides, of methodological import to scholars in England, because England too has an important tradition of commemorative documents and other material rich in early medieval personal names.

# I. The Prosopographical Background

Karl Schmid and Joachim Wollasch both became interested in commemorative records while members of the 'Freiburger Arbeitskreis' directed by the historian