1. Introduction

While editing the Helsinki Corpus of British English Dialects, I came across combinations such as eyeman, meaning a professional who inspected the informant’s eyes to provide him with spectacles. This very handy compound is much more transparent than the loanword ophthalmologist, which the dialect speaker probably did not even know. He only followed the familiar word-formation practice of adding the element MAN to a word indicating the thing the person referred to is (professionally) occupied with (cf. pigman, ditchman, toadsman, etc.). When studying the linguistic features of the above corpus and comparing them to word lists and examples of other spoken varieties of English, for example Pidgins, I noticed frequent use of the same types of compound.

The inevitable question that arose was whether MAN-compounds are a particularly common type of word-formation in dialectal English and spoken language in general. They could also be supposed to have been common in earlier English when compounding and suffixation still relied on native linguistic elements. The aim of this study is thus to chart the uses and frequency of MAN-compounds, both diachronically and synchronically, to see whether the above suppositions can be confirmed.

The study was conducted by browsing through thirteen easily accessible computer corpora, using the Word Cruncher program that facilitates search by the end string *MAN. In addition, two dictionaries were searched through and several others consulted. The study was restricted to MAN-compounds with human referents, excluding the ethnic ones, such as Englishman, etc., which have been used since the Old English period (Marchand 1969: 64) and were found to occur in all types of text. Combinations with non-human referents, such as merchant-man, which was very frequent in some corpora, or dodman ‘snail’ and other compounds referring to animals, were also excluded.

2. Formal and Semantic Types of MAN-Compounds

The following division of MAN-compounds into formal and semantic types is based chiefly on Adams, Jespersen, Marchand, and Quirk et al.

MAN-compounds can be divided into four formal types in which MAN is preceded by a noun (chairman), an adjective (madman), a verb (hangman) or a noun + s (craftsman). These formal types represent various semantic types, as given below. The capital letters N and V refer to noun and verb respectively.

TYPE 1: N1 + N2 (= MAN), where N2 is occupied with the referent of N1 in different ways:
(a) N2 controls or works N1 (chairman, fireman, gasman, postman, businessman);
   N2 sells, delivers or manufactures N1 (milkman, newspaper man);
   N2 belongs to a specified professional group (clergyman, policeman, seaman)
(b) N2 is (like) N1 (frogman, spiderman). The implication is originally figurative, though in modern popular fiction the referent of the compound may also have a more obvious resemblance to N1 in appearance and function.

---

1 The research reported here was supported by the Academy of Finland Centre of Excellence funding for the Research Unit for Variation and Change in English at the Department of English, University of Helsinki.

(c) N1 indicates a place which is the ‘natural habitat’ of N2 (bushman, caveman, countryman).
(d) N1 + N2 = N1. In this ‘appositional type’ N2 is redundant, because N1 alone indicates an agent (washerman, engineer). According to Marchand (1969: 62), this type is avoided in educated speech as tautological.

TYPE 2: Adj. + N. In this ‘copulative combination’ the relationship of the elements is that of a complement and a subject (handyman, madman, viz., the man is ‘handy/mad’). The reference is not to a regular occupation but to a characteristic feature. This type includes combinations referring to social conditions (gentleman) or colour (blackman), as well as the ethnic combinations (Englishman, etc.), excluded from this study. TYPE 2 has been attested since Old English but is not considered productive by Marchand (1969: 64).

TYPE 3: V + N. In present-day English, the verbal element is often identical in form with the corresponding noun, so that classifying these combinations as TYPE 3 instead of TYPE 1 is rather a matter of interpretation. In this combination MAN functions as an agent who acts in the way expressed by the verbal element. The action may be transitive or intransitive. Thus a workman works and a hangman hangs (somebody).

TYPE 4: N1 + s + N2. The first element may indicate a particular skill or a solidarity circle (craftsman, oarsman, backwoodsman, clansman). We can ignore the origin of s, whether genitival or plural. (Various explanations for different cases are given, e.g., by Adams 1973: 70–71 and Marchand 1969: 65–66.)

3. Frequency of MAN-Compounds as Evidenced in Computer Corpora

The computer corpora used in this study will be referred to by their common abbreviated names. Further information on each is given in the References section under Corpora. The main material consisted of twelve corpora, four of which were historical (HC, divided into the chronological sub-corpora HCO, HCM and HCE; CEECS1; LAMP; and NEWD). Eight were 20th-century corpora of written (FLOB, FROWN, KOLH, ACE, WSC) and spoken (SEC, COLT, HD) English, representing different varieties of English (for details, see References). In addition, I consulted the original BROWN corpus, the Collins Cobuild Dictionary, which is based on the Bank of English corpus, and various dialect dictionaries, glossaries and vocabularies.

My method was to find all compounds in -MAN in the above corpora and compare their frequencies in different periods and varieties by counting their relative proportions of the given numbers of unique words in the corpora. Simple as it sounds, unexpected difficulties did occur. One was idiosyncratic English spelling. Logically, a compound word should be spelled as one word (cf. blackbird vs. a black bird), but this is not necessarily so in English. The spelling instructions in modern dictionaries often give us three possibilities (‘can also be spelled as two words or with a hyphen’). For the sake of consistency I decided to include only combinations in -MAN which were spelt as one word (including those combined with a hyphen), reasoning that, if a combination was established well enough, it was probably spelt as one word. I am fully aware that in this way some of the occurrences may have been lost, particularly in corpora of Early English, where the spellings may vary. I am also aware that the spellings in the corpora may also be editorial and thus do not give a truthful picture of the language.

Another difficulty was that the given numbers of unique words in individual corpora are affected by the fact that the same words occur in different grammatical forms, which are more numerous the earlier the text. In addition, the spellings of the same words may vary considerably, particularly in texts predating the standardization of spelling in the later Modern English period. Thus, for example, the word alderman occurs in 31 different spellings and forms in the HCO (198 tokens) and gentleman in twelve different spellings and forms in CEECS1 (129 tokens). These different forms and spellings...
count as different ‘unique words’, which renders a reliable quantitative study of relative frequencies in
different periods difficult. This to a lesser extent applies to modern corpora with fewer grammatical
forms and standard spelling, but considerably distorts the figures of early historical corpora. Rectifying
the numbers of unique words given would involve an immense amount of manual work so far, to my
knowledge, not undertaken by anybody.

In spite of all these reservations, I will nevertheless proceed to give some relative frequencies,
since there seems to be some evidence to support my thesis. Further research is, however, needed to
clarify the details. In Table 1 the frequencies of MAN-compounds with human reference in the corpora
mentioned above have been normalized to one thousand unique words.

Table 1. MAN-compounds: frequencies of types/1,000 unique words.

<table>
<thead>
<tr>
<th>Historical corpora:</th>
<th>HCO</th>
<th>HCM</th>
<th>CEECS1</th>
<th>HCE</th>
<th>LAMP</th>
<th>NEWD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.75</td>
<td>0.92</td>
<td>1.12</td>
<td>1.42</td>
<td>1.78</td>
<td>1.35</td>
</tr>
<tr>
<td>20th-century corpora:</td>
<td>SEC</td>
<td>FLOB</td>
<td>COLT</td>
<td>FROWN</td>
<td>KOLH</td>
<td>ACE</td>
</tr>
<tr>
<td></td>
<td>1.92</td>
<td>1.48</td>
<td>1.28</td>
<td>1.62</td>
<td>1.34</td>
<td>1.63</td>
</tr>
<tr>
<td>Dialect corpus (HD):</td>
<td>SOM</td>
<td>DEV</td>
<td>CAM</td>
<td>ELY</td>
<td>SUF</td>
<td>ESS</td>
</tr>
<tr>
<td></td>
<td>2.26</td>
<td>3.16</td>
<td>2.64</td>
<td>2.98</td>
<td>3.86</td>
<td>0.77</td>
</tr>
</tbody>
</table>

I further counted the average frequency of MAN-compounds for the Collins Cobuild English Language
Dictionary, basing it on the 75,000 references the dictionary is said to have, which gave 1.44 types per
one thousand words.

For the reasons given above, the frequencies of the pre-1500 corpora (HCO, HCM, CEECS1) could be expected to be somewhat higher than shown in Table 1. The average proportion of unique
words in the modern corpora is about 4.5 per cent of the total word count. If the numbers of unique
words in the three pre-1500 corpora are corrected to the same percentage, the figures of relative
frequency for MAN-compounds will come closer to those of the dialect corpora: 2.15 (HCO), 2.04
(HCM) and 2.26 (CEECS1). It is interesting to notice that the general frequency of MAN-compounds
from Early Modern English onwards is between c. 1.40 and 1.70, but the dialect corpora (HD) tend to
give a figure higher than 2.00. The marked frequency in the educated spoken corpus SEC (1.92) should
also be noted.

The HCE is divided into different text types and the greatest frequencies of MAN-compounds are
seen in statutes, diaries, travelogues, trials and fiction, then in private correspondence and biographies.
More official text types, such as philosophical, educational and scientific texts and official
 correspondence show the least occurrences of MAN-compounds.

4. From Gender-Neutral MAN to a Sexist Element

In Early English MAN was gender neutral. For indicating a particular gender it had to be marked
with a gender noun (the examples below were taken from the HCE and are translated by me):

OE: þær wearþ Scíþia xx M ofslagen & gefangen wifmonna & wepnedmonna. ‘there
were 20,000 Scythians killed and captured, both women and men’
(Orosius 116)

Ure Hælend Crist cydde, þæt he lufode þa clænnysse on his þeowum swuteliche, Þa
þa he mædenmann him to meder geceas. ‘Our Saviour Christ said that he loved
chastity in his servants so that he chose a virgin to be his mother’
(Aelfric Let Sigef. 14)

EME: þa namen hi ða men þe hi wenden ðat ani god hefden, bathe be nihtes & be deies,
carlmen & wimmen. ‘There they took the people who they thought might have any
property, both by night and by day, both men and women.’
(Peterborough Chronicle 1137: 55)
MAN was also used in such indefinite expressions as no man, any man and every man, in which it was gradually replaced by one or body in Middle and Early Modern English when the meaning of MAN as referring to a ‘human being of male sex’ gained more currency than the parallel neutral meaning ‘human being’ (see Raumolin-Brunberg and Kahlas-Tarkka 1997: 72–73).

As the second element in compound nouns, -MAN seems to have been well on the way to developing into a gender-neutral suffix, with its vowel weakening into a schwa. This is said to happen particularly in combinations of type 1a (fireman, etc.), which indicate a member of a specific professional group and which are said to be very productive (Quirk et al. 1986: 1574). This applies particularly to combinations of early origin which have become more established. As a suffix, -MAN would have rivalled other agent suffixes of native and foreign origin (-er, -or, -ist, etc.), cf. in combinations with a native N1 (fishman vs. fisher, etc.).

The development was halted when the feminist movement started to pay extra attention to the masculinity of the element -MAN around the middle of the 20th century. This movement has been occupied particularly with trying to establish female counterparts and gender-neutral equivalents to MAN-compounds. Some of these, both earnest and jocular, are listed in Table 2, as given by Beard and Cerf for American English.

Table 2. Non-sexist substitutes for MAN-compounds in American English (Beard & Cerf 1993). The asterisks indicate the frequencies of the words as given in the Collins Cobuild Dictionary.

chairman****/chairwoman > chairperson > chair
doorman > doorman, access controller
fireman* > firefighter > fireperson
fisherman** > fisher
freshmen* > first year students, freshpersons, freshpeople > frosh
gingerbread man > gingerbread person, gingerbread creation
henchman > henchperson, person of hench
human***** > hufem, humyn
milkman > milkperson, stolen non-human-animal product deliverer
snowman > snowhuman, snowhufem, snow person, person of snow, snow icon
woman***** > woperson, person of gender, wofem, wombant
womon—wimmin, womyn—wimyn

According to the Collins Cobuild Dictionary frequency symbols, none of the above novelties have become established in wider use than the originals. In fact, only chairperson, chair and firefighter are recognized by the dictionary. In some cases, there is a female counterpart with another element (e.g., milkmaid) or the female counterpart has a particular connotation not shown by the male compound (e.g., fishwife).

The use of ‘sexist’ names for professionals was not a problem in earlier centuries when men and women had mostly different occupations. There were also particular gender suffixes that could be used (cf. Early English webba ‘a male weaver’ and webbestere ‘a female weaver’). The types of WOMAN-counterparts in the corpora being studied are given in Table 3 (counterparts with -WIFE proved to be too infrequent for consideration).
Table 3. WOMAN-combinations as counterparts to MAN-compounds (spellings modernized). The figures give the absolute numbers of WOMAN/MAN-compounds and the percentages of WOMAN-compounds.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>WOMAN/MAN</th>
<th>WOMAN-compounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCO</td>
<td>-/49</td>
<td>only 2 without male counterparts: heofonwoman, hildewoman</td>
</tr>
<tr>
<td>HCM</td>
<td>1/58 (2%)</td>
<td>gentlewoman (+ housewife)</td>
</tr>
<tr>
<td>HCE</td>
<td>2/57 (3.7%)</td>
<td>gentlewoman, kinswoman</td>
</tr>
<tr>
<td>CEECS1</td>
<td>3/26 (11%)</td>
<td>bidswoman, gentlewoman, kinswoman</td>
</tr>
<tr>
<td>LAMPETER</td>
<td>3/75 (4%)</td>
<td>gentlewoman, kinswoman, noblewoman</td>
</tr>
<tr>
<td>NEWDIGATE</td>
<td>2/50 (4%)</td>
<td>gentlewoman, kinswoman</td>
</tr>
<tr>
<td>SEC</td>
<td>-/15</td>
<td></td>
</tr>
<tr>
<td>FLOB</td>
<td>5/78 (6%)</td>
<td>chairwoman, countrywoman, gentlewoman, kinswoman. spokeswoman + 3 without male counterparts: horsewoman, needlewoman, spiderwoman</td>
</tr>
<tr>
<td>COLT</td>
<td>-/16</td>
<td></td>
</tr>
<tr>
<td>FROWN</td>
<td>7/86 (8%)</td>
<td>businesswoman, congresswoman, lawwoman, ombudswoman, servicewoman, spokeswoman, townswoman + 1 without a male counterpart: Catwoman</td>
</tr>
<tr>
<td>KOLH</td>
<td>3/64 (4%)</td>
<td>fisherwoman, saleswoman, washerwoman</td>
</tr>
<tr>
<td>ACE</td>
<td>6/83 (7%)</td>
<td>churchwoman, gentlewoman, policewoman, spokeswoman, sportswoman, superwoman + 3 without male counterparts: girl-woman, handywoman, washerwoman</td>
</tr>
<tr>
<td>WSC</td>
<td>2/42 (5%)</td>
<td>spokeswoman, sportswoman</td>
</tr>
<tr>
<td>HD</td>
<td>2/67 (3%)</td>
<td>charwoman (SOM), postwoman (CAM, ELY)</td>
</tr>
<tr>
<td>Collins Cob.</td>
<td>18/108 (13%)</td>
<td>airwoman, anchorwoman, businesswoman, chairwoman, congresswoman, councilwoman, countrywoman, craftswoman, gentlewoman, horsewoman, kinswoman, madwoman, noblewoman, policewoman*, saleswoman, spokeswoman, sportswoman, yachtswoman</td>
</tr>
</tbody>
</table>

The percentages of WOMAN-counterparts to MAN-compounds are very small in the early corpora, the combinations in WOMAN mainly referring to social class (gentlewoman, kinswoman, noblewoman), with the exception of some particular terms, such as bidswoman in early letters or poetic Old English heofonwoman and hildewoman. It is only in the 20th-century corpora that other WOMAN-combinations begin to appear. They often relate to legal or governmental professions of a certain social status or skill. It is to be noted that only one of the 19 WOMAN-compounds in the Collins Cobuild Dictionary is frequent enough to be provided with a frequency symbol (policewoman), while as many as 39/108 (or 36 per cent) of the MAN-compounds are given a frequency symbol (from one to four asterisks).

A comparison between the original BROWN corpus from the 1960s and the later FROWN version (1991) reveals another interesting detail. The relative frequency of MAN-compounds has been reduced from 2.08 in BROWN to 1.62 in FROWN, i.e., by a fifth, at the same time as the total number of unique words has increased by nearly 4000 or c. 8 per cent. What has happened is that 49 MAN-compounds included in BROWN do not occur in FROWN, while the latter contains 31 new compounds that did not occur in the earlier version (of the 102 types in BROWN and 83 types in FROWN, 52 are shared by both corpora). Naturally, certain words are bound to their time and give way to new expressions along with the changes in society that inevitably occur in a period of thirty years (e.g., highwayman or radioman disappearing, anchorman, Quayleman, etc. appearing). It is, however, not impossible that the feminist aversion to what is seen as the masculine element -MAN may have contributed to the reduction of MAN-compounds in the latter half of the 20th century.
5. Favourite MAN-Compounds

Judging by the numbers of tokens, the favourite types of MAN-compounds show remarkable tenacity. Those with the greatest frequency of occurrence are listed in Table 4:

Table 4. Common types of MAN-compounds in the corpora studied.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Types/tokens</th>
<th>Most frequent type and other common types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Historical corpora</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HCO 49/338</td>
<td>alderman 198 (58% or tokens), headman 15, worldman 10, lemmnan 53, alderman 44, gentleman 20, yeoman 20</td>
<td></td>
</tr>
<tr>
<td>HCM 58/343</td>
<td>gentlelman 139 (48% of tokens), kinsman 24, footman 23, horseman 22, bidsman 21</td>
<td></td>
</tr>
<tr>
<td>CEECS1 26/290</td>
<td>gentlelman 226 (44% of tokens), alderman 38, nobleman 22, horseman 19, maltman 19</td>
<td></td>
</tr>
<tr>
<td>HCE 57/511</td>
<td>gentlelman 430 (39% of tokens), seaman 111, freeman 75, tradesman 62, clergyman 37, juryman 35, countryman 33</td>
<td></td>
</tr>
<tr>
<td>LAMP 75/1100</td>
<td>alderman 511 (43% of tokens), gentlelman 235 (20% of tokens), seaman 101, foreman 33, nobleman 31, waterman 31</td>
<td></td>
</tr>
<tr>
<td>NEWD 50/1180</td>
<td>gentlelman 57 (21% of tokens), gentlelman 57, spokesman 50, policeman 39, fisherman 30, businessman 23</td>
<td></td>
</tr>
<tr>
<td>20th-century corpora</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEC 15/43</td>
<td>fireman 14</td>
<td></td>
</tr>
<tr>
<td>FLOB 78/538</td>
<td>chairman 112 (21% of tokens), gentlelman 57, spokesman 50, policeman 39, fisherman 30, businessman 23</td>
<td></td>
</tr>
<tr>
<td>COLT 16/82</td>
<td>snowman 10</td>
<td></td>
</tr>
<tr>
<td>FROWN 86/427</td>
<td>chairman 69, businessman 38, gentlelman 34, spokesman 24, congressman 21</td>
<td></td>
</tr>
<tr>
<td>KOLH 64/567</td>
<td>chairman 115 (20% of tokens), gentlelman 54, policeman 44, craftsman 32, spokesman 29, businessman 25, congressman 21, headman 21</td>
<td></td>
</tr>
<tr>
<td>ACE 83/430</td>
<td>chairman 52, spokesman 50, policeman 33, businessman 32, gentlelman 30</td>
<td></td>
</tr>
<tr>
<td>WSC 42/166</td>
<td>policeman 30, chairman 20, fisherman 16</td>
<td></td>
</tr>
<tr>
<td>Dialect corpus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HD 67/323</td>
<td>horseman 111 (34% of tokens), policeman 29, postman 18</td>
<td></td>
</tr>
</tbody>
</table>

The Collins Cobuild Dictionary shows 108 types of MAN-compounds and conveniently provides them with asterisks indicating frequency of use. The following were provided with more than one *: chairman ****, spokesman ****, businessman ***, gentlelman ***, batsman **, congressman **, fisherman **, gunman **, salesmen **, statesman**.

The distribution and frequency of MAN-compounds naturally reflect the text types and topics in the respective corpora (cf. the frequency of formulaic bidsman in the early modern letters of the CEECS and the predominance of horseman in the dialectal corpus collected from rural people). The combinations in the early corpora reflect class distinctions (yeoman, nobleman, freeman, gentleman), while those in the Modern English period tend to refer more to various professions (tradesman, seaman, fisherman, etc.). A particular group of MAN-compounds represents governmental officials. Here alderman, which features as an important element in the early corpora is replaced by policeman, chairman and spokesman in the twentieth-century corpora. Gentleman retains its position remarkably well, but its meaning changes. Having referred to social class in the pre-20th-century corpora, it now appears only as a positive term of description. The choice of types in the small spoken corpora (SEC and COLT) is rather limited. What is remarkable in the speech of London teenagers (COLT) is the fact that many of the MAN-compounds refer to imaginary characters (Batman, Candyman, Spiderman, Superman) that are common in popular culture today.

Type 2, Adj. + MAN, used to be more common in the early periods, as the percentages in the HCM (24%) and CEECS1 (19%) indicate. In the other corpora this percentage is around 10 per cent. In the dialect corpus HD this type is uncommon (only blackman, freeman and gentleman). Type 4, with an -s
(craftsman), is less common in the early corpora (HCO 8%; HCM 9%), but increases in frequency in the Early Modern English period (HCE 17%) and is shown in about a fifth of the types in the Modern English period, particularly in Southern Hemisphere varieties (ACE 22%; WSC 31%). The appositional type (1d) is uncommon, even in dialects (HD shows only fisherman, side by side with fishman, and slaughterman).

6. MAN-Compounds in Dialectal Speech

The MAN-compounds in the HD are mainly of types 1a and 1b, in which MAN works or controls N1. These seem to be formed freely with any kind of material or object as a handy way of escaping the ‘regular’ expressions of profession. Thus we have bargeman, boat(s)man, coachman, butterman, maltman, milkman, cowman, horseman, pigman, stockman, toadman, clayman, coalman, gasman, ditchman, dockman, harvestman, roadman, woodman, yardman, taxman, eyeman, and so on. A quick browse through dialect dictionaries and vocabularies reveals an abundance of additional combinations. MAN-compounds are also used with reference to animals (see Appendix). The productivity of the MAN-element is particularly prominent in pidgins, where it has acquired a suffix-like status and seems largely to replace other ways of forming agent nouns. For further examples of dialectal MAN-compounds, see Appendix.

7. Conclusions

The occurrence of a linguistic element often depends on the types of text and topics discussed. Therefore, quantitative results of linguistic phenomena should at best be taken as possible tendencies of use rather than a basis for definite conclusions.

With certain reservations, it can be stated that the findings from the corpora corroborate the continuation of MAN-compounds throughout the history of English. They seem to have been more common in earlier and dialectal English (relative frequency more than 2.00/1,000 unique words as against 1.40–1.70/1,000 in other corpora) and have been preferred in text types that are considered to be closer to the spoken tradition, while the least occurrences are found in philosophical, educational and scientific texts. In more official text types MAN-compounds are represented mainly by expressions referring to social status or governmental and legal professions. Their homely character is proved by their productive use in dialects. Their transparency renders them particularly suitable to new and nonce-formations. A further study of MAN-compounds in new varieties of English, for example pidgins, might contribute to the results reported in this paper.

The semantic types 1a and 1b, consisting of MAN as N2 dealing with N1 in various ways, are particularly favoured in dialectal speech. Type 2, Adj. + MAN, occurs mainly in a number of combinations referring to social status. Type 4 with an inserted s seems to have grown in popularity after the earliest periods and to be favoured somewhat in some Southern Hemisphere Englishes.

The development of MAN into a completely gender-neutral suffix was seemingly interrupted in the middle of the 20th century, though only a few counterparts with WOMAN or alternative neutral expressions have become really frequent in general use.

Appendix

Some examples of MAN-compounds in non-standard varieties of English, as found by browsing various dialect dictionaries and vocabularies (see the list of references).

Suffolk (Claxton): DODMAN/HODMAN ‘snail’ (also HODMEDOD) (in Essex www. DODMAN also ‘a slow moving horse or man’)
PIorman/PIPMAN ‘the smallest pig in the litter’, ‘anything diminutive’
PETERMAN ‘Dutch fishing vessel’
PETER-MAN ‘a familiar term for a fisherman on the Thames’
East Anglia (Forby): CUSHION-MAN ‘the chairman at the Quarter Sessions; or at any other public meeting, where there is the same distinguishing mark of presidency’
FRENCHMAN ‘any man, of any country, who cannot speak English’
SHIREMAN/SHERE-MAN ‘any one who doesn’t understand East Anglian’; ‘not born in one of the sister counties or in Essex’, ‘a sort of foreigner to us’
STATESMAN ‘the proprietor of an estate’
TRENCHERMAN ‘a hearty feeder’ (TRENCHER = ‘a wooden platter’)

Various BE dialects according to the Survey of English Dialects (Upton, Parry, Widdowson):
ALL-ROUND-MAN/CASUAL MAN ‘a farm labourer’
BLACKMAN ‘bogey’, ‘devil’
BLIND-PLOUGHMAN ‘a worm’
BOGEYMAN ‘scarecrow’
BOMAN ‘bogey’
BULLOCK-MAN, BYREMAN, CATTLEMAN, etc. tend the animals concerned
CADMAN/CAGMAN/CAGMEG-MAN/CAT’S-MEAT-MAN ‘knacker’
BULLY-MAN ‘forker on a wagon who unloads sheaves in a stackyard’
etc.

Geordie (www):
BANKSMAN ‘the man in control at the top of the shaft of a pit’
CANDYMAN ‘a bum bailiff, the man who serves notice of ejectment’
CHECK-WEIGHMAN ‘a representative of the collier who checks the weight of coal at the surface on behalf of his men’
FOYBOATMAN ‘a boatman who watches for boats coming into the Tyne in the hope of getting employment in mooring them’ (FOY = ‘fee/reward’)
HOASTMEN ‘a coalshipper at Newcastle’ obsolete word

Pidgin (Barhorst & O’Dell-Barhorst):
BAMAN ‘bartender’
BIKPELAMAN ‘adult’
BULSITMAN ‘conman’
DAIMAN ‘dead person’
DIDIMAN ‘agricultural officer’
DRIPMAN/WELMAN ‘squatter’, ‘wanderer’
GIAMANMAN ‘liar’
HAMBAKMAN ‘annoying/cheeky person’, ‘nuisance’
KALABUSMAN ‘prisoner’
KUSAIMAN ‘deceiver’
LAIPMAN ‘living person’
LONGLONGMAN ‘a mad person’
MARITMAN ‘husband’
OLGETAMAN ‘everybody’
OL WAITMAN ‘European’
PANIMAN ‘funnyman’, ‘clown’
POROMAN ‘acquaintance’
PRETMAN ‘coward’
RABISMAN ‘poor man’
SIKMAN ‘patient’
SPAKMAN ‘drunkard’
STILMAN ‘thief’
RABELMAN ‘troublemaker’
WOKMAN ‘worker’
References

Corpora (in chronological order, with dates and approximate word counts)

HC = The Helsinki Corpus of English Texts: HCO (–1100; 505,000 words); HCM (1100–1500; 630,000 words);
HCE (1500–1710; 570,000 words). A diachronic corpus of written British English (mainly prose).

CEEPS = The Corpus of Early English Correspondence Sampler, part one (1418–1638; 255,000 words), early
English letters.

LAMP(ETER) (1640–1740; 1.1M words), written (printed) Early Modern tracts.

NEWD(IGATE) (1674–1692; 1M words), written (unprinted) Newdigate Newsletters.

BROWN (1961; 1M words), written (printed) American English.

HD = The Helsinki Corpus of British English Dialects (collected in the 1970s; c. 900,000 words; unpublished,
transcription and edition in progress), spoken rural British English. The informants were chosen mainly
following the principles used in the Leeds Survey of English Dialects.

Sub-corpora:
- the Ihalainen Somerset Corpus (c. 169,000 words, 23 informants, 15 localities);
- the Ojanen-Vasko Cambridgeshire Corpus (c. 184,000 words, 30 inf., 26 loc.);
- the Tammivaara-Balaam Isle of Ely Corpus (c. 87,000 words, 30 inf., 17 loc.);
- the Pasanen Suffolk Corpus (c. 240,000 words, 38 inf., 14 loc.);
- the Kerman Corpora, collected in the 1980s in Essex (c. 59,000 words, 5 inf., 3 loc.) and Lancashire (c.
62,000 words, 5 inf., 3 loc.).

For more information on the HD, see

KOLH(AIREF) (1978; 1M words), written (printed) Indian English.

ACE = The Australian Corpus of English (1986; 1M words), written (printed) Australian English.

SEC = The Lancaster/IBM SEC Corpus, the Machine-Readable Corpus of Spoken English (1984–1985; 52,600
words), spoken British English.

WSC = The Wellington Corpus of Spoken New Zealand English (1988–1994; 1M words), spoken New Zealand
English.

FLOB (Freiburg-LOB) (1991; 1M words), written British English.


COLT = The Bergen Corpus of London Teenage Language (1993; 450,000 words), spontaneous spoken London
teenager English.

Other Sources

& New York: Longman.

Villard Books.

Grammar of Spoken and Written English. London: Longman.


HarperCollins.


C. H. Beck’sche Verlagshandlung.


2 Based on the Bank of English Corpus of over 200M words. Frequencies of the headwords marked with 1 to 5
‘diamonds’ (replaced by asterisks in this paper). No symbol means less common but still worth including. Priority
given to the English of most general utility worldwide. Dialect words or language of small social groups or
specialists not featured (pp. viii–ix).


Wright, Joseph (1923 repr.) *The English Dialect Dictionary*. London, etc.: OUP.


**Web Pages Visited**

<http://www.anvil.clara.net/essexdialectguide.htm>.

<http://www.geordiepride.demon.co.uk/dictionary.htm>.
