

REPORT OF THE PILOT SURVEY FOR THE PROPOSED 'DICTIONARY OF THE
ENGLISH-SPEAKING PEOPLES'

DAVID CRYSTAL

1. After some months of discussion between the editorial board and the publishers over the financial and administrative arrangements for the above-named dictionary, it was agreed in the autumn of 1966 to carry out a pilot survey to determine the feasibility of the scheme. Accordingly, I wrote a Prospectus, which appeared in December 1966, in which the nature of the project was outlined and the specific aims of the pilot survey stated. This was supplemented by two papers, read to the Linguistics Association of Great Britain and the Colloquium of British Academic Phoneticians in Easter of this year, in which valuable comment and support was obtained. Before presenting the results of the now completed survey, it would seem useful to recapitulate what the working principles of the project were envisaged as being by summarising parts of the Prospectus.

2. Essentially, the survey would determine how far a dictionary based on synchronic linguistic principles, and incorporating regional information about English lexis on a world-wide scale, could be satisfactorily written from both a scholarly and a commercial point of view, within a specified time-scale (approximately seven years) and size (approximately 180,000 entries).

2.1. Audience. The specification of the audience was seen to be the primary commercial consideration, and the following groups of users, whose needs would seem to be uncatered for by currently available dictionaries, were outlined in the Prospectus. Three kinds of audience were primary.

a) Those seriously interested in understanding or actively commanding the current vocabulary of other English-speaking areas than the one(s) they naturally belong to. This would cover a number of spheres of interest: the literary (one notes the increase in university courses on American literature in this country, the tendency for new universities to foster broader courses in comparative literature, the develop-

ment of new regional literature in English from all parts of the English-speaking world, and so on); the 'occasional' (referring to the growing interest in magazines, newspapers, radio and television programmes from other areas than one's own, the increase in tourism, etc.); and the 'professional' - with the development of improved techniques of international communication, people from different areas are being brought into more frequent contact, and this has implications at the business, diplomatic, religious and other levels. In particular, one could point to the acceptance of English as a working language for communication at professional and technical levels in most Commonwealth countries. It is the case, however, that the existence of regional lexical variations between these countries can and does cause misunderstandings and other communicational difficulties, which are all the more serious because frequently unanticipated and unidentified. The proposed dictionary would thus provide a handbook which could be used in the education of Commonwealth and other citizens for the purpose of fostering a greater awareness of linguistic differences and similarities, and developing a more profound and efficient mutual understanding.

For all such people, a source-book of information about regional lexical variation would seem to be invaluable. Moreover, the dictionary would be usable by people from the different dialects on the same terms: users from different areas would be equally 'at home' with it, as information about any user's own dialect is given, as well as information about the dialects of others.

b) There is also an important audience comprising people seriously interested in studying the nature of language - and especially the English language - as an end in itself (as opposed to as a means of improving specific points of usage), and needing reliable, linguistically-orientated information about lexical structure and usage in English. This is nowadays an important group, comprising large numbers of university and training college students following courses in linguistics, English language and related fields; American college freshmen; sixth formers (one notes the demand for a more general awareness of English language usage in exams which have been introduced at this level); and others. There is no commercial dictionary written on wholly linguistic principles, incorporating a systematic approach to such features as pronunciation variation, grammatically relevant information, the semantic relationships

between words, stylistic variations, and so on: the demand for one such (bearing in mind the rate of increase in students reading language courses in Great Britain alone) will be severe in 6-8 years time.

(c) While the proposed dictionary will cater for relatively serious students, it should not be thought that it is thereby going to be over the head of the 'family' user. On the contrary, it is hoped that the information normally provided by a 'concise' family dictionary would also be given here - but in a more meaningful and assimilatable form. To depart from traditional lexicographical conventions does not make things more difficult for the everyday user, though this argument is sometimes advanced. For one thing, there is no single set of lexicographical conventions - they vary widely. Moreover, the everyday user is rarely aware of what conventions there are, and tends not to take the trouble to learn to use them (the printing of a pronunciation guide at the foot of every page in a dictionary, as sometimes happens, is in effect an admission of failure in this respect). Also, much of the traditional methodology is unsatisfactory, often providing redundant, vague and inadequate information.

To take one example, precise information about semantic structure would be invaluable for the 'average' user of a dictionary, but it is not available. There is an assumption that the meaning of a word can be determined by explaining that word in isolation, without reference to other words of different degrees of similarity or difference in meaning. But this is a naive and falsifying view. Words exist in clusters, or sets, some of which are tighter-knit than others, but all of which are inter-related in various ways - the way the terms for colours mutually define each other is a standard example. It should be one of the functions of a dictionary to expound at least the most important of these inter-relationships systematically, but to do this properly, a system of cross-referencing has to be introduced. The introduction of cross-reference into the proposed dictionary, then, is designed to make the user's job easier, by providing him with all the information he needs about a particular lexical item at a single place within the dictionary (cf. 3.2.6, 4.2.2.6, 5.7).

3. In short, the Dictionary of English-speaking Peoples (DESP) is intended to be an adult dictionary for adult users, providing comprehensive information not available in convenient form elsewhere about the current state of the English lexicon, insofar as this is known, and avoiding unnecessary simplification. This information is of two kinds, regional and linguistic.

3.1. Regional. Here the aim is to collect a specific part (to be defined below) of the distinctive lexicon of those parts of the world where English is used as a first or second language - that lexicon, in other words, which is restricted in use to a given area, and not shared by other parts of the English-speaking world. As this by no means amounts to the total size of the envisaged dictionary (the pilot survey suggests that this will be approximately one-fifth of the whole), the remaining items will consist of vocabulary common to all the areas being examined.

The following areas were originally considered relevant for inclusion: England, Ireland, Scotland, Wales, United States, Canada, Australia & New Zealand, West Indies, South Africa, India-Pakistan-Ceylon, West Africa. Since this time, it has also been agreed to include the Philippines; and the possible inclusion of Malaya and East Africa is still undergoing investigation. Consultant editors interested in taking part in the project have been found for all areas except Ireland, Malaya and East Africa, as follows: Professor W.F. Bolton, University of Reading (United States); Professor W.S. Avis, Kingston (Canada); Professor R.D. Le Page, York (West Indies); Professor W.S. Mackie, Cape Town (South Africa); Mr. J. Spencer, Leeds (West Africa, India-Pakistan-Ceylon); Mr. G.W. Turner, Adelaide (Australia & New Zealand); Mr. D. Murison, Edinburgh (Scotland); Professor B.P. Sibayan, Manila (Philippines); Dr. D. Crystal, Reading (England, Wales). I am still trying to find editors for the outstanding areas.

The editorial board has the following constitution: a general editor (myself), responsible for the planning and compilation of the dictionary as a whole, and whose particular role will be to integrate the information provided by the other members of the editorial board; an associate editor (Prof. Bolton), to assist the general editor in compilation; and the consultant editors, whose task is primarily descriptive. Other staff will comprise research assistants and a phonetician, based at Reading; fieldwork may well be carried out by assistants in the areas concerned - this would be decided by the local consultant editor in consultation with the general editor.

The aim is to describe all items which would be generally acceptable and comprehensible throughout the whole (or the greater part) of any English-speaking area, primarily to the educated or literate section of the population. In many areas, there would be a realistic notion of 'standard' vocabulary available which could be used as a basis. In other areas, no such notion would exist,

but it was hoped that a selection of the more frequent variations in usage from British or American English could be obtained. The emphasis is on regional generality: the local divisions of traditional dialectology are felt to be outside the scope of DESP. Cases of difficulty in deciding between degrees of local usage would be handled as they arise.

The results of the survey for this field will be presented in 4.1.

3.2. Linguistic. While any suspect information would be checked through informant reaction and other tests, it was not the primary purpose of this dictionary to do original research into fields where reasonably satisfactory basic information already existed. Whenever a source-book contained reliable semantic or phonetic information, for example, DESP would use this as a starting-point, although the information would then normally be treated in different ways. (Cf. 4.2. for further details of the methods used.) The total information contained in the dictionary could be broken down into a number of headings, as follows.

3.2.1. Word-selection within all regions. It was agreed to exclude highly specialist vocabulary and encyclopedic items and senses, unless the latter were prerequisite for the synchronic explanation of a word (e.g. babel)¹. 'Grammatical' words would be included in the alphabetic list, but only their pronunciation, grammatical status and information about the system of words in which they occurred would be given (e.g. my would be referred to the whole possessive pronoun system). No attempt would be made to provide a comprehensive or illustrated grammatical statement, this not being a lexicographical task.

3.2.2. Pronunciation. To indicate phonological variations between dialects, we would experiment with an IPA-based 'meta-transcription': ideally, this would be a general transcription, interpretable via a specific set of rules explained in the introduction, into different phonemic realisations for the different dialects.

3.2.3. Grammar. Information would be presented as systematically and realistically as possible, which in the first instance would mean a different treatment for parts of speech and affixation.

3.2.4. Etymology. It was proposed in the Prospectus that this information be excluded, a decision which has since been revised.

3.2.5. Usage. Labels of various kinds indicating variations in the level of usage of words would be introduced in as systematic a way as possible, to include stylistic and important connotative

associations of a word.

3.2.6. Meanings. The survey would experiment with the introduction of structural relationships between lexical items, using the notion of meaning-relations (e.g. synonymy, antonymy) suggested by Lyons (Structural Semantics, 1963).

Ordering the meanings of an item would be done not on the basis of historical development or frequency, but in terms of a principle of semantic generality: wherever possible, the most general meaning would be put first, and more specific meanings later.

3.2.7. Idioms. Where it would be useful to do so, idioms would be discussed in terms of restricted collocability.

4. Method of the pilot survey.

4.1. Regional information. In May 1967 a letter was sent to all consultant editors asking them to provide tentative information about the relevant lexicon of their areas in order to ascertain as clearly as possible the size of the dictionary and the proportion of non-British and U.S. entries. Editors were asked to list lexical items (including idioms) - in the first instance, those having a as a second letter - and to add any estimate they might be able to make about overall numbers. Three lists were asked for, under the headings of (a) distinctive lexical items, (b) distinctive senses, and (c) items whose educated usage or general status was in doubt. Most editors sent in lists, which were used as a sample to compute the overall size of this component of the dictionary. This information was supplemented by items already available in the reference-dictionaries we used. Results are presented below, 5.1.

One thing which did emerge was the extent to which regional information of the kind we require was not available for certain areas, and would demand further research in order to obtain it. In particular, Professor Le Page felt that considerable work was still needed to clarify the notion of West Indian English which one might want to include; and it is equally clear that some of the remaining areas pose problems. The question of norms for Irish English, for example, remains to be decided at present.

4.2. Linguistic information.

4.2.1. An attempt was made to see how many words we should like to include, apart from the highly technical and encyclopedic items, which were omitted by definition. To this end, a conflation of the lexicons of the following dictionaries was made:

Webster's Third New International Dictionary (W3), Thorndike-Barnhart's World Book Encyclopedia Dictionary, Funk & Wagnall's Standard Dictionary, The Oxford English Dictionary, the recent edition of the Concise Oxford English Dictionary, and (for United States English) Mitford Mathews' Dictionary of Americanisms. It was very quickly found that no one dictionary provided a comprehensive word-list for British or American English: even W3 omitted a number of important items, which were covered in other dictionaries, and certain senses were to be found in none.

The section of the alphabet BA-DAL was taken as a sample. Using the above method, we took W3 as a base, selected items from this, and supplemented the list from the other dictionaries and from our own intuitions. It emerged that a dictionary written on this basis would amount to approximately 250,000 items,² which far exceeded the terms of reference of the project, both in terms of space and time. (We estimated that one full-time research assistant would take some 25 years to go through the available material in this way!) Consequently, a reorientation was made. The collation technique was in principle indispensable; therefore further restrictions were made on the type of vocabulary selected. A more stringent notion of specialist terms was introduced; and fauna and flora (a potentially very large section of a dictionary) were only taken if there was no generic indication of the object within the morphological structure of the item, e.g. baby primrose was not taken, because the term primrose satisfactorily characterised it; babies' feet was taken, because there is no way of telling that this is a kind of flower from the constituent morphemes themselves, and to bring this information into the dictionary under either baby or foot, or both, would be too complex and uneconomic. Finally, as a procedural measure, it proved extremely time-saving to use the Seventh Webster New Collegiate Dictionary instead of W3 as a base reference dictionary - though of course comparison of all items in W3 continued to be made.

4.2.2. Once an item was included, it was entered, along with its syllable divisions, on a 8 x 5 card, and processed in the following way. (Details of any abbreviations, etc. used will be given below.)

4.2.2.1. All graphological variants would be noted, if any.

4.2.2.2. Pronunciation, with variants, if any, would be given in phonological transcription, factual queries being checked on an ad hoc basis.

4.2.2.3. Regional restriction of the type described above would be noted, if any.

4.2.2.4. Grammatical information would be given: word-class, and indication of irregularity in any important respect, e.g. in number or tense.

4.2.2.5. Etymology would be given (for which additional information was obtained from standard reference-works).

4.2.2.6. The senses would be established. The definitions of a given item in all the dictionaries which included it were compared using a 'lexeme selection' technique, i.e. the defining lexical items in a definition were isolated (along with any crucial relational grammatical information) and listed on a card. For example, a gloss for babble was 'a murmur or continuity of confused sounds' (W3).

This would be noted by us as

 sounds confused murmur/continuity

Alternative glosses from different dictionaries would all be recorded in the same way - for example, where a dictionary had 'noises' for 'sounds', this would be written down under 'sounds' on the card - and an overall picture of past definitions of a word would be built up, in terms of range of glosses used, and degrees of frequency for any particular gloss. This collation was then used as the stimulus-material for the framing of new definitions ordered on the lines discussed below. Often our definitions used words chosen from this accumulated data; occasionally, it was necessary to find new glosses where the old proved inadequate in some respect.

This procedure was simultaneously carried on with a second procedure designed to accumulate information about semantic structure in the sense described above. Other items were brought together in terms of whether they showed types of similarity or oppositeness of meaning, and related senses of these items were noted. Additional information on these points was obtained from the Webster Dictionary of Synonyms, Roget's Thesaurus, and elsewhere. All items obtained in this way were then classified in terms of the simplified model of semantic structure which DESP is using (see 5.7.).

4.2.2.7. Other information about usage, e.g. stylistic restriction, would next be established.

4.2.2.8. Finally, items grammatically related through affixation would be listed. The entire range of affixes for any item was always given, and the relevant senses duly noted.

On this basis, a number of entries were gone through, and these are attached with this report. We shall now summarise the

theoretical findings and procedural decisions made before illustrating the kinds of information provided in articles.

5. Conclusions.

5.1. Scope. The picture of regional information is not as complete as I would have liked, though it was adequate to assess the complexity of the situation. Naturally, it would not have been in order to ask the consultant editors to give up too much of their time before being able to assure them that the dictionary as a whole would go forward and that their time would not be wasted. In fact, not all editors provided material, and consequently the sample piece of alphabet is very incomplete, and should not be taken as a final presentation. Also, some of the items and senses given have to be checked with the consultant editors concerned.

As far as size is concerned, it would seem that an overall total of 180,000 entries is a realistic and useful target. B-BAL is approximately 1/350 of the total dictionary, and just over 500 entries were included under this section. Of these, about one-fifth comprise distinctive items in the sense of 4.1 - far more than was originally anticipated.

5.2. Graphology. All head-entries indicate syllable-division. The convention on the enclosed cards is to leave half a space between each syllable, e.g. babaylan. Alternative spellings are also given in full, after the word also. If in such cases there is no clear most frequent form, the head entry is taken to be that occurring earliest in alphabetical order, e.g. judgement ... also judgment. Where capitalisation is relevant, this is indicated after the definition of an item through some convention, e.g. caps or usu caps, as very often only one sense of an item is used in this way - see babel, for example. All head-entries are consequently in lower-case.

5.3. Phonological component. After a great deal of discussion, the original idea of a comprehensive meta-transcription, to cover all varieties, was found to be possible in theory but lexicographically impracticable. However, a reduced transcription of the same kind could be made to handle the differences between US and English (R.P.) pronunciations, and this was introduced. The table on p. 10 illustrates the metatranscription and its equivalents in R.P. and General American. Illustration of these systems, along with comments on the exceptions, would be given in the Introduction to DESP. Exceptions would in fact be handled through separate citation within an article, an item being given two transcriptions (one G.A., one R.P.) instead of one.

The following range of consonants are indicated in the metatranscription: they are largely the same for both varieties. /p,t,k,b,d,g,f,v,s,z,š,ž,č,ǰ,m,n,l,r,j,w,θ,ð,ŋ,h/

Regular differences (e.g. /t/ v. /tj/ in words like 'substitution') would be indicated in the introductory essay. Allophonic differences of importance (e.g. the use of a voiced flap articulation for intervocalic -t-, as in 'latter') would be discussed in a short essay ^(in this case) on US English pronunciation.

The vowel systems are as follows. Ultimately, this has to be filled out to specify the kinds of quality difference which occur in unstressed syllables, and is to that extent incomplete.

| Meta. | R.P. | G.A. | Meta. | R.P. | G.A. |
|-------|------|---------|-------|---------------------|-------|
| i | i | i | au | au | au |
| ɪ | ɪ | ɪ | əu | əu | əu |
| e | e | e | ier | ie, ie ^r | i(e)r |
| æ | æ | æ | eər | eə, eə ^r | e(ə)r |
| a | a | {æ a | uər | ue, ue ^r | u(ə)r |
| ɒ | ɒ | a | er | e, e ^r | (ə)r |
| ɔ | ɔ | ɔ | ar | a, a ^r | ar |
| ʊ | ʊ | ʊ | ɔr | ɔ, ɔ ^r | ɔr |
| u | u | u | | | |
| ə | ə | ə | | | |
| ʌ | ʌ | ə | | | |
| ei | ei | ei | | | |
| ai | ai | ai | | | |
| ɔi | ɔi | ɔi | | | |

The primary accent in a word is indicated through the use of ' above and before the syllable in question. Occasionally it is necessary to indicate a secondary accent, especially on compounds, in which case , below and before the relevant syllable is used. Where primary and secondary accent are transposable on a word - e.g. babassu, which may be both 'babas,su and ,babas'su - the W3 technique of using both indications simultaneously is used, viz. !babas!su.

Three other conventions have been made use of so far:

- (a) ~ indicates a syllable in which the transcription is identical with the one given immediately preceding it, either in the same or succeeding entries.
- (b) " indicates that the pronunciation of entry X is the same as that of entry W given immediately above.
- (c) _ beneath a symbol in transcription indicates that this symbol may be replaced by another, to give an alternative possible pronunciation; the alternative symbol is then given, and also underlined, e.g. /babiɪna, a/

Dialects other than the two just described will be treated through a system of introductory essays, in which the general

characteristics of the pronunciation of an area, as well as the systematic differences between the area and R.P. would be discussed and outlined. (A similar idea can be found at the beginning of the Funk & Wagnall Dictionary, but this only deals with Canadian English and Australian English, and treats them in very general terms, saying little about their phonology.)

5.4. Grammatical component. Grammatical information emerges at four places within an entry.

5.4.1. Word-classes. We have specified the following word-classes in the material so far examined, these being abbreviated as stated.

| | | | | | |
|-----|-------------------|------|------------------|------|-----------------|
| n. | countable noun | un. | uncountable noun | tv. | transitive verb |
| iv. | intransitive verb | adj. | adjective | adv. | adverb |

But we are anxious not to force words into traditional categories which are inappropriate: consequently, whenever a word is indeterminate in respect of the standard labels, this is noted, e.g. by ?adv (in the case of however, for example), ?adj (for asleep). The point of the query would be explained in the Introduction. Adj/adv would be explained as being a word-class where the words contained some of the properties of adjectives and some of adverbs. The full range of problem cases, along with the other labels not so far mentioned, will be established as the survey proceeds.

5.4.2. Grammatical irregularity. Where nouns have irregular plurals, no singulars, etc., verbs have irregular principal parts, adjectives have irregular comparative forms, and so on, this information is given immediately after the grammatical designation above. Irregular spellings relating to such points would also be given. No indication is given if a form is regular, e.g. adds an -s or -es to a noun.

5.4.3. Closed systems. Whenever an item functions as part of a restricted system of words, then the other members of this system are given, e.g. articles, pronouns.

5.4.4. Affixation. It was felt that a great deal of wasted space is due to lexicographers' specifying in full the meaning of an affixed morpheme, e.g. babbling

babblingly 'in a babbling manner'.

Information of this kind can largely be taken for granted by the native speaker, who is supposed to 'know' the grammatical rules of the language that he needs to interpret suffixes like 'ish', 'ation', and so on. What he may not know is which affix any

given word may take; for example, does one say 'baalite' or 'baalist' if one wants to refer to a follower of baal? Consequently, I feel it is important to specify the entire range of affixes that could be used with any root (noting in the case of suffixes what word-class the root-plus-affix combination has produced); but I do not feel it necessary to say overtly what the combination means, this usually being deducible from the components. On the other hand, it is going to be the case that occasionally a person will want to ascertain precisely what nuance a given affix has got when used in one word as opposed to in some other. To get the best of both worlds, therefore, the following methods have been used. Each affix is analysed into its distinctive senses, which are listed under a separate entry; each sense is numbered 1, 2... n. For example, -ish: 1. of or belonging to ~~to a specific age or time.~~ 2. of the nature or character of ~~to a specific age or time.~~ 3. ^{having a certain quality of} ~~approaching the nature of~~ 4. ^{approximating} ~~adj.~~ ~~A. somewhat adj.~~

Examples of each sense would be given. Whenever an affix is used with a root, then, it is listed after the relevant word or sense, and its specific sense given by means of a superscript figure, e.g. baal ~ ish¹ adj. This should read: 'if you want to form an adjective from this noun, then the suffix you want is -ish. You should know what the whole word means, as all you have to do is add the meaning of the noun (given above) to the meaning of the suffix, which you should already know. If, however, you are not clear as to precisely what the suffix means, then its sense(s) will be found under -ish number 1'.

In this way a list of affixes can be given economically and precisely: see babbitt¹, for example. If further information needs to be interpolated after any given derived form, then this is readily done: see baal, baalism, where synonym information is introduced. Also, it makes it convenient to display the situation where different senses of a given item do not allow the same range of affixes, e.g. baboon in sense 1 allows baboonery, and in sense 2 allows both this and baboonish. Moreover, if the sense of the affix varies from sense to sense, then this too can be easily noted; see again baboon where the sense of -ery required for sense 1 is ery^{4,6}, for sense 2 is ery¹. This latter kind of information is particularly neglected in dictionaries.

Where an affix applies to the whole of a polysemantic item, and not just to one or other of the senses, it is given at the end of the entry, and distinguished from the last sense through some typographical convention (to be decided) - for the present, this has been indicated via spacing, as in baal.

Where an affixed form is sufficiently different in meaning to make the above system difficult to apply economically, then the derivative is given a separate entry, as in baalism. This is rare.

Only three examples of senses not derivable in an additive way were found in the sample.

Whenever a spelling change causes a problem (e.g. babe-babish), then the whole word is given, but the suffix is underlined, and the number of the relevant sense(s) given as above, e.g. babish¹. If an irregular (i.e. unpredictable) pronunciation change occurs through affixation, then this is given in transcription. Regular stress shifts, e.g. through the addition of -istic, will not normally be given.

5.5. Regional dialect. An indication of regional restriction is given using one of a set of abbreviations, e.g. Can Canada, US United States, S Af South Africa, Mal Malaya, Ind India and Pakistan, Phil Philippines, Scot Scotland, Wal Wales, esp. especially. Other abbreviations will be introduced as need arises. Items with no regional specification are assumed to be common to all areas (an assumption which will of course have to be checked with consultant editors in due course).

5.6. Etymology. The language or state of the language immediately preceding the present is given, where relevant. In the case of Middle English, the antecedent language is also given. Also, where a non-linguistic origin is clear (as in babble), this is given: specification of the nature of this origin is made through a descriptive label, which is abbreviated (see below). No dates are given. Where words have closely related etymologies (such as pairs of nouns and verbs, as in baa), the etymology is given only for the first word. The following abbreviations are used in the accompanying material: this list will of course be supplemented.

/ used to separate two languages, e.g. ME/MF = Middle English from earlier Middle French.

Afrik Afrikaans, Fr French, Heb Hebrew, L Latin, OE Old English, ME Middle English, MF Middle French, MLG Middle Low German, Malag Malagasy, Sp Spanish, Port Portuguese, Russ Russian
Desc Descriptive, i.e. an attempt to describe through some obvious image the nature of a phenomenon, e.g. babies'-feet, so called because of the appearance of the flower's blossom.

Echoic, refers to the origin of a word in imitation of some sound.
Pers Person, i.e. the word derives from the name of some person, such as an inventor, or fictitious character. More specific detail may be given within the article concerned, e.g. babbitt.

5.7. Semantic component. Two principles have been introduced to handle the information obtained. First, senses are ordered in terms of decreasing generality, wherever a relationship of this kind can be shown to exist: the most general sense precedes the

more specific, in a kind of diminishing semantic concentricity. See babel and baal for examples of this, where in other dictionaries the senses are given in reverse order, on historical grounds. Where no such link exists between senses, then a criterion of relative frequency is introduced, wherever this is clear. Where no frequential criterion can be adduced, then the ordering is historical.

Second, selective semantic structure information is presented. It quickly became clear that it is not possible to present the reader with the entire system of meaning-relations in English, primarily because the alphabetical restriction destroyed the multi-dimensionality of the relationships. However, it is still possible to give clear clues as to the nature of the system - signposts towards the overall structural field of an item, so to say. With this in mind, we have incorporated the following conventions.

After a definition is given, if there are other words which stand in a clear relationship of synonymy, then these are given, and underlined in red. If the relationship seems to be wholly synonymous, then it is given with no additional mark; if the relationship is evidently approximate, then it is preceded by the abbreviation Cf. (for compare), along with an indication of the relevant sense, where appropriate, and the reader is expected to draw his own conclusions from the comparison.

If there are words which are important to the definition of an item, and which show a clearly hyponymous relationship, then such words may already have been used in framing the basic definition (e.g. when one defines X as a 'kind of' Y); if not, further hyponyms are given and indicated by green underlining. Both more general and more particular terms can be referred to in this way, so long as the list of potentially relevant items is not too open-ended.

Where one has a word which is part of an ordered set of incompatible items which split up a continuum in some way, e.g. parts of the body, colours, ranks, then only the adjacent terms are given, for example in ranks the definition of sergeant will involve mention of corporal and sergeant-major in this way. Incompatible terms are separated by an oblique and indicated with blue underlining.

Where there are words of opposite meaning, these are preceded by the abbreviation Con (contrast). As there are many kinds of oppositeness in language, however, a further indication of the kind of opposition is necessary when a clearly definable logical relationship is involved, e.g. ant (for antonymous relations, as in good/bad), conv (for converse terms, as buy/sell), comp (for complementary terms as single/married). This sub-classification of types is not

final, however. As more material is processed, I expect to find that further specification of certain kinds of opposition will be necessary. Where these have already occurred, an ad hoc label has been used to indicate the kind of information the dictionary should ultimately provide, e.g. babuina.

Whenever a word is cited not all of whose senses are relevant to the meaning-relation, then the specific sense of the related item is indicated via a superscript, e.g. Cf. idol¹. As few of these items have as yet been analysed, however, an asterisk has been used to indicate likely lacunae in this respect. If there is no superscript or asterisk, then all senses of the compared item are to be taken as being potentially-relevant to the understanding of the meaning-relation.

In those cases where a large number of relational terms are required for an item (e.g. in babel), then ultimately we shall avoid unnecessary redundancy by restricting the joint specification of these items to one entry in the dictionary, e.g. all the items might be listed under uproar, and under babel there would be a cross-reference to uproar for further information - a procedure adopted in part by the Funk & Wagnall Dictionary. To illustrate the kind of complexity likely to arise, however, items in the sample have been written out as if they were the pivotal items in such complexes (see babel and babble, for instance).³

5.8. Usage. Where a whole sense is distinguished from the remainder through a stylistic variation, then this is indicated with the abbreviation Styl preceding (e.g. babe). Language variety labels are given where necessary, e.g. spec (= specialist term). Standard connotations are indicated using the labels Disp (disparaging) and Laud (laudatory), see babbitt and babble. Other labels will be added as occasion arises.

5.9. Idioms. No idioms in the pilot survey presented a problem which collocational information could assist with. Idioms are simply listed at the end of an article, with the headword replaced by ~, e.g. ~ in the woods (q.v. babe). Collocational information is of course introduced into the definitions of items where necessary, e.g. babble.

5.10. Finally, an Introduction will be written, giving guidance as to how DESP might best be used, and explaining the motivation for the word-selection and the information.

6. Summary.

6.1. There remain a number of problems of detail which have to be resolved, e.g. certain questions of pronunciation norms, senses, usages, regional restrictions. These have either been left as lacunae or given a temporary answer. A list of such queries has been drawn up and will ultimately be checked, once the project proper takes place. The pilot survey should be able to be assessed without reference to this.

6.2. The pilot survey without doubt achieved its main aim, to determine the extent of the problems involved. It seems clear that the restricted linguistic aims outlined in 4 above are both feasible and useful, and that the project as a whole poses no insuperable theoretical problems. But the size of the project has caused a certain amount of rethinking as to general organisation. It now seems clear that two full-time research assistants, not one as originally envisaged, will be necessary for the first four or five years of the project, and this has financial implications. The task of collation, in my view prerequisite for this project, is too vast for one person. Also, the problems of certain consultant editors demand special attention, and it will undoubtedly be necessary to finance a certain amount of research work in some areas.

Bearing these points in mind, I feel that the original time-scale can be adhered to: this means that if the project were begun in January 1968 (the earliest time by which research assistants could be found), an end-point in August 1974, with first delivery of copy in August 1971, would be realistic.

Footnotes

1. (p.5) Lexical items cited thus refer to the attached cards, containing a sample of the entries.
2. (p.7) Interestingly, this figure corresponds to Professor Ramsay of Missouri's estimate of what the passive working vocabulary of an individual should be.
3. (p.15) A similar system of indicating structure between items is used by the Dictionnaire Robert, though its cross-referencing technique does not recognise as many types of meaning-relations as here.

Dept. of Linguistic Sciences,
University of Reading.

October 1967.

ADDENDA

5.2. Spelling differences in head-words (normally between Great Britain and United States English) are handled by always printing the larger form and underlining the omissible item, e.g. colour. Unless otherwise stated, the 'full' form is British and the 'reduced' form American. No examples occur in the enclosed cards.

5.3. Further phonological conventions. a) If an item's pronunciation can be wholly deduced from its components (e.g. in compounds), then no separate indication is given. If only the stress is important, then this is indicated using the ~ convention described above. (b) In multi-word items, where a linking r (optional in R.P.) is possible, this is indicated via a superscript, e.g. /baba^r əu r^m/. (c) Where a word may contain a glide between two consonants or a consonant and a vowel, this is indicated via a superscript ə, e.g. bei^əl, bæb^əlær.

5.4.4. Ø is the convention which indicates a 'zero' suffix, i.e. a convenient way of indicating that an item, hitherto described as belonging to word-class X, may also belong to word-class Y with no further morphological change, e.g. baby n. -- ~ Ø adj.

5.8. / is also used to separate alternative items in a given context, e.g. babassu has the spec. reference *Orbignya speciosa/martiana*, which should be read as *Orbignya speciosa*, *Orbignya martiana*.