An experiment in filing

Phyllis Healey
SIL-UND

AN EXPERIMENT IN FILING

Phyllis Healey

Two methods of linguistic file-keeping have been advanced to date, each with its peculiar advantages and disadvantages. This paper is primarily concerned with dictionary files.

(1) The 3" x 5" slip system. [1]

The great advantage of this system is that if the file is carefully kept, the entries are almost in correct alphabetical order from the beginning. "Almost", because certain alterations due to revision of phonetics, phonemicisation, and greater enlightenment as to meaning are inevitable in time. Affixed forms can be entered together on one slip, and thus stems can eventually be identified, provided they do not get lost in the file through a wrong first guess as to the phonetic form of the stem.

However, several serious disadvantages have caused this system to become bugbear to workers, especially those with children:

(a) The first and greatest disadvantage is the danger of thousands of small pieces of paper being accidentally disordered. A chance fall of the box containing them, a strong wind, or a small child may work havoc with hundreds of hours of hard work.

(b) A suitable container is another problem. The file expands quickly, outgrowing containers. Containers must be substantial to obviate the danger of the slips being scattered through getting too heavy for it.

(c) That raises a third problem, weight and bulkiness. A 3" x 5" system and its container is heavy and awkward in shape thus providing headaches for mobile persons, who must often travel, yet want to take their materials with them.

(d) It is only too easy, as many 3" x 5" users have testified, for useful words to be "lost" in the file for months or even years because they are not frequently seen. In fact, only one slip is visible at a time in this system, and thus valuable phonemic and morphemic clues that a comparison of words can afford are often missed.

(2) The Page File. [2]

The page file has hardly any of the disadvantages of the 3" x 5" system. If it is kept in book form, disordering through accident is impossible. If a loose-leaf book is used, considerable

---

1. Suggested and elaborated by E. A. Nida in "Morphology", Ann Arbor 1948, pp.195-199 and by V. Pickett "An Intruduction to the Study of Grammatical Structure", Glendale, 1956, pp 209-211.

2. I am indebted to R. S. Pittman for first suggesting this method, and to him and H. B. Kerr for some modifications that have been incorporated.

expansion is possible without inordinate weightiness, and the book can be of the same size and shape as data books, grammar files, and other analytical materials. Thus all one's work materials can be uniform, and therefore very compact for purposes of packing and transportation. The danger of entries being "lost" in a page file is greatly reduced, for each time an entry is made, a whole page or more of phonetically similiar entries is presented to the eye. Phonemic and morphemic clues can strike one even when they are not being deliberately sought. And in addition, a more or less unconscious vocabulary-learning process is going on every time entries are sighted while others are being made. Comparison of related forms is comparatively easy, and the basic forms of stems and subtleties of meaning can be obtained readily. Depending on how the page file is handled, one disadvantage waxes or wanes in importance.

The entries cannot be in correct alphabetical order, although careful forethought can reduce the necessity of arduous re-ordering to a minimum. Thus, some time spent in re-alphabetising is inevitable.

(3) Trying the Page File.

After weighing the pros and cons of the two alternatives, it was decided that the page file was worth fair trial. The disadvantage of this type of file seemed more comparative than categorical in nature, so an attempt was made to reduce them to a minimum. Some traps were encountered, but the system does seem very workable.

(a) Materials Used. All stationery brought to the field was uniform in size (quarto). Punched paper of light-medium thickness and loose-leaf files with fairly substantial covers have been used for all linguistic work, including separate data booke for each member of the team, for grammar and dictionary. Linguistic reports and papers are typed on the same size paper, and copies included in the relevant files. Apart from such reports, all entries in all files are hand written. Re-alphabetised pages are typed, however. It is submitted that non-loose-leaf books are not at all practicable for dictionary work. Any expense involved in having adequate stationery is well worth it.

Slightly different approaches proved necessary for English and language sections of the dictionary:

(b) English Section. A Standard English Dictionary was taken, and the relative space taken by each initial CV and VC combination was roughly assessed. Blank paper was proportionately headed alphabetically. It was suggested that about 300-400 pages be used. This seems inordinately wasteful at first sight, but the pages soon fill up. It is very necessary to start big. It was found that starting with too few pages resulted before long in the necessity for a 4-fold expansion, with consequent re-writing. It was soon found too that extra space was required for certain predominant cultural items, such as species of birds and plants, rice, arrows, etc. Such a contingency could not, of course, have been predicted by reference to the English Dictionary.

(c) Language Section. As each new initial CV or VC was discovered, a new page was headed with the first two phonemes of the word, and it was entered thereunder. Thus the number of pages continued to expand until th e possible initial CV combinations were exhausted. One bad oversight was made, however, for it was known that /a/ is the most frequent vowel (55%) in the language being studied. The Ca pages (i.e. ba, ka, da, etc.) soon became unwieldy. The solution of the problem proved to be the dividing of the Ca's into about 5 CaC pages in alphabetical order (i.g. bab-bak, bad-bag, bah-bal, etc.)

(4) Traps discovered and ways of eliminating them.

(a) The lists on each page could have been much closer to true alphabetical order had the words been arranged on the page instead of just listed from the top downwards. This was true for both English and Language sections. For example, one page was enough in the English section for A to AG, but it would have been better had words starting with ab, ac, ad, ae, ag been appropriately spaced at intervals down the page.

(b) Centres of interest for both tribesfolk and worker may result in some pages filling up much quicker than could have been anticipated, e.g. the frequency of occurrence of natural species, forms of rice, bamboo, rattan, etc. A suggested solution is the setting aside of a whole page in the English section for such things as 'bird', 'plant', 'rice', etc. as soon as the beginning of such a situation is observed.

(c) Some other initial CV combinations besides Ca might turn out to be unexpectedly frequent in occurrence (e.g. /si/ was in our language). The solution in such cases is the same as for Ca, namely successive pages of CVC's in alphabetical order.

(d) The biggest difficulty with the page file is the necessity ultimately at least for re-alphabetisation. Ways of reducing this work to a minimum have been suggested. In comparing the two methods of filing, it should not be forgotten that the 3" x 5" file probably requires two drafts to get it into true alphabetical order, because of revisions mentioned above. It is submitted that two drafts can be sufficient for the page file also, and that the initial draft need not be such a burden as some have argued. A suggested method of re-alphabetising is to take some sheets of scrap paper, and make a list of stems of each initial CVC (or VCV, etc.) combination as you come to it in the file. The stems are then numbered (1, 2, 3, 4, etc.) in alphabetical order on the scratch list. This need take very little time. A dictionary in correct alphabetical order can then be made by reference to the original file. If several pages of original file are involved in one re-alphabetisation, it is suggested that they be indicated (e.g. by a line between stems for each page break) on the scratch list for easier reference back to the file. E.g. while alphabetising the SI page (s), part of the scratch list may read:

sikig (3)
sikal (1)
siksik (4)
---
sikan (2)

(5) References and Examples.

It is admitted that the page file cannot provide room for examples in each case as can the 3" x 5". However a minority of words only will require examples in the ultimate dictionary, and these can be obtained from the references. It has proved absolutely essential to record references to every significent occurrence in the data (e.g. A125.9 - Alan's data book, page 125, 9/10 of the way down the page). Several references are often needed, especially in a monolingual situation, to establish the exact area of meaning of a word. It has also proved necessary to record each different occurrence of stem plus verb affix, partly to establish the basic form of the stem, partly to discover which set of verb affixes a particular stem takes, partly to ascertain the meanings of the affixes and the basic meaning of the stem (which often yields surprises). It is suggested that, in order to make a dictionary useable, the set of affixes a particular stem can take should be indicated in some way.

(6) The Grammar File.

(a) Alphabetical List of Lateral Morphemes. Because of the greater difficulty experienced in finding the meaning and usage of lateral morphemes, these were filed separately from the stems, which comprised the dictionary. One page was taken for each "function word" (i.e. non-affixable root), and one page for each affix, amd all examples of its occurrence were recorded in full along with meaning and page reference. This eventually enabled the meaning of the word or affix to be determined, and also supplied examples of its syntactic usage.

(b) File of Grammatical Constructions. For the first week or two no book was employed. Pages were strung out vertically so that the heading of each could be seen easily, and all were clamped together at the side with a bulldog clip. One page was taken for each apparent construction type and suitably headed. Examples were listed below in full with page references.

After a week or two it was noticed that very few new constructions were being discovered, that the total number that had been discovered was very limited, and it was then possible to arrange the pages systematically in a book according to the method of analysis, preceded by an index.

Re-ordering of this file many times is inevitable as understanding of the true structure of the language increases, and some re-classification of the examples is probably necessary.