The Trukese-English Dictionary

Recording a Language on the Computer

WARD H. GOODENOUGH

The creation of a dictionary often strikes people as an extraordinary undertaking, although it is more of a commonplace at The University Museum than elsewhere because of the Sumerian and Aramaic dictionaries now being compiled there. Putting together the Trukese-English dictionary called for a somewhat different approach, for the initial recording of the language had not been made by the people who spoke it. Trukese, which belongs to the Austronesian language family (see box), is spoken on the islands of Truk, an atoll in Micronesia in the western Pacific (Fig. 1).

Compilation of the many words, their meanings, and other necessary information was a complicated process greatly aided by computerization. Having all the data computerized also means that further work on the dictionary can continue easily and indefinitely. The present product—the dictionary itself—is a two-volume work. The first volume, published in 1980, contains 55 pages of introductory material describing alphabetic representation of sounds and the gram-

1 Truk, the southwest shore of Romonum.
The Austronesian Language Family

The Austronesian family includes the following (Fig. 2): the aboriginal languages of Taiwan; all languages of the Philippines; with a few exceptions all languages of Indonesia and Malaysia; a few language enclaves in Vietnam and Malagasy (in Madagascar); most languages of island Melanesia; many languages of the north, north-west, and south-east coasts of New Guinea; and all languages of Micronesia and Polynesia. The Austronesian group includes many more mutually unintelligible languages than any other known language family. It is peculiar in having developed a prehylotic distribution more than halfway around the world by sea, a situation unparalleled until the post-Columbian era.

Gilbertese, Marshallese, Kosraean, Nauruan, Fonualei, and Trukese form a subdivision of the Oceanic branch of the Austronesian language family (Fig. 2). They are classified as Nuclear Micronesian.

Glossary

gloss: the word(s) in one language that can appropriately translate the concept in another language.

infection: when a word takes a series of prefixes or suffixes, the set of resulting words is said to be inflected, as Latin nouns are. In Trukese, possessive suffixes are added to nouns and objective pronouns suffixes to verbs.

labil consonants: those consonants made with the lips, such as p, b, and m.

phonology: the study of sound systems; the phonology of a particular language is the sound system of that language and how it works.

semivowels: vowel-like sounds that function as consonants like y and w in English.
Early Processing

Over several years, I tried to assemble the material brought back from Truk, putting it on 5" by 8" sheets, one sheet for an entry, thus to compose a manuscript for the dictionary. Because so many words were compounded of elements that would be separate entries in themselves (Trukese is much like German in this respect), I decided to show the constituent roots from which each complex word was built and to have an entry for each of those roots. Roots that appeared only in complex words and never as words by themselves had to have their own separate entries. What happened was that as I discovered the need to set up new entries for such roots as I went along, I had no easy way to go back and find words already listed in which these roots appeared. Retrieving roots and adding other necessary items were becoming increasingly more difficult as the data accumulated. I had thus worked my way through several letters of the alphabet, learning as I went about what I should be trying to do, when I heard that the University of Hawaii had obtained a contract from the U.S. Trust Territory of the Pacific Islands to prepare a series of dictionaries and grammars of the languages spoken in Micronesia, including the language of Truk. I offered to join the Hawaii project, and was accepted. Thus began my Trukese-English volume, and much of my year as a visiting professor there in 1982-83 was devoted to doing the same for the supplementary volume.

Alphabetic Representation

Toward the end of the last century, missionaries had developed an alphabet for Trukese and the dialects of neighboring atolls that had become well established in schools (Fig. 8) and in ordinary usage before the middle of this century. It was, however, inadequate: it used five vowel symbols to represent nine different vowels, it represented semi-vowels as if they were vowels, it failed to distinguish adequately between two different sets of labial consonants, and it took no account of the crucial distinction between long (double) and short vowels and between consonants that were held for a syllabic beat (now written as double consonants) and those that were not. Thus what the dictionary shows as mūnak ("man, male"), mūnakāp ("wrong"), mūnakā ("sour"), and mūnak ("first, ahead") were all written the same way as mūn. Clearly needed were an alphabet and writing conventions that did justice to Trukese phonology.

The contract the University of Hawaii had with the Trust Territory stipulated that the orthography used in each dictionary be agreed upon by a representative committee of educators, public administrators, political representatives, and traditional language experts who were native speakers of the language in question. Hiroshi Sugita and I met for two weeks in the summer of 1972 with such a group in Truk. They concluded that the dictionary should employ a mode of alphabetic representation that took account of all the meaningful distinctions of sound in Trukese in a way that required the least modification of existing usage. Having exhausted the symbols already in use and determined what unit of sound each would unambiguously represent, the group went on, after examining alternative possibilities, to decide how to represent units of sound not already symbolized. Use of an acute accent mark to distinguish the remaining vowels and of the digraphs pce, mce, ch, and ge accomplished that. It was also decided to retain the mode of alphabetic ordering already customary in Truk, which listed the vowels first and the consonants thereafter (e.g., the consonant s comes where it would be in the English alphabet).

Computer Processing

The University of Hawaii project was planned to be computerized from the outset, which had several advantages. First, the computer could realphabetize entries as corrections and additions were made, and these changes would not necessitate retyping whole pages. Most important was the possibility of coding all the different kinds of information for each entry so the computer could readily retrieve every instance of a particular formation, such as all roots of a particular combining form type, all words that were names of trees, or all words that used the same construction for making a possessive. This prospect made computerization worthwhile the initial efforts, for it would save all kinds of time in the end. Fortunately, I had already determined the types of material that should be marked for independent retrieval. With the help of Robert Has, the computer programming expert in the University of Hawaii's department of linguistics, I was able to set up a
The Computer Codes

The computerization can be illustrated by two homophone entries and their final appearance in the published dictionary (Fig. 9). Under the coding scheme, HW marks a new headword or entry in the dictionary. The other codes—such as CC, EE, ITM, and IG—mark different kinds of information that pertain to a headword. There were two levels of information coded. The first level, represented by double letters, contains information that relates directly to the headword, such as which of several possible combining forms it belonged to (CC) and, if it was a complex word, what other words it was composed of (FF). Each entry of reference had a "base form" representing the root designated by code EE.

The second level of information relates to glosses. First is information on grammatical function, such as noun, verb, etc., together with its inflection type (ITM). Next is the actual gloss or definition in English (1A), followed by useful information about the subject matter of the gloss, such as medicinal uses in the case of plant names (1D), necessary grammatical information (1E, IF, IG), and sample phrases and sentences and idioms (II). This is followed by instances of use with directional suffixes with glosses when needed (IV), synonyms and antonyms (1Y), and other entries to which the dictionary user should be alerted (IZ). The whole set of data at this level could then be begun again under code 2T if the same headword occurred as another grammatical form, as when it might function as both a noun and a verb.

a. system of information marking that served us well.

With the help of a grant from the National Science Foundation, I was able to employ graduate student assistants to code the material in accordance with the system thus created. I reviewed everything they did for accuracy and for further editing. The edited information was then typed onto scrolls, later cut up into convenient working lengths. The scroll paper made an automatic carbon for my own file, and the original was mailed to Hawaii, where a student entered the material into the mainframe computer. (Personal computers were not available yet.) Each summer in Hawaii, I proofread computer printouts of what had been entered. (1 must have proofread the whole dictionary five or six times before we were done.) We were also able to use the computer effectively to highlight typographical errors in format. Nevertheless, some errors appear in the publication, but few serious ones.

b. Of particular importance was our marking with an asterisk (*) every English gloss that was to serve as an entry in an English to Trukese turn-around or index of the Trukese to English. This enabled us to get the thesaurus to take a particular English gloss, list it as a headword, and gather under it all the Trukese words for which it was given as a gloss. The computer did this for all such asterisk-marked glosses, alphabetized them, and printed out what was, in effect, the first draft of the English-Trukese part of the supplementary volume.

The coding system enabled us to have the computer compile an alphabetic list of all the base forms, followed by the entry of reference, the gloss for that entry, and then a listing of all the other entries in the dictionary in which the word element with that base form also appeared (Fig. 9). This provided, in effect, the concordance of roots that constitutes the second part of the supplementary volume.

Thus the computer greatly simplified the task of making the English-Trukese volume, without it the concordance of roots would have been beyond serious consideration. The computer cannot, however, do things that it has not been told to do in advance. Planning the types of information to be separately coded or otherwise marked was essential. Such planning could not have been properly undertaken without a prior knowledge of the language (and of what more one would like to learn about it) sufficient to insure appropriate coding of the material and attention to getting the information for which the codes asked. We tried to fill in the missing bits of information later, but circumstances prevented doing it completely, although the dictionary project at Hawaii included native speakers of Trukese as assistants.

The result of our efforts is a dictionary that is far from complete, containing some inevitable errors, but it is an important advance over what was available for both practical and scholarly purposes before. We hope that someday it will be superseded by an expanded and upgraded edition; but that will probably have to be someone else's task. Whoever undertakes it will find the present dictionary easy to build on, thanks to its availability on diskette for use on a personal computer.

Bibliography

Dow, Washburn

Elbert, S.H.

Goodenough, W.H., and Hiroshi Sugita

Goodenough, Ward B.
1969 Curator Emeritus for Micronesian Ethnology at The University Museum. He recently retired as University Professor of Anthropology at the University of Pennsylvania. His publications include Property, Kin and Community on Truk: Native Astronomy in the Central Caroline; Cooperation in Change; Explorations in Cultural Anthropology; Description and Comparison in Cultural Anthropology; and Culture, Language, and Society.

What the Codes Made Possible

Of particular importance was our marking with an asterisk (*) every English gloss that was to serve as an entry in an English to Trukese turn-around or index of the Trukese to English. This enabled us to get the thesaurus to take a particular English gloss, list it as a headword, and gather under it all the Trukese words for which it was given as a gloss. The computer did this for all such asterisk-marked glosses, alphabetized them, and printed out what was, in effect, the first draft of the English-Trukese part of the supplementary volume.

The coding system enabled us to have the computer compile an alphabetic list of all the base forms, followed by the entry of reference, the gloss for that entry, and then a listing of all the other entries in the dictionary in which the word element with that base form also appeared (Fig. 9). This provided, in effect, the concordance of roots that constitutes the second part of the supplementary volume.

Thus the computer greatly simplified the task of making the English-Trukese volume, without it the concordance of roots would have been beyond serious consideration. The computer cannot, however, do things that it has not been told to do in advance. Planning the types of information to be separately coded or otherwise marked was essential. Such planning could not have been properly undertaken without a prior knowledge of the language (and of what more one would like to learn about it) sufficient to insure appropriate coding of the material and attention to getting the information for which the codes asked. We tried to fill in the missing bits of information later, but circumstances prevented doing it completely, although the dictionary project at Hawaii included native speakers of Trukese as assistants.

The result of our efforts is a dictionary that is far from complete, containing some inevitable errors, but it is an important advance over what was available for both practical and scholarly purposes before. We hope that someday it will be superseded by an expanded and upgraded edition; but that will probably have to be someone else's task. Whoever undertakes it will find the present dictionary easy to build on, thanks to its availability on diskette for use on a personal computer.

School on Truk is conducted, when weather permits, in the open air. The Goodenough's younger son, Garrick, attended this school when he was 5 and 6 years old.