

# Computer software to assist linguistic field work

E. Clay JOHNSTON\*

## INTRODUCTION\*\*

For more than sixty years, the Summer Institute of Linguistics (SIL) has specialized in the study of vernacular languages in many parts of the world. During that time, SIL field workers have contributed to the development of the science of descriptive linguistics and collected linguistic data on more than 1000 different languages. In most cases, the results of linguistic analysis have been applied immediately to translation and literacy efforts.

Along with its work in the field of descriptive and applied linguistics, SIL has gradually developed a library of specialized computer software to assist in linguistic field work.

Although there had been some early use of mainframe and desktop computers for linguistic data processing, the computer age for SIL field workers began with the commercial introduction of battery-powered portable computers that could be easily taken into the field. Until recently, most programs were written for IBM<sup>1</sup>-compatible machines using the MS-DOS<sup>2</sup> operating system. Now similar programs are available

\* *Applications Consultant/Trainer, Summer Institute of Linguistics, 7500 West Camp Wisdom Road, Dallas, TX 75236.*

\*\* *Several colleagues in SIL helped in the preparation of this article. Geoffrey HUNT in England and Andrew BLACK in California provided software information. Here in Dallas, Don SMITH and Karen WHITE helped edit the article, giving special attention to the section on special characters. Evan ANTWORTH helped in outlining the article, gathering information, and overall editing. Doyle PETERSON in Waxhaw, North Carolina, provided editing help with the first draft. Russell RICHARDS in Washington, DC, gave helpful advice and overall editing help. Also, my insights about the software were based on unpublished notes, program manuals, and personal comments by many other SIL workers.*

<sup>1</sup> IBM is a registered trademark of International Business Machines Corporation.

<sup>2</sup> MS-DOS is a registered trademark of the Microsoft Corporation.

for Macintosh<sup>3</sup>-based and UNIX<sup>4</sup>-based computers. New programs are being developed to take advantage of the features of MS-WINDOWS<sup>5</sup>.

The purpose of this article is to describe some currently available SIL computer programs in terms of how they help the field linguist. See 9 for a list and index of programs discussed here. These programs are available for purchase from the sources listed in the bibliography. For information about ongoing development work in linguistic software by SIL, see SIMONS, THOMPSON and DEROSE (1988). For a more general listing of current software used in linguistics and other humanities, see LANCASHIRE (1991).

## COMPARING LANGUAGES AND DIALECTS

A basic concern of SIL is identifying and cataloging the languages of the world. The result of years of effort in this regard have been published in the *Ethnologue* (GRIMES, 1988).

In surveying various geographical areas of the world, SIL has developed techniques for preliminary evaluation of the extent to which languages are similar and the extent to which speakers of one language may be able to comprehend speakers of surrounding languages.

In a given geographical area, one part of this process is a statistical comparison of the lexical forms shared by people in various communities and an evaluation of the phonetic differences between communities. This involves collecting and comparing many word lists. If done manually, such comparisons can involve a large and overwhelming amount of work.

WIMBISH (1989) developed a program called WORDSURV to facilitate entering in the computer comparative word lists from many communities. WORDSURV provides a framework for the linguist to enter judgments about what word forms are phonetically similar and the degree of their phonetic similarity. It also provides for entering judgments about the reliability of the word list data for each language.

The current version of WORDSURV is limited in that each phone must be represented by a single character for correct processing of the word list correlations. This imposes a 256 characters limit on the data and forces the use of non-standard character-phone correlations. Work is

<sup>3</sup> Macintosh is a registered trademark of Apple Computer, Inc.

<sup>4</sup> UNIX is a registered trademark of AT&T Corporation.

<sup>5</sup> MS-WINDOWS is a registered trademark of the Microsoft Corporation.

underway to produce a version of WORDSURV that can correctly process data containing multiple keystroke character designators for the International Phonetic Alphabet.

With input provided by the linguist, the program calculates and displays the number of shared lexical items between all possible pairs of communities. It also gives a numerical estimate of how phonetically similar the languages of any two communities are.

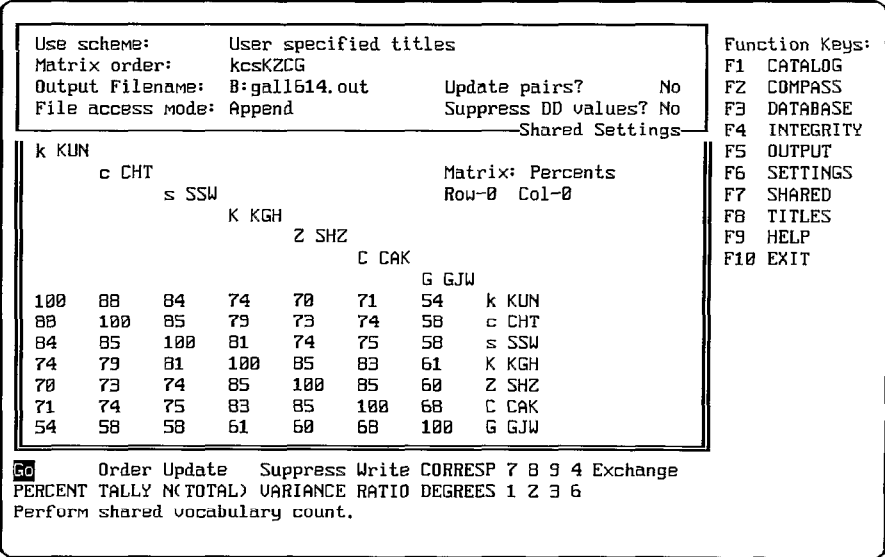


FIG. 1. — WORDSURV display showing percent variance between dialects.

The output of WORDSURV provides a starting point for identifying communities that may be able to understand the language spoken by other communities in the area. For example, the WORDSURV display (fig. 1) indicates that the word list for the community labeled “k KUN” is 84% the same as the list for the community labeled “s SSW.” Shared vocabulary above 80% indicates possible understanding between dialect communities. Other kinds of testing must be used to verify whether or not this is true in any particular case.

In this way WORDSURV helps identify the language or languages in an area that are best understood by the greatest number of people. WORDSURV also provides a framework for linguistic mapping of geographic areas and for historic studies based on linguistic reconstructions.

## PHONOLOGICAL ANALYSIS

Studying an unwritten language begins with the collection of language data including tape recorded utterances and word lists written with phonetic symbols. This phonetic data is the basis of a phonological analysis of the language.

### The FIND PHONE program

The program FIND PHONE (BEVAN, 1991) uses keyboard character combinations based on the International Phonetic Alphabet to facilitate the keyboarding of phonetic data and to provide search capability for displaying the distribution of phones in the data. For example, an aspirated voiceless alveolar stop is typed th<sup>^</sup>.

Searches may be defined for single phones in specified contexts or for classes of phones in generalized contexts. This distribution data forms the basis for formulating and testing phonological hypotheses about the language under study. The lists produced by such searches may be printed or output to disk files.

In the near future FIND PHONE will be enhanced to automate some additional aspects of phonological analysis that help identify the phonemes of a language, i.e., those contrastive sounds that make a difference in meaning and form the basis of a written alphabet.

### The CECIL program

For the field linguist, the identification of most of the sounds in a language is fairly simple. In some languages, however, there are complex interactions of loudness, pitch, and duration that defy analysis without some special analytical tools.

The CECIL (1992) program with an accompanying acoustic-to-digital converter that connects to the computer, provides the linguist with a portable phonetics laboratory for the field. CECIL (Computerized Extraction of Components of Intonation in Language) permits the linguist to capture and store short utterances in computer files. In other words, the acoustic data is digitized so it can be processed by the computer. The utterances can originate from a microphone or tape recording.

For each utterance, CECIL calculates and displays in graphical form the change of acoustic intensity (loudness) as a function of time. Also, CECIL extracts pitch frequencies from the utterance data and displays a graph of the calculated pitch pattern (fig. 2).

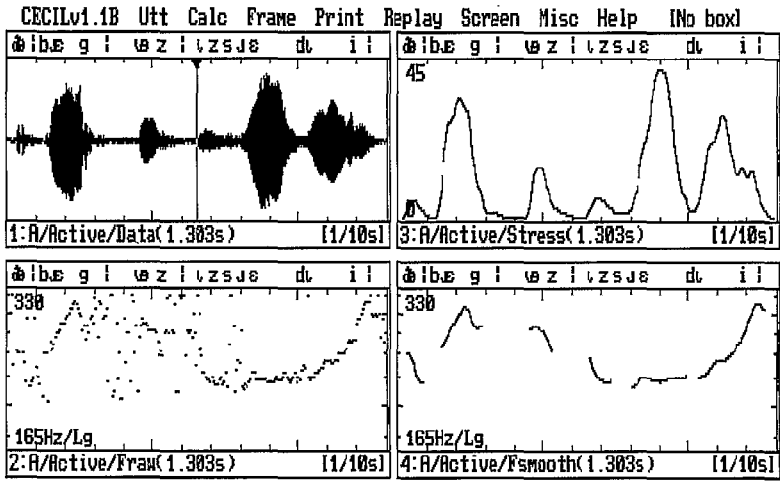


FIG. 2. — CECIL utterance display.

With CECIL the user may choose several different types of graphs for displaying the utterance data and may modify the graphic display parameters. Also, selected graphs for two utterances may be displayed at the same time for comparative purposes. A set of characters based on the International Phonetic Alphabet is provided with the program and phonetic annotations may be added to the graphs from the keyboard.

CECIL permits the digitized utterances to be played back acoustically through the converter box at normal and reduced speeds to help the user correlate the graphs with the original utterance.

A companion program named SPECTRUM converts and displays the utterance data of CECIL in the form of speech spectrograms. If needed, the graphical displays of CECIL and SPECTRUM can be printed.

CECIL comes with a tutorial manual, a program reference manual, a hardware attachment manual, and an introduction to acoustic phonetic analysis that discusses in a general way the gathering and interpreting of acoustic phonetic data. The menus and help messages in CECIL and SPECTRUM can be translated into other languages without modifying the programs. A French version is available.

TEXT ANALYSIS

In the field, recording and transcribing vernacular texts is another essential aspect of linguistic analysis. Since the early 1980's most transcription has been done on computers using word processing programs.

Computer programs have been designed to facilitate two different approaches to analyzing the data in this kind of vernacular texts: 1. the word-by-word annotation of the collected texts with output in aligned interlinear form (fig. 3); and 2. the generation of concordances from the texts, that is, a display of any or all linguistic units in their various contexts in those texts (fig. 4).

### Text annotation programs

Programs which facilitate the annotation of texts help the linguist in the following ways: 1. the program presents the words of the text and a user-designed set of empty annotation fields as a reminder of the glosses, grammatical categories, and other notations needed; 2. the program provides the editing function necessary for typing in the annotations; 3. the program automatically keeps the annotations lined up in an orderly interlinear display; and 4. the program stores the annotations so that they are automatically recalled and inserted when the same word is encountered again in the texts.

The process of annotating vernacular texts is analytical in nature. In moving from word to word through the texts, the linguist is forced to formulate hypotheses about word meanings, word classifications, morpheme breaks, morphophonemic variations, and other features.

```

\rf kac024 1      kac024.db 001

\lx Bag+o mabuat      ta      ikam primiro anay magpangdan      ka      anay.
\mr Bag+o ma -buat    ta      ikam primiro anay mag -pangdan    ka      anay
\gl before AF.NB-make FMP,II mat first first AF.NB-plant Z,S,I first
\ps ADU pre -Ut      FMP      N      ADU      suf pre -N      PRON suf

\ft Before you make a mat first look for pangdan.
\dt Z5/Mar/91

\rf kac024 2      kac024.db 002

\lx Miling      ka      tengnged ya      patio      bag-o mamangdan.
\mr m -iling    ka      tengnged ya      patio      bag-o mang -pangdan
\gl AF.NB-go    Z,S,I close FMP,I cementary before AF.NB-plant
\ps pre -UI     PRON ADJ FMP N      ADU pre -N

\ft Go close to the graveyard before you look for pangdan.
\dt Z5/Mar/91

```

StS View: <Sp.bar for Hex/ASC> <F2 Prints> <F3 FFeeds> 06/03/92 02:01:55 pm  
 file selected: KAC024.ITX Press [Esc.] to exit.

FIG. 3. — A sample of annotated interlinear text.

## The IT program

The IT (Interlinear Text) program is a specialized tool for the linguist that facilitates the word-by-word annotation of text material in an interlinear display (SIMONS and VERSAW, 1987, SIMONS and THOMSON, 1988). IT automatically maintains alignment of a series of different annotating items under the words or morphemes to which they apply. Once an item is annotated by the user, IT automatically recalls the annotation information the next time that item appears. Multiple annotations of homonyms are supported by the program.

IT permits the user to define the interlinear model, that is, the set of interlinear annotation lines that present themselves for input from the keyboard. One may define several different interlinear models and modify them as needed. A single model can implement more than one layer of parsing, i.e., the dividing of words into their smaller parts.

The annotations entered in IT are stored in a lexical database disk file that is not accessible to the user while the annotating process is underway. The lexical database, however, can be extracted by IT as a plain ASCII file<sup>6</sup> and serve as the basis for developing a more detailed lexicon. A single lexical database can support several interlinear models.

In regard to editing, IT is structured to prevent changes in the text and annotations that are not reflected in the lexical file.

IT is the program of choice if one is interested only in the preparation of annotated interlinear text. It is available in DOS and Macintosh versions.

The IT manual not only discusses program functions, but also helps the linguist decide what data needs to be included in the annotation of the texts. The manual suggests a staged approach to text annotation in which more detail is added as one gains more understanding of the language.

## The SHOEBOX program

The SHOEBOX program (WIMBISH, 1990) is a more general data management program for lexical data and field notes but it also provides the same kind of annotating and interlinearizing function as IT (see the above "IT program").

<sup>6</sup> The term "plain ASCII file" and "plain file" is used here and elsewhere for a generic computer file that contains only printable characters. With most word processors, this type of file is created when a text is saved as "text only" or "unformatted." Plain files contrast with program-specific files that contain hidden codes that implement screen and printer formatting. ASCII stands for the American Standard Code for Information Interchange, the character set standard for the microcomputer industry.

In terms of limitations, SHOEBOX permits only one parsed line in any interlinear model and requires duplicate sets of text and lexical data for different interlinear models.

Also, with SHOEBOX, it is possible to edit the interlinear text display independently of the annotation data in the lexical files. That is, it is less constrained to maintain a one-for-one correlation between the interlinear display and the lexical data.

The primary advantage of SHOEBOX over IT is that the lexical files for the annotations are always immediately accessible inside SHOEBOX and may contain any amount of additional lexical information for each word. For the purpose of adding to or editing the lexical records, the user may interrupt and resume the text annotation process at will. This is especially helpful if the primary reason for annotating the texts is to build a lexicon for the language.

### The ITF program

To help solve the special problems involved in professional printing of interlinear texts with alignment, KEW and MCCONNEL (1990) have developed a set of small programs called Interlinear Text Formatter (ITF) to work with the TEX<sup>7</sup> typesetting program. Among typesetting programs, TEX is especially suited for aligned interlinear text because it processes text units in boxes rather than lines.

ITF is used to prepare the files output by IT (and the AMPLE program discussed below) to be typeset correctly by TEX. With some preliminary editing of the files, ITF can also facilitate the formatting of interlinear texts produced by SHOEBOX. Versions of ITF are available for DOS, Macintosh, and UNIX operating systems.

The effective use of TEX with ITF requires some specialized typesetting knowledge.

### Text concordance programs

Concordance programs focus the attention of the linguist on the distribution of linguistic elements in multiple contexts. These programs can display any kind of unit from a single phoneme to a phrase in all its contexts in a collection of texts.

<sup>7</sup> TEX is a professional typesetting program developed by Stanford University and used widely in Universities and business. It has been adapted for use with non-roman writing systems such as Japanese and Hebrew. It is distributed in both commercial and public domain versions for DOS, Macintosh, and UNIX operating systems.



The distribution of a linguistic unit in all contexts is the ultimate test of the analytical assumptions made about it. While the linguist cannot possibly collect text data for all distributions of a unit, the more samples one can collect and view for analytical insights, the more confident one can be about the accuracy of the analysis.

In this regard, concordance programs complement text annotation programs. When annotating, the attention of the linguist is focused on words and morphemes in a restricted context. The analytical conclusions reflected in the annotations need to be verified by the distribution data visible with concordances.

## The FIESTA program

FIESTA (Fast Interactive Editor of Scripture and Text Analysis) is a high-speed text editing program (ALSOP and JOHNSTON, 1990) which includes a versatile concordance function. It permits the user to apply concordance searches to hundreds of pages of vernacular texts as a single database. It can process annotated and non-annotated texts.

With FIESTA, the linguist can display almost instantly a concordance of any linguistic unit from a single phoneme to a string of ten words. In figure 4, for example, one screen of a text concordance for the prefix kine is displayed.

In any concordance display the search item is highlighted on the screen and shown in all its distributions in the texts. The user may view a single line display of the concordance with the item in each context aligned for comparison (fig. 4), or, one may choose to display the entire sentence in which the search item occurs.

FIESTA permits the user to specify constraints on the concordance search. For example, one might display a particular phrase of three words only if it does not occur at the end of a sentence. Or, one might display a particular morpheme only if it occurs at the beginning of a word, or only in a specified context of other phonemes or morphemes. There are many other types of constraints on the concordances that can be applied individually or in combination with others.

This flexibility, coupled with the fact that the concordance displays are almost instantly visible on the computer screen, makes FIESTA a very powerful analytical tool for formulating and testing hypotheses about linguistic units.

Because of FIESTA's speed, it is interactive. That is, the rapid display of text data in response to concordance searches permits essentially instant evaluation of the search criteria. The linguist may then modify and repeat the searches as many times as necessary for maximum insight

```

atext1 18 18 Na, ini sa tulon denu sa kinelesut i H'esus Kelistu. Si Maliya
atext3 6 a Haldan an'i tandaan da sa kineeked da sa sal'a da. \p
atext14 12 lon diy'a si H'esus denu sa kinepatay i Huwan.
atext14 13 h i H'esus migdineg denu sa kinepatay i Huwan, eglegkang dahiya ow
atext14 19 en igbegay i Nemula. Agul'e kineb'eng-keb'eng di sa epan, owoy igb
atext15 36 'o kaenen igbegay i Nemula. Kineb'eng-keb'eng di owoy igbegay di d
atext21 35 da migtebow, medaet polo sa kinebael sa etaw tegeipat, en'u ka sin
atext21 36 Dodoo, hediya ma medaet sa kinebael da kenagda.
atext26 2 pa meuma sa pista sa agdaw kinetal'a sa kepigtamay i Nemula. Agul
atext26 17 igkaen Da Sa Pista Sa Agdaw Kinetal'a Sa Kepigtamay I Nemula \r (M
atext26 17 kaenen ta egpista sa Agdaw Kinetal'a Sa Kepigtamay I Nemula?>> \p
atext27 8 an'a sa pilak igbayad da sa kinepatay i H'esus. Huenan di, pinenga
atext27 15 man meuma sa Pista Sa Agdaw Kinetal'a Sa Kepigtamay I Nemula, iya
atext27 55 an'i eghauwen da sa langun kinebael da si H'esus. Kagda sa mig-un
atext27 59 e diy'a sa kayu bugsud owoy kinedked di sa magtu ginis mebul'a.
atext27 60 g sa lawa i H'esus. Agul'e, kinelid di sa dakel batu an'i mesagpen
atext28 2 tuntun kedu langit d'o owoy kinelid di sa batu igsagpeng lebeng, o
btext6 29 ek i Huwan migdineg denu sa kinepatay di, mig-angay da diy'a sa bi
btext6 41 n igbegay i Nemula. Agul'e, kineb'eng-keb'eng di sa epan owoy igbe
btext6 41 a langun etaw, Hediya ma sa kinebael di sa duwa sed'a.
btext7 25 busaw. Egoh di migdineg sa kinetebow i H'esus, medelamet eg-angay
btext7 33 angk'a diy'a sa medoo etaw. Kinesek di sa tosong telinga sa bisuwe
btext8 6 'o kaenen igbegay i Nemula. Kineb'eng-keb'eng di owoy igbegay di d

```

Total lines displayed are 23. Press ENTER to continue or FB to EXIT.

FIG. 4. — FIESTA concordance display.

about the units being studied. The resulting concordances may be copied to file or printed, but the strength of the program is user interaction with the data displayed on the screen.

## The CONC program

The CONC (Concordance) program provides fast and versatile concordance functions for the Macintosh computer (THOMSON, 1992). It is specifically designed to work with the annotated texts created by the Macintosh version of IT but can also process non-annotated texts.

CONC does not have editing capability but in all other ways is comparable to FIESTA (cf. the above "FIESTA program") in the concordance functions it provides.

The CONC program takes advantage of the windowing capability of Macintosh and displays the concordance and the text simultaneously in separate windows. Optionally, the index of words or phrases on which the concordance is sorted may be displayed in a third window. When an item is selected in either window, the other windows automatically adjust to show the same item.

CONC will produce phrase, word, morpheme, or character concordances and, like FIESTA, it permits the user to constrain concordance searches in almost any conceivable way.

At the present time, CONC is available as a pre-release test version.

## The IC program

The older IC (Interactive Concordance) program is a task-specific concordance program for DOS machines that provides some of the same capability as FIESTA. Its major drawback is that it processes the concordance searches in a linear fashion, making it much slower than FIESTA, especially with large databases. The speed comparison is on the order of one minute with IC to one second with FIESTA.

IC has several advantages, however. It requires a relatively small amount of disk space, it can process files of any size with any number of word forms, and it uses the special character handling features of the DTS package of programs (see paragraph "Displaying and printing special characters in the DOS environment" in this article) that permit the use of an unlimited number of special characters.

From an efficiency standpoint, IC is a good choice only if the linguist must process a large number of special characters or is prevented from using FIESTA by very large texts or disk space limitations in his or her computer.

IC is one of several programs incorporated in a package of programs under the name Text Analysis (TA-DOC 1992). All use the special character handling features of the DTS package. Other programs in the TA package generate word lists from texts (WDL), automate word segmenting in texts (WS), sort, filter, and format lexical files (SRT), and make reversals of lexical files (DICR). All of these functions are duplicated in the newer and more generalized FIESTA and SHOEBOX programs. These programs, however, are more restricted than the TA programs in terms of the number of special characters they can use (see below).

## LEXICONS

In studying a unwritten language, every linguist collects words of the language and creates some kind of lexicon. Often the lexicons are published in some form. From the early days of computers, lexical data was managed in a text file using a word processing program. More recently linguists have been using database programs to facilitate the retrieval and editing of lexical entries.

## Managing lexical data, the SHOEBOX program

The SHOEBOX database program (WIMBISH, 1990) facilitates the entry, storage, and rapid retrieval of lexical records. As the name SHOEBOX suggests, each record in the lexicon appears as a single screen and the records are manipulated like cards in a file box.

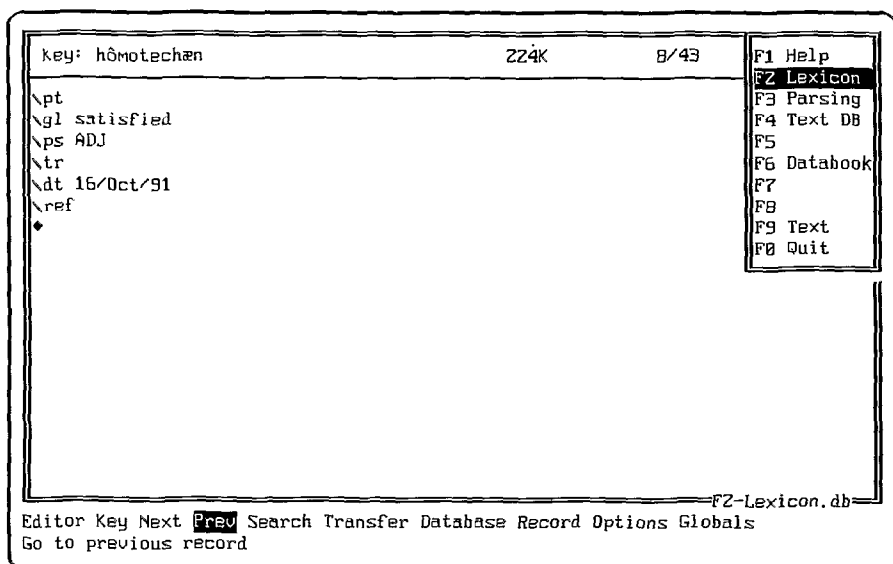


FIG. 5. — SHOEBOX main screen and sample record.

With SHOEBOX, the user may search for and retrieve lexical entries almost instantly. The number of entries in a single database is restricted primarily by how much disk space is available for the files. Also, unlike many commercial database programs, SHOEBOX is a free-form database that permits up to 5 K of data in a single record. It does not impose other restrictions on the number of fields in a single lexical record.

With SHOEBOX, the linguist may work with up to seven different databases at one time and jump from one to another as needed. This facilitates cross referencing. Data can be copied without restriction. Multiple copy buffers permit several items to be copied from different records before they are inserted elsewhere.

SHOEBOX permits grammatical and ethnographic outlines to be intermingled with lexical databases. Illustrative words, phrases, or sentences can be copied from lexical records or texts to the outlines without re-typing. In this regard SHOEBOX is a general data management program for entering and retrieving all kinds of field notes.

SHOEBOX databases can be created from the keyboard with one-by-one entry of lexical items or they can be created from lexical files created with a word processor or from annotated word lists such as those produced by the IT program.

SHOEBOX entries in the lexicon can be arranged in any alphabetical order specified by the user. Any database can be reordered at any time. The older program SRT in the TA package provides this same reordering function with lexical data in text files.

SHOEBOX has an internal filter function that enables the user to restrict the display to those records that meet a specified condition. For example, the user may design a filter that displays only those records marked "noun" in the part-of-speech field. More complicated filters involving multiple conditions are also possible. SHOEBOX filters may also be used to restrict data that is output to a file.

SHOEBOX permits the user to reverse the fields of a lexical database so that the data in a gloss field becomes the primary entry and the original lexical item becomes the gloss. For example, in a reversal, a lexicon of German words with French glosses becomes a French word list with German glosses. The older program DICR in the TA package provides this same function with lexical data in text files.

The SHOEBOX manual not only discusses program functions, but also provides help for the linguist in deciding what needs to be included in his or her annotation of lexical items and how the lexicon and field notes should be organized.

While FIESTA is primarily a text processing program, it can also process lexical files with records and fields when building its database. In contrast to SHOEBOX it does not lend itself to the addition of new records, that is, it is probably not the best choice for day-by-day entry of new lexical items. It does, however, permit unrestricted editing of existing records and has limited formatting capability for the draft-quality printing of the lexicon. It is also capable of a lexicon reversal with customized reordering of all fields in the records. FIESTA permits sorting of the lexicon on any field.

With lexical databases, the word list feature of FIESTA provides spell checking capability and rapid random searching of all fields of the lexicon. For these reasons FIESTA is a good choice for final editing of a lexicon for publication.

## Building lexicons and word lists from texts

There are several programs that generate word lists containing all the different word forms found in a set of computerized texts. In some cases the lists are annotated. This kind of list can be used to start or supplement a lexical database.

As text is annotated by the IT program (see p. 109), the words in the text and the annotations entered by the user are stored together in a special file created by the program. The lexical data in this file can be converted by IT to a plain file for use with other programs like SHOEBOX.

By design, SHOEBOX creates lexical entries any time it is being used to annotate texts. These lexical entries may be accessed by the user at any time and edited at will.

FIESTA maintains a word list for all the word forms in any database of texts. This word list is the primary element in the FIESTA program design. This word list may be output as a disk file for use with other programs such as SHOEBOX.

The WDL program which is part of the TA package will scan one or more text files and create a file containing a list of all the word forms that occur in the texts.

## TESTING A MORPHOLOGICAL ANALYSIS

The study of words and how they are formed from the combination of morphemes is basic to descriptive linguistics and to the study of any particular language. Computers make it possible to encode a language-specific morphological analysis and test it by applying it to a large inventory of words in the language.

Programs that implement morphological analysis in this way are called parsers or parsing programs because they break down words into their component parts. Depending on the analytical model of the parser, it may also offer a generating option in which phonologically correct surface forms of words are derived from their underlying morphemes.

Parsing programs do not formulate the analysis for the linguist but provide a means of testing analytical models and rules. The process of encoding the analysis for the parser, however, often produces new insights about the morphology. Then, as the program parses or generates a number of words in the language being studied, wrong parses or the generation of unacceptable words alerts the linguist to weaknesses in his or her analysis.

The goal, of course, in using a parsing program for field work is to arrive at the best possible morphological analysis for the language under study.

## **The AMPLE program**

AMPLE (A Morphological Parser for Linguistic Exploration) is a morphological parsing program based on a linear item and arrangement approach to word description (WEBER, BLACK and McCONNEL, 1988).

AMPLE facilitates the encoding of the linguist's morphological analysis based entirely on the surface forms of morphemes. The encoding data provided by the user takes the form of dictionaries of root and affix morphemes, their variants, and coded occurrence constraints for these morphemes. Constraints include such things as the order in which morphemes can occur in a word, the categories of roots with which affixes can occur, and restrictions of co-occurrence with other morphemes in the same word.

When encoding is complete, AMPLE implements this analysis by parsing the words of a text into their component morphemes. The extent to which the words of the text are correctly parsed by AMPLE indicates the accuracy of the morphological analysis. Also, the parsed text can be used as the starting point for the dialect adaptation process described below.

## **The PC-KIMMO program**

PC-KIMMO is a word parsing and generating program, that is, it can be used to break words into morphemes and to generate surface forms of words from underlying morphemes (ANTWORTH, 1990).

PC-KIMMO is an implementation for microcomputers of a program invented by Kimmo Koskeniemi, a Finnish computational linguist. He developed his word structure model and parsing program because other approaches failed to address adequately the complex morphology of the Finnish language.

PC-KIMMO is based on Koskeniemi's two-level model of word structure that encodes a phonological correspondence between underlying lexical forms and surface forms of morphemes.

With PC-KIMMO the linguist encodes his analysis in the form of a lexicon of morphemes with constraints on their order of occurrence and a table of rules related to phonology and spelling. In addition to the obvious uses of the program, the encoding process forms a framework to help the linguist formulate his analysis more precisely.

PC-KIMMO can be used to parse one word at a time entered from the keyboard or all the words in a given test list. It does not include facility for processing running text but a supplementary program KTEXT (ANTWORTH and McCONNEL, 1991) allows the PC-KIMMO model to be applied to texts with output files very similar to those produced by AMPLE.

### **AMPLE and PC-KIMMO compared**

AMPLE requires the cataloging of all surface forms of morphemes in the lexicon while the phonological rules in PC-KIMMO can recognize multiple surface forms for one underlying morpheme form. In languages with many morpheme variations, PC-KIMMO is more efficient in that its morpheme lexicon needs many fewer entries.

On the other hand, the PC-KIMMO model is limited to specifying morpheme co-occurrence constraints in terms of their linear order in words and does not handle well co-occurrence constraints for morphemes that are separated by other morphemes.

In general, PC-KIMMO is best suited for languages with complex phonological variations in morphemes while AMPLE is better suited for languages with complex combinations of morphemes and less phonological variation.

### **DIALECT ADAPTATION OF VERNACULAR TEXTS**

Linguists in the field who help produce reading materials in vernacular languages are always concerned about getting maximum output for their efforts.

In many parts of the world there are languages and dialects that are phonetically and (or) grammatically similar in predictable ways, yet different enough that a reader of one language cannot readily understand texts produced in the other. In such situations, linguists have seen the potential of using the computer to apply a set of transforming rules to a text in one language to automatically produce a readable text in a second language.

In SIL, this process is called CADA (Computer Assisted Dialect Adaptation).

Most applications of CADA have involved languages where the predictable differences were primarily at the word level. In the adaptation process, the linguist uses a morphological parsing program such as AMPLE or KTEXT to break down the words of the source language



text into their component morphemes and produce a parsed text. Then he or she uses a supplemental program called STAMP (WEBER, MCCONNEL, BLACK and BUSEMAN, 1990) to apply to the parsed text a set of transforming rules and a table of morpheme equivalences for the two languages. STAMP stands for Synthesis and Transfer for AMPLE. The output of STAMP is an adapted text in the related language.

The CADA process assumes that the original text is well-formed in the first language. It requires a careful morphological analysis of both languages and produces only a rough working draft in the second language. This draft must be carefully edited with the help of speakers of the second language to arrive at a text that is correct and natural.

The use of CADA assumes that manual adaptation of the texts would be less efficient than the effort required to encode in the AMPLE and STAMP programs the morphological analyses of the related languages. The advantage of using CADA is greater when the number of texts to be adapted is large and when the process can be applied to more than one target language.

The CADA process is well documented (CADA-T, 1992) in the manuals provided with AMPLE and STAMP but the documentation and the process assume more than an entry-level understanding of descriptive linguistics and the use of a computer. A serious application of CADA in the field is likely to require consultant help.

## GRAMMATICAL DESCRIPTIONS

The linguist in the field will normally describe in a written document the result of his or her analysis of the grammar and other features of the language under study. The computer not only helps with the composition and formatting of such documents but permits the user to access and copy illustrative material from existing text and lexical files.

The SHOEBOX program is especially suited for this purpose. It can be used in two different ways.

The grammar description can be outlined and written in draft form inside SHOEBOX. This allows easy access to other SHOEBOX databases for copying annotated sentences to illustrate various aspects of the analysis. This document can then be output from SHOEBOX and formatted for publication with a word processing program.

Alternately, the grammatical description can be written and formatted with a commercial word processor. The filter function of SHOEBOX can be used to find annotated sentences that illustrate certain aspects of the analysis. These sentences may be output from SHOEBOX to a

plain file and inserted at the appropriate places in the grammatical description. This is a simple cut and paste function.

FIESTA has similar facility to search for and output to file selected sentences in annotated or plain texts.

ROOK is a program for the Macintosh computer that facilitates the writing of a grammatical description with the commercial HyperCard<sup>8</sup> program (VALENTINE, 1991). ROOK takes advantage of the windowing, linking, and indexing functions of HyperCard to integrate the description with the data in annotated texts prepared by IT. ROOK is available in pre-release form and the user must expect uncertainties in its use.

ROOK, IT and CONC complement each other to facilitate linguistic work with Macintosh computers.

#### DISPLAYING AND PRINTING SPECIAL CHARACTERS IN THE DOS ENVIRONMENT

Software has been developed by SIL to implement the keyboarding, displaying and printing of special characters for computers that use the DOS operating system. These special characters include characters from the phonetic alphabet and custom-designed alphabetic characters for roman and non-roman writing systems.

Regardless of the approach used, the steps include using the following kinds of programs: 1. a program for designing the shapes or dot patterns of the special characters, 2. a program for converting (compiling) these character shape files to a file form which can be interpreted by the screen and (or) printer, and 3. a program for instructing the screen and/or printer to use the special characters.

The use of special characters and character sets is implemented in three different ways.

First, a group of special SIL programs permits the user to define type, display, edit, and print special characters in a way that requires no special hardware compatibility in the screen or printer. This approach is embodied in the program files and documentation of the DTS (Direct Translator Support) package (DTS-DOC, 1992). Unlike other approaches to special characters with MS-DOS, the DTS package has no limit other than hardware constraints on the number of special characters that can be used at one time.

<sup>8</sup> HyperCard is a registered trademark of Apple Computer, Inc., licensed to Claris Software Corporation.

Secondly, a group of SIL programs permits the user to define special characters and “download” them to the screen and (or) printer for use with programs like IT, SHOEBOX, and FIESTA that work with plain text files. In this approach the printer must be capable of accepting and implementing downloaded fonts. The number of special characters is constrained by the 256 limit on available ASCII codes.

Finally, a group of SIL programs that work in connection with the MS-WORD<sup>9</sup> program permit the display and printing of special characters. This approach utilizes the same kind of downloaded characters as above but inserts and correlates them in special ways with MS-WORD screen and printer files. As above, the total number of characters displayed and printed is limited to 256.

### **The DTS approach to displaying and printing special characters**

The original SIL approach to special characters in the DOS world is incorporated in the DTS package (DTS-DOC) and based on the ED program, an editor for plain ASCII files. With appropriate auxiliary files, ED can display an unlimited number of special characters, it can accept various keystroke combinations to access and display these characters, and it can process text data in a right-to-left mode to accommodate other writing systems.

The companion program MS (Manuscripter) is a text formatting and printing program designed to use the text files created by ED and to implement the printing of special characters designed for ED. Together ED and MS provide a complete editing and printing capability for complicated writing systems.

The major disadvantage of this approach is that text files with special characters produced by ED are not compatible with other programs such as IT, SHOEBOX, FIESTA, or MS-WORD. That is, the data can be used with these programs but sets of multiple keyboard characters will appear in place of each special character.

<sup>9</sup> MS-WORD is a registered trademark of the Microsoft Corporation. In SIL, the MS-WORD program has become something of a standard as a word processing and formatting program for linguistic articles and other writing for publication. The reason is that MS-WORD stylesheet formatting uses a text unit marking system that parallels closely the standardized way vernacular language texts are marked throughout SIL. The SIL system uses Standard Format Markers (SFM) with plain ASCII files. Utility programs have been developed to convert texts back and forth between the MS-WORD and SIL formats. For a full discussion of this subject, see KEW and SIMONS (1989).

The DTS approach to special characters, however, is fully incorporated in the concordance, word list, word segmenting, and dictionary programs included in the TA package (see "The IC program").

The following utility programs in the DTS package provide support for the ED and MS approach to special characters: 1. the DESIGN program is an editor for creating dot patterns (character shapes) for special characters and storing them in a computer file; 2. the SHAPES program uses the file created by DESIGN to make (compile) a character screen file that ED can access whenever these characters are called from the keyboard by a designated sequence of keystrokes; and 3. the FONT-CO program uses another kind of output file from SHAPES to make a file that MS uses to interpret the special characters for the printer.

For typing special characters, ED utilizes a built-in function to correlate multiple keystrokes with the special characters as defined in the ED character file.

### **The downloading approach to displaying and printing special characters**

The programs FIESTA, IT and SHOEBOX require a different approach to displaying and printing special characters. These programs are designed to work only with characters accessed by the 256 ASCII codes accessible in the DOS environment.

If the linguist wishes to display and print special characters, he must use utility programs to get the computer screen and printer to recognize his or her characters and correlate them with ASCII codes he enters from the keyboard. This process is called downloading.

#### **Downloading special characters to the screen**

In addition to the standard characters that appear on the keyboard, the DOS operating system provides an additional 128 characters in what is called the extended IBM set. These characters do not appear on the keyboard but can be accessed and displayed on the screen by typing their ASCII code numbers on the numeric keypad. In some cases these extra characters provide all the special characters needed by a language.

More commonly, however, languages require that all or some of the characters accessed by the ASCII codes from 128 to 254 be redefined for the screen by a process called "downloading" so that special characters may be displayed.

An obvious disadvantage of this approach is that the shape of each special character must be assigned to just one ASCII code number. This means codes for base characters and diacritics cannot be combined and every character variation must have its own ASCII number. This limits the number of special characters to 128 in addition to the standard keyboard characters.

The DESIGN and SHAPES programs can be used to create the special characters and produce a file for downloading the characters to the screen. The download file correlates the new character shapes with assigned ASCII codes.

Depending upon the type of screen, the programs CGALOAD or LDVEGA use the screen download file to redefine for the screen the dot patterns that are displayed for the assigned ASCII code numbers. When the download is implemented, the original screen characters accessed by these ASCII codes are temporarily replaced by the new characters.

### Simplifying the typing of special characters

To simplify the typing of downloaded special characters or other characters from the extended set, the KEYDEF program permits the user to define multiple keystroke combinations and correlate each one with a selected ASCII code. For example, in a keyboard file created by the user, he or she might decide to correlate the key sequence 'a (apostrophe followed by lower case a) with ASCII code 160 which displays and prints as á.

The program KEYSWAP uses the keyboard file created with KEYDEF to translate keystroke combinations into a single ASCII code for the computer and the screen. Using the keyboard file described in the example above, the user sends ASCII code 160 to the computer whenever he or she types 'a. The ' and the a do not appear on the screen. Rather, the character á appears on the screen.

In this way the user can type any of the special characters on the screen by typing the corresponding sequence of keystrokes. Using a special keystroke combination, the user can turn off the special keyboard file if it is not being used. It can be turned on again when needed. Also, when KEYSWAP is used, it can access up to nine different keyboard files. The user can switch from one to another as needed.

### Downloading special characters to the printer

When programs like FIESTA, IT and SHOEBOX are used with spe-

cial characters downloaded to the screen, they output plain ASCII files in which the special characters appear as ASCII codes from the extended IBM set. The printer must be programmed to correlate these ASCII codes with the special characters designed by the user.

For the screen and printer to work together correctly, the correlations of special character shapes and ASCII codes in the screen download file must be matched by identical correlations in a printer download file. Otherwise, the printer will not produce a text comparable with what is displayed on the computer screen. Also, the printer must have a memory buffer that is capable of accepting the downloaded fonts. Most newer printers have this capability.

The printer downloading process is similar to that used for the screen. The special characters are created with the DESIGN program. The SHAPES program then compiles and creates a printer download file for the special characters which correlates the special character shapes and ASCII codes. This download file can be designed for Toshiba<sup>10</sup> and Epson<sup>11</sup> dot matrix printers or for ink jet and laser printers.

In the case of dot matrix printers, no special printing program is needed to implement the download of special characters for the printer. In the case of ink jet and laser printers, however, the printing program PRINTF must be used. First, a font set which includes all characters to be printed must be created. The font set is produced by the SHAPES and FONTCO programs in the form of a compiled font file and a printer download file. The program PRINTF uses these two files to download the font and print the text containing the special characters.

### **The MS-WORD approach to displaying and printing special characters**

In SIL, the commercial MS-WORD program (versions 5.0 and 5.5) has become something of a standard as a word processing and formatting program for linguistic articles and other writing for publication. The reason is that MS-WORD stylesheet formatting uses a text unit marking system that parallels closely the standardized way vernacular language texts are marked throughout SIL. Utility programs have been developed to convert texts back and forth between the MS-WORD and SIL formats. For a full discussion of this subject, see KEW and SIMONS (1989).

The DESIGN and SHAPES programs discussed above can be used to prepare special character definition files for use in MS-WORD. The

<sup>10</sup> Toshiba is a registered trademark of Toshiba Corporation.

<sup>11</sup> Epson is a registered trademark of Epson America, Inc.

KEYDEF and KEYSWAP programs also work effectively with MS-WORD. The downloading of special characters to EGA and VGA screens for use in MS-WORD is the same as described above for FIESTA, IT and SHOEBOX.

For using MS-WORD with CGA and Hercules<sup>12</sup> video cards, as well as EGA and VGA, the FONTSHOP and IMPLANT programs (Loos, 1989, 1990) provide an alternate approach to the design and display of special characters. Also, with these video cards, the MODIFY program (MATHESON, 1991) implements special characters created with the DESIGN and SHAPES programs of the DTS package.

Printing special characters with MS-WORD, however, involves modifying the printer files (PRD files) that are used by MS-WORD. For users with more than average programming knowledge, this can be done manually with the MS-WORD auxiliary program MAKEPRD.

To print special characters the PRD file for MS-WORD can be modified in the following ways: 1. each special character can be defined with one or two characters and printer commands to cause the printer to backspace; 2. each special character can be defined as a printer code that correlates with characters that have been downloaded to the printer as described above; and (or) 3. each special character can be defined with a printer command sequence that switches the printer to graphics mode, sends the specified dot pattern (graphic image) of the character to the printer, and turns off the graphics mode so that the printing of built-in printer characters is resumed.

MATHESON (1991) has created a set of batch files for automating the modification of PRD files for Epson LQ Series 24-pin printers.

## SUMMARY AND INDEX OF SIL PROGRAMS FOR LINGUISTIC FIELD WORK

### Programs for DOS-based computers

#### **CADA (Computer Assisted Dialect Adaptation) package**

AMPLE	Morphological parser for texts (5.1, 6.0)
STAMP	Implementing morphological dialect adaptation (6.0)

<sup>12</sup> Hercules is a registered trademark of Hercules Computer Technology, Inc.

**DTS (Direct Translator Support) package**

ED	Editor with features for keyboarding and display of special characters and right to left text flow (8.1)
MS	Formatting and printing plain ASCII files created by ED and implementing special characters (8.1)
PRINTF	Printing plain ASCII files with special characters but without formatting (8.2.3)
DESIGN	Creating special character shapes (8.1, 8.2.1, 8.2.3, 8.3)
SHAPES	Compiling screen and printer download files for special characters (8.1, 8.2.3, 8.3)
FONTCO	Compiling special characters for printing with MS, PRINTF, and programs in the TA package (8.1, 8.2.3)
KEYDEF	Creating keyboard files for typing special characters (8.2.2)
KEYSWAP	Implementing keyboard files (8.2.2)

**TA (Text Analysis) package (3.2.3)**

ANSQ	User interface
DICR	Reversing dictionaries
IC	Interactive concordance program (3.2.3)
SRT	Sorting lexical files
WDL	Creating word lists from texts (4.2)
WS	Segmenting words in texts

**Other SIL programs**

CECIL	Speech analysis program and acoustic-digital converter box, includes the SPECTRUM program (2.2)
CGALOAD	Downloading to CGA screens (8.2.1)
EPSON	Batch files for modifying Epson LQ printer files for special characters with MS-WORD (8.3)
FIESTA	High speed editor and concordance program (3.2.1)
FINDPHONE	Search and display program for phonetic word lists—includes editor and IPA character designators (2.1)
FONTSHOP	A set of files for creating and downloading special character shapes for use with MS-WORD and CGA, Hercules, EGA, and VGA screens (8.3)
IMPLANT	A program for downloading special characters to CGA and Hercules screens with MS-WORD (8.3)
IT	Produces annotated interlinear texts (3.1.1)
ITF	Formats interlinear text for printing with TEX (3.1.3)



KTEXT	Implements PC-KIMMO for parsing text files (5.2, 6.0)
LDVEGA	Downloading special characters to VGA and EGA screens (8.2.1)
MODIFY	A set of files for modifying MS-WORD SCREEN.VID files for display of special characters on CGA and Hercules monitors (8.3)
PC-KIMMO	Morphological word parser and generator (5.2) for displaying special characters (8.3)
SHOEBOX	Data management program for lexicons and field notes— also produces annotated interlinear texts (3.1.2)
WORDSURV	Statistical comparing of word lists between languages and dialects (1.0)

### Programs for Macintosh-based computers

AMPLE	Morphological parsing of texts (5.1, 6.0)
CONC	Text concordances (3.2.2)
IT	Annotating interlinear texts (3.1.1)
ITF	Formatting interlinear text for printing with TEX (3.1.3)
KTEXT	Implements PC-KIMMO for parsing text files (5.2, 6.0)
PC-KIMMO	Morphological word parser and generator (5.2)
ROOK	Grammatical outline for HyperCard database (7.0)
STAMP	Implementing morphological dialect adaptation (6.0)

### Programs for UNIX-based computers

For UNIX applications only, address inquiries to:

SIL	Academic Computing, 7500 W. Camp Wisdom Rd, Dallas TX 75236
AMPLE	Morphological parsing of texts (5.1, 6.0)
ITF	Formatting interlinear text for printing with TEX (3.1.3)
KTEXT	Implements PC-KIMMO for parsing text files (5.2, 6.0)
PC-KIMMO	Morphological word parser and generator (5.2)
STAMP	Implementing morphological dialect adaptation (6.0)

## BIBLIOGRAPHY

- ALSOP (J.) and JOHNSTON (C.), 1990. — *FIESTA: Fast Interactive Editor of Scripture and Text Analysis, a program for translators and linguists (version 4.5)*. Dallas, Summer Institute of Linguistics. 141 p. [with program disk, also available from Waxhaw].
- ANTWORTH (E.), 1990. — *PC-KIMMO: A Two-level Processor for Morphological Analysis*. [Occasional Publications in Academic Computing 16], Dallas, Summer Institute of Linguistics. xii, 273 p. [with program disk, specify DOS, Macintosh or UNIX—DOS and Macintosh versions also available from Waxhaw].
- ANTWORTH (E.), and McCONNEL (S.), 1991. — *KTEXT User's Guide*, [Documentation file on disk with program specify DOS or Macintosh—UNIX version available at Dallas], Waxhaw.
- BEVAN (D.), 1991. — *FIND PHONE programs and documentation* (Cat.no.DOS-SIL054A), Waxhaw.
- CADA-T. 1992. — *Tutorial/Manual for Ample and Stamp*, Waxhaw.
- CECIL. 1992. — *Speech Analysis System*, [Hardware and software including the SPECTRUM program], Waxhaw.
- CGALOAD. — Program for downloading special characters to CGA screens. Waxhaw.
- DESIGN. — Program for creating special character shapes, supplied with the DTS package of programs (see DTS-DOC).
- DTS-DOC, 1992. — *Direct Translator Support*. Programs and documentation, Waxhaw.
- Ed. Editor for plain ASCII files, supplied with the DTS package of programs. (see DTS-DOC).
- IC. — Interactive concordance program, supplied with the TA package of programs, (see TA-DOC).
- KEW (P.) and SIMONS (G.), ed., 1989. — *Laptop Publishing for the Field Linguist: an approach based on Microsoft Word*, [Occasional Publications in Academic Computing 14], Dallas, Summer Institute of Linguistics. viii, 147 p. [with program disk, also available from Waxhaw].
- KEW (J.) and McCONNEL (S.), 1990. — *Formatting Interlinear Text*, [Occasional Publications in Academic Computing 17], Dallas, Summer Institute of Linguistics. x, 153 p. [with program disk, specify DOS, Macintosh, or UNIX, requires additional TEX program—DOS and Macintosh versions also available from Waxhaw].
- KEYDEF. Program for creating keyboard files for typing special characters, supplied with the DTS package of programs. (see DTS-DOC).
- KEYSWAP. Program for implementing keyboard files for special characters, supplied with the DTS package of programs. (see DTS-DOC).
- GRIMES (B.), ed., 1988. — *Ethnologue: Languages of the World, Eleventh Edition*, Dallas, Summer Institute of Linguistics, 758 p. [1992 edition available soon].
- LANCASHIRE (I.), ed., 1991. — *The Humanities Computing Yearbook, 1989-90*, Clarendon Press, Oxford. 700 p.
- LDVEGA. Program for downloading special characters to VGA and EGA screens. Waxhaw.
- LOOS (E.), 1989. — *FSCGA—Fontshop for MS-WORD and CGA cards and HERCULES cards in CGA mode*, Programs for creating and downloading special screen characters. Waxhaw.
- LOOS (E.), 1989. — *HGC—Fontshop for HERCULES monochrome and HERCULES Graphic Card Plus*, Programs for creating and downloading special screen characters including EGA and VGA formats. Waxhaw, (Updated 1992 and included with IT version 1.2).

- LOOS (E.), 1990. — *IMPLANT.EXE*, Downloads special characters to CGA and Hercules screens with MS-WORD. Waxhaw.
- MATHESON (C.), 1991. — *MODIFY and MODIFY55*, Modifies the MS-WORD 5.0 or 5.5 SCREEN.VID file so that special characters designed for downloading can be used with CGA and HERCULES screens. Waxhaw.
- MS. — Manuscripter program for formatting and printing, supplied with the DTS package of programs (see DTS-DOC).
- PRINTF. — Printing program for special characters in plain ASCII files but without formatting, supplied with the DTS package of programs (see DTS-DOC).
- SHAPES. — Program for compiling font files for downloading to the screen and/or printer. Uses characters created with the DESIGN program. Supplied with the DTS package of programs (see DTS-DOC).
- SIMONS (G.) and VERSAW (L.), 1987. *How to use IT: A Guide to Interlinear Text Processing*, Dallas, Summer Institute of Linguistics. 372 p. [with DOS program disk].
- SIMONS (G.) and THOMSON (J.), 1988. — *How to use IT: Interlinear Text Processing on the Macintosh*, Edmonds, WA: Linguist's Software. 363 p. [with program disk].
- SIMONS (G.), THOMSON (J.) and DeROSE (S.), 1988. — *Computing Environment for Linguistic, Literacy, and Anthropological Research, a project proposal*. Manuscript, available from SIL Academic Computing Dept., Dallas.
- SPECTRUM. — Program for displaying utterance data as speech spectrograms, supplied with CECIL package (see CECIL).
- TA-DOC, 1992. — *Text Analysis*, Package of programs and documentation. Waxhaw.
- THOMSON (J.), 1992. — *CONC*, Concordance program for the Macintosh. Waxhaw and Dallas.
- VALENTINE (R.), 1991. — *ROOK*, Grammatical outline for the Macintosh Hypercard program. Waxhaw.
- WEBER (D.), BLACK (A.) and McCONNEL (S.), 1988. — *AMPLE: A Tool for Exploring Morphology*. [Occasional Publications in Academic Computing 12], Dallas, Summer Institute of Linguistics. xi, 252 p. [with program disk, specify DOS, Macintosh, or UNIX, DOS and Macintosh versions also available from Waxhaw].
- WEBER (D.), McCONNEL (S.), BLACK (A.) and BUSEMAN (A.), 1990. — *STAMP: A Tool for Dialect Adaptation* [Occasional Publications in Academic Computing 15], Dallas, Summer Institute of Linguistics. xiv, 215 p. [with program disk, specify DOS, Macintosh, or UNIX, DOS and Macintosh versions also available from Waxhaw].
- WDL. — Word list program, supplied with the TA package of programs (see TA-DOC).
- WIMBISH (J.), 1989. — *WORDSURV: A Program for Analyzing Language Survey Word Lists*, [Occasional Publications in Academic Computing 13], Dallas, Summer Institute of Linguistics. vii, 108 p. [with program disk, also available from Waxhaw].
- WIMBISH (J.), 1990. — *SHOEBOX: A Data Management Program for the Field Linguist (version 1.2)*, Ambon, Summer Institute of Linguistics and Pattimura University. 208 p. [with program disk, available from Dallas or Waxhaw].

*Sources for listed materials:*

- Dallas: Summer Institute of Linguistics, International Academic Bookstore, 7500 W. Camp Wisdom Road, Dallas TX 75236 (214)709-2400.
- Waxhaw: SIL Software Library, JAARS, Box 248, Waxhaw NC 28173 (704)843-6000.
- Edmonds: Linguist's Software, Inc., Box 580, Edmonds WA 98020-0580 (206)775-1130.