FieldWorks Language Explorer 1.0

*Ronald Moe**

SIL International released version 1.0 of the new FieldWorks Language Explorer in November 2006. Language Explorer is the latest in a series of SIL programs (Ed, CC, Shoebox, MDF, LinguaLinks, Toolbox, DDP, Lexique Pro) designed to assist linguists in collecting, managing and publishing linguistic data. Language Explorer is part of the FieldWorks suite of programs, which provides an integrated system for doing linguistic and anthropological field work. The November 2006 version of FieldWorks as a whole is 4.0. Language Explorer is designed to create and manage a dictionary, create and maintain a text corpus, interlinearize texts, and study morphology. Other components enable you to do other tasks such as maintain anthropological field notes or analyze phonology.

Language Explorer is an entirely new program with new and improved features, but builds on the functionality and look of earlier programs. It has a user interface that is similar to Shoebox and Toolbox with an edit view for editing a single entry and a browse view for displaying the database in a chart format. Language Explorer is similar to LinguaLinks in that it is an object-oriented, relational database. This allows such features as bi-directional links between entries. By way of contrast Shoebox only allows one directional pointers in cross-reference fields. In Shoebox, you can add a cross-reference field to an entry, but there is no guarantee that the other entry actually exists. If you want to create reciprocal pointers, you have to add a field to both entries. In Language Explorer the user links the two entries together, thus ensuring that both entries exist and creating a cross-reference in both entries in a single step. There are many other advantages to object-oriented, relational databases, such as data normalization. For instance you can change the abbreviation for ‘pronoun’ from ‘pron’ to ‘prn’ in one place and it will be consistently changed everywhere in the project.

Developing and editing a dictionary often involves processes that must be repeated for each entry. So it is an advantage to be able to efficiently make a change to all the entries in a single step. For instance with Shoebox you could exit the program, load the database into a text editor or run a CC table on it, then load the database back into Shoebox. This was risky, because you could easily corrupt the database. It was also clumsy and required good computing skills. But many users liked this power and freedom. Unfortunately one disadvantage to an object-oriented,

---

* Ron Moe and his wife Beth spent several years in the Philippines working with the Maguindanaon language project. He now serves as a linguistics consultant with the Uganda Tanzania Group.

SIL Forum for Language Fieldwork 2008-011, September 2008

relational database such as FieldWorks is that the user cannot directly manipulate the database outside of the program.

To compensate for this lack, Language Explorer has very powerful built-in editing tools. You can edit some fields in Browse View. You can sort fields from the beginning or the end of the words. You have powerful filtering tools, including the ability to use Regular Expressions in filters. The editing tools enable you to make systematic changes to a field throughout the database. You have a Find/Replace tool. You can copy the contents of one field to another and then modify the copy in the second field. You can copy a single word from one field to another with a single click of the mouse, for instance to create a single word gloss from your definition field. For the more technically oriented, you can also run a CC table, an ICU Unicode to Unicode transducer, TECkit, or Regular Expression on a field. For instance these tools enable you to copy the contents of the Lexeme Form field into the Pronunciation field and convert the orthographic characters into IPA, all in a single step. The editing tools have a preview feature that enables you to see what changes are going to be made and enables you to confirm the change on a field by field basis.

The advantage of these tools is that you have the same power that you used to have, but without the risk of mangling your database. In fact these tools are so powerful and easy to use that you can develop your database in a fraction of the time you could in other dictionary software programs, and do so with far greater consistency and accuracy. (I will admit that I may be a bit prejudiced in my opinion, since I helped design the tools.)

One of the best features of Language Explorer is the distinction it makes between data and presentation. Language Explorer is designed to capture lexical data in a standardized structure, but to present it for view or publication in a variety of formats. The structure of lexical data has been hard-wired into the program. In contrast Shoebox enforced no structure on the data (other than Standard Format Markers), resulting in databases with a variety of ad hoc structures and ad hoc fields. MDF was created to help solve this problem. Although it provided a standard, the user was not forced to maintain the standard. In fact few people are rigorous enough to maintain the standard even when they try their hardest. We all make typing mistakes. This problem is now a thing of the past, because Language Explorer provides a built-in structure. The user no longer needs to worry about keeping the fields in the right order and is free to concentrate on the data itself. All fields are available in all entries. If you don't want to use a particular field, you can hide it from view. But it is always there in case you need it.

The designers of the program have worked hard to create a theoretically sound structure. Does this mean that you have no freedom to create specialized dictionaries? Yes and no. By distinguishing between data and presentation, Language Explorer enables you to customize which fields get printed and how they will look. You can also create user-defined fields to handle special needs. This enables you to publish a variety of products from a single database. For instance you can create a monolingual dictionary, bilingual dictionary, and a trilingual vocabulary, all from the same database. The program achieves this by enabling you to describe the vernacular words in more than one language. For instance you can enter a definition in the vernacular in order to produce a monolingual dictionary. But you can also enter a definition in the national language and in one or more international languages in order to produce a bilingual (or multilingual) dictionary.

However Language Explorer is not an unconstrained database manager like Shoebox. You cannot maintain your address list in Language Explorer like you can in Shoebox. Language Explorer is designed to describe a single language. You can describe the vernacular language in multiple analysis languages, but you cannot describe more than one vernacular language in a single project. This means you cannot use Language Explorer to create a comparative dictionary in which you describe many languages in a single work. For special data structures you need to use an unconstrained database manager like Shoebox. You also cannot create a full bilingual dictionary, for instance a Japanese-English and English-Japanese dictionary in which the Japanese words are described in English and the English words are described in Japanese. To do this you would have to create two projects in Language Explorer, one for Japanese and another for English. Language Explorer can handle multiple dialects of a single language, but only by treating words from the other dialects as variants. Language Explorer is restricted to a single language because only by restricting it in this way can the data structure be standardized. By standardizing the data structure the program is able to achieve far more power than it could if the data were unconstrained. Paradoxically, constraints result in increased freedom and power.

Although Language Explorer standardizes the data itself, it enables you to view the data in multiple ways:

- The Lexicon Edit view (that you would use to edit a single entry) actually combines three different views. It has an edit pane similar to the Shoebox edit view, a browse pane similar to the Shoebox browse view, and a dictionary preview pane that shows how the entry will look in print.
- There are four full browse views. One is simply a full screen browse view, the second enables you to use the powerful editing tools to work on entry level fields, the third enables you to work on sense level fields, and the fourth enables you to work on reversal index entries.
- There are three publication views (similar to a print preview). One shows how the dictionary will look as an alphabetized dictionary. (You can modify how the data will actually print, but this gives an idea how it could look.) One shows how the Reversal Index (finder list) will look. (You can produce an index for each analysis language.) And one shows how the database will look as a classified dictionary. (You must first classify each sense according to semantic domain.)
- There is also a special Categorized Entry tool that enables you to efficiently enter words that have been collected using the Dictionary Development Process (DDP) list of semantic domains. In fact DDP has been integrated into Language Explorer for those wishing to use that approach.

Some languages are written using more than one script. So the program enables you to record vernacular language data in multiple scripts. All data is Unicode. The program supports left to right and right to left writing systems, including complex non-Roman scripts. Keyman works automatically with FieldWorks to switch keyboards.

Language Explorer enables you to create a text corpus, either by typing texts directly into the program, or by cutting and pasting them into an empty text record. The program automatically generates a list of unique word forms from the text corpus. There is a concordance feature that displays each occurrence of a word form from the text corpus in column view, along with the wider context of a selected occurrence. There is a built-in parser that can be used to analyze a

single word form or to facilitate interlinearizing texts. The interlinearizer handles the analysis of single words and has a rudimentary way to handle idioms. (This feature of the program is being developed.) The parser enables you to build up your understanding of the morphology of the language, including allomorphs and conditioning environments. The program will also generate a grammar sketch by gleaning information from the database.

Importing data from Shoebox/Toolbox or LinguaLinks is a bit tricky. It is necessary to get your Shoebox data into MDF format before importing it. This can be a problem if you have not stuck with MDF. Finding and fixing the errors in a Shoebox database can take anywhere from several hours to several days, depending on the number and nature of the errors and your computer skills. (These errors should be fixed anyway, since they will cause errors when printing with MDF or LexiquePro.) Since FieldWorks requires Unicode, you may also need to run some kind of conversion tool on your data to convert it to Unicode. Fortunately FieldWorks comes with some standard converters.

Currently Language Explorer only has rudimentary print functions, but you can export your data in several formats. Exporting your data to Lexique Pro requires you to go through the Lexique Pro import process as you would with any database. This took me several minutes. You only need to do this once. After that you can continue editing in Language Explorer. It is then a simple step to re-export your dictionary from Language Explorer and load it into Lexique Pro.

Many additional features have been planned for Language Explorer, but have not yet been implemented. However the program is robust and stable enough to warrant release as version 1. Because Language Explorer is still under development, you will encounter an occasional crash. However the program saves everything you do, so you do not lose data when the occasional crash does occur. In my opinion, Language Explorer has many advantages over Shoebox, LinguaLinks, or LexiquePro for developing a dictionary. As the programmers continue to add new features, it can only get better.

Language Explorer is currently a bit slow in switching from one view to another. This problem is gradually diminishing as the programmers solve problems and optimize performance. This slowness is more than compensated for by the speed of the editing tools. The tools enable you to add, develop, and edit fields at speeds 10 times, 100 times, and sometimes 1,000 times faster than you could in any other program. As the programmers have time, more tools, features, and options are being added.

In recent months I've taught three lexicography workshops using FieldWorks. (The last workshop was videoed and will be available soon.) I received some very positive feedback concerning the program. One lexicography consultant said, "This is the way to do dictionaries." A linguistics consultant said, "There is no way I would go back to LinguaLinks or Shoebox."

The best feature of FieldWorks is that it is absolutely free! You can learn more (and download the program) on the SIL website: http://www.sil.org/computing/fieldworks/index.html

Ron Moe

SIL Linguistics Consultant

(Developer of the Dictionary Development Process and one of the lexicography consultants for the FieldWorks team)■