# Matapuna Dictionary Writing System

## from Thinktank Consulting Limited

Reviewed by Oumar Bah, *Linguist and lexicographer, Vienna*

**1. OVERVIEW.** Matapuna, from a Māori word meaning 'source', is an open-source Dictionary Writing System (DWS) developed by Dave Moskovitz of Thinktank Consulting Limited in collaboration with the Māori Language Commission of New Zealand. It is a DWS much like Toolbox, iLex, or TshwaneLex, but unlike the others, it is entirely web-based. This means that its interface can be opened with any usual browser, which is a huge advantage if we consider that most users are now familiar with the Internet. Furthermore, its network feature makes it well-suited for collaborative dictionary projects, either in a Local Area Network or on the Internet. In the latter case, you and your team can access the project and do the work from any part of the world. Above all, Matapuna is a multi-user DWS, meaning that different users can log in simultaneously and work on the same project. Thanks to the self-archiving function of the system, all changes to the entries are recorded with the username of each author and the date of change. Because the system is highly flexible, changes in the data structure that are made in a later stage of a project will apply to the whole database. This is not true of some dictionary software. For example, in Toolbox, changes in the template apply only to subsequent entries, not the preceding ones.

Matapuna has few technical requirements. If you run your project locally on your computer, you will need to install a Linux server as your operating system and at least 500 MB to 1 GB of RAM memory (today even most cheap computers have at least 1 GB RAM). Please note that typically you can download Linux-based server software free of charge.[1] If your project is hosted on another machine, you need nothing but a web browser. In such a case, the current operating system on your computer should be irrelevant.

Matapuna can be used to compile bilingual as well as monolingual dictionaries. My own project is a monolingual dictionary of Fula (also called Fulfulde in the literature), a West African language. At present, there are only two interface languages—Māori and English. However, the system can easily be translated into any other language (Dave Moskovitz, pers. comm.).

**2. DOWNLOAD AND INSTALLATION.** Matapuna can be downloaded from a repository such as the one linked by the SourceForge URL in the summary. For more details, please go to the homepage as also indicated below in the summary. The program can be installed easily by executing the downloaded file. If, for any reason, you cannot run Matapuna on your own computer, you can get in touch with Dave Moskovitz and check if it is possible for him to host your project. Details are provided on the Matapuna website.

---

[1] For example, Ubuntu is a distribution of Debian Linux which is available free of charge at <http://www.ubuntu.com/>

**3. CONFIGURATION.** Since Matapuna was designed for the Māori language, the configuration you get when you have downloaded and installed the software may not immediately meet your needs, unless you are dealing with a comparable Polynesian language. In my case, customization was necessary, since Fula is a rather different type of language, having a complex noun class system and an agglutinative verbal morphology, like most languages of the Niger Congo phylum. Support will be needed for customization unless the lexicographer has programming skills in PERL, the base on which Matapuna was built. My first tip is that you should plan, on paper, a "typical" entry for your dictionary; the simpler the microstructure, the better, if you plan a collaborative project. The reason is that, although Matapuna is flexible enough to support a complex microstructure (for instance, with many subentries, markers, symbols, etc.), your collaborators, or some of them, may not be able to handle it if everyone is not in the same location. This is what happened to me: I created, for example, an additional button (which was not available in the first template of my database) for adding subentries or compounds. Now I realize that my collaborators, who are neither linguists nor lexicographers, are reluctant to use it. For the same reason, I highly recommend that you involve your team in the preparation of the project from an early stage. Of course, I am aware that the structure of a dictionary entry is intimately connected with the morphology and the syntax of a language, and that ready-made solutions are hard to come by.

If your plan is ready, you should get in touch with the developer for customization. This step is necessary because, at the present stage, you cannot start your dictionary from scratch, as in Toolbox for example, unless you have programming skills. The huge advantage of this is that you will not need to go through complex set-up procedures (users of Toolbox will know what I mean), so that you can concentrate on the essentials—doing your lexicographical work. As noted above, if members of your team are residing in different locations, the network feature of this DWS should be used for collaborative work over the Internet. I recommend asking the developers about hosting facilities.

In my experience, Matapuna is ideal for small-scale projects involving typically a linguist or lexicographer leading a team of collaborators working to document an endangered or an under-resourced language. The menu of the program is mostly, as we will see, easy and self-explanatory—there is no user manual or help file!

Since Matapuna is Unicode compliant, if the language you are documenting uses special characters, displaying them will not be a problem. For data input, any keyboard mapping tool may be used. In my case, I am using Keyman to enter the extended Latin characters of Fula, but I have also successfully tested MSKLC, a free tool for Windows.

**4. MATAPUNA'S TOOLBAR.** When you log on in Matapuna, you will see a webpage with a menu bar at the top and the bottom (Figure 1).

FIGURE 1. Opening a webpage

Matapuna's menu elements are:
- *Home:* a welcome page with information on the current state of the work
- *Add:* for adding a new entry
- *My editing:* basically the same as *Home*
- *Search:* allows you to search for entries in the database. Several search options (Figure 2) exist, including the ability to filter searches by dialect.
- *Reports:* several report options in two formats: HTML or CSV (Figure 3)
- *Print:* actually a preview function

If you click on the "Print" option, your database will not be exported to any other format, but Matapuna will show you how your dictionary will look. The print tab will allow you to select options like "Start letter or word," "End letter or word" (Figure 4), or if your dictionary has a multidialectal approach, you can select the dialect(s) you want to be printed. After you have selected your options, you can view your dictionary by confirming with the "Print detail" button.

Since the printout is in HTML format, it naturally allows you to display your dictionary in a website. According to the Matapuna homepage, the entries can be exported as XML. An export of the database to a word processor has not yet been implemented, at least to my knowledge. An export to LIFT is currently planned (Dave Moskovitz, pers. comm.). LIFT (Lexicon Interchange FormaT) is an XML format in development that is devised to allow the exchange of lexical data between different programs.[2] The LIFT export is appealing, particularly if one thinks that through LexiquePro, which supports LIFT as well, a connection could be established between Matapuna and Toolbox, a DWS that is still very popular among linguists.

---

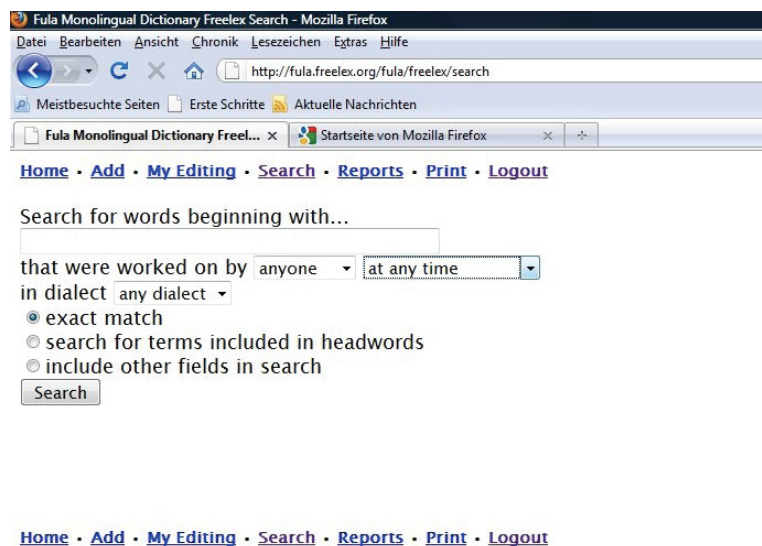[2] For more details about LIFT, please visit <http://code.google.com/p/lift-standard/>.
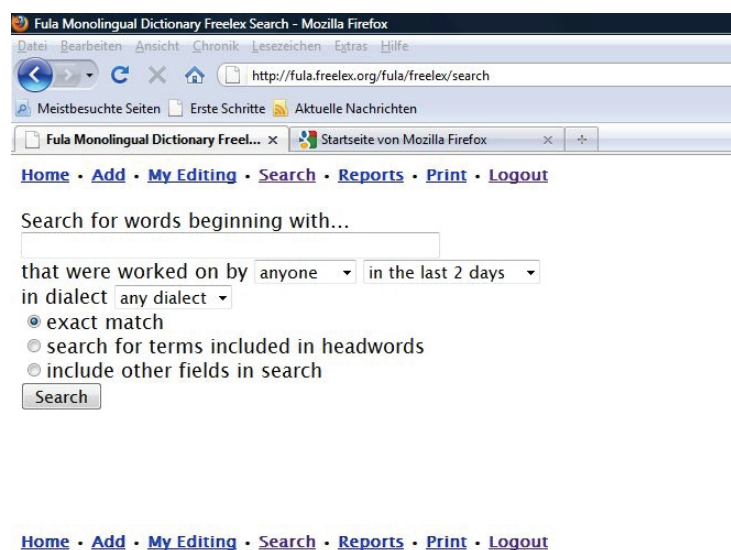
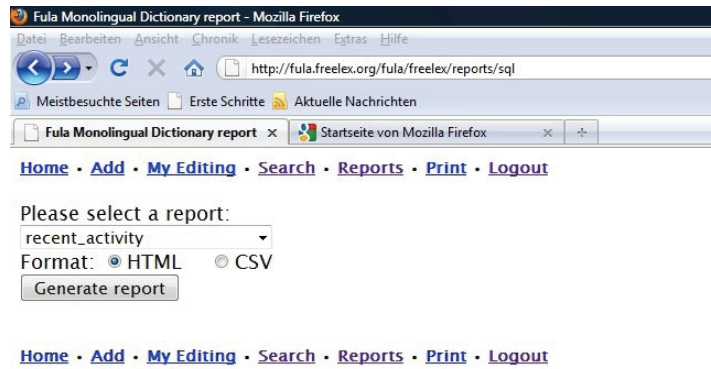Figure 2. One search option



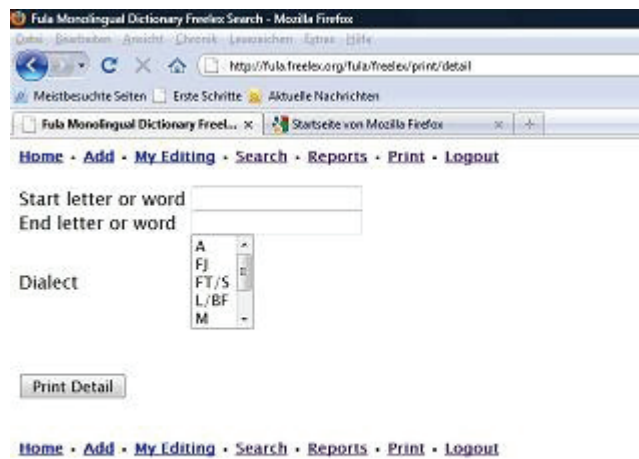Figure 3. Another search option

FIGURE 4. Reports



FIGURE 5. Print

Clicking the "Add" tab opens a new page, and you can start creating a new entry by filling in its fields. The headword is written in the "Entry" field.

**4.1 SAVING THE NEW ENTRY.** Clicking on "Save" saves the entry in the database. An interesting problem arises when two users who log in to edit the same entry and attempt to save the changes simultaneously. The database will save only the changes made by one of the two editors. I do not know which criterion is decisive, but I suppose it depends on who has the faster connection or machine. As soon as the entry is saved, an ID number will be automatically assigned to it.

**4.2 DELETING AN ENTRY.** With "Delete," the entry can be deleted. Before that operation is executed, you will be asked to confirm.

**4.3 SEPARATING SENSES.** Under the label "Sense," two boxes are used for separating homonymous and polysemic lexemes. The first box automatically creates two or more homonymous entries, which appear as superscripted headwords in the print view. The second box automatically creates a polysemic entry, which appears in the print view as a single headword with different numbers separating the different meanings. Thus, the basic principle is that Matapuna always opens a new page for each sense, irrespective of homonymous or polysemic headwords, but it handles the output differently. The numbers need to be added manually in the boxes (only Arabic numerals seem to be permitted).

**4.4 SUBENTRY.** The subentry field is connected with the third box in the "Sense" field. It was not part of the basic program, but I added it with the support of the developer in order to be able to include compounds, idioms, and proverbs. Clicking on this box allows such subentries to be inserted as needed. Each new subentry requires its own page, but Matapuna will handle all subentries as part of the same main headword in the print view. For this process too, the numbering must be done manually.

**4.5 MASTER SYNONYM.** This feature allows automatic bi-directional cross-referencing. This means that the referencing headword points to the referenced one, and vice versa, and all synonyms are arranged under a so-called master synonym in the print view. This will save you a lot of work, especially if you are working on a language with several dialects and numerous dialectal synonyms.

**4.6 PRIVATE COMMENT.** The editor can leave a comment of any type in this field. It is intended for internal use only and will not appear in the print view.

**4.7 DATA CONSISTENCY.** Like every good DWS, Matapuna offers the possibility of defining a set of field values (like the "Range Set" in Toolbox) which can be easily selected, if required, either from a roll-down menu or from check boxes. In my template, field values are defined for the following fields: "Dialect," "Word class" (part of speech), and "Usage."

There is, in this respect, a self-checking feature announced on the Matapuna homepage that allows one to check whether the defining vocabulary item (in the case of a monolingual dictionary) is itself defined. It is not implemented in my own template, so I have no experience with it. I assume this feature would be useful only for an isolating language, because a highly flexional language like Fula would have many flexion forms in definitions or examples which would be falsely identified as undefined vocabulary.

Another data consistency checking function sends you a warning whenever you try to save an entry that contains a letter or a symbol that is not included in your alphabet. However, you always have the option of ignoring the warning and saving the entry.



FIGURE 6. Entry

**5. SUMMARY.** Matapuna is a dictionary tool I warmly recommend, especially for collaborative small or medium scale projects with little funding. Thanks to its network features and the browser-based, easy-to-use interface, it is ideal for linguists or lexicographers working with local language communities. Furthermore, the system has a high level of stability, so that it seldom crashes. Therefore, once the setup of the template is done, the team can focus on the essentials, such as adding new entries to the dictionary.

The planned LIFT support will enhance interoperability with other applications like Toolbox and LexiquePro, which many linguists and researchers currently use. One could add a multimedia capability to the wish list since sounds and images, when added to texts, can considerably support language conservation.

| | |
|---|---|
| Primary function: | Web-based dictionary making and publishing system |
| Pros: | Web-based, multi-user, XML export, ease of use, flexible, very stable |
| Cons: | Direct export to a word editor currently not possible; however export to the LIFT standard is planned in future upgrades |
| Platform: | Linux server but platform-independent for the client PC |
| Open source: | Yes |
| Proprietary: | No |
| Available from: | http://sourceforge.net/projects/matapuna/ |
| Application size: | 74.3 KB |
| Documentation: | No user manual but description with screenshots and demo system available at the homepage: www.matapuna.org |

Oumar Bah
oumar.bah@chello.at