Software Demonstration: The *TshwaneLex Electronic Dictionary* System

David Joffe
Malcolm MacLeod
TshwaneDJe Human Language Technology

Gilles-Maurice de Schryver
Ghent University

In this presentation, use of the TshwaneLex Electronic Dictionary *software module will be demonstrated. This module provides a complete, customisable solution for the publication of a dictionary electronically-as a CD-ROM, or for sale or download on the Internet, or both. The "base package" provides all the functionality of a modern, user-friendly Electronic Dictionary, and can be fully customised for the desired "look and feel", branding, dictionary content and language(s) of a publisher's product. This provides a highly cost-effective solution for creating a professional Electronic Dictionary product, obviating the need for the kind of expensive custom-developed solutions that have traditionally been required. The system can be used for the immediate publication of dictionaries already in TshwaneLex, or in a comparable structured format-such as XML.*

1. Localisation and dynamic metalanguage customisation

The entire interface of the Electronic Dictionary can be translated to, and made available in, any language. The default language may also be changed. Thus for example, the interface of a bilingual Spanish-English dictionary could be available in both Spanish and English, with Spanish being the default. Instant "real-time" switching of the language allows the interface language to be changed by the end-user without even restarting the application. Taking localisation further, the metalanguage of the dictionary *content* may also be localised. The end-user may opt to see labels like "noun" or "verb", or words such as "SEE" in cross-references, in the language of their choice. Screenshots of this are shown in the Addendum (figure 2).

2. Multimedia

TshwaneLex allows images and sound recordings (e.g. pronunciation recordings) to be included in dictionary entries; this functionality carries through to the Electronic Dictionary software.

3. Microsoft word integration

The Electronic Dictionary integrates into Microsoft Word, and will automatically perform a search for the word at the current cursor position; if found, it displays a miniature version of the article (headword plus Translation Equivalents) in an overlay window. When these "mini" results are clicked on, the full article is opened in the main Electronic Dictionary window. This functionality also extends to Outlook when using Word as the e-mail editor. The screenshot below (figure 1) shows this functionality in a Swahili - English Electronic Dictionary.
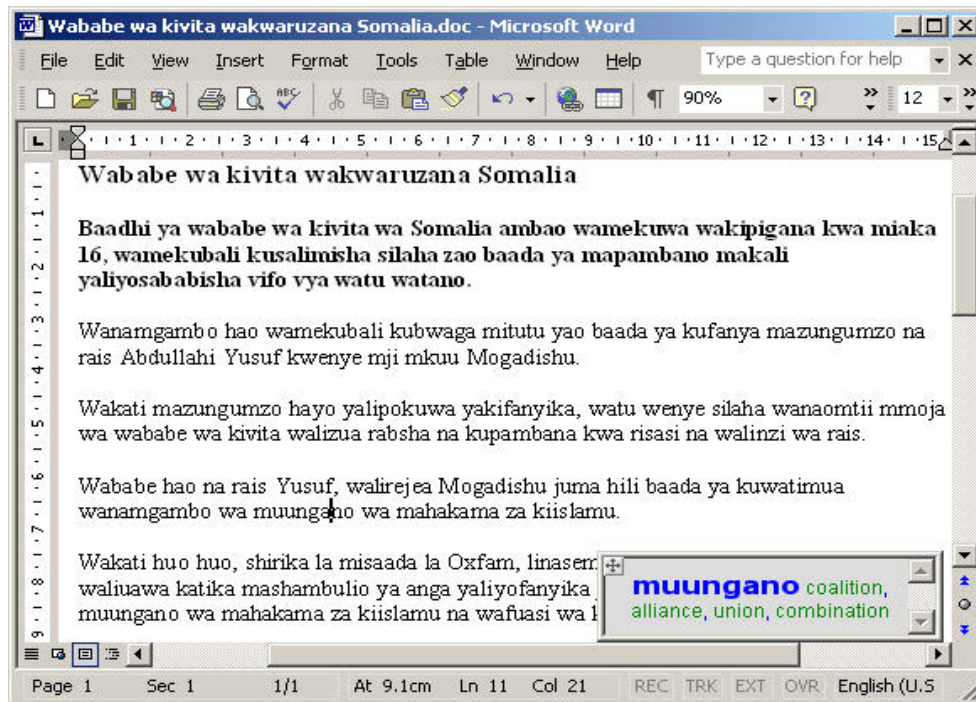
Figure 1. Screenshot showing *TshwaneLex Electronic Dictionary* integration with Microsoft Word, in this case for Swahili-English. A clickable overlay window automatically displays search results for the current word at the cursor position.

The *TshwaneLex Electronic Dictionary* thus starts to become a true Computer Assisted Translation (CAT) tool.

4. Fast and customisable searching

Smart pre-indexing allows for high-speed searching even with very large dictionaries. Further, the ability to configure a "search script" using the built-in scripting language of TshwaneLex allows the search behaviour to be potentially customised for language-specific search functionality, if desired. The use of smart indexing for searches combined with "memory-mapped I/O" for data accessing allows even large dictionaries to load quickly and to be searched quickly.

5. Automatic / downloadable updates and encryption for data security

In addition to a traditional application installer, the *TshwaneLex Electronic Dictionary* software also optionally includes the facility to be able to automatically update itself over the Internet. The system distinguishes between two "levels": Updates for the software itself, and—optionally—updates for the dictionary content. In a *TshwaneLex Electronic Dictionary*, dictionary data is also protected by default with encryption, thereby protecting it from potential "intellectual property thieves".

6. View mode options and customisable styles/colours

For a bilingual dictionary, the "Linked View" feature optionally allows the end-user to automatically display entries on the other side of the dictionary that are related to the current one by means of a Translation Equivalent treated as a headword on that side. For a bilingual dictionary, the "Expanded View" mode allows the end-user to choose to see both sections of the dictionary simultaneously in a split "side-by-side" view, or to "expand" a specific section, giving more space.

The Styles system of TshwaneLex allows the appearance (e.g. colour, font, surrounding punctuation) of each field in a dictionary to be configured; in the Electronic Dictionary, preconfigured "sets" of styles, or "views" of the data, may also be prepared for the end-user to choose between, e.g.

"advanced" or "novice" views. Different fields may also be visible or hidden in different views. This overall provides for potentially highly flexible presentation of content based on end-user preference.

7. Customisable popup help

Each field (type of information) in the dictionary may optionally be linked to a popup "help screen" that is displayed to the user when clicking on that field. This behaviour is scriptable using the scripting language built into TshwaneLex, allowing the popup help screen to exhibit some "intelligence"—e.g. different information can be displayed depending on the type of field clicked on, as well as the value of a particular field (e.g. "noun" vs. "verb").

8. "Taskbar / System Tray icon" (Windows)

An icon in the System Tray provides convenient access to the Electronic Dictionary for end-users. A configuration option also allows the Electronic Dictionary software to be loaded into memory at start-up time, allowing the dictionary to be immediately available with no delay.

9. System requirements

The *TshwaneLex Electronic Dictionary* system will run on essentially any modern Windows PC or Mac.
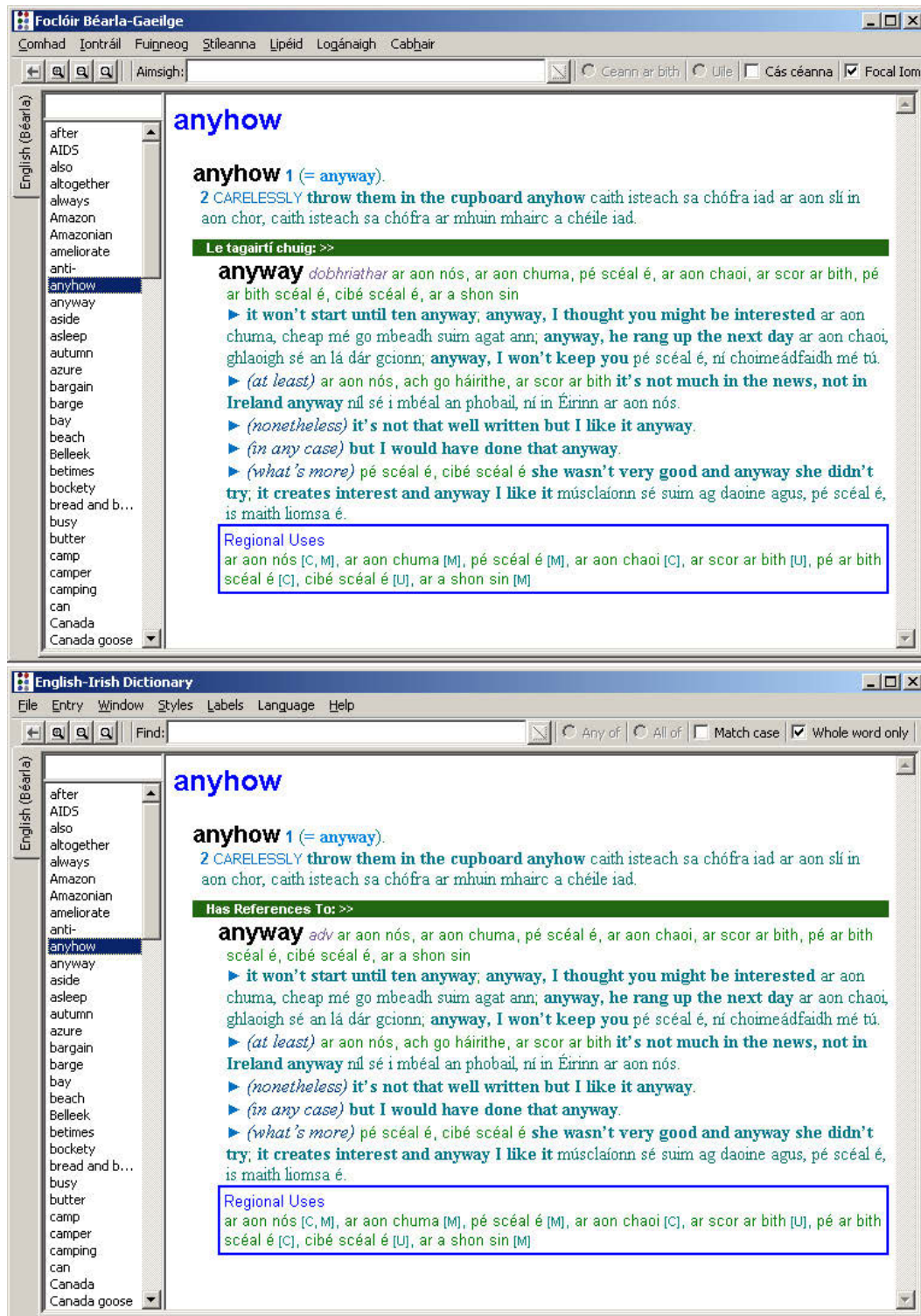
Addendum

Figure 2. Localisation and Dynamic Metalanguage Customisation: Side-by-side screenshots showing the same English-Irish content presented with both the user interface and part of speech labels localised into Irish and English respectively. Data Copyright Foras na Gaeilge.