ID: 21

David Joffe & Gilles-Maurice de Schryver

TshwaneLex – Professional off-the-shelf lexicography software

0. Introduction: On user-friendliness

TshwaneLex is a software program for compiling monolingual, bilingual or semi-bilingual dictionaries. TshwaneLex contains various innovative features designed to optimise the process of producing dictionaries, and to improve consistency and quality of the final dictionary product. The research question that drove the development was whether an off-the-shelf and language-independent dictionary writing system (DWS) could be designed – for output to paper or an electronic medium (e.g. CD-ROM or the Web) – that any lexicographers could customise to their own taste for the production of any type of (mainstream) dictionary. In this presentation we show that this bold aim was indeed achieved, and we will illustrate this with data from some of the current TshwaneLex users, which include amongst others the South African National Lexicography Units under PanSALB (Pan South African Language Board), Macmillan, and Oxford University Press.

From a designers' point of view the aim implied that a DWS had to be created with (1) full Unicode support, and (2) customisable DTDs, and which (3) ensures and enforces cross-reference integrity, (4) allows for various visualisations of the data distribution structure on micro- and macro-level (through e.g. tree views, linked view modes and the use of Rulers), and (5) is user-friendly. Regarding the latter, strong emphasis has indeed been placed during the design of TshwaneLex on producing a **user-friendly** tool, to reduce required training time, and also based on the principle that lexicographers *should not need* advanced computer literacy skills in order to compile dictionaries. Another major underlying design principle of TshwaneLex has been that the software should **automate** as much as possible for the lexicographer.

A selection of TshwaneLex features and accompanying screenshots is presented below.

1. Full Unicode Support [cf. Figure 1]

- **Unicode**, the international character set standard, is fully supported for *every* aspect throughout TshwaneLex.
- **Windows IMEs**: Data can be entered directly into TshwaneLex using any of the IMEs (Input Method Editors) available in Microsoft Windows 2000 or Windows XP, such as those for Chinese, Japanese, Korean or Arabic.

2. Customisable DTD (Document Type Definition) [cf. Figure 2]

- A DTD (Document Type Definition) is used to describe the structure of lemmas for a particular dictionary project. TshwaneLex allows the DTD to be **fully customised** by the user on a per dictionary project basis.
- A **user-friendly** interface design, supplemented with detailed documentation, allows end-users to be able to configure the DTD without the necessity of assistance from an IT expert. TshwaneLex also creates a sensible default DTD for new dictionary projects, allowing a user to get "up and running" within minutes.
- **Template DTDs** may be created, allowing a DTD to be easily re-used for other dictionary projects, or allowing new projects to be initially based on standard DTDs.
- TshwaneLex enforces **DTD constraints**, preventing lexicographers from creating invalid entries and thus ensuring consistency throughout the dictionary.

- Compatible with **XML DTDs** (supports **elements**, **attributes**, **entities**, and relational constraints such as "one child only", "one or more", "zero or more", etc.).
- The DTD allows certain fields to be restricted to a **selection from a list**, such as for instance from a list of parts of speech or labels.
- A powerful **Styles** system allows all aspects of the visual output of every element type to be configured.

3. Cross-Reference System [cf. Figure 3]
- **Show related cross-references**: Whenever one views or works on a lemma in TshwaneLex, all lemmas with cross-references *to* the current lemma as well as all lemmas cross-referenced *by* the current lemma are immediately shown in the lemma preview area.
- **Automatic homonym and sense number updating**: TshwaneLex *automatically updates* the cross-reference target homonym and sense numbers when these change on the cross-reference target lemma.
- Ensures and enforces **cross-reference integrity** throughout the dictionary editing process. There is no need to keep track of cross-references manually.

4. Bilingual Editing Features [cf. Figure 4]
- **Side by side editing window layout** allows the lexicographer to view or work on both sides of the dictionary simultaneously.
- When **Linked View** mode is selected, TshwaneLex automatically shows all lemmas on the *other* side of the dictionary related to the currently selected lemma.
- **Automated Lemma Reversal** functions save the lexicographer valuable time when creating the reverse side of the dictionary.

5. Miscellaneous Other Features
- Direct export to RTF (Microsoft Word), **XML**, HTML or HTML/CSS.
- **Sound files** can be attached to any field.
- The **Filter** function allows the lexicographer to define criteria for viewing a subset of the data, for example "show all homonyms" or "show all cross-references". More advanced filters may also be defined through the use of **Boolean operators**.
- **Customisation of the language of the meta-language**: TshwaneLex allows multiple translated sets of labels to be defined for displaying information such as cross-reference type, part of speech and usage information.
- The **Full Dictionary Search** tool allows fast text searches on the entire dictionary, with options such as case-sensitivity or whole-word/partial-word matching. Advanced users may also use **regular expressions**.
- The **Compare/Merge** tool allows different versions of a database to be visually compared with one another, allowing changes to be merged into the current database. Changes made by a lexicographer working at home, or by lexicographers that are not connected directly to the main database, can be easily merged back into the main database via a user-friendly interface.
- An **online** (Web) dictionary module and an **electronic** (CD-ROM) dictionary module are also available for TshwaneLex.

For more information about TshwaneLex, please visit the TshwaneDJe HLT website (http://tshwanedje.com/). Low-cost licenses are available for academic/non-profit use, as well as for use for endangered languages. TshwaneDJe HLT also provides consulting and training services for all aspects of the dictionary compilation process.
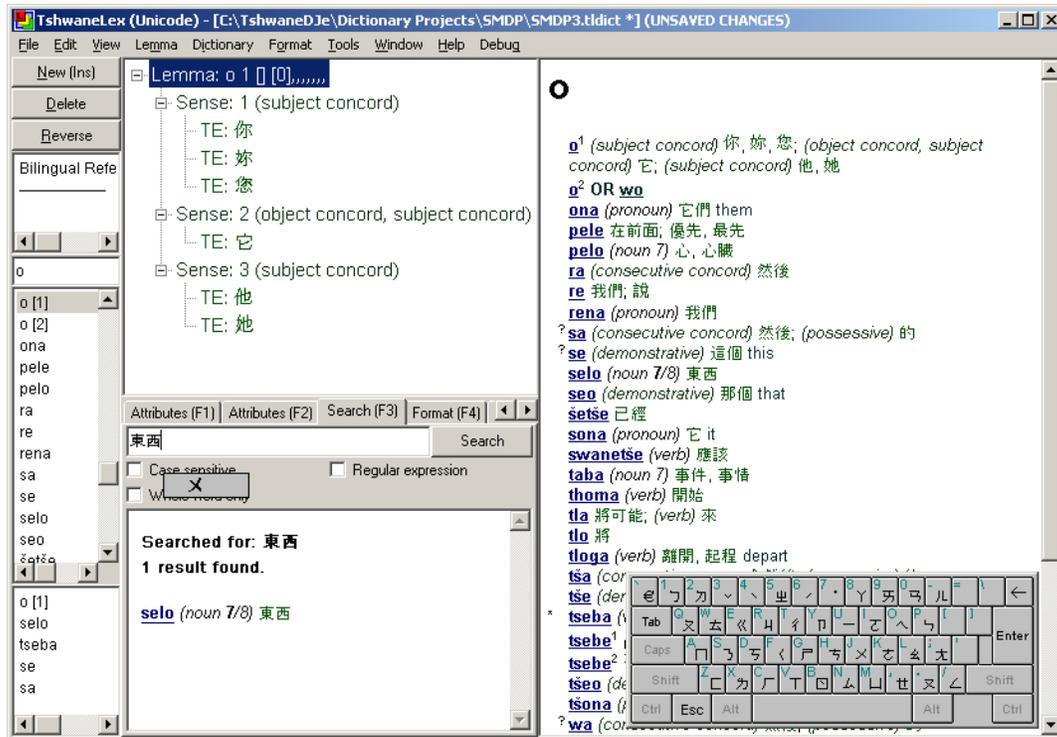
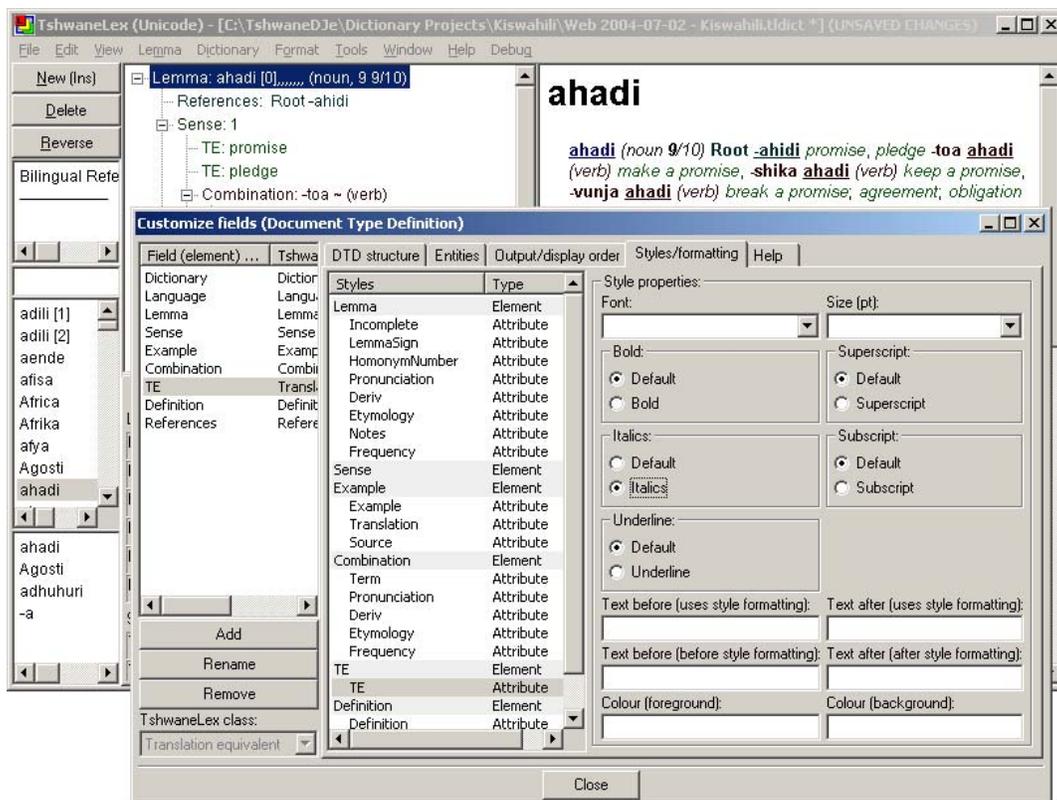**Figure 1**: Screenshot illustrating Unicode support

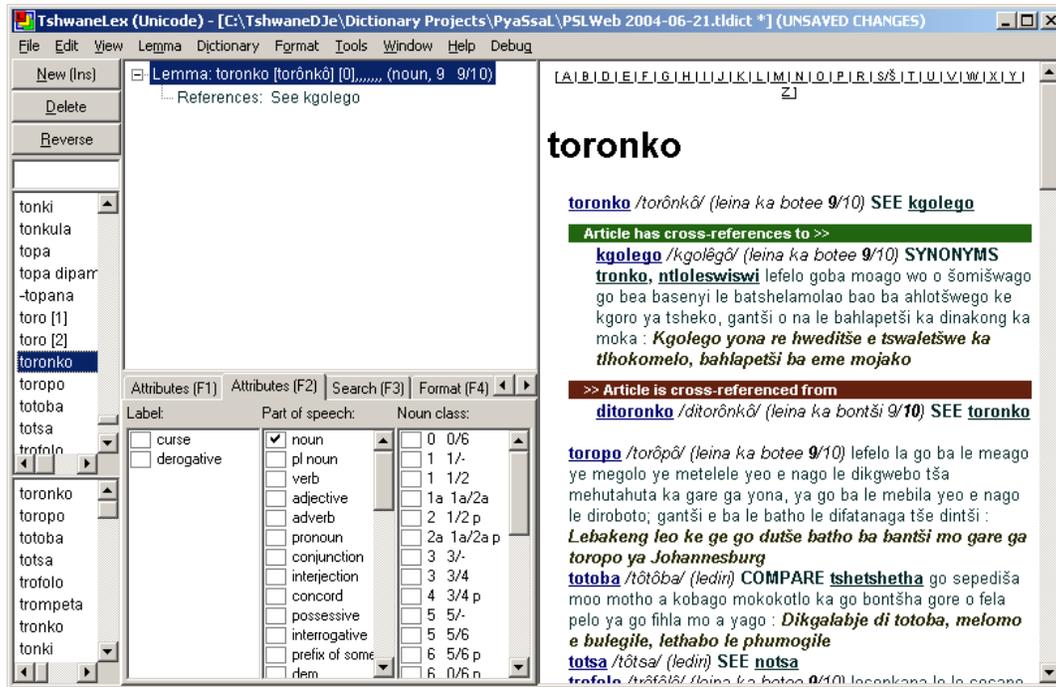**Figure 2**: Screenshot illustrating the use of a DTD
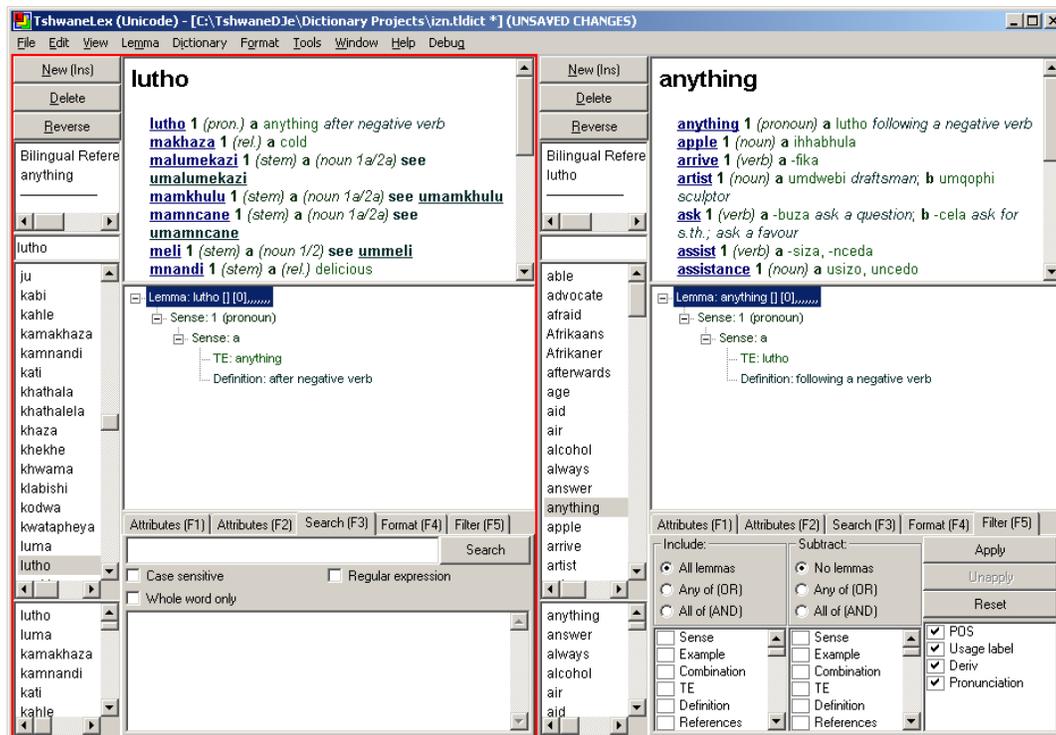
Figure 3: Screenshot illustrating the cross-reference system

Figure 4: Screenshot illustrating bilingual editing features