From *TshwaneLex* to *TshwaneTerm*: Tailoring Terminology Management for South Africa[*]

David Joffe, *TshwaneDJe HLT, Pretoria, Republic of South Africa (david.joffe@tshwanedje.com)*, and
Gilles-Maurice de Schryver, *Department of African Languages and Cultures, Ghent University, Ghent, Belgium and TshwaneDJe HLT, Pretoria, Republic of South Africa (gillesmaurice.deschryver@UGent.be)*

Abstract: The addition of a restricted number of features to the dictionary (compilation) software *TshwaneLex* suffices to turn this application into a terminology management system. This article gives a brief overview of those extra features, using the *Department of Arts and Culture* (DAC) AIDS list as case study.

Keywords: LEXICOGRAPHY, TERMINOLOGY LIST, TERMINOLOGY MANAGEMENT SYSTEM, SOFTWARE, TSHWANELEX, TSHWANETERM, DAC AIDS LIST

Samenvatting: Van *TshwaneLex* naar *TshwaneTerm*: terminologiebeheer op maat voor Zuid-Afrika. De toevoeging van een beperkt aantal uitbreidingen aan de woordenboek(aanmaak)software *TshwaneLex* is voldoende om dit programma met succes als terminologiebeheersysteem in te zetten. Dit artikel geeft een kort overzicht van die extra uitbreidingen, en gebruikt de *Department of Arts and Culture* (DAC) AIDS lijst als illustratie.

Sleutelwoorden: LEXICOGRAFIE, TERMINOLOGIELIJST, TERMINOLOGIEBEHEER-SYSTEEM, SOFTWARE, TSHWANELEX, TSHWANETERM, DAC AIDS LIJST

1. The need for a South African terminology management system

This article is a sequel to De Schryver and Joffe's (2005a) 'From *TshwaneLex* to *TshwanePedia*', in which it is pointed out that the high degree of customisability found in the dictionary (compilation) software TshwaneLex may be exploited for the production of a wide variety of reference works. The variety under discussion here are terminology lists, and by extension the move from an environment to compile dictionaries with, Tshwane*Lex*, to a terminology management system, Tshwane*Term*.

---

[*] An earlier version of this article was presented at the Tenth International Conference of the African Association for Lexicography, organised by the Sesiu sa Sesotho Lexicography Unit, University of the Free State, Bloemfontein, Republic of South Africa, 13–15 July 2005.

When Mariëtta Alberts, the 'Focus Area Manager: Standardisation and Terminology Development' of the *Pan South African Language Board* (PanSALB), pointed out that the South African terminographers who work at the *Department of Arts and Culture* (DAC) and elsewhere in the country do not have adequate tools at their disposal, *TshwaneDJe HLT* decided to produce a solution for them. In analogy with the design of TshwanePedia, and in order to create a successful application, the extra features that were added to the TshwaneLex base revolve around the 'window layout', 'multimedia' and 'export'. For the illustrations below a draft AIDS terminology list compiled by DAC, which was parsed and imported into TshwaneTerm, is used.

2.   From TshwaneLex to TshwaneTerm

To begin with one can point out that terminology databases typically contain a rather large number of languages in parallel. Creating a Document Type Definition (DTD) in TshwaneLex that can handle multilingual or interlingual data is straightforward, as the customisability already allows for the addition of fields to entries for more languages, and in fact *any number* of languages. In South Africa this usually means that one will want to work with an eleven-lingual database, and that each terminographic entry will contain data in all eleven official languages. Compared to dictionary articles, terminology entries are less complex, in the sense that, at least in theory, only one equivalent is supposed to be provided per language. To terminographers, seeing the 'structure' of the entries (in the Tree View) is thus a matter of lesser importance, as that structure is repetitive and basically the same throughout. Instead, more (vertical) space is needed to see more languages in parallel. For that reason, a so-called optional 'High Tools window layout' was implemented, which is accessible with a single 'hotkey'. Addendum 1 shows a screenshot of the TshwaneTerm interface with the high view enabled. (Note that when the high view is not enabled, the top section is taken up by the Tree View. If one for example wants to add a second translation equivalent for a particular language, as is the case for siSwati in the screenshot, then this can simply be done by adding the extra field(s) in the Tree View first.)

Secondly, a multimedia feature available in TshwaneTerm is enhanced support for audio files, through a data type 'Sound file', that can be added at *any level* in the DTD. Sound files are thus not restricted to the head-term (and, say, its morphological derivations) but can truly appear anywhere in the DTD, and they can be of any length. For the terminology list shown in Addendum 1, an audio recording was added of the pronunciation of the equivalents in each of the eleven languages. Just as for the images, sound files are stored centrally, and the compiler simply uses 'Browse …' to select one.

The third and most important feature, however, is the ability to be able to generate multiple 'editions' from a single database (cf. De Schryver and Joffe 2005: 56-59). This is done by creating different sets of styles, each of which basi-

cally 'lets through' different parts of the data. In the screenshot shown below, for example, four style sets have been defined: a default style (which corresponds with the view in Addendum 1), and then three 'masks' to output only the isiZulu, or only the Tshivenda, or only the Sotho-group languages.
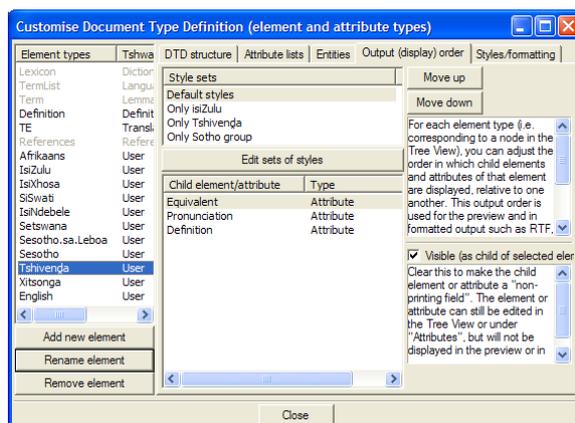

In Addendum 2 screenshots of samples of the 'isiZulu mask' and the 'Sotho mask' are shown. One can easily toggle between the different editions, and thus views, with a single key press.

The use of sets of styles within a single database is very powerful indeed. Not only can subsets of the data be output in various formatted ways, as illustrated in Addendum 2, but one can also successfully prepare, say, both a pocket edition and a full edition of the same data simultaneously. Or one can generate different views of the same data, such as a print view and an electronic/online view, or one can generate monolingual, bilingual and semi-bilingual reference works from the same database, etc. All of these possibilities are especially useful when one deals with large amounts of data, as in multilingual terminography — South Africa's richness situation par excellence.

References

DAC. 2003–2005. Department of Arts and Culture [online]. Available: <http://www.dac.gov.za/>.

De Schryver, Gilles-Maurice and David Joffe. 2005. One database, many dictionaries — varying co(n)text with the dictionary application TshwaneLex. Ooi, V.B.Y., A. Pakir, I. Talib, L. Tan, P.K.W. Tan and Y.Y. Tan (Eds.). 2005. *Words in Asian Cultural Contexts, Proceedings of the Fourth Asialex Conference, 1–3 June 2005, M Hotel, Singapore*: 54-59. Singapore: Department of English Language and Literature & Asia Research Institute, National University of Singapore.

De Schryver, Gilles-Maurice and David Joffe. 2005a. From *TshwaneLex* to *TshwanePedia*: Creating and Flexibly Maintaining Online Encyclopaedias. *Lexikos* 15: 306-311.

TshwaneDJe HLT. 2003–2005. TshwaneDJe Human Language Technology [online]. Available: <http://tshwanedje.com/>.

Addendum 1: Screenshot of TshwaneTerm, in 'High Tools window layout'
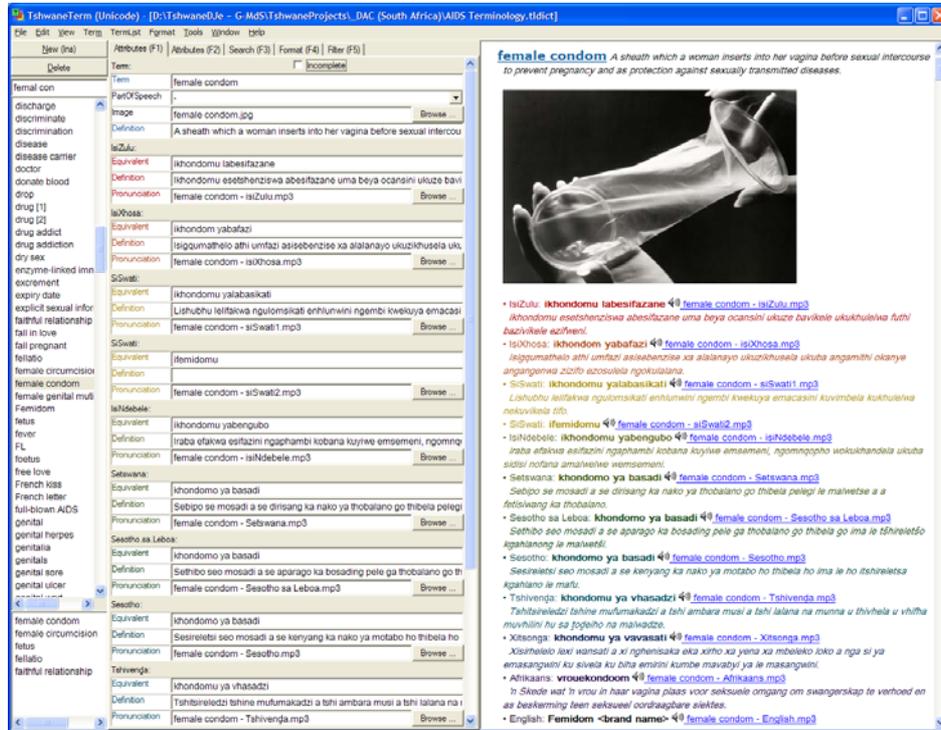
Addendum 2: Generating multiple 'editions' (views) from a single eleven-lingual terminology database with TshwaneTerm. Left: extracting isiZulu (and English). Right: extracting the three Sotho equivalents (*with a nod to D.J. Prinsloo*).