

# Developing a Multi-functional e-Spelling-Dictionary – a South African Perspective

Elsabé Taljard, D J Prinsloo

Department of African Languages, University of Pretoria  
University of Pretoria, Private Bag X20, Hatfield 0028.

[elsabe.taljard@up.ac.za](mailto:elsabe.taljard@up.ac.za), [danie.prinsloo@up.ac.za](mailto:danie.prinsloo@up.ac.za)

**Keywords:** e-spelling-dictionary, online lexicographic support, lesser-resourced languages

## Abstract

For any language to attain or retain the status of a fully-fledged, official language of higher functions, the availability of sophisticated online lexicographic support tools is essential. In this paper we discuss the design and development of a multi-functional e-spelling-dictionary which is linked to the spelling rules of a particular language. The tool specifically caters for the on-the-fly user's spelling needs but also provides for cognitive information. It can be accessed departing from its main menu with clickable options to the different functions of the word list, e.g. illustration of a particular spelling rule, or as plug-in to an existing e-dictionary or word processor. For illustrative purposes, we refer to Afrikaans (AFR) and Northern Sotho (NSO), two of the official languages of South Africa. In terms of lexicographic e-resources, both languages can be described as lesser-resourced, furthermore, existing electronic resources catering specifically for spelling guidance can at best be described as rudimentary, not providing adequate provision for users' needs. Although spell checkers exist for both languages, these are stand-alone tools which provide no additional guidance with regard to broader spelling issues. The definitive spelling guide for Afrikaans, i.e. the *Afrikaanse Woordelys en Spelreëls* ('Afrikaans word list and spelling rules') is currently only available online for paying clients, which severely restricts its accessibility. A further shortcoming of the online spelling guide for Afrikaans is that it provides no additional cognitive support to the user; there is for example no clickable link guiding the user from the look-up to the actual spelling rule that determines the spelling of a particular word. For Northern Sotho, a spelling guide is available in paper format, but due to external factors, this guide is not generally available to the Northern Sotho linguistic community. It also does not address the issue of standardized versus non-standardized forms. No online spelling resource except the spell checker referred to above, is available. The spell checker for Northern Sotho is only available on CD ([https://spel.co.za/en/product/african\\_spelling\\_checkers/](https://spel.co.za/en/product/african_spelling_checkers/)), which renders it obsolete for many potential users, since the latest generation of PCs and laptops are no longer fitted with a CD-ROM drive. This spell checker is furthermore a static tool, which does not make provision for regular updates. Consequently, having been developed in 2009, the wordlist against which spelling is checked is already outdated. One of the advantages of having an e-tool is the fact that it can be updated in real time. Northern Sotho furthermore carries the additional burden of grappling with standardization issues, leading to a proliferation of spelling variants – an aspect which needs to be considered in the design of the tool.

The design of the proposed tool is approached from the perspective of the user, one of the prevalent approaches in modern day metalexigraphy. This perspective compels compilers of any lexicographic product to compile their dictionaries according to the information needs and research skills of a well-defined target user group.

“The user-perspective, so prevalent in modern-day metalexigraphy, compels lexicographers to compile their dictionaries according to the needs and research skills of well-defined target user groups. The dominant role of the user has had a definite effect on the compilation of dictionaries as well as on the evaluation of their quality. Good dictionaries do not only display a linguistically sound treatment of a specific selection of lexical items. Good dictionaries are products that can be used as linguistic instruments by their respective target user groups. The better they can be used, the better dictionaries they are.” (Gouws and Prinsloo 2005:39).

Designing and developing a lexicographic tool of high quality therefore implies an analysis of the target users' information needs. According to Tarp (2009), analysis of log files of online dictionaries can give a good indication of users' needs. Having access to the log files of an earlier online version of the Afrikaans word list enabled us to identify a number of pertinent user's needs. The most obvious advantage of log file analysis is the identification of gaps in the lemma selection of the word list. Secondly, it provided us with valuable information on frequently misspelled and frequently confused words. Lastly, the analysis clearly indicated the need for additional cognitive information – users would often look for information regarding a specific orthographic issue, e.g. the use of circumflexes or hyphens. This implies that users need direct access to the rule that determines the use of these orthographical signs.

Based on the analysis of users' needs, the proposed tool has four functions, i.e. confirmation of correct spelling, direct spelling guidance, provision of cognitive information and guidance on frequently confused words. These functions far supersede the sophistication of currently available spelling checkers for these two languages.

Should users type in a correctly spelled search word (e.g. *galesome* 'ten times' (NSO), *lêer* 'file' (AFR)) and find it in the word list without any additional comments, it serves as **confirmation** that (a) the search word is correctly spelled and (b) the search item is regarded as part of the standard language. In the case of Northern Sotho a further indication as to the formally standardized status (or not) of

the search word is to be provided. An optional link to additional cognitive information is envisaged, e.g. to the relevant spelling rule pertaining to a specific category e.g. the spelling of derived adverbs (NSO), or the use of the circumflex (AFR). Users are also able to search for specific categories within the e-spelling rules, e.g. spelling of derived adverbs (NSO), or spelling of words with circumflexes (AFR).

Should the user type in a search word which is incorrectly spelled, (e.g. *ga lesome* ‘ten times’ (NSO), *leêr* ‘file’ (AFR)) **guidance on the correct spelling** is provided in the form of “Did you mean (*galesome* (NSO), *lêer* (AFR))?”, or by means of the typical spellchecker function of a pop-up box containing a list of related correct spellings. Again, a clickable option giving access to further cognitive information in the form of the relevant spelling rule is provided.

Lastly, the tool also provides direct guidance with regard to words the spelling of which is **often confused**. Should users type in a search word which is correctly spelled (*wyer* ‘wider’ (AFR)), but which has shown itself to be frequently confused with another word (*weier* ‘refuse’), they are alerted to the possibility that they may have spelled the search word correctly, but that the spelling and the intended meaning may not match.

Having access to (free) online spelling resources will not only provide adequate support to users, but will also contribute to strengthening the status of these languages as languages of higher functions.

The tool is still in the design phase, but most of the required components have been built, such as corpora and frequency lists for the two languages, pop-up boxes, cognitive information screens, spelling rules, typically misspelt words and alternative spell checkers for both languages.

### **Bibliographical References**

Tarp, S. (2009). Reflections on lexicographic user research. *Lexikos*, 19: pp. 257 – 296.  
Gouws, R.H and Prinsloo, D.J. (2005). *Principles and practice of South African lexicography*. Stellenbosch: African Sun Media.

### **Acknowledgements**

Financial support by South African Centre for Digital Language Resources (SADiLaR) is hereby acknowledged.