Fuzzy SF: Towards the ultimate customised dictionary

Gilles-Maurice de Schryver*
DJ Prinsloo**

Abstract

Fuzzy SF, a novel concept for an electronic-dictionary package, is presented. In Fuzzy SF, log-file based Artificial Intelligence components enable the implicit retrieval of personalised user feedback with which the package customises each user's own and unique dictionary. To that end, all the data in both the databases and the multimedia (sub)corpora are graded using Fuzzy Sets, so that the package only answers queries on the user's (current) level.

1. Simultaneous Feedback (SF) & Electronic Corpora

Since 1997, the quick compilation of dictionaries within a sound

* Department of African languages and cultures, Ghent University, Rozier 44, 9000 Ghent, Belgium, gillesmaurice. / deschryver@rug.ac.be
G-M de Schryver is Research Assistant of the Fund for Scientific Research - Flanders (Belgium) and received the Laurence Urdang Award 2000, administered by Euralex, for implementing Fuzzy SF.

**Department of African languages, University of Pretoria, Pretoria 0002, South Africa / prinsloo@postino.up.ac.za
framework has been our main area of research. This work resulted in the introduction of the theoretical concept of Simultaneous Feedback (SF) (cf. e.g. De Schryver 1999, De Schryver & Prinsloo 2000a, 2000b) and bilingual dictionaries for Cilubà and Sepedi compiled within the framework of this methodology (De Schryver & Kabuta 1997, 1998, Prinsloo & De Schryver 2000a, De Schryver & Prinsloo forthcoming). In a nutshell, Simultaneous Feedback (SF) can be understood as entailing a dictionary-making method in terms of which the release of several small-scale parallel dictionaries triggers off feedback that is instantly channelled back into the compilation process of a main dictionary. As such the target users continuously guide the compilers during the entire compilation process, and the unabated retrieval of feedback can be considered as the main pillar of the methodology. So far, this retrieval of feedback followed the channels of such standard approaches as (natural) participant observation, formal and informal discussions, anonymous mail survey questionnaires, controlled tests, etc. Through a cross-comparison of the results of the various types of feedback, we attempted to arrive at a representative body of users' desires for each of our particular target user groups. Still, the realisation that none of the employed feedback methods is devoid of problems, and that even the balancing out of different types of feedback is only approximate, prompted us to seek a straightforward, automatic, neutral and invisible arbiter. We found that arbiter in the form of electronic dictionaries' (EDs') log files.

According to Moon, the use of electronic corpora "consolidated into standard dictionary praxis [...] over the period 1986-1996" (2000: 4). In our own research and our own dictionaries, corpora have always played a crucial role (cf. e.g. De Schryver & Prinsloo 2000c, 2000d, Prinsloo & De Schryver 2000b, 2001, forthcoming).