

Looking back: the first three years of the Lexicology Research Group*

Free University of Amsterdam,
(September 1989–September 1992)

Willy MARTIN

1. Introduction

The title above is not meant as an announcement of a new theatrical play, nor as an adaptation of an existing one, rather, it is purely prosaic and down-to-earth. When being asked to write about our “travaux récents dans le domaine de l’informatique et de la statistique (dans les sciences humaines)”, it seemed appropriate to me to take the opportunity to present what we have done up till now as a whole, thus providing a frame-work for the particular projects having been or being carried out. As a corollary then I will restrict myself to factual data of the projects in question, after having situated them in the general frame-work. What follows therefore will be a brief description of some computational projects situated within a particular research context with the intention to make a first exchange of ideas possible.

* The title of this article was inspired by a brochure distributed by the Knowledge-based Systems Group, University of Twente, announcing their first lustrum.

✉ Lexicology Subdepartment; Free University of Amsterdam (The Netherlands); 1105 De Boelelaan; 1081 HV Amsterdam.

Fax: + 31 20 661 30 54

E-mail: lexico@let.vu.nl

KEYWORDS: Lexicon, NLP-lexicon, lexicology, standardization of terminological description, machine-aided translation, indexing, information retrieval, re-usability of (bilingual) dictionaries, lexical knowledge base, parsing of definitions.

2. Research Approach

The approach to research taken by our group has been influenced by several factors. The major one being the fact that the Research Group developed from the teaching staff. In 1986 a Chair in Lexicology was founded at the Free University of Amsterdam, followed by the appointment of the undersigned. At the time only one collaborator was added, mainly for teaching purposes (terminology). The curriculum itself (a four-year-study) was called 'Lexicology' and had to integrate both theoretical and practical aspects of a study leading to such professions as lexicographer, lexical knowledge engineer and terminographer. The stance taken with regard to the lexicon in the curriculum development has had a direct bearing on the ensuing research. Summarizing and somewhat simplifying one could state that 'lexicon' as a research object, such as in teaching, has been interpreted:

- not as "a set of existing words such as in a dictionary",
- nor as "a set of existing idiosyncratic lexical items (morphemes and morpheme combinations)",
- neither as "a set of existing idiosyncratic lexical items (morphemes and morpheme combinations) together with a set of morphological rules (to form possible non-idiosyncratic morpheme combinations)",
- but as "an organized lexical knowledge bank needed by users so to be able to understand and produce language".

Next to this "internal" factor, a quite different, "external" one, has contributed to our group's research agenda. The fact viz. that the demand for *computerlexicons* became more and more stringent in the 80's, stressing such aspects as *extension* (the fact that computerlexicons should no longer be "toy" lexicons but able to cope with real life situations and so be large), *intelligence* (the fact that computerlexicons, next to knowledge w.r.t. "concrete" phenomena such as orthography and pronunciation, should also contain more "abstract" phenomena such as morphology, syntax and semantics/pragmatics), and *re-usability* (the fact that computerlexicons should be constructed in such a way that they can serve various users and lead to various applications).

From the above it can be inferred that our research is to be situated within the field of *computational lexicology/lexicography*, implying that

- *the lexical component of a natural language processing system* has played/plays a *central* role in it;
- such a component is seen as a *lexical knowledge base* entailing such aspects as *lexical knowledge acquisition, lexical knowledge representation* and *lexical knowledge application*;

