

SESSION 6: CHAIRMAN'S INTRODUCTION

Professor Juan C. Sager

UMIST, Manchester, United Kingdom

The title of this session suggests topics concerned both with lexicography and terminology, and with dictionaries, databases and their interaction.

We are therefore talking about applications of information technology for the purpose of dictionary production, terminology processing and retrieval, and, at the same time, about particular aspects of applied computational linguistics. From an academic point of view it is interesting to note how linguistics and information management increasingly find common ground in the exploitation of computer applications.

The four topics of this session grouped by the organisers under the question 'Dictionary or Database?', are less a dichotomy than points on a scale of access to and use of lexical reference tools. They contain consideration of the internal organisation of lexical material as well as the preferred modes of consultation of such material in either the conventional form of printed books or the various alternatives offered by computer output.

The whole problem area can be described as an increased involvement of the computer in lexicographic work with a progressive sophistication of the utilisation of computational linguistic techniques.

As a first step the computer can be used as a simple ordering and storage device in traditional lexicography. The computer is then used for the compilation of lexical entries and the printing of dictionaries. The user may be totally unaffected as he handles a computer type-set dictionary. Dictionaries are, however, dynamic collections of knowledge and the computer can reduce the effort of updating. Entire entries can be added or deleted, thus reducing the cost of new editions. When the individual entries themselves are considered as structured entities we require more sophisticated data structures; this allows us to speak of lexical databases and new modes of usage become immediately available. We can then provide new techniques of quality control and cross-referencing for the lexicographer and new modes of access for the dictionary user. A complex lexical database permits a publisher to combine data for separate dictionaries, e.g. a pronouncing dictionary, a dictionary of synonyms, a dictionary of scientific terms or smaller or

larger editions of reference tools for various levels of users. Online access to such a database can permit a user to retrieve information selectively according to pre-established search procedures or an ad hoc constructed user profile. He can search in various depths, e.g. translation equivalent only, or this plus a definition, to make sure of the right choice of expression. The difference between a lexical database and a term bank is then only a matter of different content, data structures and combination of data elements in retrieval. The observations made by the various speakers in this panel session are therefore directly relevant to the organisation of term banks.

Lexical or terminological databases and the associated technology now open up possibilities for research and applications.

As increasing amounts of machine-readable text collections are available it is now practicable to base dictionaries and glossaries on text corpora rather than on the individual choice of the lexicographer or terminologist.

Compilation techniques can be assisted by question-and-answer systems, thus ensuring greater consistency. The interactive mode of dictionary consultation can also benefit from question-and-answer systems to guide a searcher through the full range of information available. The most exciting new developments are, however, likely to occur in the modes of representing terminological relationships. Systematically structured glossaries are not new but their realisation on a larger scale has always been hampered by cost and the sheer complexity of the task. The representation of terminological fields and networks has advanced considerably in recent years. Specialist translators in particular are likely to welcome the possibility of diagonal searches of databases which is, of course, also the most reliable test of the quality of a terminological database.