[*Translating and the Computer 9*. Proceedings of a conference… 12-13 November 1987, ed. Catriona Picken (London: Aslib, 1988)]

New tasks for terminology

Professor Juan Sager

University of Manchester Institute of Science and Technology, UK

As we are finding solutions to the terminological needs of translators and technical writers, new problems arise which require all the accumulated experience and skills of automated terminology processing. Previous conferences have demonstrated the crucial importance of effective terminology management for interlingual communication, especially in any environment depending on automatic mediation. The speakers during this session bring us up to date and present new aspects of existing services.

Terminology, like translation conundrums, is always with us. In fact, with every new method of communication new challenges are offered in the processing and retrieval of terminology which we have to address. As factual databases grow and as a greater diversity of people seek direct access to databases, we have to re-examine our methods of storage and retrieval of information held in the computer. Information in databases is particularly important in this respect because users have higher quality expectations than they had of manual services.

Individual information search in conventional sources was not error-free, quite the contrary; but there was a self-correcting mechanism at work. As many information searches led to duplication of information, there was a checking mechanism built into information supply and the user was in the habit of looking out for confirmation or contradiction.

Here we make a startling discovery: we note our ability to understand a great diversity of expressions which the computer does not. We attribute this to the natural redundancy of language and to our ability to deal with human errors of coding.

The computer, on the other hand, cannot distinguish correct data from erroneous data of whatever kind, unless it has been instructed in how to do this. The machine takes a hyphen as one character and a stop as another character, whereas we are able to distinguish between several functions of punctuation sign.

We therefore find examples like the following in the indexes of databases:

CROSS HEAD DISPLACEM	CONST STRAIN RATE
CROSSHEAD DISPLACEM	CONST STRAIN-RATE
CROSSHEAD DISPLACEM.	STRAIN RATE
CROSS HEAD DISPL	STRAIN-RATE
CROSS HEAD DISPL.	
CROSSHEAD DISPL	
CROSSHEAD DISPL.	

In this way the database in question insures itself against spelling variations while providing maximum retrieval. The database user gets the information he or she wants but is further confused in his or her appreciation of the spelling vagaries of English compounds.

As readers of running text we ignore the differences; only in interrogating a database are we made aware of such differences. As writers of database input we are inconsistent in precisely this area of writing. The quality of proofreading of database language clearly leaves much to be desired.

Here are some real examples of database language:

Chemical formulae:
CO for Co
H2O for water
S102 for silicon dioxide

adjectives:
foamable
polishable
ippable
spbayable
elect. depos. machinable

(Some of these words are exclusive to the database; users who do not know or cannot interpret their meaning have no possible assistance since these words are certainly not in a dictionary.)

abbreviations:
MFR – manftr – manuf. – manf.
subsidrs – subsids.
industs – industrs

(all these were found in the same database)

cntrl – contr.
str for strain
strength
stress
const mean tens stre

It is sometimes impossible to tell whether a term has been shortened, e.g. is ‘long transverse maximum’ an abbreviation in a set of terms including ‘longitudinal maximum value’ and ‘transversal value’?

It is clearly not yet understood that the restrictions that we impose upon language in order to provide more concise information also entail a responsibility of greater accuracy. The reduction of redundancy can only be achieved by greater prescription of usage.

If these problems lead to ineffective communication in one language, it is readily appreciated how much greater the problem is when communicating across languages.

The problem terminological advisors to databases must tackle is a triple one: teaching to write for the database, teaching to interrogate or interact with the database and teaching to read the output.

In writing we must examine how far we can assist input by overt controls in the form of interactive editing or to what extent it is preferable to covertly rectify input errors and only bring them to the attention of the writer when there is an error or ambiguity the machine cannot resolve. Our enquiries have, however, also shown that data input is frequently carried out by totally untrained personnel and that the basic precautions of proofreading are not taken.

Providing assistance with interrogation will probably vary considerably with the user-friendliness of the database and the tools available. The existence of upper and lower case alphabets and other typographical devices in both the database and the datacarrier, the use and availability of characters of other languages, all contribute to the success of information retrieval. User manuals and user guidance systems must be available in all user languages, even if the database is not, because it is at the level of particular conventions that so many searches for information are frustrated or aborted.

Reading database files should be made easier if database input is at least as carefully controlled as material for quality publication. In the meantime and until this happens the only help that can be provided is an appendix to the user manual which explains the abbreviations and other idiosyncratic spelling devices used in the database.