

## PRESENTATION AND DISCUSSION OF PAPER 26

### PRESENTATION BY M. LECERF

M. Lecerf emphasised three points.

Ambiguity exists in translating. Its resolution is perhaps a matter of choice, but if one decides to resolve, then adequate information for resolution must be fed into the machine. He instanced the English translation of his own paper wherein "normal" was the equivalent given to **normative** (Fr.) and "imperative" the equivalent for **imperatif** (Fr.) instead of the preferred "normative" (surely an acceptable English word) and "requirement" respectively.

By "practically infinite" grammar he meant a grammar that he couldn't afford to pay for - a grammar that would take a million years to compile by asking questions of a native speaker.

The "price" to be paid for an item of grammatical information in a machine depends on three things:-

- (i) cost of address of the first component linked in the item - word, syntagm or group of words
- (ii) cost of address of the second component linked in the item
- (iii) cost of stating the relation between the two linked components.

The cost goes up if the components are groups of words for then a more complex address is needed - the addresses of the first and last words of groups are needed. Also, in general, the cost goes up as the distance between the linked components goes up. Dodges such as drawing diagrams like those of Tesniere, Hays or Ceccato bring linked components artificially closer together, and rules are thus cheaper, but more machine time is taken. There must be a compromise between expensive rules and expensive machine time.

### DISCUSSION

DR. BOOTH asked what is M. Lecerf's precise application of intrinsic curves of geometry to machine translation.

M. LECERF answered that several programmes using intrinsic addressing had been written and were in use at Euratom. Dr. Hays and Mme. Hirschberg are doing their own experiments with it, as also is Prof. Ceccato, whose needs for cost saving are greatest of all, as he is writing a semantic grammar. M. Lecerf offered the colourful analogy of an iceberg to a language. The visible part of the iceberg is likened to the normal grammar of a language; the vast bulk of the iceberg below the water line is likened to the semantics of the language, which Prof. Ceccato is attempting to describe.

PROF. CECCATO agreed and, to emphasise the point, considered the following sentence (Fr.):-

J'ai acheté un piano et un violon que je vous ai apporté immédiatement sous mon bras.

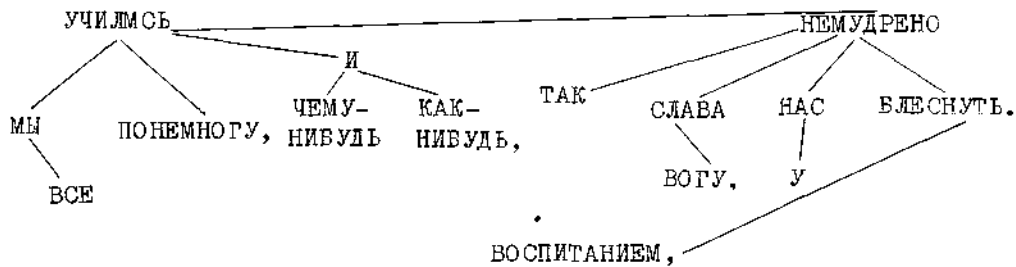
When the pronoun (que) is used in this way there is no possibility of inflection (in English, French or Italian), and it could refer to "piano", "violon", or both. So imagine the amount of information which must be inserted into a dictionary to indicate that it is only the violin which can be carried under the arm. All physical objects would need a "carryable-under-the-arm" indicator, or perhaps, more generally, their weight range would need recording. In other than scientific writing, three or four instances of this sort of ambiguity are met with on each page. Whilst thus emphasising the huge problem of semantic description, Prof. Ceccato was optimistic that it could be possible to solve it if circularity could be introduced into the semantic links described.

M. LECERF added that he thought that people doing high-quality machine translation and people doing low quality machine translation are not dealing with the same problem, and so cannot help one another much (the seemingly-inevitable analogy from M. Lecerf this time took the form of a refrigerator engineer talking with a cryogenics physicist used to working near 0°K). The yardstick here is how much error is allowed. To get a near error-free system will involve spending lots of money.

MME. HIRSCHBERG gave an example of a non-projective sentence stemma which she thought illustrated M. Lecerf's ideas. The sentence is:-

МЫ ВСЕ УЧИЛИСЬ ПОНЕМНОГУ,  
ЧЕМУ-НИБУДЬ И КАК-НИБУДЬ,  
ТАК ВОСПИТАНИЕМ, СЛАВА БОГУ,  
У НАС НЕМУДРЕНО БЛЕСНУЪ.

and the non-projective stemma for this sentence is:-



It is a quotation from Pushkin. The hypothesis of productivity\* is used to reduce the search for various bonds between words in a sentence. With non-projective sentences, however, one has to define the criteria by which non-projectivity may be recognised, and then find rules which will minimise the search for the components of non-projective links<sup>†</sup>. Specifically, with regard to the sentence above, ВОСПИТАНИЕМ is the non-projective "intruder". Mme. Hirschberg's rules for finding the governor of this word begin by eliminating the sub-tree of the stemma deriving from the word (in this case, only the word itself) and then completing the non-projective stemma. The governor of ВОСПИТАНИЕМ is searched for amongst the words in the sub-tree deriving from the "highest" word in the tree which can have a projective link with ВОСПИТАНИЕМ when it is inserted into its correct linear position in the stemma, ignoring grammatical requirements. In this case, НЕМУДРЕНО is the "highest" word, so the governor of ВОСПИТАНИЕМ should be searched for in the sub-tree:- words 8-15. The preferred *order* of search amongst these words may be to consider the nearer (in sentence order) words first, or to consider first the first level descendants of the "highest" word, then the second-level descendants and so on. This preferred order is currently being investigated with many examples of non-projective sentences. Mme. Hirschberg appealed to delegates to send her such examples to help her in her work.

---

\* D. G. HAYS, Basic principles in sentence-structure determination, RAND Corp. report - P1984, May, 1960.

Y. LECERF, P. IHM, Euratom report Grisa No. 1, April, 1960.

† L. HIRSCHBERG, I. LYNCH, Discussions sur l'hypothese de projectivite, Euratom report Cetus No. 35, October, 1961.

DR. HAYS pointed out that intrinsic is simpler than other forms of addressing because dependent bonds connect pairs of words and these can be listed in a table, but it is not practical to tabulate pairs of syntagms.

*J. MCDANIEL*