

Session 2: CURRENT RESEARCH

SUMMATION BY CHAIRMAN

CANNON: I am confronted with the situation which was well described by Professor Dostert this morning. I do wish to thank the speakers for their presentations. If one attempts to classify approaches to a machine translation or avenues of attack to problems, he at least might place at the two extremes what I will call the student approach and the teacher approach. I think both have been in evidence. The student approach, I would say, is the approach of an investigator who pretends he knows nothing whatever about the rules of the source language, and who uses the computer to begin trying to translate. He develops rules which he thinks will be peculiarly applicable to computer manipulation on an ad hoc basis; he expects those rules to appear less and less of ad hoc nature; and he hopes that eventually he will have a body of rules which will be rather widely applicable to machine translation. In the teacher approach, one presumes that he knows enough about the rules of a language to prepare, in advance of any experimentation, a body of rules incorporated in the program which will yield at least a workable first approximation to translation; and he presumes he will go on from there.

Whatever the approach, I think that investigators in all the MT groups recognize five main phases: dictionary lookup, morphology, syntax, semantics, and programming. Of course, the emphasis of different MT groups is different. We have heard almost every speaker today devote a considerable time to syntax, and have heard some mention of polysemia. I will use the five phases as an outline for some brief remarks.

With respect to dictionary lookup I shall say very little except to observe that we have heard different approaches. Some investigators prefer the stems to be stored in the dictionary. Some would store complete words and groups of words. Both approaches have merit and I shall not attempt to compare them, except to say that I think they are intertwined with the methodology that the particular MT group is employing to obtain machine translation.

Everyone realizes that fair attention must be given to morphology, whether the translator is a machine or a human being; and every group

Session 2: CURRENT RESEARCH

uses morphological analysis to some degree and to some extent as a basic element of its method.

With respect to syntax, we have seen different approaches. We have observed the sequential, or systematic, approach to syntactical analysis, taking words one by one as they come, with use of hindsight and predictions. Then we have heard of the fascinating and different analysis of a sentence by analyzing the articles, the verbs, and the location of the adjectives. We have also heard of analysis on the basis of groups of words rather than on a word-for-word basis. There was the consideration of sentence-by-sentence transfer, which I might call a mapping procedure, mapping the source sentence into some descriptive framework. Perhaps one could call it descriptive morphology or grammar associated with a source sentence, which is then interplaned against the same part of the map of all possible target sentences. Then we have heard a consideration of natural order. The investigator feels that by assigning codes indicating priority in sequence to the classes of occurrences he can achieve the result that this code, plus other features in the process, can be used to break the sentence into segments which can be analyzed separately and successfully. Some MT groups have indicated they have considered only the phases up to syntax.

With respect to the polysemia problem, I feel that we are in the position of preparing to deploy our forces to attack the problem. There was a description of the polysemia arising from subject ambiguity or from environmental ambiguity. With respect to the latter, the investigator stated that his basic tool for coping with it was a code pattern consisting of a boundary number and semantic and morphological codes. It seems to me that this could work out very well in the case of prepositional phrases. There are other places that might have this difficulty with polysemia. I feel that the difficulty in semantics transcends those of the phases that I mentioned before.

With respect to methods of programming, I was certainly pleased to hear mention of automatic programming techniques applied in linguistic research and in attempts to achieve machine translation. I feel that one of the things that has been acting as a brake to retard progress in the field in the past has been the lack of articulateness in communication between programmers on the one hand and linguists on the other. The two groups must collaborate on this problem.

Session 2: CURRENT RESEARCH

I feel that effort in this direction is well spent. Certainly it will tend to eliminate possible duplication of effort and will channel our efforts into productive goals.

I do hope that people will accept the challenges which were issued this morning. Again, I think that if we could compare methods on the same samples this would tend to accelerate progress, because it would certainly be one of the best ways of arranging or achieving fruitful communication among the groups.