QUESTIONS AND DISCUSSION

DOSTERT: In order that the discussions be properly recorded, members of the audience are asked to formulate their questions lucidly and succinctly. I can repeat the questions if they are succinct, but if they are more elaborate. I invite the speaker to come to the platform. The meeting is now open for discussion on the papers and reports which have been presented this morning.

LAMB: I had a specific question to ask Dr. Hays. When can we expect that the book will include a grammar and a dictionary?

HAYS: The MT project at RAND will never create an adequate dictionary and grammar for any language. We are limited to the kind of research that can be done with a small professional staff. We believe the grammar books and dictionaries that are going to be adequate for automatic language-data processing will be based on extensive empirical work, which requires not only the best available professional guidance--whether we have that or not, I could not say-but also a sizable staff of sub-professional help. Although I cannot speak for RAND on this, I can predict that The RAND Corporation does not want to engage in such a task. We regard this as a matter of the greatest importance, but not a task for ourselves; so we are not going to produce dictionaries and grammar books. The intermediate results that we have are available to anyone who can make use of them, but their publication calls for more work than we can do.

DOSTERT: The reason I am interested in this question is that at the beginning you said you had originally intended to include in your book a section on grammar and another section on the dictionary. I did not mean to imply that it should be total and transcendant on the first printing, but I want to know whether you are pursuing this or whether you have definitely and permanently abandoned it. I think that was what Dr. Lamb was seeking to find out.

HAYS: What we plan to do falls into a sequence of categories. We certainly will continue processing text as we have done in the past as long as we are engaged in this kind of research. We also intend, as we process more text, to improve the descriptive grammar of Russian that is embodied in tables that are used in our translation system. We intend to elaborate our sentence-structure-determination routine, but without emphasizing the output system, the construction system for English sentences. We hope to produce machine routines for performing transformations and transformation-like operation sentences whose structures in the first language are known (that is, have been obtained by a computer routine). We hope to make empirical studies of transformations and distributional semantics. I said only a few words about those two topics; I think they will occupy most of our time, but I am unable to say in detail how we shall carry out those studies, except by looking for pairs of types of connections (pairs of syntagmas is the Russian term--or pairs of configurations) that are used with equivalent pairs of words. We hope to find empirically equivalent syntagmas or equivalent configurations by empirical studies of what nouns are used for subjects with what verbs, what nouns are used in the genitive case with what other nouns, what adjectives are used with what nouns, and so on. We hope to begin establishing semi-semantic classifications.

HUNT: My question is directed to Dr. Hays. What is the relation if any between your "dependencies" and the "foresight-hindsight" routine that Mrs. Rhodes speaks about?

HAYS: It appears to me that each of her predictions is equivalent to a potential dependency connection, but that a prediction may be a prediction of a governor or a prediction of a dependent. As she said, when you see an adjective you expect it to have a governing noun, and when you see a noun you expect it perhaps to have a dependent participle, say, following it. The connections that are established by these predictions appear to me to be dependency connections. 1 think that Mrs. Rhodes' syntactic theory is very close to ours and thus different to some degree from the more frequent phrase-structure conception of grammar, and that her algorithm is considerably more like the one that we have in operation than like any other in the field.

HARPER: I would like to ask Mrs. Rhodes why, in view of the fantastic machines that will be available in 10 years, it is necessary to devote so much work and effort to the building of stem glossaries and to the analysis of glossary entries? The amount of storage available is apparently one of the reasons that you do not store glossary entries but we are going to have them sometime.

RHODES: I have to have a glossary to work my sample questions on. Why could I not do two things at one time? I might just as well make some experiments to see how much I could save. In the future I shall probably do better, but it did not take me any more time to do the right thing than to do the wrong thing. Why do the wrong thing when it does not take any time to do the right thing? So while I knew it was going to be temporary, at the same time I tried to do the very best I could while I had the opportunity.

RICHINS: I would like to ask Mr. Mersel what he meant by the term "computer transliterations" when he was referring to his input words.

MERSEL: We are translating from text systems printed in Cyrillic. Unfortunately our keypunch does not have Cyrillic characters on it, nor does our printer, nor does the computer internally have anything in it but bits. We find ourselves going through a series of transliterations. First of all, a keypunching is done not directly on an 026 keypunch but on an 826, which is a regular IBM typewriter hooked to the 026. On the typewriter we put Cyrillic caps and Cyrillic type. The keypuncher types on this typewriter; and since we have put it in the regular UN Russian typewriter format, and this is connected by the ordinary English typewriter format to the keypunch, we get a completely unconnected set of Hollerith characters for our Cyrillic characters. Taking these into the computer, we then change them into bit patterns; and since Russian has only 32 basic characters, we use the octal numbers (bioctal numbers really) from 40 to 77 to represent the alphabet, and all the earlier 32 octal combinations to represent numerals and special symbols. Coming to the printer, we go through still another transliteration, for we attempted, as much as we could, to make the Russian letter look like the English sound;

that is, when we have a Cyrillic С , it does not appear as a "C" , it appears as an "S" . This is not carried through completely, since we are still trying to retain a one-to-one correspondence, so we have such things as a "9" standing for “Ф” .

MELKANOFF: I would like to ask the speakers if they have any special opinion regarding the use of an interlingua, or whether any research has been done in the field or is currently going on.

RHODES: You have noticed, I hope, that I am the only person who never uses the word "algorithm", because I am not always sure what it means. Nor do I know much about linguistics. Now, I have always wondered why we should be bothered translating from Russian into English, from Russian into French, from French into English, and so on. Why can we not have a beautiful language--I do not want to use an interlingua unless we mean by that some kind of a new language. It cannot be Interlingua or any of those that have been already made, because they were made for human consumption. I mean there is something that an interlingua will not do for the things that I had in mind. It has to be an unutterable and invisible language, which is so perfect and so marvelous that no human being would want to bother with it, but the machine does not care.