[From: Allen Kent (ed.) *Information retrieval and machine translation*
(New York: Interscience, 1959)]

CHAPTER 38

# Discussions

As reported in Chapter 22 (Volume I), the conference was organized in such a way that little or no time was available for discussions following each paper. Rather, a period of time was devoted in the evening to discussion sessions. The ready availability of microphones, and the fervor of the discussions sometimes made it impossible to identify the discussant. When this material has been retained here, it is identified as (Floor).

The following comments were made by participants at the end of the second day of papers (Chapters 23-37). Discussion following individual papers had not been permitted. The Chairman of this session was Dr. A. D. Booth, Director, Numerical Analysis Laboratory, Birkbeck College, London, England.

PARKER-RHODES: If I might make one or two comments on the work of the Russian workers that we heard about today. One point which I would particularly like to mention is that several of these groups appear to be working with the technique of an intermediate language. This is one which interested me quite a lot because its technique has also been used by the Cambridge Language Research Unit—and I think in the same way as it is used by the Russians. I would like to ask one question which perhaps Miss Martynova would be able to answer.

There are, I think, three types of intermediate languages which could be used and I'm not sure which one they are using. There is first of all the possibility of having as an intermediate language a purely arbitrary code in which all the ideas or the word uses of either language—of any language if you make the thing completely thorough—is simply represented by a number. I don't think this is practical—I don't think anyone thinks it is practical, because the number of units you would require would be pretty colossal and it wouldn't really help very much—you might just as well use the alphabetic code or the input word or the output word and leave it at that. But at least this is a possible way of doing it. The second, slightly more refined, method would be to use, not an arbitrary encoding, but to reduce the meanings of all the words, if this could be done (it has been tried and to quite a large extent it could be done), to some structure based on a comparatively small number of key words—something of the order of 50, 100, 200, perhaps, which could be listed and each word of your input would have to be represented in your intermediate language by a thought—a thought formula—consisting of a number of key words connected by

appropriate connecting links. The difficulty here is that the connection links would be fairly numerous; you have a rather complicated type of formula and the whole system would be rather impractical mathematically. An example of that would be (I don't know whether the word is current in American)... in Britain, we have a thing called a percolator which is used for making coffee—and if you don't happen to know what the word percolator means—to replace this by coffee making apparatus would be understandable... you would have for coffee something like alimentary liquid no. 1792. The making and apparatus could both very well be in your vocabulary. In addition to that you ought to have some sign, some indication that the apparatus made the coffee and not the coffee which made the apparatus. This may be obvious if you have got all the meanings pegged, but it wouldn't be obvious if you were to use a simple code. So we have got a complication.

There is a third possibility which is the one which rather intrigues me. It has never actually been done, but I think it could be done. That is to devise a system in which there is only one connecting link in which the mathematical structure of the system was simple enough to be the basis of a straightforward algorithm in the machine program, but which nevertheless would carry the meaning adequately. For this purpose you would have to cut up the units of meaning to give you more—you wouldn't be able to say coffee making apparatus—you would have to have a separate word for coffee as the subject of the operation, and coffee as the object of the operation and similarly a good many of your terms would have to be duplicated and you probably have to have a great many supplementary terms added to your formula, and some of your formula would become quite complex, and coding problems would arise as a result of that. But it might be possible to devise such a system whereby this intermediate language could be the basis of a straightforward machinable algorithm for carrying across the meaning of the input sentence into the output. Would it be possible for our Russian colleagues, if they are here, to tell us which of these three levels they are working on when they use their intermediate languages.

A. D. BOOTH: I am very interested in that contribution. I am not so sure I wouldn't like to claim paternity for it, for I suspect that I suggested the fallacy that the best idea is to use a meta-language because this reduces the p(p-l) language pairs which one normally wants to 2p. I noticed the old fallacy coming up again this morning. It is a fallacy because if you use any one of the base languages as your meta-language then you reduce your 2p to 2p - 2. This reduction of two may indeed be trivial but it is probably worth doing. As a purely philosophical point, I suppose, it would be rather nice to invent a meta-language but I would like to hear what the Russian workers have to say in answer to Parker-Rhodes' remarks on this.

JOHN MELTON: Perhaps since some of this seems to have been my fault in presenting Dr. Andreyev's report, I might be able to clear up a little bit about this for the people in general. I have already had

one or two questions about just what I meant when I was talking about the para-language, the meta-language, and the ortho-language, in regard to the symbolic language. Apparently I didn't quite make that clear. That might be some help to everyone here. By the para-language, as I understand it, subject of course to Miss Martynova's correction —Professor Andreyev refers actually to what we might think of as an analogy to the punched card. It is the form in which the extra language, as he says—the input or output language is handled rather directly in terms of the machines. The meta-language refers, I believe, to what Mr. Parker-Rhodes is thinking of as the true meta-language—a kind of informational code as it were in which the extra language is handled in the machine while the ortho-language refers to the information about the algorithm with which the machine is programmed. I don't know whether that clarifies it for the rest of the group or not, but I feel that I should do something about clarifying Dr. Andreyev's paper.

de GROLIER: It seems to me that there is a very important point in the two reports of our Russian colleagues, and that is—it seems that their efforts, or part of their efforts—are in the direction of a common language or intermediate language for machine translation and information retrieval. So they are just on the way for point 2 of the outline guide. And I would be very desirous to know more exactly how their efforts are going from the intermediate language for machine translation to the information retrieval language, as it is said in the last part of Andreyev's report.

The second question I would ask is that it seems that there are two different schools in Russia, or more exactly 2 series of schools: A series of schools which are promoters of the bilingual systems of mechanical translation and the other series of schools who recommend the use of an intermediate language. I would like to know what is the situation between these two series of schools and what is the tendency. In the West it seems to me that the first efforts, Reifler, etc. were on the direction of an intermediate language and after that we went to the bilingual system. It seems that in Russia there is exactly the contrary. It would be interesting to know why and how.

BOOTH: I might perhaps make one remark myself in this connection. I suspect that the difference between these two approaches is more illusory than real. Perhaps it is again in the form of words in which they were put. You can describe something as an intermediate language—whereas in fact it is merely a computer coding. I was surprised to find two groups which appear to me to very much overlap. We are accustomed to this in the West but I did think things were much better organized in Russia.

SOLOMONOFF: I think I have a tentative answer to your last question. I think that in Russia the emphasis is on trying to make intertranslations between perhaps 10-20 different languages and there the difference between $n^2$-n and 2n is a lot larger than it is in other countries.

GLEITMAN: I would like to raise a somewhat different kind of question. Mr. Booth, when he raised the question a few minutes ago

of why we might need a detailed computable recognized grammar of English, answered himself by saying that English teachers might make use of it. This is not our answer. It is the opinion of some people here with the notable exception of the Russian speakers and the exception of the people from Georgetown University—that mechanical translation is possible without detailed syntactic knowledge of what you call, unfortunately, the source and target languages. I am not willing to say at this point whether I think mechanical translation is possible at all. But I am, along with most linguists, convinced that it cannot be done by any word by word analysis—no matter how finely the semantic concepts are defined. Without syntactic analysis a semantically clear dog biting a semantically clear man is still liable to be translated as a man biting a dog. Therefore, we would claim that if you are seriously concerned with mechanical translation, a very practical though difficult question you must first solve is the syntactic structure of any languages that you are interested in translating.

BOOTH: That is a very interesting remark. It is the usual fallacy into which linguists fall when they discuss machine translation. I might perhaps remark—as I have had this question I should think some hundreds of times before—that there are two levels. In the first place, if one wants perfect translation I don't think that any people working in this field would dispute that you want the syntactic analysis before you start. I certainly wouldn't. Neither do I think that the more level-headed of us would maintain that you could made "perfect" translation by machine anyway. There is a theorem in mathematics which casts a certain doubt on the matter. On the other hand, to dismiss the usefulness of word-for-word translation is doing it an injustice because many scientists are quite content with word-for-word translation which they find extremely useful. In a survey which was made of this—many people in England maintained that word-for-word translations if they were available would be very, very welcome.

FAIRTHORNE: I was slightly surprised to find that an expert in mechanical translation doubted whether it existed. I, who am not an expert in any kind of translation—am in fact, the world's worst linguist—the only language I know is mathematics, and that not too well. I have, of course, used mechanical translation devices with great success as most of you have. In other words, a phrase book. That is entirely an absolutely mechanical—and works very well over the particular field it is meant to apply to. I agree that in its current form it probably isn't suitable for a larger field, but it is definitely mechanical translation. The headings in a phrase book can be pure numbers—they have no meaning. A man could look at one column and indicate it to somebody speaking in another language, if that is the column he is interested in; and from this sentence you may read off the other sentence. As I say I don't think it is quite practical to list all sentences in a language, but nevertheless, in that particular context it works, so that certainly there is nothing intrinsically impossible about it. So I am a little distressed to find an expert doubting the fact that you can mechanically translate, because you can, and you do.

HIŻ: It would be very profitable if we can read details of the linguistic analysis behind various proposals which we hear now described in general terms. One really does not learn—from our colleagues of Soviet Union—by listing how the linguistic laboratories are organized. We invite you cordially to present detailed grammatical analysis of the translated text so that we can examine them and profit and learn from them. The same applied to Mr. Booth's book on "Mechanical Resolutions of Linguistic Problems." There are not enough technical details presented in this otherwise very interesting book. But this book as well as papers which we hear from other colleagues stimulates our curiosity without really satisfying it. This can be solved, in my opinion, only if we can examine all of the details of the analysis of concrete texts. Second, it is very regrettable that many people who work on mechanical translation use military terminology, like "target language" or "strategy." These terms are neither necessary nor pleasant. More estatic terminology should be worked out with very little difficulties.

BOOTH: All I can say in the idiom is: "Them's fighting words." I can't, of course, answer for the Russians, who can answer for themselves, but as far as the book is concerned, it is a fairly long book and I think the publishers were a little bit well fed up with the length of it anyway. If you can do something about getting a publisher who will publish a book of let us say 1,000 pages on the subject and make a considerable loss, then no doubt you can have the details. The only other thing I can suggest is that perhaps you can spend time in our laboratory and see details of the methods yourself. We always make this offer, but are seldom taken up on it; but we would be prepared to provide anyone who wants it with a program tape for doing the type of things which we do—and with a description of the program. This contains at least the syntactical processing. As far as the dictionaries themselves are concerned, as I mentioned in my talk this morning, the ones which we have made are limited because we have no intention at any time of doing practical translation. This is a job for a service organization. The making of a dictionary is not interesting research work. There is a fair sample of a dictionary in the chapter on German in our book; it shows the type of thing which one is going to do. A dictionary in French you should be able to make out for yourself.

SOLOMONOFF: I do have one comment. At the recent Paris conference, there was one Russian paper on MT that did give detailed analyses of some of the algorithms they used. I don't know whether they were complete or not, but they were very specific.

JOHN MELTON: I seem to be the apologist for the Soviet Union, tonight. There is one point I can't answer, of course, which Dr. Hiż requests in detail, but on the point of the description of the laboratory process as being the only thing that they. Soviet Union, has to offer, I assure Dr. Hiż that tomorrow there will be a paper presenting the MT in the Soviet Union in much greater detail. On the second report, that of source, target language, and so on, I might point out that my

understanding is that there is a committee of some kind to be organized talking about terminology and nomenclature in general, and perhaps such abhorrent phrases as that can be cleared up before the group departs to its various countries, cities, towns, for good.

MARTYNOVA: I want to thank my colleagues who replied so cordially on what has been said about our work. I want to mention that I think it is clear to all of us that in this brief report we cannot present satisfactory examples of the practical results that have been achieved in our work. But I want to say that the first thing that I kept in mind when I was speaking about all these problems was that they were more or less known to linguists present at the conference and, as I have heard just now (and I was going to reply the same thing), one of these practical examples was presented at the Paris conference this summer. It is up to those who have spoken to believe them or not—to believe whether they are completed or not. But they are practical and the work has been done, and I want to remark that they were done in our Laboratory, as a field of precision mechanics.

Another thing I want to speak about is that we welcome greatly the idea that was expressed and the discussion there to exchange information between our scientific workers, between our linguists and mathematicians. We have many publications in the Soviet Union that might be of interest to the people who work in fields of mechanical translation. You probably all well know the problems of cybernetics, where there have been published some papers by Kulagina et al., that describe the algorithms in the full form—not approaches to the problem, but the complete work that was done; the translation that was achieved; the results. And another thing, I would be very glad if people who are interested in the practical results of our work should first express their wish to get those papers that were published already. By the way, they can write direct to the people whom I mentioned and I have with me the papers on one of the conferences held in the Soviet Union and this paper represents all the addresses and all the possible available publications. You can get whatever information, and I'll stress the complete information, from those people and from those publications. I want to once more thank those of my colleagues who expressed interest in the work that has been done at the Academy.

BOOTH: I think that is a very generous offer. I would think maybe that one of the objectives of this conference—apart from our formal one—might be—if possible—to distribute a list of addresses of the workers in both the Soviet Union and in the Western World. I thing this would be an easy thing for us to do at this conference—and it might be a very worthwhile one for the future.

MARTYNOVA: We have already discussed the problems of interchangeability, so to speak, between linguistics and mathematics—the most complicated problem in the strategy of mechanical translation, and I am not a specialist in mathematical problems (as you may already have noticed I mentioned very often the name of Miss Kulagina who suggested her highly valuable theory of conception of such a theoretical approach.) Her work in this aspect is fully described in

many publications, so everybody who is interested in it can be informed. I must apologize that I am strictly a linguist and it may sound profane if I dwell on mathematical problems—I'll better restrict myself within linguistical limits.

BOOTH: I think that's a fair enough comment for Miss Martynova. We can't really expect her to answer for work which she is not personally conducting. I think I can make one observation on this matter myself which might be a contribution to international comity. Recently in the University of London we have had 2 Russian guests; they stayed with us for a period of 6 months; they worked with us and we were able to find out a lot of the work which they were doing. This was in the field of electronics, not my own subject, but I can say that the people concerned were very happy with the arrangement. I rather wonder whether something like this can't be done in the field of machine translation. Would it not be that one of the groups—Parker-Rhodes, for example, of Cambridge, could invite one or more of the Russian workers to work with his group. This would be a practical means of fostering the exchange of information. I don't know whether the Soviet Union can do the same thing but I wouldn't think it was past the bounds of probability.

S. ROME: In the interests of international comity I would like to observe that the term "source" is not military, and that the term "strategy" I would guess was originated by either Cliff Shaw or Al Newal of the RAND Corporation, and applied to strategy as used in chess—in chess playing machines. This leaves "target" and I don't like the word "target" for "target language."

BOOTH: I rather wonder about that because actually, in 'The Pirates of Penzance' one of the Gilbert and Sullivan operas, the major general sings about strategy, so its obviously before the RAND Corporation.

YNGVE: As far as I can recall, the history of these terms in the field of mechanical translation started at the MIT conference in 1952. A number of us here were at that conference. We discussed the matter of the terms "source" and "target" and we didn't like them. We spent perhaps 10 minutes trying to think of better terms but had no success. The terms had been used in one or two mimeographed papers. Later, the terms "input" and "output" were suggested as good alternatives, and we have been suggesting these editorially in the journal "Mechanical Translation," which we publish. Much to my horror, however, the terms "source" and "target" started cropping up in the linguistic literature, and here we are with them again.

ALEXANDER: In the enumeration of the work of the Soviet Union, there are references to work in other locations in the Soviet Union. I wonder if Miss Martynova might be able to expand a bit on the nature of the activities that are going on at the other organizations that were mentioned in the paper.

[Miss Martynova needed an opportunity to formulate a response and the question was repeated later, at which time she answered.]

ALEXANDER: May I add a comment to an earlier discussion? The

request for a great deal of detail—the kind of details that is necessary for a deep understanding and perhaps to profit by the activities of other organizations—is posed all through the field of using automatic machines. The question remains of how to create a library of completed programs and better yet a library of flow diagrams. Most journals will not accept the endless pages of such things. I know that most of these programs seem to run into the thousands of instructions or more. There is a pattern that is little used that I might recommend as a mechanism for handling this. In the physical sciences there is occasional need, in order to buttress the conclusions of an investigator, to refer to hundreds of entries of technical data which are not appropriately represented as graphs. There is an arrangement, I believe, through the Library of Congress to deposit as many pages of data as the author feels is essential to justify the conclusions that he has drawn. He is then free to reference this item in a way similar to open literature, except that only those who wish the gory details need bother to request a photocopy at the usual expense, in order to satisfy themselves of the details. Perhaps this is the same path that might be used for depositing complete programs and complete flow diagrams of the work that we have so lovingly executed and for which most journals seem to be unprepared to accept as legitimate text. I know that the Library of Congress people may be horrified at the prospect of the extra work that might come, but, perhaps, this pattern can be expanded and a depository for such things can be created. Since we are looking for standardizations of all sorts, may I recommend that we include this sort of idea in our deliberations at this time. I think it is unfair to depend upon personal contact between researchers in order to provide one or other access to the full detail. A young investigator coming up half a generation behind us may have equal justification for such detail and may not know the address of the particular individual to whom to write.

BOOTH: I think that is a very valuable contribution from Dr. Alexander. As a sort of justification for the lack of detail, I mentioned the difficulty of formal publication. We are at the moment in the course of attempting to publish the full program for our “MT 6” as we call the sixth version of our French procedure which is considerably later than the one given in the book and very much more detailed. But I may say that the editor of the journal concerned decided that he does not want to publish a dictionary of a 1,000 words and says, “isn’t it enough to publish the first 6 and the last 2 or 3?” He is not going to publish the rest of it. It is all very well to blame the unfortunate author, but it isn’t his fault. As far as the Library of Congress idea is concerned, it is an excellent one. It is as Dr. Alexander says, in common use in this country, just as our own Royal Society in England acts as a repository for unpublished mathematical tables and masses of data which are otherwise unpublishable.

There is another mechanism which is used, as Dr. Alexander knows, in the mathematical tables field, which is to announce that the existence of these data in a journal, usually ‘MTAC’—Mathematical

Tables and Aids to Computation—and then people who want the information can write in and get it at the usual price.

HIŻ: I would like to clarify a point or two. My appeal for more details was especially in the field of linguistic analysis of the texts which are translated or attempted to be translated. The feeling behind this appeal was first of all that I would like to learn how more experienced people attack the problem and secondly, because I think that the linguistic problem in mechanical translation is the essence of the difficulty and not the machine problem. We are solving on high speed computers differential equations only because for the last 200 years mathematicians solved differential equations on blackboards. I want to know the linguistic analysis on which the mechanical translation is based, first of all on the blackboard and only later on, to put it on the machine. This point presumably is the essence of our discussion and I would like to make the appeal that we are not asking necessarily for details of machine work, but we are asking for the syntactical analysis of language from which you translate and language into which you translate and for the grammatical comparison of the sentence structure. This is not a question of going into the details of codes and programming.

BOOTH: But as a mathematician might I remark that the fact that mathematicians have been working on differential equations for 200 years and that therefore we can solve them might lead us to suppose that, because classicists and other linguists have been working on language for 2,000 years, the results of their labor are apparent in the fact that we can't translate language on machines. I leave that with you as a thought.

GULL: May I add too, my understanding of Dr. Alexander's remarks before they are obscured in translation again. It is my understanding that the service of depositing additional information to articles in the Library of Congress is one which is sponsored by the American Documentation Institute, and has been underway roughly since 1935. There is, I believe, one difference in understanding between us, however, and this is that I believe it is at the option of the editor that the extra pages shall be deposited with the American Documentation Institute. However, it might be even more desirable if the authors had this option of their own origin.

BOOTH: I can't think of objections to that. One of the functions of an editor—and particularly of an editorial board—is that of refereeing papers which frankly would plug up the works of any organization that tried to store them in an uncritical form. It might make things a little bit difficult if you said any author might write to the Library of Congress and deposit his material irrespective of merit.

MARTYNOVA: I want to comment once more on the fact that all I was speaking about is the practical work that has been done at the Academy of Sciences. When I speak about our Institute I refer to one of the various approaches; and as to the practical examples after the linguistical schemes that can be shown on the blackboard, I want to reply that they were not only shown on the blackboard, but they were

used in practical work. They were put into the machine and that's where lies the achievement of this linguistical analysis; and I would be glad to present the papers I have in my possession now, where those examples are shown, and I'll recommend that my colleagues refer to those publications I mentioned before.

BOOTH: I think that is as fair an offer as we can have at a time like this. It is quite obviously impossible to present this information at an open conference.

PARKER-RHODES: It has always been regarded an essential part of the scientific method in scientific work that people's experiments should be repeated, or repeatable at any rate, by others in the field. Therefore, I think it desirable in principle that precise detailed accounts of what has been done, however lengthy, should be in some form available. But I think it is also a rather theoretical point, not quite factual at the moment, because people are not in fact willing to repeat other people's experiments. It's because these experiments are liable to be very expensive and laborious and most of us find it difficult enough to get our own experiments and tests through on machines and we just haven't got the resources to do other people's as well and we attempt to take them as read or as unreadable. I would suggest therefore, that this isn't quite so urgent.

As regards linguistic analysis, this is, perhaps, a different thing. As a mathematician rather than a linguist, I can't help thinking that whatever linguistic analysis is used must be implicit in the way in which the procedure has been programmed. I daresay the linguists will not agree with this, but if one did have a complete account of the programs, the linguistic analysis would be there, too.

BOOTH: I agree with Parker-Rhodes on both counts here. I particularly reinforce his first point, where he talks about the difficulties encountered by workers preparing this information—the expense of it. In England we don't often have the means of justifying any expense—instead we work slowly with small groups of unpaid people. If someone wants a detailed description of a program it means that the unfortunate person who is doing the work does no more work for a year while he writes it up. Many of us prefer to push our work on rather than write some of this detail. Of course, if workers on this side of the Atlantic really want this information, no doubt they could put up an adequate secretariat of British workers so that we can document our work properly.

ALEXANDER: My first question, I would like to re-address, as there were references to quite a few other research groups in the Soviet Union at the beginning of the paper and I wondered if we might get an idea of the nature of their work; whether there are some special features of the work in the other organizations.

MARTYNOVA: I would be glad to answer this question and repeat a bit what has been already been said in the report. Among those groups that are busy with the work on mechanical translation they are not only our group at the Institute of Precision Mechanics, but a group that I've already mentioned, at the Steklov Institute of Mathematics, where Miss Kulagina works; and another group at the Institute of

Information; and the whole staff of linguists working at mechanical translation at the Institute of Linguistics of the Academy of Sciences; at the University—at the Moscow State University—at the Institute of Foreign Languages; and so on. This list can easily be prolonged and I will be glad to give it to those who want to get it. As to a full description of their work, I would like to remark that this description will take twice the time and maybe more than my brief report on the whole work of the Academy of Sciences, because it concerns the details of their practical work, that's why I speak about the exchange of information about the publications that can easily be available. And you'll find all sort of information there, and if you want some information just now, I will be glad to give it to you on further discussion.

COOPER: I would just like to say that the details of machine program no matter how elaborate, don't allow me, at any rate, to infer what kind or what amount of linguistic analysis has preceded them. And I and some others would like to see more detail in the area of linguistic analysis. I might say that in times past, at any rate, some of our own Western colleagues have, I think, been as remiss in this matter as people elsewhere.

BOOTH: That's an interesting point; I'm not sure that I entirely agree. If you are given a computer program, together with the list of words, you can in fact find the degree of linguistic analysis by running it on the machine. This is a good test of a program. We usually take the line that we don't attempt to cover all the linguistic points, as mentioned this morning. One just hasn't the time or the effort, at least in a group of the size that I look after. What one in fact does is take the program and see how good it is on text. Then, if you find the particular examples arise with any frequency in which your methods are inadequate, or perhaps deliberately misleading—you do something to plug up the holes. Now this is, I am afraid, what I would describe as an "earthy" approach. It is certainly not the pure approach which the true linguist would adopt.

COOPER: It may very well be earthy; it is also quite expensive. Perhaps the issue is whether or not a linguistic analysis has any kind of primary spot in the whole business. And if it does, it may be that we will save a good deal of money. We don't have to put a program on a machine to test it if we can work out the linguistic analysis which is supposedly programmed. If this is deficient it conceivably means that we have saved ourselves the trouble of going on the machine.

BOOTH: Yes, this is an extremely interesting remark. But I find myself again speaking as a scientist in great difficulty to understand what it means. You say "linguistic analysis." Now what I would define to be linguistic analysis (perhaps linguists wouldn't agree with me) is a statement that a particular configuration of words or of parts of speech occurs in a certain proportion of cases. The only way in which you can really make an exact science of this is to analyze all of the possible sentences which can be constructed. This will, of course, take an infinite time and is, therefore, impossible. So what do you mean by it?

COOPER: Well, I think this much, at least, needs to be said. That

I for one do not know what kind of light statistical analysis will bring to bear on the problem. I feel that this is not a matter of statistics. It is not a question of in what proportion of a given sample such and such phenomena is present. In this respect, linguistic phenomena differ from most kinds of phenomena which are handled by physical science. This is extremely difficult for a non-linguist to accept. I think we all know this, since we are all linguists. We have all been speaking all our lives. Like the woman who applied for a job as a dietician and gave as her qualifications the fact that she had been eating for 40 years.

BOOTH: That is a very nice point. I don't mind in the least bit giving the source of our linguistic information. I'll tell you precisely what we did about French and mention the degree of complexity or the lack of complexity of the French program which we have written. I went in the first place to our department of French in the college. I asked the professor of French in what way French sentence structure differed most radically from English. And he said, "Well, I have to go and think about it." After about a week, he said, "Well, I think it would be better if you talked to Dr. X," which I did and he referred me to Mrs. Y on a lower level and Mrs. Y eventually came up with what she considered to be the most obvious differences between French and English. I mentioned these this morning—noun-adjective order, definite article and pronoun structures. These were obvious differences. Now I was prepared to accept this as a reasonable approximation about the 90% level, so we went away and programmed it. It is quite obvious that this is not going to lead to 100% translation of French literary text. We did try it on a variety of scientific text and it doesn't seem to make bad sense at all. If as a result of these trials we discover in the long run that some important French scientist or mathematician writes with different idiom and different constructions, we shall plug the holes in the program. This is practical and earthy machine translation as distinct from trying to codify a couple of thousand years' knowledge before you start doing anything. I said this at the MIT conference some years ago and I'll say it again, that machine translation will make progress not by everybody sitting down and waiting until they've got a perfect set of rules that will cover any given language but by somebody sitting down and writing a program which will work for the simplest cases of some languages. As Mr. Fairthorne said earlier on, one USES phrase books. One doesn't start by someone trying to rewrite the whole language. Linguists have been trying to do that for 2,000 years and they haven't succeeded.

PARKER-RHODES: I would like to challenge this last remark. I don't think its quite true to say that the alternative to stopping the holes in a not very good program is to sit down for several years and get out a set of perfect rules. You won't get out a perfect set of rules. But I do think that an important part of machine translation is to have an adequate theoretical basis. An adequate theoretical basis must be to a large extent mathematical in form and linguistic in context. But that is rather a difficult combination to produce. Nevertheless, I do

think that some years' work is in fact necessary to produce such a theory and I don't think that a machine translation program can expect to be sufficiently smooth and free flowing to form the basis of a commercially successful procedure unless it has a theoretical basis which is a good deal more closely articulated and more mathematically respectable than we have yet attained. And I do, therefore, hope that not everybody will go around stopping up holes in bad programs, but will try and sit down a bit and think about theory—I wouldn't say for several years, but at least for 12 months.

BOOTH: I would have thought that this was a most unscientific remark because it is very difficult to say whether a program is bad without seeing it. On the other hand, one might say that if you are forming an adequate mathematical theory, it should be a theory which is in touch with what you can do on computing machines. At least one of the theories which I have seen and which I won't bother to mention in detail doesn't appear to me to be very closely in touch with the realities of machine use. One has to tailor one's mathematics, to what you can do on a machine. You can do wonderful things in mathematics, you can have various orders of infinity that can be handled very easily on an abstract mathematical basis. You can't handle them on any computing machine with which I am familiar.

YNGVE: In the field of mathematics, when you want to solve an equation, you don't make yourself a program that pulls out a random number. You have a theory. You know what you are doing. This is the sort of thing that some people are trying to provide for language. When we speak, there is implicit in what we say a grammar. Now what we are trying to do is to make it explicit. This is what we need. I think there are some people here in this room, and in several other places, who are trying to do it. It's a very difficult job, a long term job. It won't be done overnight. It probably won't be done in a year. Maybe not in 10 years, I don't know. Those of you who are less patient, those of you who want to do something quick, will have to fall back on something less precise. You will have to fall back on a patchwork sort of a program, and if such a makeshift is worthwhile in the interim—I say more power to you. But there are some of us who are working with a long term goal and trying to do the job right.

BOOTH: I don't want Professor Yngve to take my remarks as a personal criticism of what he is doing at MIT, it wasn't intended in this sense at all. I said earlier on, if one talks about MT for a number of years without doing anything, the subject will fall into disrepute. I think these were roughly my own words, at MIT a few years ago. But, on the other hand, he has now brought up a mathematical argument in which he said one doesn't select a random number and then use it to try to solve an equation. Of course, as he knows as well as I do that this is precisely what you may do. If in fact you want to solve an equation, let's say in the degree 3, you may choose a number and then use this as a trial for successive approximation. And this is notwithstanding that we do have a formal solution of cubic equations.

YNGVE: But you have a theory that tells you this will converge.

BOOTH: This is certainly true in the case of some equations, but you don't always have a theory which tells you that your first approximation will converge. I may mention the case of trying to find a complex root, in which a method which doesn't always converge is frequently used on computers.

ALEXANDER: I am vastly encouraged by this controversy—it looks like we computing machine designers will be busy until our retirement.

SOLOMONOFF: There was one remark he made about these people in linguistics not being completely aware of the true limitations of machines and so forth. Well, I don't think there are any limitations in machines that we are really interested in at the present time. The difficulties in mathematics about infinity and such that really don't have much to do with machines—well, I think that people in linguistics don't get involved with that sort of thing; and all the work that I know of in MT is not of the sort where you would.

BOOTH: I think as a matter of fact that comment was probably based on the misunderstanding of what I meant. It's difficult to clarify it in detail without going into personalities. Perhaps I can quote a personal story from my own lab which at least keeps the fight in the family, so to speak. Brandwood, who was until recently a colleague of mine—and whom I hold in the greatest esteem—had the most odd ideas about what one could do on machines. He would say: Ah, yes, you do this if "sense permits." On an even more elementary plane, in the transcription of English into Braille, where linguists aren't concerned, the people at our National Institute for the Blind seriously suggested that we could mechanize a Braille rule which said that one makes the certain contraction in the event that the "sense permits." And they saw nothing in the least bit uncomputable about this. I am quite prepared to believe, with some persuasion, that linguists in America are all very enlightened men who never make this sort of statement, but in that case they are quite different from most of the linguists I have met in England. And these include some of the systematic ones.

HIŻ: I would like to make two comments. First, that I would like to clarify once more that at the University of Pennsylvania there was a program performed on Univac which analyzes the grammatical structure, in global terms, of every English sentence which is met in practical scientific tests. We think that this kind of approach—not necessarily the work which we have done, but at least in this style, is a necessary step for any mechanical translation.

BOOTH: Could I say something on that before you go on? You said that you have analyzed every sentence which could occur in any possible English scientific text. This means that you have analyzed all English scientific texts that have been written since the beginning of time?

HIŻ: We have tested about 15 pages of biochemical text; we are quite aware of several shortcomings of our program—and of our analysis. We were working only for a year, and we cannot, of course,

produce now a dictionary of sizable nature. However, our grammatical analysis seems to work for the majority of sentences which we met, and the sentences which were not susceptible to our analysis are, in practice, met extremely rarely. At any time we meet such a sentence we try to adjust our analysis, at least in principle, so as to include it in the program of the machine.

I would like to stress a second point which is of considerable importance. One studies a language because of the interest which one has in the language. Translation is interesting because presumably it will reveal something about the structure of the language. We are waiting to see whether for mechanical translation we will learn something about the fundamentals of the language. And not only to perform actual practical rough translation. Therefore, my appeals to people who have practice with mechanical translation is to try to draw the moral out of your experience. Try to learn about the essential features of the language; after all, this seems to me the essence of linguistic enterprise.

BOOTH: That is beautiful. I think it illustrated my point rather well. When I talked about the difference in the approach to linguists and mathematicians and scientists. You see here we had the statement that because certain rules apply sometimes to 15 pages of biochemistry they are universally true!

GLEITMAN: It is certainly true that even taking a linguistic point of view, the entire job cannot be done in one year—and since the program was very recently completed—it is very difficult for us to have run a great deal of material through the machine. Now we also know from running small amounts of material through the machine and running a huge amount of material through hand computation that the program as it stands handles most of the English sentences and that a very few linguistics changes in the program handle almost all of the rest of the linguistic problem. It is not the case that we expect at the end of the second year to find that our program does not cover say, 1 out of 25 sentences; this is presumably a very rapidly converging process, since the structures in any natural language, although they are many, are not that many, and we have certainly taken care of the majority of them already,

BOOTH: I find it very interesting that we—earthy exponents—of MT should be criticized for merely approaching linguists of say 40 years experience and asking what their experience was; and that this should be rejected as a useless procedure. Whereas a trial of a year leading to what amounts to your procedure, with modifications of the program in the light of experience, is apparently OK. Wonderful!

de GROLIER: I will return to a question I asked an hour ago on another point of view. The reason why we information retrieval specialists and mechanical translation specialists are here together is that our conference, as shown by the title of our conference, is an international conference for standards on a common language for machine searching and translation. The important word being the little "and." It seems that there are different opinions of that point—and

different methods. A paper of my friend Vickery, which was distributed on a limited edition, says that although there are similarities (Vickery will excuse me if I cite his text) between the retrieval and the translation problems and although each study can aid the other, Vickery does not believe that either will be had by a shotgun union—let us consider that common language problem separately. I made some remarks on the text and I am not exactly of this opinion. I must mention that this morning we had Mr. Melton—on a special kind of union between mechanical translation and information retrieval—using a special form of information retrieval methods with the scope of making a sort of second-hand translation. On the other hand—if I understand well—the method of Andreyev is exactly the contrary. They begin by making translations—mechanical translations—and elaborating an intermediate language; and from that intermediate language they hope they are doing effectively the work of passing to information retrieval. I had from our Russian colleague some interesting comments she made to me in private, and I hope she will make them in public, about the difference between two schools of Russian people about the questions of an intermediate language. But after all, the intermediate language of Andreyev is leading to a series of specialized information retrieval languages. There is another school of people—from the French Atomic Energy Commission who think that their work for information retrieval is a sort of a linguistic transformation in the sense of Harris or Chomsky, but a special kind of transformation—that is a "reduction." That would mean that the information problem is a sort of second-hand mechanical translation. I would like to see the discussion going on that point which seems to me that we are not a conference on mechanical translation but on mechanical translation and information retrieval for the scope of a common language between the two.

S. ROME: This is again an outsider's comment. The two of us have been having quite a difficulty trying to determine what problems are being solved. On the basis of about 20 seconds' reflection, I find at least 54 problems. The way you generate these 54 is as follows: you have 4 sets; 3 of the sets contain 3 elements; that makes 27; the 4th set contains 2 elements; that makes a total of 54. Set No. 1; are we talking language, formulation, or machines? Set No. 2; are we talking clerical details, engineering questions or theoretical a priori and salient questions? Set No. 3; are we talking para-language, meta-language, ortho-language? Set No. 4; are we talking the path to the machine or the path from the machine ? I suspect that much of the disagreement is only illusory—that different people are addressing themselves to different parts of the problem.

B. ROME: At the risk of appearing controversial with my respected colleague, who is also my husband, I don't find 54 problems at all. I rather see it as an old conflict between a purist, elegant, mathematical, a priori approach, on the one hand, which in the long run is really a far more efficient and productive proceeding, we can afford to wait for the results, and on the other hand—the age-old British,

empirical, Baconian approach which, while paying service to the fruits of light and placing mathematics on a high pinnacle, nevertheless seeks far more in the direction of the ant to get immediate results, because we cannot wait for the spider to complete its beautiful work. The empirical approach is always beguiling because it promises fruits of profit that are immediate. Unfortunately, in our present economic state of affairs, it requires a great deal of money to finance the empirical approach because it tends to be patchwork. It looks as though you get immediate results and you do—you will solve some problems. But if we are dealing with such complicated and expensive gadgets (and I heard a good deal of economics yesterday, and today), then I do think there is much to be said for the slower a priori method —with all due courtesy to our British colleagues for their demands for immediate pragmatic results; and by golly, I would like them too. If we have the money, let us go ahead, produce machine translations and get a program started. I happen to be in an organization where that is often possible; and when it is possible, I keep telling my colleague, who is my husband: let's get the program started. If it doesn't work, we'll patch it up. However, side by side with this economic approach, I would persuade, if I have to do so, the colleagues who are the pure linguists and the pure mathematicians to keep their eyes on that long, hard, elaborate job which really is not at all 2,000 years old—it's very recent, and will produce the rewards far more efficiently in the long run and outstrip our noted empiricists.

IBERALL: As a relatively innocent bystander, I would like to punctuate the previous remarks in a somewhat different fashion. The basic difficulty as I see it is not one between the purist and the empiricist, which is perhaps one characterization, but much more realistically the same sort of battle that exists all the time between the mathematician and the physical scientist. I am a physical scientist, so the battle is an old one. The point of issue is this: there are in the terms of the previous reference, purist problems that exist of a theoretical nature in mathematics; these are the problems of topology and analysis and differential equations and other problems of this sort. Some of these problems represent closed problems and others not; and these are certainly the problems for the purist mathematician. The only trouble is that the physical scientist has another class of problems: problems in a real universe. What he is concerned with theoretically is the strategy of game playing and what he does is to seek out a solution among all possible games that he can learn—how many games he has learned is mostly a question of his own experience and so on. Most often at various times he finds some games that are rather attractive and he says all problems are solved with Laplace transforms, or everything has to be cast in the terms of the electrical engineer or a block diagram or something of that sort; but the more profound physical scientist recognizes that he usually has a more open game. The only piece of solace which the chairman mentioned that the physical scientist has—is once in awhile to tweak the mathematician's nose and to remind him that we're not quite certain about

the strategy value of any long game, so that we play these out as we can. This is no criticism of the mathematician, but we play out whatever particular problem we have.

The discussion here is very much of that sort. The protagonists on one side—linguists, logicians and so on, who have specific areas in which they must develop closed sciences or open sciences—depending on their own problem as well as they can, but do not and can not have a basic right to discourage or even comment on what is needed. They can sneer all the time in passing, but in the games that are being played for the one who must fight the battles. He must use whatever strategy suits him, and the end result all the time is the same thing. If they can produce models which are reasonably accurate and consistent with the physical world that they have dealt with—they have developed good strategy.

BOOTH: I think that was a most level-headed comment. Calculated to reduce general over-all temperature. I can't help remarking in this connection that some years ago, the late Johnny von Neumann showed that chess is a determinate game. Having learned of this I was most impressed and I decided to read the "Mathematical Theory of Games and Economic Systems." I read and I read and I read right on till about 3:20 A.M. I learned a lot about all sorts of games until upon page 125; after much labor von Neumann had analyzed the game of chess and he said "as a result of the foregoing analysis I find that one may play a game of chess and the result is determinate; it may be either A a win, B a lose, or C a draw." I regret to say that although the book has about 500 pages in it, as a mathematical physicist I didn't read any further.

If you want to see the impiety of mathematicians I can refer you to the review which appeared in my last month's copy of the "Bulletin of the Proceedings of the American Mathematical Society." It is a review by Paul Burr of Margenau and Murphy's well-known and valuable book of Mathematics of Chemistry and Physics which I believe is well thought of in many institutions in this country. Burr has condemned it out of hand as a dreadful book. In fact he said that the authors were nothing better than "Pigin Mathematicians." I know nothing either about Messrs. Margenau and Murphy, and they may be "Pigin Mathematicians," but quite a lot of our modern technology is based on this pigeon mathematics. It may not be rigorous, but it does produce results.

FAIRTHORNE: In this connection, there are two things I would like to clear up. But in the first instance I might say that Margenau and Murphy sit on my shelves—I am a mathematician—and my copy is extremely well worn. It is indeed a dreadful book, but I use it a lot and I find it very useful. The word "dreadful" does not apply to its applicability—it probably applies to those who use it. But there does seem to be some curious idea that the devices known as computers have something to do with mathematics. Mathematics has nothing to do with computers, mathematics is the way you talk about things; and therefore devices have no say in mathematics. Computers have never

solved a differential equation—differential equations are not in fact accessible to computers, because they happen to deal with a continuum which is of course not a physically observable matter at all but a concept, and you cannot under any circumstances whatsoever make a physical model of a continuum. I may point out for instance a librarian need never use the number Pi as a serial number for a book. He can't. What a computing machine does is to push signals about according to rules. It so happens that, under certain conditions, the results are convenient to the mathematician. Speaking as one who has used computing machines for over a quarter of a century—there are many cases where they are not convenient to mathematicians. But of course you keep rather quiet about that. If people would just remember that when you are talking about computers—you are talking about a device for pushing physical configurations into other physical configurations according to rules, and sensible people try to make the configurations into which they are pushed have some practical application. But please don't talk about computers solving differential equations—they can't touch differential equations! At the best they can show the constancies of applying to a series of numbers represented as cogwheels or something of the sort, the constant in applying a law which we happen to call a differential equation; which is another kettle of fish altogether. And even then it takes somebody outside who knows what a differential equation is to know that the thing does represent it. There is an awful lot of difference between the interpretation of something and the thing itself. Otherwise you are like the electrical engineers—I hope we never shall sink as low as that. We talk about "information" and all they mean is a "signal."

BOOTH: I can't help telling my second elephant story after that. There was a young gentleman who was travelling up to Oxford by train. The young gentleman was sitting next to an elderly gentleman who every five minutes took from his pocket a peanut, wrapped it up in a piece of paper, and threw it out of the window. The young man from Oxford, unlike some Englishmen, was not standoffish, and was quite prepared to talk to anybody. He said to the elderly gentleman: "Excuse me sir—I don't want to appear rude—would you mind telling me why it is you take a peanut out of your pocket, wrap it in a piece of paper and throw it out of the window?" The elderly gentleman turned to him and said: "I don't mind telling you—a valuable piece of information for you. I do it to keep elephants off the track." And the young man said: "but there aren't any elephants on the track." and the elderly gentleman said: "yes, that proves my point."

(FLOOR):* I would like to make some statements—but before I make them I have to say that I agree with Dr. Fairthorne. When somebody says that we must have a theoretical study of linguistic structure, the same thing is implied there—the linguists have been studying languages for many years, as Dr. Booth pointed out, and their aim was just a description of the language as they observe it and it comes

*This notation is used to represent unidentified discussants.

under the title descriptive linguistics. The object is just to describe the language as it is seen. Now we want to construct a grammar for an entirely different purpose: a grammar which can be handled by machine; and therefore we are now asking for a different type of description of the same language, and the point which Dr. Booth made —that linguists, although they have been working for several years, haven't produced anything—is not quite satisfactory, because their object was quite different—their object was just to describe the language. Here our object is just to construct the grammar which can be used in a machine, either for translation or for any other purpose. So this is a new theoretical slant on the old problem and it is not that linguists have worked for so many years and haven't come up with anything.

BOOTH: There is a very good answer to this—because I didn't say anything of the sort. I think for the remaining half hour we ought to address ourselves to the subject of this conference, which is standardization, as you remember, not argumentation. We've had some jolly good fun so far as I have aimed to get you all at one another's throats, but I seemed to have succeeded more in getting everybody at my throat, but this doesn't matter. This is the misfortune of a chairman, and this chairman at least doesn't mind—in fact, enjoys it. And now on the subject of standardization. I'm sure some of you at least, may have remarks to make in this field.

SHINER: I would like to hear some comments about standardization as to input to an information machine. In other words—what is this difficulty in inserting the information into your machine. Punched paper tape, punched card, magnetic tape, whatever it may be, also what is the difficulty of the machine itself in performing your research in doing language translation. I can't help but feel that the present-day data processors are ill-adapted for doing literature searching or machine translation. For instance, Western Reserve University has built this special machine for searching literature. I just wonder if we need a special machine for research in language translation. I'd like to hear some comments.

BOOTH: I might make a comment on that as a machine designer. There is no doubt whatever in my mind that if we ever do machine translation, or if we do literature searching on a large scale, then a general purpose computer isn't the way to do it. On the other hand, what a general purpose computer is, is a device which is capable of carrying out any operation, however apparently complicated, which can't be reduced to the rules of arithmetic and to certain simple operations with data. The great strength of this device is that by means of it you can simulate many processes which are apparently complex, which would take a very long time in special apparatus design before you can get a trial in any other way. We intend to use general purpose computers with all of their limitations of nonsuitability to do these prototype experiments. I mentioned in my character recognition talk this afternoon that we were using a computer to find out the best system, whereby some one might eventually build a special purpose de-

vice for recognizing characters. Nobody would think that the computer, which takes a minute to recognize an ordinary alphanumeric symbol, is a good way to do it, especially when you consider the cost of the computer. On the other hand, I think we've done a good job and this is the reason why we are using computers, not because we think they are suitable, but simply because they are a powerful laboratory tool. We use an oscilloscope to investigate phenomena. We might, if we were trying to invent a television receiver, use an oscilloscope to receive a television broadcast, we would hardly be likely to take the various apparatus into our homes, hitch up some sort of lash up with the laboratory oscilloscope and view our television entertainment. No, at this point, we in fact produce the special purpose device. I don't know whether that goes some way to answering your questions, but that is probably the essence of the sort of answer that you will get.

PARKER-RHODES: I would like to take up this point of the desirability of having special machines to do language process and, in particular, research on a language. It isn't, I think, strictly the whole truth of what Dr. Booth said, that what the computers really do is elementary arithmetic. What the computers really do, is Boolean algebra and it requires circuits which, at least to a person like me who is not an expert in these things, looks thoroughly complicated to get what is basically a series of Boolean operations turned into arithmetical operations. I would have said that a general purpose computer would be, in fact, more complicated than the type of machine that would be required to do simple Boolean operations. Well, of course, what the language research machine has to do is not quite that, but it certainly wouldn't have, for example, to have a divider. It wouldn't even, I think, have to have a multiplier. It would have to be able to add and count; but beyond that, the form of operations one requires are much more closely akin to the Boolean operations which are the kind of basic workings which the machine is based on. I would have thought that there was a point in having a machine specially built for the job, but I shouldn't have thought it would have been a particularly difficult job for computer engineers to turn out such a machine. In fact, it might even be easier in absolute terms, than the kind of machines they do turn out.

BOOTH: I'm afraid Dr. Parker-Rhodes must have been not listening to what I said. What I said was that one should certainly produce a machine eventually. It was convenient to use the machines that which we already have in the first place. As far as a computing machine being a device for doing Boolean algebra—as a designer of these machines I think I know something about them and this is not correct—you can do Boolean algebra on machines, but they are not designed to do it.

JOHNSON: In line with the standards bit, it seems to me that when we relate this to special purpose or general purpose computers, we will have to specify quite closely "what digital computer" in every aspect we are referring to. At the risk of bringing up the completely hellish argument of stem-ending to section vs. complete entry opera-

tions in the dictionary, we, at the University of Washington, have been quite sensitive from the knees down every time we get out in public, because we have been quite practically the lone component of the complete entry operations. This, however, was done at the request of a sponsor, who assumed that we had a multiple disk memory, and under this operating basis, one standard dictionary becomes completely legitimate, completely understandable and completely practical. So where do our standards go if we are to say: "let us standardize our operations between … standardize first our dictionaries then so forth and so on.… on up the line;" unless we are to specify precisely what digital computer we are to be working with and my feeling is that as soon as we have specified precisely what digital computer we will be working with, we've reached the ultimate in immediate stagnation.

BOOTH: I'm sorry that the University of Washington has got out on a limb over this. With the usual British tendency to compromise, I would never have maintained that the stem ending decomposition was the only way. In fact, stem ending decomposition is not universally applicable by any manner of means and one has to use complete words in a number of cases. I quote merely the treatment of irregularly formed words which require the whole thing. Now I would agree entirely that if you have a device by and large you must tailor your methods to the device, and this I think might give us cause for discussion later on in this meeting. The fact that at the moment there exists a variety of computers in this country, in Europe, in my own country, and presumably in the Soviet Union makes it seem to me rather unlikely at the present state of the art that any particular computer will attain a general enough acceptance that we can standardize on it, and this may in fact make the task of this conference if not impossible, at least extremely difficult. This is one of the things that we have to think about.

(FLOOR): I'd like to go on too on this subject of general purpose and special purpose computers and especially because it would give me a golden opportunity to preach the gospel according to the Army; but I would like to ask your indulgence to let me put myself down to a schoolboy level just for a bit and ask a very, very interesting couple of questions. Sometimes we find that putting ourselves down on the childish level will get us answers where complicated discussions will not. A couple of weeks ago at Chapel Hill I sat in on a symposium on the topic called machine perception and learning. It impressed me there that these experts didn't quite know what they were talking about. I don't by this mean to imply that you gentlemen didn't quite know what you are talking about, but I want to get down now to the very innocent little questions. What are we trying to standardize and what do we mean by standardization?

YNGVE: There has been very little said about standardization in the field of mechanical translation. I think there is a very simple answer to this. The reason for the fact that very little has been said is that it is premature to standardize in a field that is as young and undeveloped as the field of mechanical translation.

BOOTH: I wrote, just a little while ago, an editorial on the question of standardization, and unfortunately, in this article, I remarked that by the time a subject has become suitable for standardization it has become ossified—so in this respect I am inclined to agree. However, I should qualify this by another remark which is this: That in many fields of human endeavor it so happens that when a subject has developed for some little time, one has a number of conflicting bodies of information. This information which exists has to a large extent to be thrown away or to be reprocessed. This means that in the future, one has a horrible mix-up. I can't help feeling that in this subject it may be impossible; but it would be a wonderful thing if we could at this stage, and before too much information accrues, do something to insure that in the long run all of our results are compatible. This does not necessarily mean standardization, but it does mean that they might be compatible with one another. Otherwise, I agree entirely with Vic Yngve.

SOLOMONOFF: I would like to discuss two aspects of standardization. The first is that in information retrieval. Say we have one group working on information retrieval, for example, the Western Reserve Group, and they've abstracted a lot of documents in their own telegraphic type abstract, and suppose there's another group doing something like this on the other side of the world. They would like to get together and divide up the work or be able to use one another's results. They're not going to be able to do this if they use slightly different vocabularies and such. Now in this sense you can, I think, easily understand what standardization would mean, and then the question is to what extent do we want to stick our necks out and say that, well, this system is just about good enough and we will standardize it now. The difficulty usually is this: We would like to standardize in a way such that we can change things a little bit in the future, and unfortunately we don't know how we want it changed in the future and it's sort of trying to hedge on something when you don't know exactly what you've hedging about.

You could hope that there are certain kinds of compensations that you can make. For instance, in many servomechanisms, you can build them to be stable against two or three kinds of things and then, surprisingly enough, they are stable against things you didn't think about. But this is only one of a very small class of devices that do this and perhaps, if we're lucky, we can do this with standardization. .. I don't know.

The other thing is in mechanical translation. Now here the problem is a little bit different. Right now the thing is very, very much in a research stage. Perhaps what we would like to do is this. We would like to have some people do some work in mechanical translation and then someone else with another computer, perhaps can take their tapes, run them through some sort of process that can convert this into their own format or into their own order structure, perhaps, and then use the same tapes to do their own mechanical translation and then start research at that point, in modifying it, and so forth. Now, I

do not know the extent to which this is do-able at the present time. There is difficulty even at the point of format change. If you have a tape for one kind of computer, for instance, magnetic tape or paper tape, we have just begun constructing machines to do this simple mechanical problem of converting one type of tape into a tape for another computer—which has to be done at first before we can do any of the other things.

Now, as for the "other things," we've begun to do something like this in algebra—algebraic notations for computers and meta-programming; this algol business is something in that direction. It may be that for much of the mechanical translation work that has been done on specific computers you can make a device for first of all, converting it into the right format and then doing this translation routine on your own computer. It might be that you can already do this and we don't have to worry too much about standardization, I don't know. I wouldn't be too surprised if between many computer pairs this was no problem.

BOOTH: That's another rather good point. I sympathize with this difficulty of the actual input medium varying from one computer to another as someone in the U.S.A. would no doubt remind me if I offered them a computer tape from my machine; because being five hole British tape it probably wouldn't go into a seven or eight hole American paper tape reader. Although, with a little ingenuity of the true Cavendish variety with two bits of chewing gum and a couple of knots—you could probably run five hole tape in a seven hole reader. Of course, we couldn't do the reverse with your tape, this is one impossible process. The other point about translation of programs is, I think, established by what I always call Turing's theorem. It is perhaps enunciated by the late Allen Turing in a paper on computable numbers and mathematical logic. He established that any well-formed computer, could, in fact, accept the code of any other well-formed computer.

SOLOMONOFF: From a theoretical point of view occasionally this is true, but from a practical point of view it may not be.

BOOTH: Yes, I realize this difficulty. It happens particularly with computers that use echelon storage, where there are peculiar difficulties in writing programs for a in terms of b. But I would say this—that for those who can't write a simple translation program, you can always write a program of some sort, and this might enable you to carry out the test. I'm not really throwing this out as a serious suggestion, I would think that if one supplied a tape or, if you like, a program, it would probably be better to re-program it in the terms of the same method, which I think you will agree you can do.

SOLOMONOFF: With regard to this point, quite often taking a program built for one computer and running it on another will take perhaps ten times as much time. Well, if you're doing something that's in research, like mechanical translation, this isn't too bad. You just want to know roughly what this mechanical translation scheme will do in this particular sentence, or a group of sentences. But you can't afford this excessive time for production jobs.

JOHN MELTON: Not being a pure linguist or pure mathematician or pure anything, I would like to point out that it seems to me that possibly we're on a wrong path here. Looking at the title page, I see that this is an International Conference for Standards on a Common Language for Machine Searching and Translation. We're not talking about standardization, and in my freshman English classes I point out that standards and standardization are not precisely the same thing.

BOOTH: Well now, having stuck your neck out, perhaps you can tell us just what the terms do mean, then. I would rather like a definition. I've looked at this through this title several times myself. Could you perhaps tell us what this does mean?

IBERALL: I think this is the roughest sense in which standards was intended. In the year 1959 there was a sufficient number of different schools of thought that were in process of going off on tangents and it was thought that a meeting to discuss a common language at least for communication purposes was warranted.

BOOTH: Does this mean that what we are here for is to exchange information and to avoid overlapping. I would think this is the best use for any conference that you can have.

FAIRTHORNE: I suggest that however many schools of thought there are, we are trying to achieve a certain performance in mechanical translation and what we're after is a standard of performance—not on "how you do it." And there are some very obvious things we might ask for minimum standards of performance. For instance, question of scope and range; scope of vocabulary, range of semantic resolutions, discrimination of topics—how far we are going to say that something is about something. There are many things like this that we can very usefully regard as a type of standards. There is a degree of crudeness or lack of discrimination in which there is obviously not much point in either doing mechanical translation or retrieval except by going to haul the books on the floor and look through them. And on other occasions where the resolution is so fine that it would mean pulling bits of pages out of a book. All that sort of thing is thoroughly practical, it's a standard of how you are to do the thing, we don't know how to do the thing anyhow.

BOOTH: If I may say so I would have thought that standards of performance were even more difficult to define than standards of anything else. Look at the critics of music and drama in the papers.

(FLOOR): I am actually a computer person, a programmer, and attending this Conference because we're attempting to take an information retrieval system and program it on a computer. But the title of the conference implies that there might be a common language for information retrieval people and a common language for machine translation people and we'd like to bring these two together. Well, at the present time there are many of what has been called meta-languages to machine translation. It is attempted to take a human mode of expression, operate on it by a finite set of, or number of instructions, form a symbolism in the machine which is then operated on by another set of finite set of instructions, to provide a second mode of

human expression. The information retrieval person on the other hand, is attempting to take a human mode of expression, operate on it by a number of instructions, and symbolize in the machine the informational content of the document. Thus we have an informational language versus the meta-language. Then we desire to take the human expression, again operated on by a number of instructions, compare the two, and retrieve the informational content or the indication of where the information may be located. It would seem to me that eventually maybe these two languages would be developed and be brought together and maybe a possible outcome of the conference could be the desires of the information retrieval people for this language, the common language, and the desires of the machine translation people for this common language. As they develop their respective languages, these things can be kept in mind and included if possible in the languages developed.

BOOTH: I imagine that if we got down to trying to decide on some acceptable terms, one of two things would happen. Either both sides would agree to disagree, or alternatively, maybe you could browbeat the few machine translators present at these meetings into agreeing with you, so then the majority of machine translators outside would refuse to conform. Then where are you?—which is always the difficulty. Now, I think that I should conclude this session. I want to wind it up by remarking that we have had, in my opinion, a most interesting and stimulating set of papers today. They've covered all sorts of topics, they've inflamed the passions, as we say, they've brought down hell and high water on my head; and they've got to squabbling with one another, which, I think is always a healthy thing at a conference. Now I'm not at all sure, in fact, I'm almost sure of the contrary, that we're in a state to provide standards of anything. This has been said by a number of people, and I think it must be quite obvious to most of us when we talk about the meta-languages of the machine translators. To the best of my knowledge, none of these meta-languages exist. We all have certain computer codes; we couldn't standardize on any meta-language if we wanted to, because of the nature of our beast; this has been brought up by one person or another, so what are we to do about it.

What I do think has come out is something else; quite different probably from that which the organizers of the conference had in mind. This is the need for cooperation between the East and the West. This, of course, comes up everywhere, but it is in an obvious form here. There's something we can do about it. We've seen just what that is—on the lowest plane, the exchange of addresses, on a better plane, the exchange of actual workers, which after all is the way to break down quite a lot of the barriers. This sounds like a philosophical and platitudinous speech, but I really believe that it's the way to operate in science. For too long has there been this distinct cleavage between workers in the two fields.

The third thing which the conference might do, possibly, is to compile lists which could be distributed. These are all easy things which

won't annoy anybody; they can include all of the points of view, and, I think, will do a considerable amount of good. Unfortunately, this is still only at the first stage of the conference so any remarks that I might make I can quite easily change over on Thursday when I see what other people are in fact thinking.