

Session 4
METHODOLOGY

Chairman:	Robert M. Hayes	Electrada Corporation
Introducer:	H.P. Edmundson	Planning Research Corporation
Discussants:	A. F. R. Brown	Georgetown University
	Vincent Giuliano	Arthur D. Little, Inc.
	David G. Hays	The RAND Corporation
	Harry H. Josselson	Wayne State University
	Gilbert King	International Business Machines Corporation
	Sydney M. Lamb	University of California, Berkeley
	Ariadne Lukjanow	C-E-I-R, Inc.
	Anthony Oettinger	Harvard University
	Alan J. Perlis	Carnegie Institute of Technology
	Erwin Reifler	University of Washington
	Ida Rhodes	National Bureau of Standards
	Don R. Swanson	Ramo-Wooldridge Laboratories
	Stanley N. Werbow	University of Texas
Victor H. Yngve	Massachusetts Institute of Technology	
Michael Zarechnak	Georgetown University	

Questions and Discussion
Summation by Chairman

Session 4
METHODOLOGY

In what ways do the methods and approaches now in use differ from one another?

To what extent do they complement each other?

To what extent do they duplicate each other?

Would closer cooperation between the various groups be desirable?

If so, how can this be achieved?

What qualities of translation are we aiming for?

In what ways are the various sources of linguistic information being exploited?

What standards and criteria for reliability are being applied?

Session 4: METHODOLOGY

METHODOLOGY

HAYES: Welcome to this afternoon's session. I would now like to call on Dr. Edmundson to introduce this very interesting panel.

EDMUNDSON: You will note in your program that this is called a panel on methodology, and specifically this is a panel on MT methodology. Professor Sebeok and others have expressed interest in the interrelations and feedbacks between machine translation and other areas of what we may call automatic linguistic analysis. I think that some of these relations will become more apparent at the conclusion of this panel in particular. I am, of course, referring to problems of automatic indexing, automatic abstracting, and information retrieval. Nevertheless, we wanted to restrict the discussions of this National Symposium to those of machine translation alone. For that reason peripheral remarks will appear only in connection with the central problem of machine translation. We are naturally concerned with the various techniques and methods employed at the research centers and the research groups -- how they approximate one another, how they complement one another, and whether there is unnecessary duplication. There is nothing wrong with confirmatory duplication. On the other hand, the MT field has been hampered somewhat by a lack of standardization. I am thinking now not just of terminology but of working materials; for example, glossaries on cards and on tapes. The question has repeatedly come up, "should there be closer cooperation" ? I am speaking now of cooperation in meetings, agreements that might be worked out to the mutual benefit of the various research centers, and how this can be achieved. Dr. Lamb defined a criteria for reliability or accuracy of translation. What is the final product going to look like and what do we mean by a rough translation or a good translation? Since figures of merit seem to be poorly defined, I hope that some of the speakers today will bring out their ideas as to how we might introduce a metric to quantize this notion of accuracy. There is also the question of sources for machine translation research. By this I mean sources of linguistic information. You have heard a great deal

Session 4: METHODOLOGY

about examining large volumes of running text. Some investigators prefer sample sizes of 10^6 , while others say that 10^3 is what they want to work with. I have been impressed by the fact that our methodology has grown in both breadth and depth. As you know, the first serious work in machine translation was done in the United States. You are also aware of the British contributions and from Professor Harper's remarks, that the Russians are expending an even larger effort constrained essentially only by machines and machine time. Even though all questions of methodology cannot be resolved today, I hope that the members of this discussion panel will attempt to state their positions as lucidly and succinctly as possible. It is very easy to think you understand what a certain research group is using as a basic method, when in fact you really do not. This may be due to the fact that perhaps they also do not. As a stimulating exercise, attempt to place the activities of the research centers in this country in a three dimensional space. Along one axis measure how theoretical, or conversely how empirical the study has proceeded. Along the second axis measure the notion of whether the method uses an iterative approach or instead a complete closed-form solution to the problem. In the third dimension of the space measure how utilitarian the goal. Let me discuss these three dimensions for a moment. Since it is fallacious to use dichotomies, the research centers are not points in a discrete three dimensional space. This is in part due to the fact that it takes a third party to explicate what someone else is doing. Perhaps the "native informants" of the research centers are not as reliable as we would like. As to the first dimension, the research heads of the centers hesitate to say that their work is more theoretical than someone else's. Some phases are; some phases are not. As you are aware, MT is distinctly an interdisciplinary field, but with very little discipline. We hope that by defining our terms and by explaining the basic features of our methods we can come to some understanding in the sense of appreciation, if not agreement. You will note from the prominent linguists on the program, that we hoped by bringing together outstanding members of the linguistic profession to expose them, and therefore their graduate students, to this very interesting problem area. As to the second dimension, that of

Session 4: METHODOLOGY

iteration, the question of whether a cyclic approach is ideal or not certainly is a meaningful question. It may not be answered satisfactorily during this session; it may never be settled. However, I hope that the discussants will reveal whether they believe (as I think quite a few linguists did believe a decade ago) that somehow you would solve "the machine translation problem" by bringing together two or three of the countries best linguists and leaving them alone for four or five years. They would come forth with a flow chart which could be programmed, then it would be put on a machine, and would somehow run well. As to the third dimension, that of utility or goal, I think it is apparent to all of us that some of the MT investigators are not really concerned with production translation. They want to remain in the research area. As we get close to machine translation in a production sense, they want to move on to other frontier areas of this general field of linguistic analysis; such as automatic abstracting and other problems of information retrieval. On the other hand, there are some who are most anxious to accelerate the effort in order to have a translation service running in the shortest possible time. With these as general background remarks I would like to turn the meeting back to Dr. Hayes.

HAYES: I believe that at this time Dr. King will probably repeat his remark of yesterday, "It is a problem for some people, but not for IBM". I will try to interpose myself as little as I can. Obviously, the people on the panel will have their own comments. I would like however, to introduce the following as topics which ought to be discussed. First, how do you react to the type of classification presented by Dr. Edmundson? Second, if your reaction is favorable, where would you place yourself in this three-dimensional space? Third, do you consider the question of criteria or figures of merit to be relevant to your work?

SWANSON: Before I react to the classification you have made Dr. Edmundson, would you clarify one word for me. Are you referring to the cyclic, or psychic, approach?

EDMUNDSON: I am not at all sure that we yet know the answer to that -- the Symposium is not over.

Session 4: METHODOLOGY

SWANSON: Let me pick up from a comment that you addressed to Dr. Lamb this morning in the brief exchange at the end of this morning's meeting, to which he did not answer at that point—what happens after you plug it in to see if it will work? I think the answer is obvious. He looks at it and if it doesn't work he tries to do better. Now isn't this the cyclic approach? Is there anyone who would quarrel with this? Regardless of whether or not you are trying to conceive in all its grandeur a complete scheme for mechanical translation out of the speculative powers of the combined minds of the linguists and the informant, the fact of the matter is that once you have so conceived the scheme the next step is to plug it in and see if it works. By definition this then becomes a cyclic approach.

LAMB: I don't know quite how to begin, but let's put it this way. The first time we at Berkeley plug in an automatic translator and have it work the plan is that we will have put much more information into it than anybody has ever accumulated, and much more even than most people are trying to accumulate. In other words it is not comparable at all to what has been done in the past by people using cyclic approaches. What they have done is to take a corpus of 20, 000 or 30, 000 running words of text, and work out rules that would be able to translate this amount of text. Then they try it on another corpus and so on. There is your cyclic approach.

SWANSON: I think that you have perhaps correctly stated the essence of what people have called the cyclic approach but in a way which reflects a very common distortion. I don't think anyone would pretend to say that the way one creates a machine-translation scheme is to take any amount of text and develop a system for translating that particular text. Whether or not it is explicitly stated, I suspect that anyone who uses the cyclic approach is doing this as a way of stimulating and extending the speculative powers of the linguist. To amplify that further, clearly no one is going to deliberately create ad hoc rules that violate linguistic intuition just in order to explain a relatively few occurrences that have come up in a limited body of text. The text does serve a useful purpose, however, in stimulating

Session 4: METHODOLOGY

the linguist and the informant to speculating on situations and occurrences that he might not otherwise have thought of. There is one other reason: the problem is not entirely one of linguistics and language, and in fact I am not altogether sure that it is even largely so, and this point was brought up by Dr. King yesterday. There are simply a lot of messy engineering problems to be exposed, and so far as we are concerned we would like to expose those at the beginning no matter how trivial they seem to be. The fact that in scientific text one has to deal in some way or another with a lot of symbols and a lot of equations, the fact that a dash is often used as a predicate nominative, and other questions that a linguist would consider completely trivial are exposed and would not necessarily come to anyone's attention unless one attempts it on large volumes of text. So I wish at once to defend the cyclic approach in the sense that it extends and stimulates our speculative powers. At the same time I want to make the observation that it is my suspicion that everyone working in the field is using the experimental approach of trying it out on bodies of text. Even in Mrs. Rhodes' case, I suspect that eventually she will try using her scheme on greater and greater amounts of text until situations are exposed that hadn't been taken into account from the beginning.

HAYS: First of all, Dr. Swanson appears to have said that there isn't really a difference between the iterative approach and the attempt to construct a complete program in advance -- that both come to the same thing in the long run. Now I would like to question whether there is a difference between the theoretical and empirical approaches. I think that at RAND we are quite theoretical and fairly empirical. I hope so because theoretical work without empirical study tends to sterility, and empirical work without theoretical underpinnings dies aborning. So that only leaves us utilitarian versus research interests and that would get us off the track.

SEE: I would like to suggest at this point, that perhaps a polar distinction between black and white, cyclic and noncyclic, and utilitarian and nonutilitarian are not quite in order; perhaps what

Session 4: METHODOLOGY

we are really looking for is what I believe was originally suggested-- the difference in degree. Each of the groups could try to find some number on the scale that would characterize themselves. We should get back to the original idea of not, cyclic or noncyclic, but if cyclic to how great an extent. That is, what is the time interval, what is the length of the period. I think at the Bureau of Standards the cycle is intended to be very long.

HAYS: It appears to me that what we want to do at RAND is to establish some basic arrangement of programs that are in general capable of handling things like languages and then using these by iterative methods to build up the size of the dictionary, the accuracy of each entry, and the accuracy of the elements of a grammar. There are some things that you have to get once and for all and you cannot create a basic program step by step. When people talk about building up a glossary, iteratively, they are not talking about gradually inventing a search procedure for looking into the glossary, they are talking about gradually increasing the number of entries, and if you don't make that distinction then arguments about iterative versus not iterative can go a long way without much being accomplished.

LUKJANOW: I am absolutely in agreement with what Dr. Hays said just now. I think that it is impossible to separate the theoretical approach from a certain amount of practice. Unless you have a basic scheme with which you get on a machine, it would be very difficult to arrive at any kind of a scheme just going from sentence to sentence and making rules. I think that one has to have some sort of a basic idea, basic approach, basic rules, and program made up before one approaches the text. As far as the text itself is concerned I think it is a proving ground. That is where you find out what mistakes you have made.

LAMB: There are several points that have come up which I need to comment on. Instead of taking them one by one, I will give the general picture. It seems that there is an opinion that some people hold that the iterative or cyclic approach is advantageous because

Session 4: METHODOLOGY

of the fact that it enables the investigator to take advantage of material as it actually occurs in text. We have a method which enables us to do text analysis without doing trial translation. We do pay attention to what is in the text. That is the main source of the data that we use for analysis, but it is possible to do this without setting up hypotheses. It is possible to do it in such a way that what you are actually doing is accumulating information. The only advantage of the cyclic approach, which is very important, is that when you test your trial translator on a new corpus it enables you to find out what it is that you do not yet know so that you can first get the additional information and then make the changes. In advance we have worked out the plans in such a way that we already know what we don't know. We don't have to build a trial translator and then try translating text in order to find out what we don't know.

A. F. R. BROWN: If one had to have a polar distinction between the cyclic and noncyclic approach one could make a dichotomy between those who expect to be pleased at the end of the first iteration and those who do not. While I think all of us at Georgetown are very cyclic and we do not expect to be terribly pleased after any given iteration, what you can foresee in the way of trouble are the problems which are really trivial. The longer your cycle, the longer your period before the first iteration ends, and the longer you are working on things which are of course important, but in the scale of difficulty, trivial.

LAMB: All I can say is that I violently disagree with that. It may be that one needs to have a lot of professional training in linguistics before he is able to point out in advance what needs to be known--but I don't know what it is but it seems to me that the things that we have foreseen are far from trivial.

JOSSELYN: First, I would like to talk about the problem of theoretical versus utilitarian. I think that the very nature of the field and the work you are doing in a sense will force you toward utilitarian. After all, if we linguists make a mistake we are confronted with it. Therefore, when you start out, certainly you need

Session 4: METHODOLOGY

repetition, reiteration or cyclical procedure. You start out with the assumption that you don't know everything about what you are going to do. You are going to be very careful, first of all, to list everything that you know. You are going to go to the dictionaries and then you are going to use yourself whenever you can as a native informant. On the basis of that you are going to draw up certain procedures, and after you get through with that you are still going to find out a lot of things you don't know.

OETTINGER: This discussion on methodology disturbs me. The linguist part of me rebels about the arteriosclerotic approach to linguistics which says that everything is done by this or that method; you are either a structuralist or Harrisonian. I say, so what? The mathematician in me says it is perhaps a little bit wrong for somebody just because he happens to be a mathematician and therefore smart, to presume that centuries of work in linguistics have led to nothing. This isn't really true. There are things in linguistics that are valuable. To my mind it doesn't really matter how you get your results. I don't care whether my students use a Ouija board or what. The proof of the pudding is in the eating. Now there are two kinds of puddings and two kinds of eating. If you have a theory and the theory stands by itself it should have certain criteria of simplicity and elegance, and if people agree on aesthetic criteria they can judge the theory. If the theory pretends to apply to something then you can test whether in fact it predicts something or it really applies. Now, you have a test of practice, either it does what it is supposed to do or it doesn't, and no questions should be asked how you got there. It seems to me that a discussion of methodology is a sign of decadence and I hope that we can get into more productive subjects.

EDMUNDSON: I would like to address this question to Dr. Oettinger. Do you think that we can perhaps profit by our MT experiences, particularly our errors, and thereby do a better job when we work on some of the other problems in the area of general linguistic analysis ?

Session 4: METHODOLOGY

OETTINGER: One can profit from one's mistakes in anything, but I think it would be extremely difficult to write down a rule book as to how thou shalt profit by thy mistakes. If one were able to do this, this would be tantamount to a program that you could put into a machine to do research. If I knew how to avoid all the mistakes I'm going to make tomorrow, I'd have my problems solved. Surely, you learn from experience, but beyond the saying so of it, I don't see that you can gain much from it.

LAMB: There is another aspect of this general problem. To a large extent the different groups are trying to do the whole job all by themselves and to a large extent this has been necessary because none of the groups has been producing the type of material which other groups would have been able to use. If you are putting your research results in the form of a hypothesis, the trial translator to be tested, then you are putting it in a form where nobody but you will be able to use it. Because nobody except one who was actually involved in the construction of the trial translator knows what kind of knowledge it is based on and how valid it is--which parts are based on good knowledge and which parts are guesses. So if we are going to be able to contribute to each other's efforts and thus speed up the work of the whole field we ought to try to put our research results in the form of information which other people can use.

HAYES: This brings up one of the questions that was raised by Dr. Edmundson, namely, is cooperation desirable ?

OETTINGER: Cooperation, like God, mother, and country, is desirable. The question is with whom and about what. My personal feeling about the discussion of cooperation is that it should end with the statement that those who want to cooperate should, those who do not should be left alone. Scientific cooperation, by and large, proceeds through the medium of lucid and informative publications in meetings such as these, and anyone who can read published material, understand it, and find that there is something in the literature that is ahead or better or different from what he is doing would be stupid if he tried to duplicate it. Other than saying that one ought to encourage lucid publication and accurate reporting and

Session 4: METHODOLOGY

honest scholarship, I don't quite see what further one can say about cooperation.

SEE: I will try to say a little further on the subject of cooperation. In particular I think publication in itself is not sufficient to convey all the information. One of the problems in this field that we have had in distributing information is that much of it is in the prepublication state. A large machine run is not something that lends itself to publication, but perhaps these things could be distributed on an informal basis or even a semi-formal basis among the groups themselves. Normally, publication includes only these finished products that we see in the corner bookstore. I think there is a real place for distribution of any and all material. My list would include dictionaries, flow charts, and any programs that are in some way compatible with the other group's computer; in fact any information that could be even partially understood by the other groups.

HAYS: There is a difficulty in this field which may apply to some others, but it doesn't seem as general. This problem appears to be a general problem in science but more or less specific to linguistics and that is, that everybody who is studying Russian syntax, morphology, and grammar in general is concerned with identically the same body of material. That is tantamount to saying that you cannot publish a summary statistic for your dictionary in this field, but you have to publish the contents of the dictionary itself. When you come down to any specific entry in the dictionary, it was probably a mistake to publish that one because the words (the items that you are describing by that entry in your dictionary) have properties which are not revealed in the body of text that you personally examined, but which came up in somebody else's corpus. So it appears to me that the intrinsic difficulty in the science of linguistics is that we are all interested in the same body of material and each of us finds out something different about it. In order to make the best progress we have somehow got to be able to exchange in minute detail information that is costly to reproduce and would be much better exchanged from one machine to another than through the medium of printed pages--if there is some suitable way of doing it.

Session 4: METHODOLOGY

EDMUNDSON: I think Mr. See's point was that trash should be collected; not published.

HAYS: I feel that, despite the hard work that is involved in exchanging material of any kind, exchange is still cheaper than each group's working only for itself from its own text. But how can we make the exchanges ?

LAMB: We cannot do anything about the past, but in the future it would help if we could, even when formulating the results of research for ourselves, put them into such a form that they can be communicated to others just as easily as to ourselves. Now it happens, that the way our methodology works, this is the way the material comes out.

SEE: I think that one answer might be that hand-in-hand with exchange of material must go exchange of personnel. I think the reason that it is difficult to understand these things, in many cases, is that we have not been together very much. If, as in some groups, personnel could be exchanged for a month or two, I think that the working materials of the other group will be usable because there will be someone accompanying the very materials that are in question.

OETTINGER: We would welcome visitors from any group, preferably of the type that would stay around for a couple of months. I just want to make clear that two-day visits are not the solution to this.

KING: To go back to the title of the session which is on methodology -- it seems to me that one of the most difficult parts of translation of languages is that part we call semantic analysis; resolution of difficulties, on the basis of semantic clues. It seems to me that this is by far the biggest problem that has to be solved and we haven't even begun to talk or think about it seriously. It seems to be a little ridiculous for us to talk about methodology since all we are doing is discussing the peripheral parts of the problem. I admit syntax has

Session 4: METHODOLOGY

to be done but it isn't a very deep subject, it is just a complex one. Whereas semantics may not be very complicated but it is a very deep problem.

A. F. R. BROWN: I can't speak to the last question. Perhaps I can cause a termination of the preceding discussion in a more definitive way. Has anybody at this table derived any substantial benefit from anybody else's experience so far. I know that certainly a benefit has been received by Dr. Oettinger from Mrs. Rhodes; this is a highly exceptional case. In my own very small way I once or twice tried to shed light, and of course it turned out that just as other people's ideas look to me too idiosyncratic or too much imbedded in their own habits of thought and work to be of any use, that the same thing seems to be true in the reverse direction. What concrete instances can anybody bring up?

EDMUNDSON: Since Dr. Hays is probably too modest to say this, and since I am no longer with The RAND Corporation, I think I would like to point out that The RAND Corporation has from the very beginning supplied and offered materials to many of the groups that are in this room today. The University of Michigan, Ramo-Wooldridge Laboratories, and Georgetown were supplied with text material, programs and results. Professor Dostert had five of his graduate students spend time with us one summer. We have also had an exchange with Professor Josselson. I think the list has grown.

JOSSELSON: We have been talking to Dr. Hays and getting a good deal of the materials. Recently we have been profiting a good deal by our experience and contacts and exchange of information with the Ramo-Wooldridge group. We have had some rather close, but brief, contact with at least two groups.

WERBOW: There are things that need to be done for better communication and cooperation. How many people had even seen the February 1960 report of Mrs. Rhodes; let alone the November 1959 one? This kind of thing is something that we could conceivably lick.

Session 4: METHODOLOGY

OETTINGER: I want to briefly comment that I agree with Mr. See and Dr. Hays. By publication I don't mean the restricted sense. Certainly microfilm or magnetic tape might be a suitable medium, but what I would like to convey is a measure of quality of the type that one achieves, say, with competent refereeing. The report literature suffers by and large from pressures of deadlines and other pressures which result in a fantastic glut on the market. Any of this contributes vastly to our information retrieval problems.

SEE: I think there is an important distinction to be made here between what we might call metareporting and reporting of data. I think I agree completely that the trash should not be distributed. I would certainly make a distinction between that and a dictionary in its working form that one wouldn't like to show to a harsh critic because it hasn't been examined for accuracy in every detail. There is an important difference here between the raw data that you are working on which constitutes your universe and the statements you have to make about this data. I think we should keep this in mind.

OETTINGER: Yes, but you see you don't want to say just dictionary; supposing you made available a reel of tape and a printout of it. This is worse than useless unless accompanied by a completely detailed description of how the thing was put together, what every code symbol means; everything about it so someone can reproduce the results. This takes a great deal of expository writing. We have had this experience with making our dictionary available to the group at the National Physics Laboratory in England. We supplied them with a complete set of our reports. They read through these and came to the point where on a subsequent visit they had mastered the thing well enough so that they had every one of our mistakes spotted. From that point we could get started on a useful, fruitful, mutually agreeable discussion. But I must say that the effort that went into preparing the material even to that point was pretty high and that there has been a tendency to say here are the results, now you figure it out. This is unacceptable.

Session 4: METHODOLOGY

OETTINGER: The Journal of Mechanical Translation is an awfully good medium—perhaps it isn't used enough and it might well be that one could channel the publication of some of these reports, not as papers but as auxiliary monographs, or reports, and use the distribution list of the journal. I think part of the difficulty is in fragmentation and in various people's handling distribution in their own way--some kind of centralized distribution might be useful. The details of this should be discussed elsewhere.

YNGVE: I think that the coming year will probably show a good deal more interchange of text, programs, and of dictionaries. There are a number of things that are reaching completion at the present time. A number of programs and dictionaries are almost finished. Some have already been offered. Professor Oettinger has offered his dictionary for example. As far as publication in MT is concerned we are taking steps to increase the speed of publication. I want to point out at this time some of the difficulties in running a journal like this. If there are not very many papers to publish we can only come out about three times a year. This means that there will be a minimum lag of several months before we can get publication. As the field grows we expect that there will be more papers worth publishing in the journal MT. When this happens, the journal will come out more often and the lag in publication will be reduced.

HAYES: Does anyone want to comment on Dr. King's somewhat controversial comment? Would you repeat it?

KING: I feel we have not even begun working on the hard part of the problem, so I don't think we want to argue about methodology. I think everybody who is working in the field undoubtedly has his own pet way of doing things, but I think everybody just has to be trusted.

JOSSERSON: I would like to kick off the discussion simply by this: I don't know what semantics is and I have been a linguist and have been working a good many years. As far as I know, there are as many schools of semantics as there are semanticists. When you get going on analysis of language for MT purposes you work out

Session 4: METHODOLOGY

rules, and then you say this is morphology, that is syntax, and then you have the wastebasket class and that is semantics. That doesn't seem to be quite right. It seems to me there should be some tentative approaches to semantics. Maybe it would be helpful if some of the investigators here spoke up.

YNGVE: I am not one of those people. I do want to say that I agree that semantics is perhaps one of the most difficult problems. I also want to underline the fact, as I see it at the moment, that syntax is also an extremely difficult problem, and I think probably 10 years from now it will still be a difficult problem. I don't think MT will be passé in 10 years, and I don't think that work on syntax will be passé in 10 years either. Semantics, I think, is perhaps more difficult than syntax.

GIULIANO: I tend to agree with Professor Yngve that syntax is not a simple problem, but I think the important point is that an understanding of syntax will probably be an essential prerequisite to the ability to handle semantics. In other words, can you really speak of how you should interpret the meaning of say a word which functions as the object of a preposition until you know it is the object of a preposition? A certain amount of the structure or content of language is conveyed more or less on the surface by syntax. This can be used to resolve a number of ambiguities. Is it not a logical approach to remove these ambiguities and then look at what is left which is considerably cleaned up?

HAYS: I don't think I agree with Dr. Giuliano. First of all, let us see what we are talking about. There is a famous partition of semiotics into syntax, semantics, and pragmatics, and I thought we got it cleared up in the Paris meeting, that we were all talking about syntax and nobody is interested in semantics. At least nobody in this field. What we are really interested in is syntax beyond the traditional range. We are interested not only in the rules connecting morphological classes and rules connecting lexical classes with morphological classes (as for example the lexical class of verbs that take the morphological class of accusative noun), but also in lexical-lexical classes. This latter category is evidently what most people

Session 4: METHODOLOGY

are referring to in the field of machine translation when they talk about semantics, but it is not semantics in the old sense; it is still syntax. The question is whether you can partition syntax, namely a body of rules all relating to the relations of items in text into several parts, and deal with lexical-lexical rules after all the other kinds of rules have been used up. I think that probably the answer is negative; that in order to determine the structure of a sentence you have to take into account from the first the lexical classes of all the items. As an example I take the prepositional phrase which in several languages is a real nuisance because the location of the item modified by, or governing, the prepositional phrase is hard to deal with. Consider a lexical classification designed around the possibility of governing certain prepositions as in Paul Garvin's new grammatic classification, which by the way we are certainly going to consider as a replacement for our present grammatic classification. Anybody who knows how much grammatic classifications cost will realize how seriously we take this contribution. It appears to me that what might be called a semantic classification because it is of value in selecting the English equivalents of Russian preposition, is also a syntactic classification because it is based on the behavior of the material in text. It is also essential in the determination of sentence structure and the recognition of grammatic structures because otherwise you just can't find the governor of a preposition.

OETTINGER: I am puzzled. From what Dr. Hays and Professor Yngve have said, one gets the impression that we have 10 year's work ahead of us. On the other hand, and this relates to the utilitarian versus research question, there are those who say they will be in production within a year. Now there is an apparent contradiction here . . . either there is no work for the next 10 years and we are wasting our time; or there is, and then production does not make sense. I wonder whether Hays and Yngve would indicate just why we have 10 years' worth of problems, and whether Zarechnak, Brown, and Dostert could explain why we are ready for production.

HAYS: I think both statements are true, that there is 10 years and more of work to be done and that we are going to have production

Session 4: METHODOLOGY

translations pretty soon. Because I don't think that the class of translations is homogeneous yet, there are all sorts of things that can be called translations.

A. F. R. BROWN: Dr. Hays has anticipated me as usual. One can be terribly pragmatic if anybody says he thinks he can be in production in a year. All this means is that there is somebody in the world who wants that output or who says he is going to want that output on the basis of the expected quality. If nobody wants output at that level of quality, then it is obvious that the level of quality must be raised. If perfection is demanded, then we can all just go out and look for other work.

LAMB: If we are talking about early production, then, obviously, it is a purely practical matter. If we are talking about a purely practical matter and if we are also talking about Russian, then we have to realize also that most Russian journalists publish abstracts in English of all articles. So nobody is going to want a translation until, we would presume, he would have first read the abstract and known what the article is about. The only kind of a translation he would want then would be one that would be very accurate and detailed.

SWANSON: I think I can account for about 15 of those 10 years ! In order for a machine to go from something roughly comparable to the worst of what passes for human translation to the best of human translation it would require about that length of time.

HAYES: May I introduce something then which is being skirted here. How are you going to determine this? How are you going to measure; what criteria do you have, or should we obtain, for measuring the quality of a translation?

OETTINGER: The only really competent work on the problem is a paper that is about three or four years old now, by Miller and B. B. Sentor in Mechanical Translation. I don't think I am being incorrect in summarizing their conclusion that this problem is tough. There are questions of intelligibility, accuracy and of economics. My feeling, on the basis of what I have observed with regard to anything we or most people here can do, is that at this stage no-one is in the

Session 4: METHODOLOGY

position to guarantee that any output of the machine is accurate. In the absence of such a guarantee, the only safe way to use such material is to have somebody around who can check its accuracy, and this somebody is going to have to know some Russian. If this somebody is going to be around, maybe he might be more usefully employed doing the translation in the first place. This is a question I would like to see discussed further -- it requires amplification.

EDMUNDSON: I think this a good point for me to say something about the use of mathematics. We have heard about the use of various algebraic structures such as trees and graphs. We are aware of the work the British have done with lattices which have been applied in the analysis portion of machine translation research and also in the synthesis portion when you approach the output language. I would like to make a point that what we would like is more serious work by mathematicians to measure the quality of output -- undoubtedly some stochastic measure is involved here. If we agree that we have iterations, we should be able to measure the residual error after each iteration. This error would in some sense be the difference between the ideal translation and the machine output. The need for a mathematical model in this area is not a critical problem, but is nevertheless, a very interesting problem. I have seen nothing in the literature which attempts to do this. Does anyone know of any methods which have attempted to assign a measure of merit to the output sentence? Finally I would like to ask, whether anyone knows of some recent successes in the use of trees and lattices?

DOSTERT: I am responding to the kind invitation that Dr. Oettinger extended to me a moment ago to deal with the apparently complex and irksome word "production". Obviously no-one, certainly not Georgetown, has made the claim that we are ready to go into production of x number of languages in y number of disciplines. All that we have tried to say is that in one discipline, that of organic chemistry, and going from Russian into English, we expect within a year to have built up a lexicon adequate to produce acceptable translation. I know that I will be challenged on the word "acceptable", and I should reassure our friends that when we produced a random text in the field of organic chemistry last June we asked a specialist in the

Session 4: METHODOLOGY

field of chemistry, Professor Summers of the University of North Dakota, who has a measure of linguistic orientation and who is competent in Russian, to make a critical analysis of this particular random sample. His objective judgment was that the information had been accurately transferred from the source data to the target data. It was not elegant; if one of the criteria is elegance, we should make it clear that we have not met the criteria of elegance. We have met the criteria of information transfer at least to the satisfaction of a specialist in the field. Now I should like to say just a brief word on this matter between practical objectives and theoretical studies. Certain words have been used this afternoon like "utilitarian" with the suggestion that there is something wrong in seeking to do something which is useful, because the contrary of "utilitarian" is "nonutilitarian" which does not necessarily mean pure. It simply means useless. Therefore I see nothing wrong at all in seeking to find out whether we can produce on a machine after four years of work with a team of about 20 people reasonably acceptable text from one language to another in one given discipline. If this be asking too much -- five years of investment in one discipline from one language to another to produce inelegant but reasonable reliable text -- then I confess that I am at loss to see what the motivation in research is.

OETTINGER: All of us who are dedicated to research have to face the prospects that our efforts may have been in vain. This has happened to many investigators. There is nothing wrong with usefulness, but to be useful something has to be right. A sample of one is hardly an adequate measure of right. We have found people one, two, and three years ago who were ready to state that the word-by-word translation produced at that time was intelligible. It probably is, but I do not think that this kind of restricted sample is a reasonable criterion.

GIULIANO: Coming back to the point of word-by-word translation as an acceptable translation, if this could be an acceptable translation then one is led to wonder exactly what type of translation will be produced in a year from now. Will it be produced by the system that was described by Dr. Zarechnak? The point was raised that a lexicon would be prepared in the next 12 months. Will this lexicon be for this system in particular? I believe that we all here are interested in

Session 4: METHODOLOGY

knowing what type of a system can be placed in production at this time. This might affect our course of research in some cases.

OETTINGER: We are ready to quit if the problem is solved.

EDMUNDSON: I would like to return again to the question of methodology since I think that perhaps this may put us on a more neutral ground, a more objective ground. I again would like to ask if anyone has any measures which have been formulated to settle this problem. Perhaps someone would care to comment on this.

ZARECHNAK: I have a general statement on my impressions from reading the papers on methodology in machine translation as described in the Soviet Union. I think that perhaps some of my impressions may be of interest. So far as I am able to follow literature on this particular problem, I can say that they have two camps of thinking; one led by mathematicians and the other led by linguists. The mathematician Kolmogorov stated in the Academy of Sciences that he doesn't believe that machine translation is necessary and he further did not believe that a linguist is able to carry out rigorous scientific research. He was stopped by the President of the Academy of Sciences, the famous chemist Vavilov, who reminded him through particular examples, of past discoveries, that the statement was not proper, and that the machine translation would be carried out under his, Vavilov's, personal guidance. When Vavilov started discussing this problem, he said that in Soviet experience the engineering aspect of any solution usually runs a little bit ahead of the theoretical explanation of that particular solution. Therefore, he made what we may call a friendly reminder to both camps that no matter where you start there will always be some edge between the theoreticians and the engineers. Now, today when we discuss Dr. Edmundson's three dichotomies of machine translation (theoretical or empirical, cyclic or non-cyclic, research or utilitarian) there is one point underlying all of them, namely, that for a linguist exactly the phenomenon, which is so trivial to a mathematician (computability) is not primitive at all to a linguist. This is a new area for him. On the other hand, what is rather primitive for a linguist is not primitive for a mathematician, no matter how capable he is. As you know, the same mathematician, Kolmogorov, when he was challenged by Vavilov

Session 4: METHODOLOGY

decided to become a linguist and wrote an article on the Russian case system. This article was not accepted by Russian linguists who were involved. Now, therefore, may I come to our own American situation. I am personally torn between two poles which of course may be described as my black moments and my rosy moments. The black pole is my intuition (not my ability to observe) that perhaps one day a mathematician or a linguist will come around with a completely lucid statement which will be published in some scientific journal that proof has been found that there are problems in language structure which never will be solved as has been done in other areas. This day will be a black day for machine translation. Now this is my intuition. As to my desire to continue research, I would be most grateful to all the participants that we should remember -- no matter where we start from, cyclic, theoretical or utilitarian -- the end point of our research is the product which will come out of the computer. This product if properly presented will actually constitute the criteria which will unite all of us because the data we are trying to adequately describe are actually the same. Therefore I express the hope that this discussion of methodology will not be stopped this afternoon, will go on later in some journal, and that we will learn from each other. I have learned much from all groups; specifically from the Harvard group (so far as the formal presentation is concerned), if you take away the way they presented the assumptions, the rules which follow from the assumptions, the data tested, and the data which failed to be described by those assumptions. I think that portion of the Harvard studies should be acceptable as a minimum requirement for methodological presentations. But this is not to say that the assumptions with which they start are acceptable and not subject to discussion.

REIFLER: I should like to comment on the question of the quality or acceptability of machine translation. Before we go into this problem we should first consider what are the standards of quality and acceptability of human translations. Many years ago Legge translated the works of the Chinese philosopher Chuang-tzu about 400 B. C. and it was quite acceptable, not only by linguists but also by Sinologists. Finally, more people went into the field, read the original and found out how much was wrong in his translations, but it was acceptable

Session 4: METHODOLOGY

for quite a time. It was in good English but often not correct. Thus the readers or the intended sponsors play quite an important role in the judgment of the quality or acceptability. When we started out with our project our sponsors told us that they would be quite satisfied if the ultimate translation system would give them the general idea of what it was all about -- what the original Russian author intended. But in the course of our work our sponsors became more and more demanding. Their appetites came with eating. I think that the same thing has happened also with many other people working in machine translation or sponsoring machine translation research.

OETTINGER: I would hate to give a student a translation of a chemistry text in which the question of whether acid should be put into water or water into acid was left as a random translation process.

GIULIANO: This isn't a question, but it's an answer to Dr. Edmundson's question with regard to reliability and ability to develop indices that measure the quality of resulting translations. To some extent the ability to have measures of this type will probably follow an ability to translate well by machine rather than precede it. More concretely, I now firmly believe that so far as syntax is concerned the scheme of predictive analysis of Mrs. Rhodes (particularly the one program used at Harvard University) does provide indices of reliability insofar as it goes. It goes through the analysis of syntax and insofar as syntactic analysis is concerned there is an index — there is a way to tell when a mistake is made. That is, if there is an essential prediction for a subject and nothing is found that satisfies this, then the particular chain of analyses that hypothesize this prediction is incorrect. There are other indices with regard to the chain number. When we can handle semantics in the same way that this system can handle syntax, perhaps there will also be indices for measuring the quality of meaning transfer.