Session 6: SYNTAX

QUESTIONS AND DISCUSSION

TOMPKINS: I would like to take it upon myself to point out that a digital computer, for example, knows completely the sequencing of its operations, the sequencing of its words, and the syntax of its sentences, if you like; that, on the other hand, as we have observed today and throughout our lives, there is some considerable ambiguity in the structure of a sentence used for communication purposes between people, and there are several obvious things a person can try to do in analyzing the sentence structure: namely, he can try to give a unique solution as to the most probable structure, or he can try to give all solutions. I will predict that a fairly high degree of entropy will be circulating around here simply because we have not described which of these extremes, or some intermediate objectives, we are addressing ourselves to. So, during the discussion, I think it would be helpful if people would state rather explicitly what their objective is in the analysis; that is, whether they are trying to seek all plausible solutions, the most probable solution, or something in between.

HARPER: Perhaps before we get into a discussion of some of the different points of view that were brought forth this morning and the last couple of days, we could have statements of the reactions--statements of a comparative nature, of a defensive nature or offensive nature from the pre-discussants before we open the session to general questions and answers. May I first call upon Miss Evelyn Bristol, University of California at Berkeley?

BRISTOL: To begin with, it might be best to say something about the approach to syntax that we at Berkeley plan to take, since our group is relatively new and our personnel have not come from other MT projects. Dr. Garvin has given a good start for me, since our approach will most closely resemble his. We will begin from the bottom of the sentence, rather than from the top, and rather than making a left-to-right analysis, we will build up in blocks toward the top. Our difference with Dr. Garvin is slight. Rather than starting at the fulcrum, we choose the dependent element as it occurs in the sentence and place it with its head; in this way we build our blocks toward the

top. When there is a complex Russian sentence that requires analysis, the result is a necessary word-order change in English; and when we have built up these blocks toward the top, we will then come on the parts which need to be transposed as units, from one part of the sentence to another. I think that should give an idea of what we are doing at Berkeley.

HARPER: Thank you. Now can I call on Dr. Micklesen from IBM?

MICKLESEN: May I discuss a few generally accepted attributes of our subject of machine translation? Language, of course, is fraught with ambiguities. We do a great deal of fretting about the ambiguities we run across in our text. Ambiguities are inherent in grammar, and in semantics. I am sure that 10 years from now we will still be confronted with very obvious ambiguities at some very advanced level of syntactic recognition. Perhaps the way to solve this would be to select some best approximation. It seems we may never be able to completely approximate language because of the manifold problems. The best approximation, of course, is very difficult to define. I do not think anybody can do that at the present time. Let us say that we might start with a reasonably adequate approximation of a model of language. This, too, is very difficult to define. It seems to me that one might define it as a basically operational and compatible scheme depending on what kind of system one selects and, of course, basically linguistically sound. I think we all would agree on the principles involved in the latter at least. I am thinking of a quick and slick method of syntactic recognition. Let us apply the criterion of simplicity first of all. This must be quite comprehensive linguistically, but it must be quick and slick by some sort of operational methods. Then try to adapt this further. This is not a simple problem, of course. It is not simple to select a quick and slick method of attack, and it must also be very flexible. We must be able to build in restraints in the coding that come from more serious grammatical studies,based probably on the most powerful grammatical operations that are being developed, which will certainly be enhanced in the future. It must be flexible so that it will be able to include very easily any semantic information as this is accumulated. When I think about the restraints to be built into the code, I am thinking of the kind of work that was described today by

Dr. Applegate and yesterday by Dr. Matthews in their work on transformational grammar.

HARPER: Perhaps, since the next two discussants have been on the program before, we have an idea of what they are doing. From Dr. A. F. R. Brown and Mr. Sherry we can get something in the nature of a comparison of their approach to the syntactic resolution problem with those presented here this morning by the speakers. Dr. Brown, would you like to set off your approach to this problem against the approaches of other people?

A.F.R. BROWN: I will do so by contrast. I thought up the trivial sentence, "To this node we attach another sub-tree". Let this be translated out of French (my particular rock pile), in which "to" and "at" are the same word. The nice, middle-of-the-road, conservative approach of Dr. Garvin, and also of the Berkeley group, is much better. The best way to handle "To this node we attach a sub-tree" is probably by a small, tailor-made routine which is brought into operation by the word "attach". Of course, not everything can be done by small, tailor-made routines, but there is so much they can do so easily which, conceivably, general schemes do faster, but with much greater coding difficulty and much greater chance of error. I am a great believer in empiricism and in tailor-making the solutions to the problem.

HARPER: Mr. Sherry, perhaps you could say something on this question too.

SHERRY: You have heard before some ideas as to how predictive analysis works. I believe there is an illusion that it might be awkward. We believe that our program is quick and slick, though we use the words "flexible and efficient." I would like to discuss the problem that Dr. Zarechnak mentioned. To begin with, I would like to suggest a nice, simple example where there is no problem whatsoever and go through that to show how predictive syntactic analysis would solve it, and then go back to the original sentence which has been put back on the board.

Let me propose the beginning of a sentence,

эффективная в этом смысле методака ...

The first word is picked up as an adjective in the nominative case and it will predict a master which will agree with it in case, number, and gender. The next word is the preposition в; this will come in from infinity and will set up a prediction for either an accusative or a locative. That is what we call a prepositional complement, этом will be locative case, and it will be the prepositional complement. It will also predict a master, since it is generally used as an adjective, and смысле will agree with it in case, number, and gender and will be picked up as the master. смысле , being a noun, will predict a noun complement which is either an adjective or a noun in the genitive case; and therefore our prediction pool at this point will have this noun complement on top, since this is the last prediction made, and underneath it we still have the unsatisfied master from эффективная . методика is not the noun complement, and so it will get picked up as the master. This analysis proceeds very nicely, and you have the nested structure.

Now, having this simple example, we proceed to Dr. Zarechnak's problem sentence or fragment of a sentence. The first two words were из получаемых . The problem here was the fact that получаемых was in the genitive case, and therefore the master predicts that получаемых is also in the genitive case. We analyze на специальных машинах , and we have here a phrase again in the locative case. After машинах we now predict a noun complement in the genitive case. It is obvious at this point that you have a noun complement at the top of the pool in the genitive case, and you have a master underneath it in the genitive case, but a word can very easily satisfy both of them. The reason why the noun complement is at the top of the pool is that most of the time we feel the noun complement will indeed be the word that follows. In this manner we try the first time to pickup the most probable occurrence. In this case the most probable occurrence is wrong, because the words that follow are not masters of the word получаемых . Now, we will make the same error that the GAT technique made; however, having selected the noun complement, the master will go into hindsight, because this was just as good an intersection and just happens to be below in the prediction pool. On the assumption that the rest of the sentence is unambiguous (we do not have the rest of the sentence at this point), we do insist that this master be fulfilled. It will not be fulfilled on the one hand; but on the

other hand, we will have listed in hindsight precisely which one and only one word could have fulfilled this master prediction, so that on a subsequent pass it will be possible to change this to the master. To do this we did not have to get adjective-noun hookups first, then get noun-noun relationships, and then worry about prepositional-noun relationships. These all come in together; it is all done in one simple pass.

I do not believe that we can hope to tackle this kind of problem by looking at examples of failure and making a note of them. It seems a rather difficult method to get this information, and you have to look at a tremendous amount of text. In any event, I just cannot see how you can hope to do this perfectly; and what I consider very important is that the approach be fail-safe. This is as essential as anything is in this field of syntactic analysis.

HARPER: Professor Garvin, could you give us any reaction to this effective sentence with regard to the predictive method versus the pass method?

GARVIN: First of all, I should like to say that the relative efficiency of the pass method over a left-to-right predictive method cannot be judged without regard to the professional preferences of the originators. I think that the pass method, which I have been suggesting and which some people were kind enough to adopt, is based on a linguist's view of language and is obviously biased toward the conception of language in terms of nesting. I prefer the term "encapsulation" because it is prettier.

On the other hand, I have an impression--and this has been strengthened by the comment made by Dr. Hayes yesterday -- that the predictive method is a computerman's dream. Since I am not a computerman, my dreams differ from theirs. This is why at the last minute I added the phrase "in my opinion" to my evaluated statements; it is merely an opinion, and I don't hold it terribly strongly. There was a time when I said to myself, "Is there any point in my doing this when Mrs. Rhodes does things so cleverly?" Then I finally found a few reasons for going on with my own work. I do not yet know whether they are conclusive, and I think this is not false modesty but is essential in research. I do not think one can conclusively tell at this

point whether any of the preferences that we hold today is absolute.

As for myself, I have, in the past, revised some of my opinions although not the basic point of view that I have held; and I do not see any fundamental difference between my own conception of languages and that represented in the predictive method, because it does in a sense operate with fulcra. The reason this is the case, I believe, is because this is simply an obvious impressionistic insight that that suggests itself to anybody who looks at language with more than a casual eye. It is the same idea that Dr. Hays proposes when he talks about dependencies, and it is obviously comparable to the kinds of codes that are being assigned to words by various people, and so on. It is merely a matter of how you handle this basic concept.

Now, if I understand it correctly, the list of predictions is a list of possible fulcra or fulcrum-oriented functions that can be encountered as you go left to right; and instead of having one pass per problem, you call upon various and sundry subroutines to solve things as they present themselves in the text. I may misunderstand it, but this is my understanding of it as a non-computerman. I think that this is essentially a matter of the difference between a linguist's and a computerman's preference.

There is one factual difference, in a sense; and that is that in a pass method you do not have to carry out everything. It is perfectly possible to do a partial analysis for a special purpose. It is, for instance, perfectly possible not to make up all the necessary packages, but simply to ignore certain things and still come out with certain information about the sentence which may be significant for semantic purposes in terms of allowing you tabulations. That is to say, it is possible to write a syntax routine which will just identify verbs and their subject and objects for certain simply constructed sentences.

When it comes to the feature of fail-safeness, I think that it is obvious that in any honest routine there should be flags of error. In the truncated syntax which was programmed at Ramo-Wooldridge, there are two kinds of such flags. One is the notice of error which indicates that a condition has been found by the computer which, by the linguist, is not considered reasonable for Russian text. The second notice is the notice of syntactic difficulty, which indicates that a condition has been uncovered by the computer for which the program,

as yet, has no adequate provisions. In addition to this, I think that the linguistic approach to the problem, if it is to be a real working method, has to be debugged in exactly the same way as the program. That is to say, the fact that a particular analytic technique fails in a number of instances is not greatly different from the fact that a particular subroutine does not work because you failed to set a switch on time. It is just a matter of going through the labor of debugging it, and in that sense I am strongly in favor of running one's flow charts against text. If it is cheap enough to do it by programming, and if you have large enough samples available for the input of the computer, you can just program your routine and then run a large sample through and examine the residue. I think this in a large measure is the sense of debugging.

Finally, I want to say that I was very pleased about the topic of Professor Zarechnak's paper, because it points out the great importance of the appropriate sequencing of passes. One of the difficulties that arose was that the noun-noun relation was put into a pass which precedes the governing-modifier packaging. That is to say, if you attempt to establish certain kinds of extensions of the nominal block without previously having examined certain of the structures that can go inside a nominal block, you will be more likely to a failure than if you do examine them. I have a much nastier example than the one which is on the board. It reads с получаемыми этими способами результатами, wherein you have two instrumental nouns after each other and either of the two can fill the conditions of being the governor of either of the two modifiers. Now, the solution that we would propose is that we first put together very obvious agreement packages. This is the test operation now in progress at Wayne University. We would package up этими способами because it is obvious that they belong together. Then when we do a pass to uncover governing-modifier packages, we would pick up подучаемыми as a governing modifier, and we would read its government code. The government code would give us the information that it governs an instrumental, and we would then record the following instrumental package. Then we would set a switch indicating that the last noun of the instrumental package is also possibly in agreement with the governing modifier; and then if we find a second noun, we disregard the switch. If we do not find a second noun, we undo everything, disband

the package, and consider that the noun where the switch has been set is in agreement with the governing modifier and therefore not part of the governed matter. In this manner, by sequencing passes appropriately, you avoid including syntactic information which, on the face of it, is relevant but upon closer examination turns out to be not pertinent. This can be done just as efficiently from a purely programming standpoint by going left to right except that I have the impression that in going left to right and predicting, you have to do everything; whereas using the pass method, you can stop at any one point and just do as much as you need for a particular objective, which may not necessarily be translation but instead, research. Remember that the computer should be used as a research tool.

HARPER: Thank you. Mr. Sherry, do you have something to comment on?

SHERRY: I will answer in inverse order. As far as having to do everything is concerned, just as you can program to go through n passes out of m, we can put in n predictions out of m; so we can accomplish the same purpose with predictive analysis. On those grounds, I do not think it makes any difference. As to the matter of putting the passes in a given order, I question where you win. You may put the passes in any order you wish, and I am sure we can find an example for which you should have put them in another order.

If I can go back and comment a little more about this technique, I wonder precisely what is solved by the various preliminary passes. Certain ones are obvious, and other ones bring some questions to mind. You state that with your verbs you provide a number of codes--for the gerund, for the infinitive, and for the indicative, and so on--so that you can determine, to some extent, how this verb form is used in the sentence. Let us take, for example, an infinitive. Just that you found an infinitive does not tell you whether this infinitive is the subject of a clause or whether it is what we call a verb master -- that is, the infinitive that modifies any preceding verb form. In syntactic analysis, if this verb master appears, it is far up in the prediction pool, because this is one of these "immediately after" situations where the subject can get pushed down if it is not fulfilled initially. On the other hand, if you start with an infinitive and there is no other noun,

you would expect that the infinitive is the subject. So, the location in the prediction pool is extremely sensitive.

Another question on preliminary passes was that of homographs. I think you used the word есть as an example. We have made a number of tests on the Harvard automatic dictionary, and one of the features that we noticed is that these genuine homographs (a word that can be more than one distinct morphological type) occur approximately 5% of the time. We used a large number of texts for this sample, so I believe that within some reasonable error this is approximately the correct percentage to expect. I wonder how you would handle this, not knowing--at the time when you are trying to eliminate one of the two homographs--how it is behaving. In predictive analysis it does not make any difference whether you have a homograph of this form or whether you have a noun in the nominative or the accusative. There are still alternative functions. You list them and look for the intersection. If you get only one you are very fortunate; you have probably picked the right thing, unless there was some error way back. If there is more than one, you pick what you believe is the most logical one because this prediction was the last one put into the prediction pool and is therefore in its deepest nest, so to say; and if you have others, you put them into hindsight. You are not throwing them away in this preliminary pass.

HARPER: I am sure Professor Garvin has several answers for each of those questions, but perhaps we should ask Dr. Hays if he has any comments of a comparative nature.

HAYS: Yes, I would like to offer one slick trick for solving this question of how to deal with several possible functions of words that come right together in a sentence. When we find the Russian sequence: noun, passive participle, which by case agreement can agree with the preceding noun, and with a case requirement for a complement in the same case that can follow it, I am sure you will immediately tell me to look for a comma. That is not what we want to do; we do not like punctuation. We feel like the colonial printer who puts a couple of pages of periods and commas at the end and said, "Sprinkle as desired through the text". So, we change the form class of a passive participle when we find that it agrees in its ending with the preceding noun; we

mark it as a dependent of the preceding noun so that its new form class permits its complementation requirement to come into play. It cannot be connected to the following noun as the governor of that following noun unless it has been previously attached as a post-modifier to a preceding noun. If a participle is compared with a following noun and the noun could either govern or complement the participle, we have only choice, in effect: it must govern the participle because the form class at that stage does not permit anything else.

Something else I would like to point out is a system that does not have a name, unless you call it the precedence-dependence system. It is a programming technique that permits a variety of additions. For example, if we wanted to distinguish syntagmatic and syntactic levels, or any other set of levels that you like, then we could use two sets of agreement rules and, during the first phase of sentence-structure determination, make connections only from one table or one package of subroutines. Then after all possible connections have been made on that basis, we turn to the second table. The first might contain adjective-noun and noun-preposition-noun combinations, and the second table might contain noun-verb combinations and relative-pronoun-verb combinations. This is one of several variations on our basic method. The basic programming structure does not really have very much influence on the kind of linguistic principles that you add on. I think that this is a common characteristic of predictive analysis and precedence-dependence analysis. They treat a relatively small part of the phenomenon of language mainly dealing with word-order rules; then they deal with everything else separately, independently, and in whatever combination is appropriate and convenient.

HARPER: I think we should give Dr. Garvin a chance to answer the questions posed by Mr. Sherry.

GARVIN: I should like to answer very briefly. First of all, the idea of an iterative program strikes me as very interesting; for it is perfectly possible that, even if you carry out a pass method, you can have an iteration and run several passes through the sentence after each other because your first solution is unsatisfactory, and then you try it the other way. This may be necessary in certain instances.

Now, there are the few questions of detail which I can answer extremely briefly. One is regarding the infinitives, gerunds, and so on. All I meant to say was that they should have a separate grammar-code word-class designation, because they function differently from finite verb forms. An infinitive differs from a finite verb form in exactly the manner indicated, and that is why it gets a different word-class bit. It will obviously not be included in a search for a clause fulcrum in the same way in which the finite form would be included. Equally, one might have to have a pass which picks up infinitive packages and then it will later have to be decided whether this is a subject package, part of the subject package, part of a predicate package, or some other thing. As regards homographs, I think the only reason I brought up есть as a special example is that I would like to think of two kinds of homographs: the ones that are extremely systematic (for instance, the cases of neuter predicative adjectives which also may be adverbs), and the ones which are accidental. The accidental cases should be treated in a manner similar to other less systematic ambiguities, perhaps in connection with semantics, whereas the ones that are systematic can be treated in passes along more systematic lines. Finally, it is perfectly reasonable to say that all alternative functions are alternative; but in linguistics when you say that things are different, it is useful to ask yourself, "How are they different?" There are some differences that are more important than others. From the standpoint of syntax, word-class ambiguities have a different status in the functioning of the system from that of case ambiguities. This does not mean that anything should be ignored; it simply means that there is an order of priority in the solution. If you ignore a case ambiguity you may mess up a small block; if you ignore a word-class ambiguity you may mess up the resolution of an entire sentence.

HARPER: I would like to get some questions from the floor. Mr. Applegate, do you have any comments to make?

APPLEGATE: I might point out that our approach to syntactic analysis differs from those that have been presented in that we assume that if we are able to find the set of rules by which a sentence has been generated, we have, in effect, the analysis of the sentence. We

know how the sentence was produced, and this tells us what the sentence is. If it turns out that there are two possible sets of rules for generating a sentence, then the sentence will be ambiguous and this ambiguity will be clearly indicated.

ZARECHNAK: I have two or three general statements to make which may be of interest to future syntactic analysis, no matter whether you prefer the predictive, the pass, or the level technique. There are basic differences of purpose when you enter into syntactic investigations. Do you study that particular sentence on the syntactic level in order to translate the sentence into a target language? Or, disregarding the target language, do you want to describe that sentence in its own terms on the syntactic level? The possible consequence is that if you describe a Russian sentence per se on the syntactic level, you can hook up a French program, a Chinese program, or an English program and translate the same syntactic model into different target languages. On the other hand, if you study a sentence in Russian only for transfer into a particular target language, you will no doubt skip all the features that are useless from the point of view of translation into the target language. If you are going from Russian into English, then you are not going to analyze the gender of adjectives; but you must do so if you are going to translate the same sentence into Czech. Therefore, we have this fine distinction in syntactic analysis: is it syntactic per se, or is it only for translative purposes? There are disadvantages and advantages to both.

I would like to point out that in predictive analysis there is one particular section which looks to me like a syntagmatic one in GAT, namely, when an adjective is encountered and the question is asked, "Which is the master?" This is exactly the same as in GAT on the syntagmatic level when an adjective is encountered and the question is asked, "Where is the noun it agrees with?" I don't see any basic difference.

I would also like to point out a similarity between Professor Garvin's passes over the sentence and our leveling over the sentence. If you disregard the difference in terminology, it will be very difficult to find the real difference--with but one exception. Professor Garvin believes that the fulcrum, if identified as the finite verb, is more predictive than a noun. My experience has led me to believe that both

are variables, and both may be very predictive, or not. To show that the verb does not predict what the subject is, I will give two sentences: уголь добывают and добывают . The second sentence is one of impersonal structure. If you have two subjects иван и мария пошли в город . then пошли is plural; but do not expect that the subject will be plural, because the conjunction is in there taking care of two nouns that are singular. As to the predictability of the subject versus predicate, I think they are on an equal basis. Therefore, I do not see a basic difference whether the fulcrum is the finite verb, or the subject in the nominative case, or its substitute.

YNGVE: I can understand Dr. Garvin's point about preferring an immediate-constituent, or the inverse of the immediate-constituent, method of analysis of a sentence from a linguistic point of view. On the other hand, I think that there are now some linguistic results concerning the question of the left-to-right versus the pass method that point in favor of the left-to-right. These I alluded to very briefly in my paper. There are certain phenomena in English syntax that can be explained only on the basis of assuming left-to-right handling of sentences by humans. You can, of course, program a computer to handle a sentence in a left-to-right or in a pass method. It remains to be seen whether we should program our computer to do it the way the human does. I think it probably will turn out in the end to be simpler. We will have simpler programs, and we will have programs more easily understood, programs that divide into subroutines or rules much better, if we try to do it the way the human does, which is left-to-right.

LAMB: Dr. Bristol's comparison of our approach with that of Dr. Garvin is quite right. She also pointed out the difference. He starts from the fulcrum and looks for the other member, which we call the peripheral one; we start from the peripheral. In terms of the language used in the Copenhagen school, the peripheral member is also the one that presupposes something else. This is the same thing again, in slightly different language, as saying that it predicts that something else has to be there. So, it looks as if her saying that we agree with Dr. Garvin in going from top to bottom is not the whole picture, because I pointed out yesterday that we go left to right. It

looks as if we are in the middle ground between Dr. Garvin and both Dr. Hays and Mrs. Rhodes.

One of the most valuable things that I have heard at this Symposium is something which was not in one of the regular sessions but was in a small meeting that was held last night. Dr. Hays said that once every month or so he was prepared to furnish a list of information of the type that he is collecting: for example, verbs which govern что . This is the type of thing that we need more than anything else in the field of syntactic analysis, namely, the actual information. It appears, from my understanding of the various approaches, that this kind of information can be used and, in fact, must be incorporated into any of these systems before they will work. What is really important is that one has to have the necessary syntactic information built into the system in order for it to work. I think what we ought to concentrate on is not these fairly minor differences in procedure but getting the information that all of them must have before they can work.

TOMPKINS: Could we have a short remark by Dr. Edmundson next?

EDMUNDSON: I would like to point out that, perhaps for the first time, we are beginning to see serious mathematical terminology creeping into MT research. I am referring more particularly to the use of the words "topology, trees, group structures, Boolean algebras, and lattices". I hope that we will not find, in this field, what has happened elsewhere when convenient words have been adopted and corrupted. I hope that fellow mathematicians who are interested in linguistics will influence the precise use of these words, not to stifle research ideas but, quite the contrary, to make them more precise. For example, in connection with the tree-structure discussion that Dr. Hays presented, we should be more precise in talking about its topology. Some linguistics are concerned only with features of connectivity, others with features of partial ordering. I hope that MT researchers will pay some attention to this matter and use these now popular, but ever tricky, words with scientific care.

TOMPKINS: There is one point I would like to mention with regard to the left-to-right and right-to-left business. On my porch several years ago Professor von Neumann was playing not syntactic games

but spelling games, like words spelled backwards. My son, who was then 5 years old, came up to him and said, "How about Oppo?" Professor von Neumann said, "Go away, we are playing grown-up games." My son kept coming back and finally von Neumann, who was a very patient man, turned around and said, "Okay what does Oppo spell?" My son said, "Poop--inside out."