Toward a Theory of Computational Linguistics

by

Jacob Mey

Linguistics Research Center

and

Department of Linguistics

The University of Texas at Austin

1. Two Kinds of Models

To begin with, I would like to assert that computational linguistics (henceforth: CL), despite its qualifying adjective, has to do with human behavior, and, in particular, with that subset of human behavioral patterns that we study in linguistics. In other words, the aim of CL as a science is to explain human behavior insofar as it avails itself of the possibilities inherent in man's faculty of speech. In this sense, CL and linguistics proper both pursue the same aim. However, there are differences, as we will see shortly; for the moment, let us just establish that CL can be considered as a subfield of linguistics, and leave the delineation of the boundaries for later.

An important notion in behavioral sciences is that of a model as a set of hypotheses and empirical assumptions leading to certain testable conclusions, called <u>predictions</u> (on this, cf., e.g. Braithwaite 1968; Šaumjan 1966). I would like to call this kind of model the <u>descriptive</u> one. "Descriptive" here is not taken in the sense that Chomsky distinguishes descriptive adequacy from explanatory adequacy: indeed, the function of the descriptive model is to explain, as will become clear below. However, there is another respect in which the descriptive model reminds one of some of the characteristics attributed to Chomskyan models: it need not be (and should not be) considered a "faithful" reproduction of reality, in the sense that to each part of the model there corresponds, by some kind of isomorphic mapping, a particular chunk of "real" life. In other words, this descriptive kind of model does not attempt to imitate the behavior of its descriptum.

The other kind of model I propose to call the simulative one. As the name indicates, we are dealing with a conscious effort to picture, point by point, the activities that we want to describe. Of course, the simulative model, in order to be scientifically interesting, must attempt to explain; a _machina loquax_, to use Ceccato's expression (1967) is no good if there is a _deus in machina_. Although the idea of building _homunculi_, robots and what else they are called is not exactly a new one, the advent of the computer made it possible to conduct these experiments on a hitherto unknown scale, both with regard to dimensions and to exactitude. In fact, one of the popular views of the computer is exactly that: a man-like machine. Interestingly, the fears connected with this kind of image (such as an impending take-over by some super-computer like HAL in the movie "2001") have their counterpart in certain objections that are sometimes voiced against the other kind of model, the descriptive one: namely, that it de-humanizes human activities (such as speech), and establishes a new kind of man, made in the machine's image: machine-like man. Below, in section 3, I will discuss some of the implications of these views for computational linguistics; but first I want to raise the question: what importance do the two kinds of models have for linguistics itself?

2. Competence and Performance

The distinction between competence and performance in linguistics has been belabored often enough to let me squeak by here with a short restatement of Chomsky's remarks in _Aspects_ (1965: 4 et pass.): competence is the speaker's knowledge of a language, performance is what he actually does with his knowledge in a given situation that involves linguistic activity. A theory of competence, Chomsky says, is not a model of the speaker-hearer; according to the distinction made in section 1, above, I would rather say that it is not a simulative model, but a descriptive one. In other words, the model that is a grammar does not attempt to explain linguistic activity on the part of the speaker or hearer by appealing to direct similarities between that activity and the rules of the grammar. Rather, the activity of the speaker (his performance) is explained by pointing to the fact that the rules give exactly the same result (if they are correct, that is) as does the performance of the speaker-hearer: the set of all possible utterances of a given language.

Although a theory of performance thus is closer to the idea of simulating an actual linguistic situation, it is by no means identical with the simulative model. Rather, simulating actual linguistic activity depends on such a theory for its success; without it, a simulative model will be of little interest to linguists. To take an example: in any concrete linguistic situation there will be a lot of "unexplained" phenomena, such as hemming and hawing, false starts, anacolouths, etc. I feel that Chomsky is wrong in ascribing all of this to what he calls performance: linguistic theory should not account for these aspects of speech (they belong more properly in what one might call "corrective linguistics"). A simulative model wanting to represent this kind of "performance" would be a waste of energy and time.

What, then, is the proper object of a theory of linguistic performance? To understand this question is to answer it: if performance by definition is actual human activity, then linguistic performance is activity exercised by humans in the form of speech acts. In terms of the restriction made in the preceding paragraph, our "ideal" performance is that activity minus irrelevant "noise". Notice that this ideal performance does not coincide with that of Chomsky's "ideal speaker-hearer" of the language: as I understand this person, he is some kind of linguistic Superman (with unlimited memory, boundless embedding facilities, etc.). In other words, Chomsky's "ideal speaker" reflects competence rather than performance (in Chomsky's sense). To take a very simple example: the set of sentences generated by a grammar is potentially infinite; this is a fact of competence. However, any actual speaker or set of speakers will always generate some finite subset of the set of all possible sentences: a fact of performance. On a more sophisticated level, consider such questions as: why is it the case that regressive embedding beyond a certain bound is unacceptable? Chomsky calls sentences such as The rat the cat the dog chased killed ate the malt "perfectly grammatical" (1963:286); true enough, if one understands by this term: generatable by a competence model. But a performance model would have to incorporate some restrictions by which these "improbable and confusing" sentences (Chomsky, ibid.) would be ruled out. Actually, much of the research in the fields of psycho-, socio-, neuro-, etc., linguistics deals with performance; it is my thesis that computational linguistics, too, is a province of the same realm.

3. Competence and Performance in CL

The next question to be answered is: how do these theoretical considerations reflect on past and current work in CL? Until recently, very little attention has been paid to the performance aspect of CL. The only really large-scale computer-aided research in performance has been concentrating on machine translation and related areas. The lack of success that characterized these efforts has been material in turning off research funds as well as researchers. The result has been that CL workers now mainly direct their attention to such questions as: how to implement grammars on the machine; and: how to let the machine take over some of the work that linguists traditionally have done by hand? An example of the first kind is the transformational grammar developed by Friedman c.s. at Ann Arbor, formerly Stanford (1968 et seqq.); work in the second category ranges all the way from fairly unsophisticated and theoretically uninteresting "book-keeping" and "fact-finding" aids to theoretically motivated work in the development of syntactic and phonological rule testers (e.g., Londe & Schoene 1968; Fraser 1969). Common to this type of research is its ancillary character: these models (descriptive) purport to be an aid in the establishing of a theory of competence. As to performance (and, by inclusion, simulation), it is interesting to note that some of the more worthwhile results of MT research fall in the area of competence, too. I am thinking here of such by-products of MT as context-free and context-sensitive recognition procedures and their theoretical foundations (as explored, e.g. by Kuno, Greibach, Griffiths, Petrick, Peters and Ritchie, and, most recently, Woods (1970)). The results obtained in this area have certainly helped to clarify the theoretical issues involved, and as such, are of great value. But (as competence theory in general) they have not stimulated research or clarified any of the problems in the area of performance (except, of course, indirectly inasmuch as any theoretical development in one sector affects the whole field).

On the other hand, computerized efforts directed at simulating human linguistic performance cannot boast of any great achievements either. The fate of MT may have acted as a deterrent, but cannot be said to be the only reason why theoretical research has shunned, to a large degree, questions of simulation. In linguistics, in particular, the domineering trend of theoretical research was, until recently, to stay clear of what goes on in the speaker-hearer. As I pointed out above, in a sense it is perfectly true that a grammar is not a model of what is going on in the speaker's head; as Chomsky told the world in Syntactic Structures, "a grammar does not tell us how to synthesize a specific utterance; it does not tell us how to analyze a particular given utterance. In fact, these two tasks. . .are both outside the scope of grammars. .. " (1957:48). It should be kept in mind, though, that the grammars discussed here are concerned with competence, and that performance, in early generative grammar, was thought of as something less than ideal. I have the feeling, however, that the Manichaean streak which accompanied the distinction competence-performance at its birth is about to lose its power, and that competence now is seen as relevant only inasmuch as it can explain performance. But why talk about a theory of performance at all, then? Would it not be possible, with people such as Bar-Hillel (1970), to abolish the distinction altogether, and say: "competence is the theory of performance" or something similar? In the following, I will attempt to show that a theory of performance serves a purpose of its own, dependent on, but distinct from a theory of competence.

4. A Tale of Two Machines

In this section, I will conduct a Gedankenexperiment. *) Let us imagine two computers (or two computer programs), one (A) with the characteristics of a competence model (e.g., a system analogous to the transformational grammar described by Joyce Friedman), the other (B) resembling more or less Ceccato's machina loquax (see above, and also Mey 1968). Let us furthermore concentrate on the accepting part of the program, and try to figure out what happens in case the machines are

---

*) The basic idea behind this experiment is due to Schank (1970). Schank makes his purpose clear as follows: ". . .the notions of acceptability and grammaticality are part of the justification and purpose of transformational grammar. Our purposes are entirely different. In terms of analysis we are concerned with assigning a conceptual realizate to a string. " (1970:41)

confronted with a sentence that does not conform to their specifications. To take a concrete example, take the sentence: Colorless green ideas sleep furiously. Suppose A has built-in restrictions that, among other things, state that the subject of sleep has to be [+Animate], that the adjective green selects a [+Concrete] noun, and so on. Since the sentence presented to A violates almost all of the given selectional restrictions, the result would predictably be that A prints out a "reject" message, possibly with the reasons for rejection attached.

What would our "Zwittermaschine" (Klee 1926) B do? Since B is a model of a human, and expressly purports to imitate human behavior, we can look towards a human hearer to obtain an answer. (Klee wouldn't lie). I think it was Arch Hill who first remarked that such deviant sentences sometimes are very well received by humans; in some of his experiments, students thought sentences like the above to be not only "modern poetry", but "good modern poetry" (Hill 1961). There is also a persistent rumor around that Dell Hymes, having read Syntactic Structures, promptly sat down and conceived a poem whose first line read: "Colorless green ideas sleep furiously, . . . ". Not to mention, of course, that all-time status symbol, the bumper sticker carrying the same text and serving to fatten the pockets of some enterprising graduate student, while providing the more well-heeled members of the trade with a convenient shibboleth. To come back to our machine B: under the given presuppositions, it would have to find some way of imitating this human behavior, so disturbing to the creators of the selectional restrictions designed to produce the ultimate impossible sentence. For let us face it: there is no sentence so impossible that some human, in some devious way, cannot assign a possible interpretation to it. A quick glance at modern poetry will convince even the most incredulous (see also an article by Joseph Featherstone in The New Republic, 11 July 1970, "On Teaching Writing", where some interesting experiments in teaching children how to write poetry are described). This is not to say that selectional restrictions are for the birds (not even the one sitting perched on the leftmost handle of Klee's machine); only that it seems to be an innate human trait always to try to make the best of seemingly impossible linguistic input. If a machine loquax (or audiens, for that matter) wants to be true to its name, it will have to imitate this kind of behavior, and by doing so, explain some or all of it. [1]) And at this point I wish to discontinue the Gedankenexperiment, since I do not know how to make my machine do all this. But I hope to have made the issue clear: a simulative model, such as the one described, is different from a descriptive model. The difference becomes even clearer when one tries to implement both models on a computer. The simulative model requires a theoretical base of its own, since the theory of competence, by its own assumptions, rules out some phenomena that were described as typical for the human-like device. Conclusion: if CL wants to address itself to problems such as the ones involved in our little experiment, it will have to provide a wider theoretical base than the one accepted by most CL workers thus far. What we need is a theory of performance with special reference to CL.

5. Some further perspectives

In this final section, I will try to briefly indicate some of the areas in which I think a performance theory will be of use to CL. I will not propose any concrete solutions to any problems raised. The only aim I have set myself here is to provide some central perspective that I think may be fruitful to those working with the actual problems.

---

[1] Of course it would take a machine both loquax and audiens. So why not audax?

As a general preamble, I would like to discuss the question: what do we want to use CL and CL methods for? If the answer is: as an ancillary to theoretical linguistics, i.e., as a practical aid in solving some of the problems that theoretical linguistics poses, then the theory of CL is simply the theory of linguistics. Applications of this theory include, on the one hand such uses as grammar testers, on the other, such purely mechanical aids as automated dictionaries, programs for finding certain morphemes in a corpus, etc. If, on the other hand, the answer is: to implement and perfect actually working models of human behavior in the area of speech production and recognition, then CL needs a theory of its own. Some of the aspects of such a theory are covered, or should be, in what one might call "general robototology" (for some ideas on this, cf. Simon (1968)): questions pertaining to the interaction between robot and man, or even the "computer use of human beings", to paraphrase Wiener. Another general question is that of the degree of fidelity in simulation of human behavior, and the best way to implement this simulation. For example, what exactly does it mean: "to achieve a point by point imitation of human behavior"? Surely we do not want to reproduce certain states of the human that we consider irrelevant to the simulated process? In actual speech production, to take one example, we may very frequently be confronted with poor performance on account of extraneous conditions (colds, objects in the mouth, drowsiness of the subject, etc.) For a linguist, there is little point in examining and wishing to simulate these conditions. True, in marginal instances abnormal conditions may throw light on certain otherwise obscured processes; but this is not usually so. But even abstracting from these cases, there are areas where the difference between a competence approach and a performance approach manifests itself in the simulative set-up. Take again the example of embedded sentences.

Despite the fact that the recursive embedding rule permits unlimited embedding, actual sentences will always be finite, hence contain a finite number of embedded clauses. Hence the question arises: can we set an upper bound for embeddings such that, for a particular sentence, the depth of embedding will not exceed that bound? And, more importantly, how can we linguistically motivate such a decision?

Certain problems in the field of information retrieval have affinity to certain linguistic performance problems. For example, given a certain input to a question-answering system, how can one minimize the number of spurious answers, especially in the case of an imperfectly formulated question? Parallel to this is the problem of perfect understanding of imperfect questions by humans: how much do we really need to identify a given question and produce the correct answer? Traditionally, computational linguists have proceeded from the assumption that one first had to decompose the structure of the sentence (the question), then assign it a semantic interpretation, which subsequently is matched with the data file and produces the correct output. However, it seems clear that humans, in their analysis of linguistic input, often bypass the syntactic part and go straight for the semantics. A very simple and inadequate illustration is found in newspaper titles; a better one is provided by the ease with which small children handle conceptual structures without having the syntax correct. My own under-fours often produce rather complicated "sentences" that are perfectly intelligible, although syntactically completely ill-formed (or non-formed). As an example, consider the following: far gå huse ikke (Norwegian), where the negation is placed at the end of the sentence: ikke ('daddy go house not', i.e., 'daddy don't go to your study'). The most interesting thing about my 3-year old daughter's negative sentences is that the negation particle invariably is placed at the end, no matter how long the sentence. Think of the savings in syntactic analysis time we would obtain if we had this kind of input to English question-answer systems! Furthermore, in a construction such as the one above, certain transformations (NEG-placement, e.g.) are clearly being omitted; but this does not affect the recognizability of the sentence by a human, or even by a computer that would be programmed to recognize deep, rather than surface, structures. Consider also the ease with which a computer could simulate such negative sentences, rather than spend costly time on rearranging the <u>not</u>'s, <u>nicht</u>'s, and so on that are the horror of freshman classes in ESL or German.

I am convinced that simulation experiments will prove to be extremely useful by pointing up phenomena about human speech use that at present are being obscured by the overly abstract approach to grammar of the last decade or so. Current research in applied linguistics as well as in the so-called "hyphenated" areas seems to confirm the trend that is apparent in theoretical linguistics proper: a greater concern for naturalness and directness in explaining the phenomena of language, with an emphasis on semantics rather than syntax, also in CL.

References

Bar-Hillel, Y., 1970, Some Reflections on the Present Outlook for High-Quality Machine Translation. Austin, Texas: Linguistic Research Center (mimeogr.)

Braithwaite, R., 1968, Scientific Explanation. Cambridge, England: The University Press.

Ceccato, S., 1967, "Correlational Analysis and Machine Translation". In Machine Translation, A.D. Booth (ed.). New York, N.Y.: John Wiley & Sons.

Chomsky, N., 1957, Syntactic Structures. The Hague: Mouton & Co.

_________, "Formal Analysis of Natural Languages". In Handbook of Mathematical Psychology, Vol. II, R. D. Luce et al. (edd.). New York, N.Y.: John Wiley & Sons.

Fraser, B. & D. Bobrow, 1968, "A Phonological Rule Tester", Comm. ACM 11:766-783.

Friedman, J. et al., 1968 et seqq., A Computer System for Transformational Grammar. Stanford, Calif.: Stanford University (mimeogr.)

Hill, A. A., 1961, "Grammatically". Word 17:1-10.

Londe, D. & W. Schoene, 1968, "TGT: Transformational Grammar Tester". In Proceedings 1968 Spring Joint Computer Conference. Washington, D.C.: Spartan Books.

Klee, P., 1926, Zwittermaschine ('Twittering Machine'), New York: Museum of Modern Art.

Mey, J., 1968, Review of Ceccato (1967). Computing Reviews 9:317-318.

Šaumjan, S.K., 1965, Strukturnaja Lingvistika. Moscow: Izdatel'stvo Nauka.

Schank, R., 1970, "Semantics" in Conceptual Analysis. Stanford, Calif.: Stanford University (mimeogr.)

Simon, H.A., 1969, The Sciences of the Artificial. Cambridge, Mass., M.I.T. Press

Woods, W., 1970, "Context-Sensitive Parsing". Comm. ACM 13:437-445.

Note added in proof:

Having completed the final redaction of this manuscript, I came across Christine Montgomery's contribution to the 1969 International Conference on Computational Linguistics (Sånga-Säby, Sweden), entitled: "Linguistics and Automated Language Processing". On p. 17 of her paper, the author advocates the necessity of a theory of performance along much the same lines as I do. Interesting is especially the fact that support for such a need is provided by some findings in the field of data retrieval rather than by linguistic considerations only. This is because the speaker-hearer of a language, in providing utterances, relates not only to his innate ability, but to "the total environment of the speech event as well". The result is that "speakers can and do process sentences which the grammar is not capable of generating; in other words, the relation between the sentences of competence and those of performance is not one of simple inclusion."

I regret having overlooked this important contribution, and herewith offer to Christine Montgomery my apologies, and to my readers, the advice to consult the paper in its entirety, as well as the references quoted there on p. 17.