

1965 International Conference
on Computational Linguistics

GENERATION, PRODUCTION, AND TRANSLATION

Petr Sgall

Malostranské n. 25

Prague 1

Czechoslovakia



Abstract: The notions of generation, production, synthesis etc. are analyzed, with the conclusion, that the description of a natural language should have ~~exactly~~, a creative and a transductive part. As a system meeting the principal known conditions, a sequence of pushdown store transducers, interpreted as a stratificational language description, is proposed.

1.1 In the works of N. Chomsky, P.M. Postal, J.J. Katz and others there are formulated three general aims of generative grammar:

- a) the grammar generates, all and only the grammatical sentences of the described language, in the sense of a mathematical enumeration of the set of grammatical sentences;
- b) the grammar automatically ascribes to each generated sentence a structural description, not contradictory to the intuitive evidence of native speakers; if the same sentence is generated by the grammar in n different ways, n different structural descriptions are ascribed to it;
- c) the grammar is a description of the mechanism the user of the language has internalized and uses in the process of language communication (as a speaker, hearer etc.).

There are, of course, other aims or conditions laid upon generative grammars, e.g. those concerning degrees of grammaticality of sentences. But our main point here concerns the three conditions cited; we assume that these are actually three different conditions, in particular that b and c are not identical, b being implied by c, so that a grammar can exist, which satisfies the conditions a and b, but not c.

1.2 If we examine the new version of transformational description ^(2;5;7;1) from that point of view, we see that there are difficulties concerning the condition c. Of course, we are now applying the conditions not only to the grammar, but to the whole description, including its semantic component. We are not certain, however, that the mutual relations of the three components of this description have been characterized quite correctly. There is one creative component, (A), whose output is interpreted by

both (B), i.e. semantics, and (C); but then - in my opinion - (A) is the basis (including lexical rules), and the transformational and phonological parts belong to (C). Whether this is the case or not, the formulation of the new shape of the transformational description is a further and important step in overcoming the one-sidedness of descriptivism. An adequate basis for fulfilling the conditions a and b may be considered to have been reached in this way.

But our question remains: can this system really be considered to describe the mechanism used in speech communication?

J.J.Katz writes (/5/, p.4; similarly in /6/): "The whole process may be pictured as follows: the speaker, for reasons that are biographically relevant but linguistically not, chooses a message he wants to convey to the hearer. ...This message is translated into syntactic form by the selection of a syntactic structure whose semantic interpretation is this message." Of course, the question, why a particular message has been chosen, is not a linguistic one. But if the mechanism used by the speaker to translate the message into syntactic form were to be described as a device for translating, not just for selection, this other description would be preferable, as to that point.

1.3 The generative description is neutral to the speaker and to the hearer, i.e. to the production and understanding of sentences. But in a sense the generation is nearer to the production, just as the recognition procedure is nearer to the understanding of sentences. That is, the sentence appears as an output string in the generative description and in the process of production, whereas in the recognition routine and in the process of understanding it appears as an input string.

N.Chomsky introduced the distinction of the competence and the performance of language users; the methodological importance of

this distinction is undeniable. But it is not evident whether the generative grammar is the same as the description of the competence of both speaker and hearer; it is not clear whether a recognition grammar belongs to the realm of performance (cf. Hays /4,p.8/), or to the realm of competence, also (cf. categorical and other grammars, formulated as recognition grammars, e.g. /1/, but not intended as models of the hearer's activity). One sees, of course, that the actual programs for random generation of sentences /17,9/ are not models of the speaker's activity, either. But that is not due, perhaps, to their being related to generative grammars.

So our questions are: Has a description of the user's competence to be neutral with regard to the speaker and the hearer? (Or, better: Are all the differences between the speaker's and the hearer's activity due only to their performance, are there no differences which should be respected in a description of their competence, of the mechanism they have internalized?) And, further: If the generative description is neutral in this respect, is the recognition description also neutral? - If we assume that the set of grammatical sentences is a recursive one, then why should we prefer generation as the form of the description of language, and not recognition?

2.1. The mechanism internalized by the user of a language enables him (I) to choose a message as such, (II) to formulate this message, i.e. to translate it from its "semantic" form to one of the corresponding phonetic forms, (III) to "interpret", or understand a sentence, i.e. to translate it from its phonetic form to any of the corresponding "semantic" forms, (IV) to choose the most appropriate form, while translating in either direction. The parts (I) and (IV) cannot be fully described by linguistic

means alone. As to parts (II) and (III), we see that they are essentially translating procedures. They may include some blocking or checking sub parts, but we should try to describe at least their main parts as transducers or in some similar way (or, alternatively, to show that this is impossible).

The vast and heterogeneous field presystematically called semantics should be analyzed from such a point of view; there are elements in it which belong to the "transductive" part of the mechanism and are purely linguistic; the relations between levels or strata of the language system can be described as "semantic" relations (relations of "representation", with the so-called "assymetrical dualism", i.e. generally many - to - many relations /⁸14/). On the other hand, there are other, quite distinct, "semantic" problems, concerning the manner of choice of a message itself, and of these only some are intrinsically linguistic (e.g. questions concerning analysis^{ficity}, contradiction).

2.2 Even if the mechanism briefly outlined above could be integrally described, it would be a description of the competence, not of the performance of a user of language. It would not describe, particularly, the relations connected with restrictions of memory and other faculties of individual speakers, nor various possible short-cuts used during the process of speech. It would hardly be possible to describe, by linguistic means, the conditions under which a construction is being built up^{by} the speaker during the act of speech, and the conditions under which it is delivered as a whole by the memory; only extreme cases, where there is only the latter~~xx~~ possibility, are distinguished explicitly by the grammar; the former constructions are, e.g. "analytical" forms of newly coined or accepted words formed "by analogy"; at the other extreme are idioms, "irregular" forms and other "exceptions", which have to be listed

by the grammar. In most cases, there are both possibilities, and it is not one of the purposes of linguistics to say whether (or, under what conditions) the speaker chooses the first or the second.

We can say, then, that the distinction between competence and performance is independent from that between speaker and hearer. If the description contains some part(s) used only by the speaker, and others used only by the hearer, it does not follow that it is no longer a description only of their competence.

2.3. If I understand correctly the words of J.J.Katz, cited in 1.2, then he defines the message as a unit of the semantic level (the message is a semantic interpretation of a sentoid). Enumeration of all possible messages (of a language) is a necessary part (or prerequisite, perhaps) of a description of a language, and it is, of course, preferable that the mechanism of this enumeration should ascribe structural descriptions to the generated messages (see a,b in 1.1). But it is not clear whether the transformational description really can serve as such an enumeration. If so (e.g. if the ~~X~~ semantic component does not, as a filter, block an infinite number of sentoids), then it would be possible to reverse the semantic component in some way (to discover its inverse machine), and use this inverted procedure as the description of the mechanism the speaker uses while translating the message to its syntactic form. If there is no such possibility, then the semantic component in its present form is not satisfactory, either.

On the other hand, it is not necessary, in a linguistic description, to generate the set of messages by a system which could serve as a model of the mechanism internalized by the users of a language (1.1.c). Bearing in mind, that the description of the complicated relations between the semantic and phonetic levels, and not the selection of the message, belongs to the classical aims of linguistics, we can model not the mechanism used by the speaker and the

only its part
 hearer, but ~~that~~ used by the translator, who has to discover the ~~the~~ meaning(s) pertaining to a certain phonetic string and vice versa, but not to choose the meaning itself. (But cf. 3.3.)^x Of course, we have to distinguish here between cases where translation proceeds only as an operation on texts, without using other than linguistics knowledge and knowledge obtained from the translated text itself, and cases, where the translator has to obtain from external sources information not contained in the text but required by the grammar of the output language (in cases as "king John's son" translated to German "ein Sohn des Königs J." or "der Sohn...", cf. /13/). We are concerned here only with translation in the narrow sense; the mechanism described does not enable the translator to select one of the possibilities in the cases where such external information is necessary.

3.1. Viewed~~x~~ in this way, the generative system should have ~~two~~ two parts: a purely generative one, serving to enumerate possible messages, or meanings of messages, and another "transductive" part, describing the translation of messages from one level to another. The message can be composed, of course, of many sentences, but, as is usual today, we take no account of the difference and speak of sentences only, in the sequel.

This system should not be more elaborate than necessary, but there are some known conditions it has to fulfil. As to weak generative power, it has to generate not only the sets of strings generated by a context-free phrase structure grammar, but also at least the set of all strings of the form xax, where a is a symbol of the output vocabulary and x is any string of such symbols not containing a; cf., in a somewhat different formulation, /12/. As to strong generative power, the system has to ascribe not only taxonomical structural descriptions, i.e. there must be a possibility of having

structural descriptions not just for sentences, but for several levels of sentence structure, such as the deep structure and the surface structure.

3.2. One possible system meeting these conditions is a sequence of pushdown store transducers /11;3/, interpreted as a stratificational description /8;4/. Its first, generative and recursive part would be a context-free phrase structure grammar or a system weakly equivalent to it, enumerating the "sentence meanings" in such a way that the output string itself serves as the description of the sentence structure at this level. This string is translated then by an other pushdown store transducer (or other *such* transducers) to the level of "surface structure" (a linearized ^{tree} version of the dependency ~~configuration~~ of the sentence), and further to the morphological level. For a ~~diagram~~ ^{rough outline} of such a system, see /15;16/.

We can formulate the first part of this system so as to generate all strings of the form \underline{xax}' and only them (where \underline{x}' is the reflection of \underline{x}), and the second part can consist of the following rules (symbolized here as done by Evey, with some simplification; \underline{b} denotes any symbol of the input alphabet other than \underline{a}):

input	p	inn.	-->	inn.	p	output
		state		state		
b		1		1		b
a		1		2		a
b		2		2	b	
∅		2		3		
	b	3		3		b
	∅	3	stop!			

The output of this translator would be xax, whenever xax' was its input string. So the first condition of 3.1. is fulfilled. It should be noted that we are not using the word "translate" in the technical sense used by R.J.Evey; the input language of our translator contains the set of the strings xax' as its proper subset. There is still a possibility that, at least for some languages, a description would be adequate where the input language of each translator is the same as the output language of its predecessor in the sequence of machines. The weak generative power of such a system would then equal that of a context-free phrase structure language (cf./13/, theorem 2:6.6).

As to the second condition, concerning strong generative power, our system, as a stratificational one, shares the main property of transformational description and cannot be held to be taxonomic. There are three levels or strata (not to speak of phonology and phonetics; but the question of the number of strata is an empirical one and it is possible that typologically different languages do not coincide in this respect; we are working with inflected languages, such as Czech or Russian): the tectogrammatical or "semantic" level (L_0), the phenogrammatical level (surface structure, L_1), and the morphological level (L_2). The sequence of representations of a sentence on each of these levels, a member of $L_0 \times L_1 \times L_2$

Fulfilling the condition that each of its elements is derivable by the transducer system (from its predecessor in the sequence, if there is any), is a structural description of that sentence. A complete derivational history is not needed here, because the symbols useful for the structural description appear in the terminal strings of the transducers, so that these strings (or some of them, as a sequence of two transducers is generally needed to translate the sentence from one level to another) themselves serve as the

representations of sentences at the corresponding level.

So, not a unique description of the syntactic structure is ascribed to a sentence (with no homonymy) but, rather, two distinct descriptions (elements of L_0 and L_1 , respectively), whose mutual relation can be compared with that of Chomsky's deep structure and surface structure. In cases such as questions, grammatically conditioned change of word order apparently disturbing the constituent structure of the sentence, nominalizations etc., the rules of our system are to a great extent parallel to those of the transformational description. But cases such as the passive or lexical synonymy and homonymy are handled in our system in such a way that the representation of the sentence at the tectogrammatical level is unique for the synonymous expressions (and distinct, of course, for the homonymous ones).

The third level of our system, the morphological, includes categories of "rection" and concord, not included in the "syntactic" levels.

3.3. The recursive component of our system, with the initial symbol \underline{S} and the gradual derivation of the terminal string first by rules of expansion and then by rules of selection, does not fulfil condition \underline{c} of 1.1. But, as stated in 2.3., it is not necessary that a linguistic description should contain a description of the mechanism used by the speaker in that part of his activity. Possibly this gap can be filled in the future by reformulation of the generation routine.

There are other difficulties in this component, concerning the condition \underline{b} . The main point here is the formulation of context restrictions relevant to the selection of terminal symbols; we think that it is possible to define various partitions on the vocabulary of this component, so that we can work with rule schemata

not only as with a conventional simplification of notation, but also as with an intrinsic part of the system; the weak equivalence with a context-free phrase structure grammar should not be lost by this.

Another remark concerns the confrontation of the tectogrammatical levels of two languages, necessarily made by a translator. An ideal case of translation would be the full coincidence of both these levels. Generally, the translator, if he is to translate correctly, is supposed to use (as a part of the mechanism internalized) some routine converting strings of the tectogrammatical level of the input language to those of the output language. This part of the mechanism is not as yet described by our system, as we consider one language only, i.e. a translator translating from his mother tongue to the same language. But elaboration of this part of the system is needed, and is directly relevant to Machine Translation as well as to general linguistic theory.

3.4. The transducers have to fulfil certain special conditions, so that the existence not only of an inverse machine for each of them but also of a recognition routine for the whole system is ensured as well. These conditions, which we shall not present here in full, are parallel to the absence of deletion rules in a phrase structure grammar under similar circumstances. We can say, informally, that no symbol read in the input is deleted by the rules of a transducer, except in cases where that symbol can be determined uniquely by the resulting output. Free deletions are confined to transduction from the morphonological to the phonological level, or from the phonological to the phonetic level. At those stages, the existence of a recognition routine is given by the fact that the corresponding transducers are finite state.

The system can serve then, in its transductive part, as a

description of a part of the translator's competence, as well as a basis for translation algorithms. As to the actual performance, as well as the actual algorithms, here various digressions from the process defined by the transducer system can be found, as well as restrictions due to the finite memory etc. A hearer performing (mostly in an unconscious way, probably) the process of understanding a sentence, does not have to follow the rules of the above mentioned inverse machines punctually. He can use various shortcuts or trial-and-error methods, checking the results (cf. Matthews, /10/). But that may also be true for the speaker, as far as we know.

3.5. With regard to terminology, it is possible to distinguish several pairs of notions concerning routines of both kinds:

the sentence (its phonetic representation) is		note
in the output	in the input	
enumeration	decision	definition of a set
generation	recognition	...combined with automatic ascribing of structural description
synthesis	analysis	...interpretable as the description of the mechanism used by the translator
production	understanding	...process of performance

4. So we arrive at the conclusion that the connection between the theory of grammar and the purposes of Machine Translation is an intrinsic one. It is of importance to practical research (where research on MT, using the findings of the theory of grammar and algebraic linguistics, makes possible through study of its empirical

questions and testing of its working hypotheses). Moreover, the relationship between the theory and the purposes of translation has a bearing on the understanding of the theoretical concepts themselves. And the importance of this aspect will increase, probably, as theory proceeds from the study of individual languages to the description of language universals and other relations between natural languages, and to the description of language in general.

L I T E R A T U R A .

- /1/ Y.Bar-Hillel, Language and Information, Jerusalem - Reading - London, 1964
- /2/ N.Chomsky, Categories and Relations in Syntactic Theory (mimeogr. for the Conference on Structural Linguistics, Magdeburg, 1964)
- /3/ R.J.Evey, The Theory and Applications of Pushdown Store Machines, Mathematical Linguistics and Automatic Translation, Report No. 10-NSF of M.I.T., Laboratory of Electronics, Cambridge, Mass., 1963
- /4/ D.G.Hays, Dependency Theory: A Formalism and Some Observations, Memorandum RM-4087/PR, The Rand Corp., Sta. Monica, 1964
- /5/ J.J.Katz, The Semantic Component of a Linguistic Description (mimeogr. for the Conference on Structural Linguistics, Magdeburg, 1964)
- /6/ J.J.Katz, In ~~the~~ Defense of Mentalism, Language⁴⁷, 1964
- /7/ J.J.Katz - P.M.Postal, An Integrated Theory of Linguistic Descriptions, Research Monograph No. 26, M.I.T., Cambridge, Mass., 1964.
- /8/ S.M.Lamb, The Semantic Approach to Semantics, Amer. Anthropologist 66, No. 3, P.2, 1964, 57-78
- /9/ R.B.Lees, Automatic Generation of Natural-Language Sentences, read at the Colloquium "Foundations of Mathematics...", Tihany, 1962 (in press)
- /10/ G.H.Matthews, Analysis by Synthesis of Sentences in the Light of Recent Developments in the Theory of Grammar (read at the Colloquium on Algebraic Linguistics and MT, Prague 1964 (in press in Kybernetika)

- /11/ A.G.Oettinger, Automatic Syntactic Analysis and the Pushdown Store, Structure of Language and Its Mathematical Aspects, Proc. of the Symp. in Appl. Math. 12, 1961, 104-129
- /12/ P.M.Postal, Constituent Structure, Indiana Univ., Bloomington (=IJAL 30/1, Part III), 1964
- /13/ I.I.Revzin - V.J.Rozencvejk, Osnovy obš^ěego i mašinnogo perevo-da, Moscow 1964
- /14/ P.Sgall, Zur Frage der Ebenen im Sprachsystem, Travaux linguistiques de Prague 1, 1964, 95-106
- /15/ P.Sgall, Generative Beschreibung und die Ebenen des Sprachsystems (read at the Conf. on Struct. Ling., Magdeburg, 1964)
- /16/ P.Sgall, Ein mehrstufiges generatives System (read at the Colloquium on Algebr. Ling. and MT, Prague, 1964)
- /17/ V.H.Yngve, Random Generation of English Sentences, in Proc. 1961 Intern. Congr. on Machine Translation of Languages and Applied Lang. Analysis, Teddington (1962)