# Building a psychologically plausible sentence generator

**Gerard Kempen**

**Department of Psychology University of Nijmegen**

**Abstract**

The psychological process of translating semantic into syntactic structures has dynamic properties such as the following. (1) The speaker is able to start pronouncing an utterance before having worked out the semantic content he wishes to express. Selection of semantic content and construction of syntactic form proceed partially in parallel. (2) The human sentence generator takes as input not only a specification of semantic content but also some indication of desired syntactic shape. Such indications, if present, do not complicate the generation process but make it easier. (3) Certain regularities of speech errors suggest a two-stage generation process. Stage I constructs the "syntactic skeleton" of an utterance; stage II provides the skeleton with morpho-phonological information.

An outline is given of the type of grammar which is used by a sentence generation system embodying these characteristics. The system is being implemented on a computer.

If you ask a psychologist to give a definition of “sentence production” you are likely to get an answer like: the process of translating semantic (conceptual) structures into syntactic structures. Such a definition may seem plausible and fairly straightforward. But when you give it a close look it appears to carry two presuppositions that are not completely harmless — to say the least. First, the definition suggests that the input to the sentence construction process is a semantic structure and only a semantic structure. Second, this semantic structure is assumed to be present at the moment sentence construction begins — in other words, the process of putting together a sentence waits until after final decisions have been made as regards its semantic content.

Both presuppositions are shared by all sentence production systems of linguistic origin which generate from some specification of semantic content. Grammars of generative semantic signature are one class of such systems; another class is artificial generators as developed by computational linguists. However, I think that both assumptions are incorrect from a psychological point of view. Let me quickly review the empirical evidence supporting this claim. After that, I will sketch the sentence generator that I am presently working on one that does not accept the two presuppositions and looks rather different from the linguistic generators just mentioned. (For details see Kempen, 1977a,b,c.)

## 1. Is semantic information the only input to the sentence generator?

In many situations of everyday speech the speaker is guided not only by the content he wishes to convey to his audience but also by information on how to express that content syntactically. I will give some examples. If a person is asked to give a definition of a word, he has available to him a limited set of syntactic possibilities for phrasing the definition sentence (e.g. An X is a Y which …) Also, there are types of connected discourse where syntactic constraints govern the shape of whole sequences of sentences. A clear case is syllogisms with obvious constraints on both content and sentence form. A less academic illustration is the well-known opening line of fairy tales.

What these examples suggest is a generator which takes as input not only some piece of semantic content but at the same time some indication of desired syntactic shape, and one which is indeed able to find a formulation showing that shape. I will make the assumption that the input to the sentence generator is a pair of specifications, one for semantic content, one for syntactic form. Thus we get rid of the usual presupposition that the syntactic specification of the to-be-constructed utterance is empty whereas its semantic specification is full and unchangeable.

But there is more. Not only should it be possible for the generator to bias its output towards certain prespecified syntactic shapes, it also has to find this easy. Many a radio reporter who is reading out the results of a series of soccer matches is subject to a "natural tendency" to use one and the same sentence frame for several successive matches. He even has to spend a little extra mental effort trying to attain the desired level of syntactic variation. Apparently, if semantic and pragmatic factors permit two successive sentences taking the same syntactic format, people find the second sentence easier to produce than the first one. However, the linguistic generators mentioned earlier probably predict just the opposite. Since in those systems there is no guarantee for the second sentence to come out the same way the first one did, extra precautions are necessary while generating the second sentence, thus making it harder to produce. (In a generative-semantics grammar this can presumably be handled by global derivational constraints.)

## 2. Is semantic content completely specified at the moment sentence construction is initiated?

The second above assumption seems wrong on both intuitive and experimental grounds. Most people will have experienced situations where they initiated overt speech production after having worked out only a fragment of the semantic content of the resulting sentence. In such cases content selection and syntactic structure formation proceed, at least partially, in parallel, not strictly serially.

This conclusion also follows from a recent series of sentence production experiments by Lindsley (1975, 1976). He had experimental subjects produce simple Subject-Verb (SV) sentences such as The boy is greeting, The girl is kicking, etc., as descriptions of pictures presented to them. The action shown in a picture was either greeting, kicking, touching, or simply standing. The actor was either a girl, a boy, a man or a woman.

The measurements collected by Lindsley were reaction times needed by the subjects to look at a picture, to recognize the actor and the action, and to initiate the utterance. That is, a reaction time was defined as the interval between the moment a picture appeared on a screen, and the onset of the vocal response by the subject.

Independent measurements showed that the actor of a picture could be identified faster than the action, the difference being some 110 ms. (This difference was measured by having the experimental subjects name the actors alone, or the actions alone.) In our terminology, the actor (the semantic information underlying the surface subject noun phrase) could be fed into the generator about 110 ms earlier than the action (the semantic information underlying the verb phrase).

Lindsley used this experimental setup in order to decide between three possible temporal alignments of content selection (here: recognizing actors and actions) and form selection (here: looking up lexical items from a mental lexicon and combining them into a SV sentence). One alignment is the serial one implied by our second presupposition: sentence construction waits until the complete content is in. Another possible alignment is strict parallelism: as soon as a semantic fragment (e.g. the actor) has arrived, it is immediately translated into a piece of sentence and pronounced. The distinguishing characteristic of this idea is the assumption that no provisions are taken to guarantee syntactic coherence between the various sentence fragments uttered consecutively. That is, the speaker behaves as if the successive sentential fragments were totally independent responses. (Parenthetically, this model comes close to the ancient doctrine of natural word order according to which order of words in a sentence mirrors accurately order of thoughts in the speaker’s mind) Needless to say that the third possible alignment, partial parallelism, is the realistic one. When a semantic fragment has become available, the generator will immediately try to convert it into a piece of sentence while taking precautions to maximize syntactic coherence between successive parts of the sentence. This may imply interrupting sentence construction for a while until more semantic information has come in; this later semantic fragment may then very well be expressed earlier than its predecessors.

Lindsley was able to demonstrate that the latter model, partial parallelism, was the best one. This may not be surprising as such. But remarkable, at any rate, is the fact that the model is valid in a situation where the arrival times of two semantic fragments were separated by an interval of only one-tenth of a second. This suggests that the phenomenon of partial parallelism between content and form selection is very widespread and must be considered rule rather than exception.

### 3. **Sentence construction as a two-stage process**.

Another constraint on possible designs for a human sentence generator stems from the study of speech errors. Some two years ago, Garrett (1975) published a very detailed analysis of more than 4000 speech errors. Apart from classifying them into various types — as other researchers had been doing before him — he computed values for a number of statistics which proved to be very revealing.

A large proportion of speech errors consists of exchanges of linguistic units — exchanges of complete words, of lexemes, morphemes, syllables, phonemes, and even single phonological distinctive features. The first relevant statistic here is distance between the words involved in an exchange. E.g. in the erroneous sentence Every time I put one of these buttons OFF, another one comas ON, the exchange is between full words (printed in capital) that are three words apart. In Bill snovels show, which arose from Bill shovels snow by a phoneme exchange, the distance between affected words is zero. Another statistic refers to word class of words involved. In the shovel example the affected words belong to different word classes; in the buttons example they are members of the same class.

Without going into further detail I just summarize Garrett's results. Exchanges between full words behave differently than all other exchange errors with respect to both statistics. In case of word exchanges the affected elements are farther apart and belong to the same word class much more often than in case of any other error type.

It seems plausible to invoke some notion of similarity and/or simultaneity in order to explain the interchanges between linguistic elements. Elements that the generator is processing simultaneously or almost simultaneously have a chance of getting mixed up, especially when they are similar. As for interchanges of full words, similarity/

simultaneity is clearly unrelated to being close together in the output sentence, but rather to their grammatical functions. All other exchange types, though, occur primarily between words which are close to each other in the utterance produced by the speaker. From this consideration it is only a small step toward positing two processing stages during sentence production. In one stage, elements of similar grammatical function are processed (almost) simultaneously; in the other stage it is adjacency within surface word order that determines simultaneity of processing.

Taking these and other speech error data into account one arrives at the following two-stage organization of the sentence construction process. During Stage I, lexical items covering the to-be-expressed semantic content are looked up from the lexicon and assigned grammatical functions. Stage II takes care of final morphological and phonological shape of the lexical material brought together in Stage I. This division of labor between stages reflects an important regularity that has been observed in all exchange errors: lexical content items never interchange with syntactic (mainly inflectional) morphemes. By assigning lexical selection and syntactic morphology to different stages, this empirical rule is nicely accounted for.

At this point I wish to introduce one new term. I call the structure that is delivered by the activity of Stage I, the syntactic skeleton of the sentence. It is a sequence of lexical items provided with a structural description which includes, for each item , a specification of word class membership and of grammatical relations to other items. The structural description enables rules of syntactic morphology to apply correctly (during Stage II).

The term "sequence of lexical items" in the definition of a syntactic skeleton must be understood as follows. First, lexical items in a syntactic skeleton aren't phonologically specified yet but only contain a pointer to a memory location where the phonological form is to be found. There is sound, experimental evidence (again, collected by Lindsley, 1976, Experiment III) for splitting up the process of lexicalization into two steps, one for lexical selection (finding a lexical item which covers a given piece of semantic information), and another one for lexical insertion (substituting a phonological form for a pointer).

Lexical selection makes part of Stage I; lexical insertion is done in Stage II. Second, the left-to-right order of lexical items in the syntactic skeleton corresponds exactly to their order in the spoken utterance and won't be changed by Stage II. This assumption follows directly if one accepts partial parallelism as a model for the temporal alignment of content and form selection. As stated before, this model says that a semantic fragment arriving at the generator will be translated into a sentence fragment and overtly pronounced as soon as a syntactically coherent formulation has been found. This implies considerably less freedom for moving lexical items back and forth than is needed for transformations, such as Passive, Extraposition and Subject-Raising, to apply correctly.

I want to conclude my discussion of the two-stage model by warning against a possible misunderstanding. It would be incorrect to assume that the generator completes all Stage I processing for a sentence before moving on to Stage II. On the contrary, the generator may already have finished up Stage II for the initial parts of an utterance while still in Stage I for later parts. This is demonstrated by so-called lexical hesitation pauses during spontaneous speech. While in the middle of a sentence or a clause, the speaker hesitates, unable to find the right word for a semantic structure he wishes to express. That such pauses are indeed reflections of lexical search is attested, among other things, by manual gestures speakers make during lexical pauses: such a gesture often means the same thing as the word following the pause (Butterworth & Beattie, 1977).

## 4. Building syntactic skeletons.

The central question, of course, is: what does the procedure which builds syntactic skeletons look like? Needless to say I haven't worked out a full solution to this problem yet. Given the time and space limitations of this paper I'll outline the grammar I developed for a sentence generator which is currently being implemented on a computer (a PDP 10, programming language MLISP). I won't discuss Stage II anymore; elsewhere (Kempen, 1977c) I have dealt with its organization in some detail.

The input to the sentence generator is a pair of a semantic and a syntactic specification. The semantic specifications in my system refer to conceptual dependency networks as developed by Schank (73, 1975). A syntactic specification is a node labeled by a list of one or more syntactic features. If a full sentence is going to be constructed (and not just a noun phrase, for instance), then its syntactic skeleton is generated from an initial node labeled by at least the features [Category: Clause; Subcategory: Main].

One part of the grammar is a set of rules for expanding feature lists. These rules are similar to the "systems" of systemic grammar but, in many cases, the expansion process is controlled by the contents of the semantic specification. E.g. a feature list like the one above needs information about mood. Which mood is chosen (declarative, imperative, or interrogative) depends on what is present in the semantic input structure.

After the initial syntactic node has been supplied with all features required for syntactic and semantic purposes, a set of replacement rules gets into operation. The initial node is replaced by a syntactic construction which is a sequence of one or more nodes. E.g. the node with feature list [Category: Clause; Subcategory: Main; Mood: Declarative] may be replaced by, among other things, a Subject-Verb-Object construction or a Clause-Conjunction-Clause construction. Which replacement will actually take place is determined by (a) the content of the semantic structure, and (b) the result of a search through the lexicon.

By way of example, suppose the semantic structure contains a causal relationship between an Event1 ("cause") and Event2 ("result"). The procedure which inspects the semantic structure will notice this relationship and communicate it to the lexical selection routine. Searching the lexicon, this routine may come up with an SVO-construction (the verb cause), a Clause-Conjunction-Clause construction (because or since), or a nominal construction (the noun cause). The replacement rules would reject the noun but accept the other constructions as legitimate replacements of the initial main-clause node. (I assume that lexical search stops as soon as a legitimate replacer has been retrieved.)

A lexical entry is not always a single word but may be a multiword construction. If an entry contains several words, they make up a syntactic dependency tree. E.g. the lexical entry for the active main verb cause consists of three nodes for subject, verb and object, respectively. In this construction the verb is marked as governor; subject and object are dependents. Each node is labeled by a syntactic feature list and is accompanied by either a phonological or a semantic pointer. Going on with the example, the node labeled with features [Category: Noun; Case: Subject] is accompanied by a notation pointing to Eventl in the semantic structure. Thus the generator knows that Eventl must be lexicalized as subject noun phrase of the sentence. The main verb node in the construction contains a pointer to the phonological form cause. (Parenthetically, since the standard form of a lexical entry is a multiword unit, speech formulae of all sorts (idioms, phraseology, standard expressions) are handled easily.)

A third type of rules may be called lexical transformations. They operate on a syntactic construction retrieved from the lexicon and convert it into the shape necessitated by syntactic features on the node it is going to replace. For example, in Dutch and German interrogative sentences, the tensed verb precedes the subject noun phrase. The generator handles this in terms of a lexical transformation. Before an SVO or SV construction is inserted in a syntactic skeleton, the order of the nodes is changed to VSO or VS (if no auxiliaries are present). (I haven't decided yet on the "basic" order of V, S. and 0; that is, their order in lexical entries for verbs. As for Dutch, one finds arguments for VSO, SVO as well as SOV. I don't know of any good evidence, linguistic or psychological, picking out one of these as the best one.) Anyhow, the point I want to stress is that transformations, if necessary, are applied to constructions right after they have been selected from the lexicon, whereafter the resulting transform is given its definitive place in the syntactic skeleton.

Once the initial node has thus been replaced, the whole process starts over again on those nodes of the replacing construction which are accompanied by a semantic pointer, going from left to right.

Presently, I only have the rudiments of what I hope will develop into a grammar which handles a reasonable subset of Dutch. This grammar will be one that, at all stages of syntactic skeleton formation, enables an efficient interplay between the ensemble of syntactic possibilities which are open at a given moment, and semantic constraints serving to choose exactly one member from that ensemble (collaborating with other constraints, perhaps, such as activation thresholds of lexical items).

In an earlier paper (Kempen, 1977 b) I have indicated how the sentence generator deals with fragmentary semantic structures (cf. Section 2) and is able to maintain syntactic coherence across sentence fragments produced one after another. Here I only repeat that it involves making a "syntactic summary" of a syntactic fragment before it is uttered, and adding the summary to the feature list which controls the generation process for the remainder of the sentence. The type of grammar I have outlined seems very well able to carry this load.

## 5. Linguistic versus psychological plausibility.

A final comment on the linguistic status of my grammatical system. Since I set it up as a psychological theory, I cannot claim much linguistic plausibility for it. However, some recent developments in syntactic theory have struck me as similar in spirit to the grammar I have proposed here. One such development is the increasingly important role that some transformational grammarians (e.g. Bresnan, 1976) assign to the lexicon, thus cutting down the influence of the transformational component of the grammar. Another development is R. Hudson's (1976) book on what he calls daughter-dependency grammar. His classification rules and daughter-dependency rules look very much like my expansion rules and replacement rules, respectively, although he has worked out many more linguistic details than I have. Hudson's sister-dependency rules have their counterpart in the dependency structure of my lexical entries. I wouldn't be surprised if, one day not far from now, linguistic and psychological researchers will advocate very similar grammatical systems even though basing their opinions on vastly different empirical evidence.

**References.**

Bresnan, J., *Towards a realistic model of transformational grammar*. Paper presented at the Convocation on Communications, M.I.T., March 1976.

Butterworth, B. & Beattie, G., Gesture and silence as indicators of planning in speech. In: Campbell, R. & Smith, P. (eds.) *Proceedings of the Psychology of Language Conference*, Stirling, June, 1976. NY: Plenum Press, 1977b.

Garrett, M., The analysis of sentence production. In: Bower, G. (ed.) *The psychology of learning and motivation*, vol.9. New York: Academic Press, 1975.

Garrett, M., Syntactic processes in sentence production. In: Wales, R. & Walker, E. (eds.), *New approaches to language mechanisms*. Amsterdam: North-Holland, 1976.

Hudson, R., *Arguments for a non-transformational grammar*. Chicago: University of Chicago Press, 1976.

Kempen, G., On conceptualizing and formulating in sentence production. In: Rosenberg, S. (ed.) *Sentence production*. Hillsdale: Erlbaum, 1977a.

Kempen, G., Sentence construction by a psychologically plausible formulator. In: Campbell, R. & Smith, P. (eds.) *Proceedings of the Psychology of Language Conference*, Stirling, June, 1976. NY: Plenum Press, 1977b.

Kempen, G., Man's sentence generator: aspects of its control structure. In: de Hey, M. et al. (eds.), *International workshop on the cognitive viewpoint*, University of Ghent, March, 1977c.

Lindsley, J., Producing simple utterances: how far ahead do we plan? *Cognitive Psychology*, 1975, 7, 1-19.

Lindsley, J., Producing simple utterances: details of the planning process. *Journal of Psycholinguistic Research*, 1976, 5, 331 — 353.

Schank, R., Conceptual dependency: a theory of natural language understanding. Cognitive Psychology, 1972, *3*, 552 — 631.

Schank, R. (Ed.), *Conceptual information processing*. Amsterdam: North-Holland, 1975.