Man's sentence generator: aspects of its control structure

Gerard Kempen

Psychological Laboratory
University of Nijmegen, The Netherlands

December 1976

Advance papers for the International Colloquium
"The cognitive viewpoint", Ghent, March 1977.

The human sentence generator faces the complex task of expressing thoughts in the form of natural language utterances. Stated differently, it has to convert conceptual into syntactic structures. (Phonetically realizing – pronouncing – syntactic structures is left out of consideration here.) In this paper I will focus on the control strucure of the sentence production system. By control structure I mean: the way the complex task is divided into subtasks; how these subtasks are didtributed over subsystems and data structures; how and when these are put to work; and how their actions are coordinated.

On the basis of recent analyses of speech errors and hesitations that have been observed during spontaneous speaking, I will argue for two levels of control. One level, which mainly uses man's "central channel of attention", works out the content and the global syntactic shape of an utterance. The second level doesn't need central attention capacity and consists of a set of "demons" that take care of the detailed phonological shape of the words in the utterance. They operate in parallel, more or less independent of each other. This proposal is a continuation, with a number of modifications, of the sentence production model developed in Kempen (1977b).

1. Some observations on the sentence generator

The argument in this paper uses four results of the empirical study of spontaneous speech production. I take them to be facts that are reasonably well established.

(a) Speakers often start pronouncing parts of an utterance before they have completely worked out the conceptual content of the utterance. This statement reflects no hard data but an intuition that is shared by many authors on language production. Let the terms conceptualizing and formulating denote the processes of, respectively, generating a conceptual structure and converting it into a syntactic structure. Apparently the formulating system accepts fragmentary conceptual structures as input and is able to verbalize successive conceptual fragments in such a way that the corresponding parts of the utterance fit together syntactically. For instance, suppose the first fragment is the conceptual structure underlying two Belgians, and only the second fragment indicates that a literary prize was awarded to them. Then the formulator must be able to select constructions such as ...were awarded a prize... or ...received a prize..., preferably without first having to reject forms which would put two Belgians in non-initial position (e.g. ...prize was awarded to... or

... awarded prize to...).

(b) Hesitation pauses during spontaneous speech are largely attributable to three factors: conceptualization processes, lexicalization processes, and surface clause boundaries. According to a recent analysis by Butterworth (1976), these factors account for roughly 1/3, 1/4 and 1/5 of the total pausing time. Conceptual pauzes are used for organizing the content of the next few sentences. Lexical pauses occur when the speaker is searching his mental lexicon for the best word or idiom to express a piece of conceptual structure he has in mind. Pauses at clause junctures probably serve a variety of functions. They may benefit the parsing process in the listener (Butterworth, 1976) or result from breathing activity; so they do not necessarily reflect time needed for syntactically planning the next clause. (Butterworth couldn't replicate Boomer's (1965) often quoted finding of high pausing frequency after the first word of a clause.)

(c) Lexical insertion occurs at a relatively late stage of the formulating process. This statement is indirectly supported by the phenomenon of lexical pauses. If, in the middle of a sentence or clause, a speaker is forced to stop in order to look up an adequate word or phrase, then he must already have decided on a large portion of the syntactic structure and the intonation contour (cf. Butterworth, 1976). Exactly the same conclusion follows from the study of speech errors. Word substitutions and word interchanges usually are "syntactically correct": nouns are replaced by nouns, verbs by verbs, etc. Also, they leave the intonation contour intact. The famous example is how bád things are → how thíngs bad are, where sentence stress is maintained on the second word. I conclude that lexical insertion takes place only after large portions of syntactic structure and intonation contour of a sentence have been worked out.

(d) Converting a (fragmentary) conceptual structure into a (fragmentary) syntactic structure is a two-stage process. On the basis of a detailed analysis of a large corpus of speech errors, Garrett (1975) distinguishes between a "functional" level of processing and a "positional" level. The first level prepares a syntactic structure specifying grammatical (functional) relations among lexical formatives. The latter are only "content words". Syntactic words/morphemes ("function words") are added in during the second processing stage which, generally speaking, is responsible for serially ordering the content words and integrating them with syntactic function words. Below I will take over the gist of Garrett's proposal but assign somewhat different tasks to the processing levels. (The modifications

are compatible with the speech error data.)

2. The first level of control: the monitor

The third empirical result discussed above (lexical insertion after global syntactic structure has been decided) entails a serious problem. How is it possible to build a syntactic structure without knowing the lexical material that is available in the mental lexicon? Suppose a to-be-expressed conceptual structure specifies a causal relationship between two events. The lexicon is entered and several possibilities are found for expressing causal relationships: (a) conjunctions such as because or since, (b) verbs like cause or be caused by, (c) nouns like in be the cause of. The syntactic environments required by these lexical items are widely different. If a conjunction is selected, then the two causally related events have to be expressed in the form of a main and a subordinate clause; these events must be verbalized as noun phrases in case a verb is chosen, however.

Such examples demonstrate that syntactic structure formation and lexicalization are highly interdependent. How can we reconcile this observation with the third empirical result which seems to indicate that syntactic structure formation is relatively independent of lexicalization? The solution that I will propose consists of two elements, both having to do with the format of lexical entries in the mental lexicon. First, an entry does not contain a specification of its phonological form but only a pointer to a memory location where the phonological specification can be found. For example, the entry for the causal conjunction doesn't contain the word forms because or since but pointers to a place where both forms are listed as possible phonological realizations of that entry. Thus, the lexicalization process is divided into two steps: one for looking up a lexical entry which adequately expresses a piece of conceptual information, and another step which replaces a pointer with a phonological form. From now on I will term these steps lexical selection and lexical insertion, respectively.

Second, an entry contains a full specification of the syntactic environment it contracts. E.g. one entry for because/since states that the conjunction is followed by a subordinate clause and a main clause, in that order. (There may be a second because/since entry with a different order.) Moreover, the entry specifies which conceptual information will have to fill this environment. Continuing the example of two causally related events, the subordinate clause must

be a verbalization of the temporally prior event (the "cause"), whereas the content of the main clause must be the "effect". Inspection of the to-be-expressed conceptual structure will reveal which parts correspond to cause and effect. These parts are then formulated in the syntactic shape necessitated by the conjunction entry, etc. Goldman (1975) and Kempen (1977b) work out concrete procedures for building syntactic structures in unisono with the lexical material available in the lexicon.

Two important points can be made now. First, lexical selection is not only guided by conceptual information ("I need a lexical entry expressing conceptual structure X") but also by syntactic requirements ("That lexical entry will have to fit in syntactic context Y"). Second, while building a syntactic structure there is no need yet for the phonological forms of the selected lexical entries. So the second lexicalization step (insertion) can be postponed till a later stage of the sentence production process (see next Section).

The procedure which controls the complicated activity of inspecting the content of a fragmentary or complete conceptual structure, of looking up adequate lexical entries, in such a way that the syntactic environments required by the various entries are properly filled, I will call the monitor. So the monitor keeps an eye on both the conceptualization process and the syntactic structure formation process. It receives the output of the conceptualizer, hands it over to the syntactic structure builder, and keeps track of any progress made by the latter. Also , the monitor may ask the conceptualizer to revise a conceptual structure so as to tune it better to current syntactic possibilities and desirabilities (Kempen, 1977a).

The conceptualization system and the monitor heavily load the central channel of attention. This is attested by the phenomena of conceptual and lexical pauses during spontaneous speech. Many authors on sentence production tend to view syntactic structure formation and lexicalization as independent processes. In combination with the low prominence of hesitations that are unambiguously attributable to syntactic factors, such a view easily leads to the inference that syntactic planning must be a highly automatic and skilled activity which doesn't occupy much attention capacity. However, as soon as one realizes that lexicalization and syntactic structure formation are heavily interdependent, it follows that lexical pauses must serve a double function: syntactic planning and lexicalization.

Which aspects of syntactic planning do belong to the realm of skilled behavior, is investigated in the next Section.

3. The second level of control: morphonological demons

A syntactic structure prepared under monitor control is a bracketed sequence of lexical items. The bracketing enables identification of noun phrases, verb phrases and the like. Each lexical item is accompanied by morphonologically relevant information, e.g. word class, case,number, tense, etc. For details see Kempen (1977b). Remember that the lexical items are not yet phonologically specified. Also, a syntactic structure delivered by the monitor need not correspond to a complete utterance but may be part of one: the first empirical result of Section 1.

In order to introduce the concept of a demon – a computer science term – I'll use the blackboard metaphor. The structures built by the monitor appear on a blackboard which is continually watched by a set of demons. A demon is a piece of program (a procedure, one that is small in comparison with the main program) which is not called "by name" but "by pattern". A demon is activated as soon as an instance of the pattern of symbols it is conditioned to appears on the blackboard. Then its program is started immediately. Usually that program causes some change in the symbol structure on the blackboard. As a consequence of such a change, some other demon may become active, etc. The main program (here: the monitor) never looses time deciding which demon to call ("by name"). Moreover, since each demon has its own little processing unit, their activity doesn't have to occupy the central processor (here: the central channel of attention).

I propose that the job of providing syntactic structures with phonetic substance is taken care of by a set of morphonological demons. This job consists of three parts: (a) lexical insertion (replacing pointers with phonological forms), (b) applying rules of inflection, and (c) sending the resulting structures off to the articulatory mechanisms. Since the monitor doesn't exert any control over these processes they seem to run "automatically" as highly skilled behavior. Finally,the demons don't care about each other's success or failure, and there is no supervising demon which overlooks and checks all the acyivity that is going on: "distributed" control, contrasting with "centralized" control in the monitor.

This proposal is in line with Garrett's (1975) distinction between two levels of processing but adds to it the flavor of different control structures. The empirical advantages of my proposal will become clear in the next Section. But before going on I wish to remark that Garrett's levels have a somewhat different stuffing from mine.

For example, a speech error like the blending of shout and yell into shell is seen by him as a slip of the first (functional) level of processing. I would blame a morphonological demon for it: while trying to replace a lexical pointer, two phonological word forms found at the same location were mixed up.

4. Empirical arguments for two levels of control

There are no hard data that unambiguously support my distinction between sentence construction processes which are controlled by a central processing mechanism, and those taking place at the level of skilled behavior. But several empirical observations become better understandable.

(a) No evidence exists, formal or informal, to the effect that highly inflected languages are spoken slower or with more hesitations than uninflected languages. So, the job of providing a sentential sequence of words with their proper inflection morphemes doesn't appear to take up space in the central channel of attention. If this observation is worth anything, it seems to refer inflection to the realm of skilled behavior.

(b) Several years ago, Goldman-Eisler concluded from her analyses of hesitational pausing during spontaneous speech that "syntactical operations ... seem to be organized at the level of skills" (1968, p.76). Recently, Butterworth (1976) confirmed her findings with new data. In Section 2 I argued that what these authors call lexical pauses also have to do with syntactic planning: syntactic structure formation and lexical selection are mutually dependent. They both occupy the central channel of attention and may cause hesitation pauses. Goldman-Eisler's result now means that whatever "syntactical operations" remain necessary after syntactic structure formation are functioning at the level of skills. Some of these I have listed in Section 3.

(c) An important category of speech errors in Garrett's (1975) sample are sound exchanges (e.g. hell of a mess → mell of a hess). His sample consists of naturally made errors. Baars, Motley & MacKay (1975) induced sound exchanges ("spoonerisms" they call them) by using an experimental technique. Their subjects didn't construct sentences but reproduced from short-term memory (STM) word material they had just finished reading.

Garrett (1976) points out an interesting inconsistency between his corpus of "natural" spoonerisms and the "artificial" ones obtained by

Baars et al. The natural spoonerisms didn't show any tendency toward "wordhood".That is, the probability for an error form (e.g. mell and hess in the above example) to be an English word was not higher than chance. However, the artificial spoonerisms did show this tendency. In short, lexical status of an error form has found to be important in artificial but not in natural spoonerisms.

Data on the lexical status of spoonerisms are relevant because they may demonstrate the existence of an editing mechanism which checks the output from the sentence generator before it is pronounced. Error forms that stay within the class of English words might be harder to detect for this editing mechanism than error forms that have no status as English words. The former would have a higher chance of getting through.

Now suppose the editing job were part of the regular duties of the monitor. This assumption is made by every sentence production theory which, implicitly or explicitly, assigns all planning/editing/monitoring to one central mechanism. Such theories will predict an effect of lexical status (the data of Baars et al.) for all situations of spontaneous speech.

The present theory, however, is able to solve Garrett's paradox. The monitor's task is controlling the conceptualization and syntactic structure formation processes; it never engages in checking lexical status of output words. At the morphonological level no such checking takes place either: it would be against the idea od distributed control by more or less independent demons. So the editing job must be carried out by some mechanism outside the sentence generator. A plausible candidate is the language comprehension system (parser). But, since the parser also uses the central channel of attention, it cannot operate if this channel is already heavily loaded by the monitor. The latter was the case when Garret's natural spoonerisms were produced: the speakers were busily constructing new sentences. Apparently no processing capacity was left for editing. The subjects of Baars et al. were engaged in a relatively simple STM task and, since they didn't make up sentences on their own, the monitor wasn't active. So, in this situation, the central channel of attention may very well have had some spare capacity for calling in the parser and doing the editing job.

5. Conclusion and summary

I have proposed a model of man's sentence generator whose functioning is controlled at two levels. First, the monitor controls the processes of conceptualizing (generating a to-be-expressed conceptual structure) and syntactic structure formation (building the syntactic skeleton of the utterance). The monitor uses the central channel of attention. The syntactic skeleton is provided with phonetic substance by a set of morphonological demons that make up the second level of control. The demons operate in parallel, more or less independent of the monitor and of each other. Thus the morphonological level is characterized by distributed control, whereas the monitor has centralized control.

This outline of the sentence production system is shown to be in good agreement with important sentence production data (speech errors and hesitation pauses).

Literature

Baars, B.J., Motley, M.T. & MacKay, D.G. Output editing for lexical status in artificially elicited slips of the tongue. Journ. Verb. Learn. Verb. Beh., 1975, 14, 382-391.

Boomer, D.S. Hesitation and grammatical encoding. Lang. Speech, 1965, 8, 148-158.

Butterworth, B. Semantic planning, lexical choice and syntactic organisation in spontaneous speech. Manuscript, U. of Cambridge, 1976.

Garrett, M.F. The analysis of sentence production. In: Bower, G. (ed.) The psychology of learning and motivation, Vol. 9. NY: Acad. Pr., 1975.

Garrett, M.F. Syntactic processes in sentence production. In: Walker, E.C.T. & Wales, R.J. (eds.) New approaches to language mechanisms, in press, 1977.

Goldman, N. Conceptual generation. In: Schank, R.C. (ed.) Conceptual information processing. Amsterdam: North-Holland, 1975.

Goldman-Eisler, F. Psycholinguistics. London: Academic Press, 1968.

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.