Cognition, 14 (1983) 185–209

The lexicalization process in sentence production and naming: Indirect election of words*

GERARD KEMPEN

PIETER HUIJBERS

University of Nijmegen

Abstract

A series of experiments is reported in which subjects describe simple visual scenes by means of both sentential and non-sentential responses. The data support the following statements about the lexicalization (word finding) process. (1) Words used by speakers in overt naming or sentence production responses are selected by a sequence of two lexical retrieval processes, the first yielding abstract pre-phonological items (L1-items), the second one adding their phonological shapes (L2-items). (2) The selection of several L1-items for a multi-word utterance can take place simultaneously. (3) A monitoring process is watching the output of L1-lexicalization to check if it is in keeping with prevailing constraints upon utterance format. (4) Retrieval of the L2-item which corresponds with a given L1-item waits until the L1-item has been checked by the monitor, and all other L1-items needed for the utterance under construction have become available.

A coherent picture of the lexicalization process begins to emerge when these characteristics are brought together with other empirical results in the area of naming and sentence production, e.g., picture naming reaction times (Seymour, 1979), speech errors (Garrett, 1980), and word order preferences (Bock, 1982).

*The work reported in this paper was supported by a grant from the Ministry of Education of The Netherlands and the Netherlands Foundation for Pure Research (ZWO) to the Max-Planck-Institut für Psycholinguistik in Nijmegen, the Catholic University of Nijmegen and the Institute for Perception Research (IPO) in Eindhoven. We thank the members of the Special Project on Descriptive Language for a number of very stimulating discussions on the subject. We are also indebted to William Levelt and two anonymous reviewers for their comments on earlier versions of the paper.

Requests for reprints should be sent to Gerard Kempen, Department of Psychology, University of Nijmegen, Montessorilaan 3, 6525 HE Nijmegen, The Netherlands.

0010–0277/83/$8.00

An important aspect of cognitive processing involved in production of linguistic utterances concerns lexicalization: retrieving from the mental lexicon word material for the sentence under construction. The process of lexical selection must be sensitive to the intention in the speaker's mind. Each word or idiom covers part of the meaning content conceptualized by the speaker while thinking or perceiving. In the psycholinguistic literature it has recently been suggested that for each content word the lexicon is consulted *twice* (Butterworth, 1980; Garrett, 1980; Kempen, 1977*a*, *b*; Levelt and Maassen, 1981). The first look-up retrieves a somewhat abstract lexical item supplied with a set of syntactic features. These enable the item to be allocated a grammatically appropriate place in a syntactic skeleton. The second look-up serves to associate with the item the morphological and phonological information necessary for guiding further articulatory processing. (For a similar linguistic proposal within the context of Transformational Grammar see Hudson, 1976.) According to these suggestions, then, lexicalization proceeds like an indirect election. The first step designates a number of lexical items which give rise to the syntactic shape of an utterance suitable for expressing the speaker's intention. These lexical items, in turn, elect the phonological forms from which the sound shape of the utterance can be computed.

Evidence for the 'double look-up' lexicalization hypothesis derives primarily from studies of speech errors in spontaneous speech. For example, the class of word substitution errors naturally divides into two groups: errors characterized by *meaning* similarity between target word and intrusion (e.g., *tomorrow* substituting for the intended *yesterday*), and so-called malapropisms (Fay and Cutler, 1977), where the sound shapes of target and intrusion are very similar (e.g., *result* instead of *resort*). The overlap between these groups is very small since sound similarity is accompanied by little or no meaning similarity, and vice-versa. Garrett assigns the former substitution type to the first lexicalization step where meaning content guides the selection process. Malapropisms are supposed to arise during the second step in which lexical items become replaced by the corresponding phonological forms. Failure of the second lexicalization step may result in a tip-of-the-tongue state (Brown and McNeill, 1966). Various mechanisms could be proposed to account for the sound similarity of malapropisms to target words, but we will not go into that issue here.[1]

[1]The 'double-lookup' lexicalization hypothesis asserts that, for each word of a sentence, two lexical entries have been retrieved and consulted: the first one specifying syntactic information, the second one (mor)phonological information. It should be distinguished from the competing assumption that a lexical entry consists of two segments containing syntactic and (mor)phonological information respectively which are inspected at two different points in time. The latter hypothesis, which is of the single-lookup type, cannot account for the existence of malapropisms because it provides no explanation for the *sound* similarity between target word and intruding word.

The double lexical look-up theory bears some similarity to current theoretical views on the process of object naming. Seymour (1979, p. 287) presents a decomposition of this process into four stages:

(1) pictorial encoding of the presented object;
(2) the retrieval of a semantic code in which attributes of the object are specified;
(3) the retrieval of a phonological representation of an object name (selected at a level of abstraction appropriate to the task); and
(4) the expression of the name as audible speech.

While Stage 3 can safely be identified with the second step of lexicalization, there is an essential difference between Stage 2 and the first lexicalization step. The semantic code as envisaged by Seymour does not exhibit the single-element character of a lexical item. It is, rather, a multi-element code comprising a list of attributes representing cognitive (predominantly perceptual) features of the object. The model put forward by Clark and Clark (1977, p. 469) is somewhat simpler but essentially the same. A perceptual stage of *object identification* is followed by a linguistic stage of *word selection*. Words are chosen on the basis of features present in the identified object. (So-called semantic procedures associated with words of the lexicon are able to establish the presence or absence of such features.)

In this paper we wish to explore the nature of the processing stage prior to retrieval of phonological word form. Is it an essentially non-lexical (but cognitive, perceptual, semantic) stage as claimed in the object naming literature? Or is it lexical in character, as the above-mentioned students of sentence production have recently proposed? We have approached these questions by means of an experimental paradigm which is a mixture of sentence production and object naming. Subjects had to describe pictures in terms of (a) a single word, (b) sequences of words, and (c) sentences. In all three conditions, latencies were measured between onset of picture presentation and onset of vocal description response. Subjects attempted to keep the latency intervals as short as possible.

The single-word condition (a) is similar to a traditional object naming task. The pictures we used contained several aspects which could be named independently; for example, a person together with an action performed by that person (a woman greeting, a girl kicking, etc.). We were therefore able to run various naming conditions for each pictorial aspect separately. In the word sequence condition (b), subjects had to combine the names for several such aspects in a multi-word response (e.g., *woman-greet, girl-kick*). The order of to-be-named aspects was designated in advance, and pausing between words was not permitted. In the sentence production condition (c),

subjects described the same pictorial aspects in terms of a syntactically structured utterance (*woman greets, girl kicks*), again avoiding pauses between words.

We begin with a brief look at the model proposed by Lindsley (1975, 1976). He was the first investigator to compare latency patterns for object naming and sentence production. Lindsley makes the assumption, typical of the object naming literature, that the stage which precedes phonological retrieval is perceptual-semantic in nature. We follow this with a report on a series of Dutch sentence production experiments demonstrating that parts of Lindsley's model are inadequate. Our data, in conjunction with those yielded by Lindsley's experiments (which we could replicate for Dutch), force us to assume a separate retrieval step prior to retrieval of phonological word form. Further experimentation employing the word sequence condition (b) shows that the double look-up model for sentence production generalizes to naming tasks.

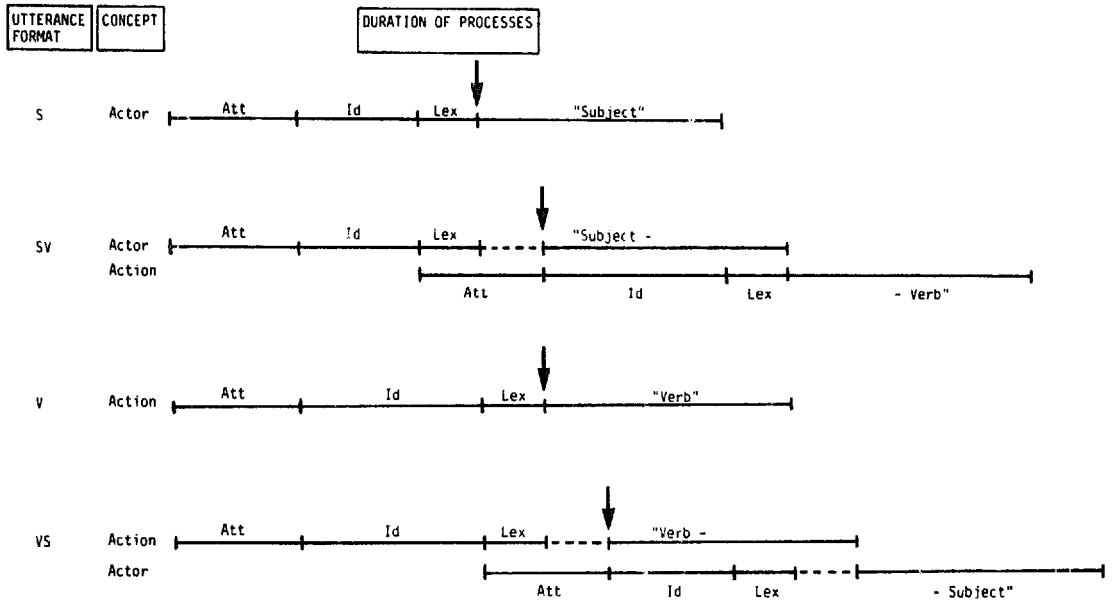
A subsidiary problem which immediately crops up when studying latency data for multi-word utterances concerns the temporal arrangement of lexicalization processes for the individual words. Are the various words looked up more or less simultaneously (parallel alignment) or one after the other (serial alignment), e.g., in their order of mention in the utterance? We will deal with this question in conjunction with our main issue: how to decompose the lexicalization process itself.

Lindsley's model for naming and sentence production

Lindsley's (1975, 1976) main concern was to predict the latency for a Subject-Verb (SV) sentence given the naming latencies for the individual subject noun (S) and main verb (V). The participants in his experiments were shown, in each trial, a picture of either a man, a woman or a boy who were either greeting, kicking or touching one of the other figures. A full description would have comprised a Subject-Verb-Object sentence such as *the boy is kicking the woman*. However, most latency comparisons were restricted to conditions S (naming the actor alone), V (naming the action alone) and SV (describing actor and action according to the format *the S is V-ing*). The V latencies were some 100 milliseconds longer than the S latencies–a difference without immediate implications since it is at least partially due to the way actors and actions had been drawn in the pictures.

The critical finding was a substantial delay in the sentential SV latencies as compared to S naming latencies. It is the length of this delay upon which Lindsley's model is focused. Various hypotheses that immediately spring to mind were rejected in control experiments. For example, the SV delay does

Figure 1. *Temporal alignment of processes according to Lindsley's (1976) original model.* Att = *attending to the picture;* Id = *identifying actor/action;* Lex = *lexical retrieval. Dotted lines represent empty intervals. Vertical arrows indicate onset of spoken response.*



not depend on utterance length *per se* (two words *versus* one word). Nor is it the case that SV 'waits for the verb'. This strategy would imply that SV latencies can never be shorter than V latencies. However, although SV and V were often in the same range, SV did become substantially shorter than V when S was made very easy to name. After a number of intricate experimental manipulations (which we cannot review here), Lindsley arrived at the model outlined in Figure 1. He divides the naming process for actor and action into four stages: (1) an attention stage which serves to extract perceptual features from the picture; (2) identification—establishing the identity of actor/action on the basis of perceptual features; (3) retrieval of a name, and (4) overt utterance of the name. In the SV condition it is the actor which is processed first. More precisely, work on the action is begun only after the actor has been identified. Name retrieval and attending to the action then proceed in parallel. The spoken response (the actor's name) is initiated only after the attentional stage for the action has come to an end. Thus SV responses incur a delay (as compared to S responses) since the attentional stage is supposed to consume more time than lexicalization of the actor. The resulting model is depicted in the S, V and SV parts of Figure 1 (the durations of the segments are arbitrary). In short, Lindsley attributes the S-SV delay to attentional processing of the action.

An important assumption is that processing for the actor and processing for the action proceed in series rather than in parallel. The first of our experiments will show that this temporal alignment fails to predict the latencies for Verb-Subject (VS) sentences, which in Dutch are synonymous with SV sentences and sound equally familiar.

Experiment 1

The aim of this study was threefold. Firstly, we wished to replicate Lindsley's findings for Dutch. Secondly, we investigated the extent to which the latency pattern would remain intact when changing from homogeneous to heterogeneous trial blocks. In Lindsley's experiments, a block contained trials of no more than one type (e.g., only S, or only SV, etc.). We mixed up the four conditions by signalling in advance which utterance type was desired for the next picture (S, V, SV or VS). Thirdly and most importantly, we wanted to explore the extensibility of the model to different sentence forms, in particular to Verb-Subject constructions. In main clauses of Dutch, two orders of subject and finite verb are possible (exactly like in German). The finite verb is always in second position. If no other constituent opens the clause, the subject phrase takes first position (SV). Any other constituent (an adverbial, for example) occupying the initial position causes an obligatory inversion and moves the subject over the finite verb (VS).

Lindsley's model does not make very firm predictions for VS sentences. The bottommost scheme of Figure 1 is a natural extrapolation, though. It is identical to the SV scheme except that actor and action processing have been interchanged. The critical component is 'attending to the actor'. If, as assumed in Figure 1, the duration of this component is identical to 'attending to the action', then it follows that any latency difference between S and V will lead to a comparable difference between SV and VS: V-S $\approx$ VS-SV. A less plausible extrapolation of Lindsley's model is based on the assumption that in the VS condition actor and action are processed in the same order as in SV, i.e., actor followed by action. Some additional (presumably syntactic) mechanism has to be invoked enabling the speaker to withhold the subject noun until the verb has been pronounced. This predicts very long VS latencies because overt responses can be initiated only after both words have been looked up in succession: V-S < VS-SV.

Method

The experiment makes use of the picture description paradigm introduced above. In each trial, the subjects were presented with a line drawing depicting an actor who is performing an action directed towards another figure. The task was to describe as quickly as possible actor, action, or both. One second in advance of a new picture, the description format (S, V, SV or VS) was signalled to the subject by means of a 'frame'. The four frames consisted of the following printed texts:

S: 'zelfst. nmw.' (abbreviation of Dutch: zelfstandig naamwoord; English: substantive)
V: 'werkwoord' (verb)
SV: 'omdat hier ...' (because here ...)
VS: 'want hier ...' (for here ...)

The S and V frames simply denote the grammatical category of the desired response: a noun naming the actor, a verb naming the action. The SV and VS frames are sentence fragments which have to be completed. *Omdat* is a conjunction which introduces a subordinate clause. In Dutch subordinate clauses, the only possible order of Subject and Verb is S-V. *Want* is a coordinating conjunction leading up to a main clause. The adverbial *hier* causes Subject-Verb inversion, so the ensuing word order is V-S. In a separate series of experiments with exactly the same method we have shown that syntactic distinctions (main clause *versus* subordinate clause) do not cause any differences in time needed to initiate utterances of the type employed in the present study (Van Wijk and Kempen, 1982). It follows that no RT effects can be attributed to differing amounts of syntactic computing in the SV and VS conditions.

The actor and action on a picture were chosen from the sets {man , woman, boy, girl} and {kicking, greeting, slapping, teasing}. (Teasing was depicted as 'thumbing one's nose'.) Each of the 16 possible actor-action pairs was combined with one fixed object figure which had to be different from the actor (e.g., man-teasing-boy, girl-kicking-man). The actor always appeared on the left-hand side of the picture, the object figure on the right. All 64 possible combinations of 16 pictures and 4 frames were used as stimuli. The morphological form of the responses was as follows:

S: singular noun (*man*, *vrouw*, *jongen*, *meisje*; man, woman, boy, girl)
V: infinitive verbs (*schoppen*, *plagen*, *slaan*, *groeten*; kick, tease, slap, greet)

SV: singular noun (without article) followed by conjugated verb
(e.g., *man plaagt, meisje schopt*;
man teases, girl kicks)
VS: inversion of SV (e.g., ... *plaagt man*; ... man teases).

All verbs can be used intransitively, so the absence of a direct object does not render the sentences ungrammatical.

Each subject participated in one session consisting of two parts. In one part, s/he went through *homogeneous* blocks of 32 trials of the same response format. In the other session there were four *randomized* blocks with different formats (frames) intermingled. In each block, homogeneous or randomized, the 16 pictures appeared twice. Their order was random except that the same actor or action was not permitted to occur more than three times in a row. In randomized blocks the latter restriction held for frames as well. Subjects were able to take a pause between blocks. Half the subjects did the homogeneous session first, the other half started with the randomized blocks. There were 16 subjects, all students of the Catholic University of Nijmegen, who participated individually.

The stimuli were displayed on a TV screen located in front of the subject at a distance of 75 centimeters. The onset of the vocal response was registered via a microphone and voice-key. The pictures had been recorded on an Ampex video disk and could be accessed very quickly. The frame texts were displayed as subtitles by means of a hardware character generator. The experiment was run under the control of a PDP 11/34 computer.

A trial was defined in terms of the following program steps.

1. Clear the screen; select a picture and a frame; wait 3 seconds.
2. Display the frame; wait 1 second.
3. Add the picture; start latency timer.
4. Wait for the voice-key to trigger; stop the timer; compute and store latency; wait 2 seconds; go to 1 for next trial.

In step 4 latencies were cut off at the maximum of 3 seconds, so that one whole trial lasted 9 seconds at most.

The subject received standard reaction time instructions, that is, to respond as fast as possible while avoiding any errors. Moreover, he was told that pauses between the words of SV and VS sentences counted as errors. In order to familiarize the subject with pictures, words, apparatus and task, he was given 20 to 32 practice trials in such a way that each picture had occurred at least once. During the sessions the Experimenter checked the accuracy of the responses given by the subject. A button was pressed to discard latencies in case of incorrect timing, i.e., when a noticeable time difference occurred

Table 1. *Average latencies (milliseconds) for homogeneous and heterogeneous blocks of trials in Experiment I*

Block type	Utterance type			
	S	SV	V	VS
Homogeneous	790	849	838	843
Randomized	754	869	856	918

between onset of the subject's response and the moment of voice-key triggering (marked by a visual signal operated by the voice-key). The push-button also served to discard latencies in the case of word choice or word order mistakes. Extremely long and extremely short latencies (above 2000 milliseconds or under 200) were discarded automatically.

Results

The experiment yielded a total of 4096 responses, of which 7.8 percent had to be discarded as erroneous. Errors were evenly distributed over conditions (most of them resulting from incorrect voice-key triggering; there were very few hesitations between words). The remaining latencies underwent an analysis of variance with Block Types (homogeneous *versus* randomized), Frames and Pictures as within-subject variables and Order of Block Types (homogeneous first *versus* randomized first) as a between-subject variable. Both Subjects and Pictures were considered random factors. Significance tests of fixed effects and of their interactions were carried out by means of quasi-F-ratios (Winer 1971, p. 375).

The critical aspects of the data are discernible in the Frames × Block Types interaction ($F(3,83) = 2.54$; MS = 411872; $p = 0.061$). The Frames × Block Types × Order of Block Types interaction did not reach significance ($F(7,76) = 0.79$; MS = 82143; $p > 0.5$). This means that the just mentioned Frames × Block Types interaction is independent of the order in which the two blocks (homogeneous first or randomized first) were presented to the subjects. Table 1 gives the corresponding average latencies. In the homogeneous blocks, S utterances are faster than any of the other utterance types whose averages are close together. The linear contrast between the S and the V frames is significant, the minimal difference for a significant t ($p = 0.025$, df = 500) being 46 milliseconds. The randomized blocks show a more dif-

ferentiated pattern of latencies. With the same minimal difference of 46 milliseconds as before, only the 13 milliseconds difference between V and SV fails to reach significance.

Discussion

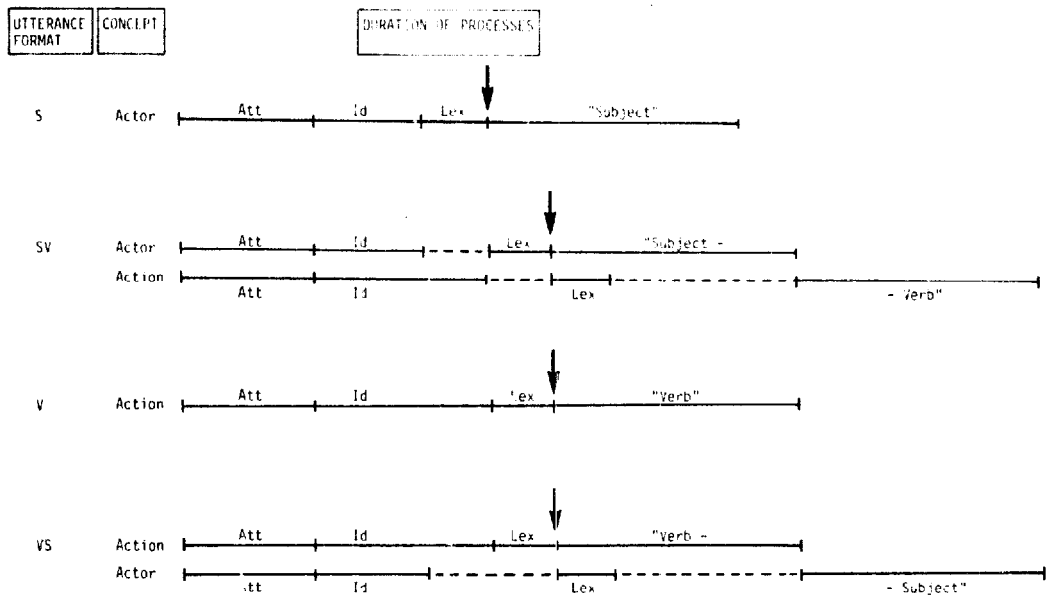
The results enable clear answers to be given to the three questions which motivated this experimental study. First of all, Lindsley's data are replicable in Dutch S, V and SV utterances. Secondly, mixing up the various frames in randomized blocks leaves the latency pattern observable in homogeneous blocks basically intact. The most salient deviation is the significant VS-V delay which amounts to 62 milliseconds. At this point we do not attempt to interpret this phenomenon; in the Discussion of Experiment III below we will put forward an explanation in terms of monitoring. We first need a better understanding of the naming and sentence planning strategy employed by our speakers. For—and this is the third and central conclusion we have to draw—it seems that a reconsideration of Lindsley's model is necessary. Neither the homogeneous nor the randomized blocks yielded any of the predicted outcomes, i.e., a difference between the SV and VS latencies which is equal to or larger than the one between S and V. On the contrary, the VS-SV difference is considerably smaller (randomized blocks) or virtually nil (homogeneous blocks). The two sentential conditions, in fact, appear to center around the most difficult of the two naming conditions: $S < V \approx VS \approx SV$. We have taken this characterization of the latency pattern as the starting point for a revision of Lindsley's model.[2]

An improved model of the lexicalization process in sentence production and naming

Figure 2 presents an alignment of processes which is in better agreement with the essential results of Experiment I. This modified Lindsley model differs from the original in two respects. First, the naming processes for actor and

[2]By choosing appropriate durations for the Att, Id and Lex components it is possible to obtain models which generate the prediction $S < V \approx VS \approx SV$. In particular, one might attribute the greater difficulty of naming actions as compared with naming actors to the attention rather than to the identification stage: Att(actor) < Att(action); Id(actor) = Id(action). However, Lindsley (1976, p. 341) argues at length that set size effects (e.g., naming one actor out of two is easier than naming one out of four) reside in the Id rather than the Att components. We found it preferable to ascribe both phenomena to the same cause (i.e., Id, when following Lindsley's argument).

Figure 2. *Temporal alignment of processes according to the modified Lindsley model.*



action start simultaneously and proceed largely in parallel. Second, lexicalization is postponed until the to-be-expressed content has been fully identified. In the SV condition, for example, the actor's name is retrieved only after both actor and action have been recognized. Lindsley explicitly rejected a model of this sort for two reasons. Part of his experimental work was devoted to exploring the effects of *set size*, that is, of the number of different actors or actions occurring in a block of trials. For example, he observed that if the number of action alternatives increases from two to four (conditions denoted as V2, V3 and V4), the V latencies increased regularly. But this trend was hardly visible in the corresponding SV latencies: the increase from S3V2 via S3V3 to S3V4 was weak and irregular. This finding should be predicted from the Figure 1 model if, as in Lindsley's model, set size effects (Hick's Law) are exclusively allocated to the Identification stage. And SV latencies are not supposed to involve action identification. The modified model of Figure 2 is rejected since it does include action identification in the SV latencies.

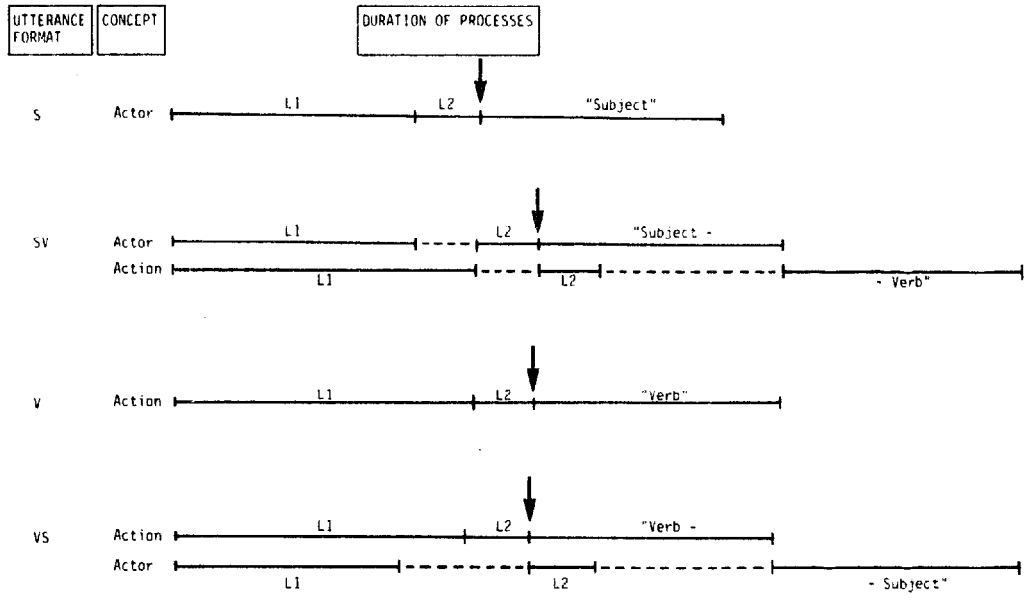
However, it is possible to attribute set size effects to the lexical retrieval stage. This alternative is even desirable in the light of recent developments in the human performance literature. Theios (1975) for example, convincingly argues that *response* factors must be held responsible for effects of set

size (see also the survey by Sanders, 1980). Lindsley (1976, p. 335) rejects the idea of set size effects due to number of alternative names in a trial block, but for unclear reasons. However, it is not only the set size data in the 1976 paper which are interpretable on that theory. Seen in retrospect, the first experiment he reported (in the 1975 paper) can also be accounted for in terms of a set size effect of lexical origin. In this study he had one condition where the actor was held constant over a whole trial block (S1V3), in addition to the normal S, V and SV conditions (denotable as S3, V3 and S3V3). The mean latencies Lindsley obtained were as follows: 567 milliseconds for S1V3, 715 for S3V3, 710 for V3, and 597 for S3. The 148 milliseconds difference between S1V3 and S3V3 must be attributed to the retrieval of one rather than three names. This time estimate fits in with a 63 milliseconds difference between S2V3 and S4V3 measured in a later study by Lindsley (1976, p. 343), if reaction time is logarithmically related to number of alternatives (Hick's Law).

The second reason for Lindsley's rejection of a model like that of Figure 2 has to do with the effect of introducing uncommon names for the actors. In his last study, some subjects were instructed to use the labels *princess, duchess, squire* and *count* instead of *girl, woman, boy* and *man*. (The pictures were left unchanged.) We assume that this manipulation affects the *typicality* of the picture-noun pairs. The pictures were certainly less typical of squires and princesses than of boys and girls, etc. The empirical effects of typicality are often brought together with those of *level of abstraction*, e.g., labeling a picture of a dog as *dog* (basic level term), *spaniel* (subordinate term) or *animal* (superordinate term). The locus of typicality and level of abstraction effects, according to the standard interpretation in the literature, is the name retrieval stage. See, for example, the quotation from Seymour (1979) in the introductory section above. Lindsley adopts this course, too. The uncommon names are assumed to prolong actor lexicalization to such an extent that its duration exceeds the 'attention to the action' stage (in terms of Figure 1: Lex(Actor) > (Att(Action)). It follows that the SV latency will not be any longer than that for S alone. This is indeed what Lindsley found. The S and SV latencies for uncommon names were both about 830 milliseconds. This result contradicts the modified model of Figure 2 if it is true that the typicality effect resides in the lexicalization process Lex(Actor). The S-SV difference for common names will then be expected to show up unaltered in the (slower) reaction times for uncommon names.

Rather than giving up the basic idea of parallel processing for actor and action embodied in Figure 2, we have worked out an alternative which leads to the double look-up hypothesis. Let us assume that the stage preceding the retrieval of phonological form delivers an 'abstract' (i.e., non-phonological) lexical item, and that factors such as typicality and level of abstraction some-

Figure 3. *Temporal alignment of processes according to the new model.* L1 = *retrieval of pre-phonological lexical items;* L2 = *retrieval of phonologically specified lexical items.*



how influence the duration of this retrieval process. Stated differently, the processing that takes place during this pre-phonological stage, accomplishes a mapping from pictorial images onto the set of abstract lexical items corresponding to permitted names. This process halts at the point where exactly one member of the set has been elected as output item. This point will be reached faster for pictorial material that is typical of the categories denoted by the permitted response names, and for names that are basic level terms.

The resulting model is depicted in Figure 3. The symbols L1 and L2 stand for the two lexicalization steps which return 'L1-items' and 'L2-items', respectively, as their results. Kempen and Hoenkamp (forthcoming) denote the lexicalization outcomes by the terms 'lemma' (L1) and 'lexeme' (L2). A separate stage of pictorial encoding (Seymour, 1979; cf. above quote) has been omitted here, since even should it exist, it plays no role in accounting for any data discussed in this paper. The effects of typicality and set size are assumed to reside in L1 and L2 respectively. This is the reverse of the allocations accompanying the modified Lindsley model in Figure 2 (set size attributed to Id, typicality to Lex).

Returning to the outcomes of Lindsley's study with uncommon actor names, we can now provide a straightforward interpretation. The atypical

combinations of actor picture and actor name made the L1 stage for the actors longer than for the actions. Hence the SV latency depended on the difficulty of S rather than of V. Since V no longer delayed the overt response, the S and SV latencies would become equally long, as was indeed the case. The two experiments to follow provide empirical support for specific details of the new model.

Experiment II

Comparison of the models in Figures 2 and 3 might give the impression that there aren't any testable differences between them: the models are identical except for the names of the components. This impression would be misleading, however. The new model states that, in the SV condition, retrieval of phonological word forms ('L2-items') is initiated after the selection of prephonological lexical items ('L1-items') for both actor and action. The alternative models deriving from Lindsley do not go as far as this, claiming only that some form of *prelexical* processing of the action is causing the delay from S to SV (attention or identification). The present experiment addresses the question of whether SV latencies do indeed reflect *lexical* processing for the verb as implied by the new model. The alternative models assume only prelexical processing of the action.

The average latency scores obtained from subjects improve considerably over successive trial blocks. Such practice effects can be reduced or even completely abolished by modifying relevant aspects of the task. In this study we have investigated the extent to which practice effects were disrupted by the introduction of a new set of verbs for the same actions. The new verbs described the actions equally well; in fact, they were synonymous with the original verbs. For example, after a long series of trials in which the verb *meppen* (slap) had been used to refer to one of the actions, the subject was instructed that henceforth only the verb *slaan* (beat) would be permitted. Other subjects started out with *slaan* and, halfway through the experiment, were transferred to *meppen*. Our goal was to compare the magnitudes of the effect that this manipulation would introduce in the SV and VS conditions. If SV responses do not involve any action processing of a lexical nature, then shifting towards a new set of action names will not be expected to disrupt performance, except for some deterioration due to temporary factors such as distraction or loss of concentration. If, however, lexical processing (at the level of L1-items) is implicated in SV responses, then the latter will become more permanently affected by the transition to new verbs. All three models predict a slowdown in VS reaction times. Both original and modified Lindsley

models would point to the name retrieval stage ('Lex' in Figures 1 and 2) as the locus of the delay. The new model, however, leads us to anticipate the VS delay to be even larger than the SV delay because, after the shift, VS reaction times involve not only a new set of L1-items but also new L2-items.

In sum, the new model predicts a substantial disruption of performance in both SV and VS reaction times following the transition to new verbs. The VS delay will be longer than the SV delay. The two models deriving from Lindsley predict that subjects will return to their pre-transition SV performance level after a temporary slowdown due to distraction.

Method

The general experimental set-up used in this study was the same as that in Experiment I. We used only sentential responses. SV responses were elicited by the frames *omdat* (because) and *want* (for). VS responses followed upon the adverbials *hier* (here) and *soms* (sometimes). Each subject participated in one session which consisted of two parts. Parts I and II were identical except for the verbs designated by the experimenter as legal descriptions of the actions. The two sets of verbs were (A) *plagen, slaan, schoppen, groeten* (tease, slap, kick, greet) as in Experiment I, and (B) *pesten, meppen, trappen* and *zwaaien*. The members of B are synonymous with the corresponding members of A, with one possible exception: *zwaaien* (wave hands) denotes a particular form of greeting. Half the subjects received the sets in the order A-B (for Parts I and II respectively), the other half in the order B-A. Part I consisted of 2 blocks of 64 trials. Each block contained all possible combinations of 16 pictures with 4 frames. The sequence of stimuli in a block was determined by the same regime as in Experiment I. Part II also comprised 128 trials, but in the reverse order of that in Part I.

Before the beginning of Part II, the subjects received 25 training trials in order to familiarize themselves with the new verbs. These trials were identical to normal experimental trials, except for the frames. The frame introducing a new trial was here replaced by the (new) name of the action shown in the ensuing picture. Subjects were asked to pronounce that name immediately upon presentation of the picture.

Twenty-two new subjects were drawn from the same population as in Experiment I. In order to obtain data on the distribution of hesitational pauses, we tape-recorded all responses produced by the subject during experimental trials. In contrast to Experiments I and III, the Experimenter did not exert any pressure on the subject to avoid hesitations between the words of a sentential response.

Table 2. *Average latencies (milliseconds) for SV and VS responses in Experiment II*

Utterance type	Pre-shift		Post-shift	
	Block 1	Block 2	Block 1	Block 2
SV	929	837	961	887
VS	957	864	1013	917

Results

Table 2 presents the average latencies for SV and VS utterances in two pre-shift and two post-shift trial blocks. Means do not include errors (5.5 percent on the average); responses containing hesitations between words were not counted as erroneous. An analysis of variance was carried out with Parts (I *versus* II), Practice (first *versus* second block of each part), Frames and Pictures as within-subject variables, and Order of Verb Sets (A-B *versus* B-A) as a between-subject variable. Both Subjects and Frames were considered random effects. All tests of significance here reported were done by means of quasi-F-ratios.

The main effect of Order and the interactions of Order with other factors were all insignificant. This implies that verb sets A and B were of comparable difficulty. All other main effects were significant. In particular: VS was harder than SV (as in Experiment I; $t = 2.48$; df = 1408; $p < 0.01$); there was a substantial practice effect ($F(1\ 32) = 42.15$; MS = 11081061; $p < 0.0001$); and post-shift reaction times were slower than pre-shift ones ($F(1,32) = 7.36$; MS = 3221368; $p < 0.001$). However, none of the interactions between Frames, Practice and Part reached significance. The non-significant Practice × Part × Frame interaction shows that the SV and VS conditions behaved roughly the same as a consequence of practice and verb shift. Both SV and VS reaction times suffered under the introduction of new action names. The SV latencies did not recover any sooner than the VS latencies. Neither the SV nor the VS responses ever recovered completey—even during the final post-shift block the participants performed less well than they did in the block immediately preceding the shift. This rules out the distraction hypothesis by which post-shift SV performance should quickly return to the level reached just before the shift, thereby implying a substantial Practice × Part × Frame interaction. Since the distraction hypothesis was the only possibility by means

of which the original and modified Lindsley models were able to explain an effect of the shift upon SV latencies, we conclude that both these models are falsified. On the other hand, the data are in keeping with the new model in so far as the effect upon SV latencies is concerned. Its prediction that VS would incur a greater delay than SV is confirmed: the VS delay does exceed the SV delay, to a statistically significant extent (25 milliseconds, when comparing the increase from the second pre-shift to the first post-shift block; $t = 2.25$; df = 704; $p < 0.01$).

In addition to the reaction times we obtained complete tape recordings of the sessions of 16 subjects (8 in A-B, 8 in B-A order). The frequency of pauses centered around 2 percent for VS and 4 percent for SV reponses. The only salient deviation concerns SV responses in the first post-shift block. There, the percentage increased sharply to 10 (whereas the corresponding VS percentage was hardly affected by the shift).

Discussion

The data of the present experiment show that SV latencies implicate *lexical* processing of the action (at the level of L1-items). This result is in agreement with the new model and contradicts the models deriving from Lindsley. One might attempt to defend the latter models by calling in the help of subsidiary factors such as distraction or fatigue to account for the slow post-shift SV reaction times. However, such hypotheses do not present a very strong case. The subjects of Experiment I underwent equally long sessions also consisting of two parts of 128 trials each. There, too, the parts were separated by special instructions. But this interruption hardly affected the practice curve which, after the first 64 trials, was essentially flat. The four consecutive groups of 64 trials averaged 872, 828, 832 and 819 milliseconds.

Another argument one might raise against a lexical interpretation of the obtained RT pattern assumes that shifting to a new verb necessitates a new perceptual parsing of the pictorial input. Thus, the post-shift delay is attributed to an identification rather than a lexicalization problem. However, this reasoning cannot rescue the original Lindsley model of Figure 1 which does not include action identification as a component of SV latency. More seriously, this argument overlooks the synonymy of the verb pairs which is very close indeed. The two members of a pair apply essentially the same perceptual parse to the pictorial input. It follows that the revised Lindsley model of Figure 2 is inadequate as well.

Experiment III

The double look-up hypothesis originates from the study of sentence production rather than naming. The empirical evidence we have offered in support of the existence of pre-phonological lexical items also derives from sentential responses. One might therefore accept our arguments that L1-items do indeed play a part in sentence production, while at the same time denying L1-items any role in non-sentential naming tasks. In this manner, one could maintain that naming models like Seymour's (1979) are not at all contradicted by the results of our sentence production studies.

In the present experiment, the subjects perform a *double naming task*. In a single integrated response, both actor and action of a picture are named in a predetermined order, without pauses between words, and by using citation forms of nouns and verbs (singular and infinitive, respectively). These responses are no longer sentences, but non-syntactic word sequences, e.g., *meisje-groeten, slaan-man* (*girl-greet, slap-man*), etc. Our prediction is that the latency pattern will be identical to the pattern we observed for sentential responses expressing the same pictorial contents. If, on the other hand, the subjects handle the double naming task as a sequence of two traditional naming tasks, then a totally different prediction follows, namely, latencies for two-word responses identical to the corresponding one-word latencies.

Method

The experimental set-up was identical to the randomized blocks condition of Experiment I. The only difference concerned the content of the frames and the morphological form of the responses. N responses (naming the actor alone, N for Noun) were elicited by a little star at the bottom left-hand corner of the TV screen, below the place were the actor was displayed. The frame for eliciting action descriptions (V for Verb) was a star in the right-hand bottom corner (the actions in the pictures were directed towards the right). The desired word order in double naming responses was signalled by doubling the star referring to that word which had to be spoken last. E.g., VN was signalled by one star at the right-hand side and two stars below the actor at the left. In sum:

*		N
	*	V
*	**	NV
**	*	VN

As already mentioned, inflection of the words was not permitted. We had 16 subjects taken from the same population as before. None had served as a subject in a similar experiment.

Results

Of the maximum of 2048 latencies, 8.4 percent had to be discarded as erroneous. The four utterance types contracted roughly the same number of errors. Table 3 shows that the pattern for non-sentential N, NV, V and VN utterances is more or less the same as that observed for sentential utterances in Experiment I. Double-word responses were considerably slower than those for single-words and the VN-V difference amounting to roughly half the NV-V difference (cf. the bottom line of Table 1).

The latencies were subjected to an analysis of variance with Subjects (16) and Pictures (16) as random factors, and Frames (4) as a fixed factor. The levels of the latter factor were significantly different with quasi-$F(3,80) = 24.07$, $p < 0.001$, and MS = 6536232. All 6 pairwise contrasts between means were significant (the minimal difference for significant $t(p = 0.025, \text{df} = 500)$ being 47 milliseconds).

Discussion

The new model depicted in Figure 3 is apparently not specific for sentence production and generalizes to non-sentential double naming responses. We conclude that abstract L1-items play a role not only in sentence planning but in naming as well.[3]

One aspect of the data which deserves serious consideration is the fairly large latency difference between both double responses (NV and VN) on the one hand, and the slowest of the two single-word responses (V) on the other. The new model predicts no difference at all. This was indeed the pattern

[3]One might argue that the participants were in fact treating the double naming task as if it were a sentence production task having a special constraint on the form of the verbs to be produced. While this alternative cannot be ruled out offhand, its plausibility is severely reduced by the results we obtained in a very similar double-naming experiment (unpublished data). In this the participants had to name an actor and an action taken from *two different* pictures displayed in opposite halves of a TV screen. The resulting actor-action pairs, moreover, would often form weird or meaningless combinations (e.g., *boy-undulate, man-glitter*). The adoption of an implicit sentence production strategy in an overt double-naming task seems farfetched under these circumstances. Nevertheless, the double-word responses turned out to lag behind the single-word ones by some 100 milliseconds.

Table 3. *Average latencies (milliseconds) in Experiment III*

Utterance type			
N	NV	V	VN
646	845	797	897

originally obtained in the homogeneous blocks of Experiment I, but the randomized blocks condition of that study had already yielded an unexplained VS-V difference of 62 milliseconds. Now, in the double naming task, the corresponding difference has increased to 100 milliseconds, and also the significant NV-V difference of 48 milliseconds is at variance with the model.

The possibility that attentional or perceptual factors are responsible for these systematic deviations from the predicted pattern is rendered unlikely by the following observation. The separation between single-word and two-word responses is greater in the present experiment than in the randomized blocks of Experiment I. This is the case, notwithstanding the fact that the pictures and the picture sequences in the two studies were exact replicas of each other. So a response planning factor must have been at work, causing a somewhat larger separation between single and double responses in the non-sentential than in the sentential task.

We propose an explanation in terms of a *monitoring* process occurring between the first and the second lexicalization steps. This process watches the output of L1-lexicalization and checks whether the retrieved L1-items fit into the utterance format imposed by the frame. Each L1-item requires a certain amount of monitoring time, which will depend on the probability of an erroneous lexicalization result (L1-items incompatible with the required format) in the experimental condition that is in force. In other words, monitoring time for an L1-item in an utterance will not be constant but vary with the subjective probability of errors attracted by that item in its utterance context. The probability of format errors will be higher in randomized blocks where the frames are changing from trial to trial, than in homogeneous blocks. In the latter condition, the need for elaborate format checking may even completely disappear after the first few trials. Moreover, the probability of errors against word order will have been higher in the present experiment which employed an entirely ad hoc word order rule, than in Experiment I where the subjects could rely on a highly automatized linguistic rule. We assume that the monitoring process does not wait until all L1-items intended

for the utterance under construction have been retrieved. On the contrary, as soon as an L1-item has arrived, it is immediately treated by the monitoring process. Consider the SV condition, for example. If the actor is much easier to lexicalize than the action, then the monitoring of the subject noun may be well underway or even completed at the moment the verb arrives. In terms of the SV part of Figure 3, the monitoring process will take place, at least partly, during the empty intervals (marked by dotted lines) between the L1 and L2 segments. It follows that monitoring the subject noun does not necessarily show up as an increased SV latency. Only if the monitoring process is complicated, will the SV responses be delayed. The VS responses in our experiments are more likely to give away the presence of a monitoring mechanism: monitoring the verb cannot fill an otherwise empty interval and is therefore bound to show up in the latencies. This not only explains the fact that there is a VS-V delay in our randomized blocks, but also that this delay is greater than the SV-V delay.

The account given so far presupposes that in our heterogeneous trial blocks the time required for monitoring an L1-item of a *two-word* utterance was longer than the monitoring time for the same L1-item in a *single-word* utterance. This difference was virtually annihilated in the homogeneous trial blocks of Experiment I, where VS and SV latencies approached the V latency. However, under certain circumstances monitoring time for L1-items in single-word utterances may exceed that for the same L1-items in two-word utterances. We observed this phenomenon in a related study which was similar to the randomized blocks part of Experiment I (unpublished data). The participants described pictures showing very familiar and invariable actor-action pairs such as birds singing, dogs barking, jewels glittering, etc. Single-word descriptions often took longer than double-word ones, presumably because the latter were more easily available and tended to intrude into trials which asked for a single-word response. By carefully checking the format of a planned utterance, the monitor is able to prevent such intrusions from occurring. It does not seem unreasonable that under such circumstances single-word utterances consume more monitoring time than double-word ones.

General discussion

The experiments reported in this paper permit us to pin down some of the operating characteristics of the lexicalization system speakers deploy in naming and sentence production. These characteristics can be summed up in the following four statements. (1) Words belonging to an overt naming or sen-

tence production response come about as the resultants of two lexical selection processes connected in series, the first one yielding abstract pre-phonological items (L1-items), the second one adding their phonological shapes (L2-items). (2) The selection of several L1-items for a multi-word utterance, sentential or otherwise, can take place simultaneously. (3) A monitoring process is watching the output of L1-lexicalization to check if it is in keeping with prevailing constraints upon utterance format. The time taken for monitoring, which may fill otherwise empty intervals after the delivery of an L1-item, is hard to predict: it depends on the probability of erroneous outputs from L1-lexicalization, the seriousness of the consequences of overt errors, etc. (4) Retrieval of that L2-item which corresponds with a given L1-item waits until the L1-item has been checked by the Monitor, and all other L1-items needed for the utterance under construction have become available. (By 'utterance under construction' we here refer to a short sentence or a fragment of a longer sentence. We assume that longer sentences typically come about as the result of incremental or piecemeal sentence production (Kempen and Hoenkamp, forthcoming), i.e., as a sequence of fragmentary sentences.)

To what extent is this set of operating characteristics compatible with what we already know about sentence production and naming mechanisms? In the introductory Section we have seen that in Seymour's (1979) model of object naming the processing stage devoted to the elaboration of a perceptual-semantic code is immediately followed by the retrieval of a phonologically specified lexical item. This seems to imply that Seymour's theoretical decomposition of the object naming process stands in need of improvement. One possibility would be to simply intercalate the new pre-phonological lexical retrieval stage between the perceptual-semantic and phonological name retrieval stages. We prefer a more elegant solution which assumes parallelism of Seymour's perceptual-semantic coding and our L1-lexicalization. Suppose there exist processing units which are able to respond to incoming patterns of perceptual and semantic features. Many lexical processing units are simultaneously active, each of them trying to establish whether the set of criteria for which it is responsible, has been fulfilled. As soon as one unit has reached a positive decision it makes available a lexical item which covers (names) the current combination of incoming features. In systems like these (e.g., Morton's 1970 Logogen Model), the lexical lookup process need not wait until the full set of perceptual-semantic codes of a to-be-named object has been worked out in detail. Instead, the lexical processing units are able to watch and respond to the evolving perceptual-semantic code while it is still being elaborated. The new assumption we are forced to make is that there are lexical processing units corresponding to our pre-phonological L1-items. This

step can save Seymour's model of object naming if one is prepared to redefine the perceptual-semantic stage as a combination of perceptual-semantic coding and retrieval of pre-phonological lexical items.

Garrett's (1975, 1980) extensive analysis of a large corpus of speech errors has led to a two-stage model for the syntactic aspects of sentence production. The stages are called the 'functional' and the 'positional' levels of processing, respectively. The first, functional stage assembles a syntactic tree subsuming all content words of the utterance under construction. This includes the first retrieval step for these words. The various branches of the tree can be computed simultaneously. During the second, positional stage a new lexical retrieval step takes place which, among other things, takes care of the phonological shape of both content and function words. The various branches of the syntactic tree are now processed from left to right, i.e., sequentially rather than simultaneously. Needless to say that these proposals parallel what we assumed when drawing up the model of Figure 3.

The distinction between L1- and L2-lexicalization might provide a way out of a recent empirical paradox. Bock (1982) cites several experiments demonstrating that the accessibility of lexical items is one of the determinants of word order in sentence production. Highly accessible words tend to occupy early positions in the sentence. For example, in one of the experiments subjects had to memorize the sentence 'A rancher sold the cowboy the horses'. When prompted with a question containing the phrase *a stallion*, this sentence was typically recalled as 'The rancher sold the horses to the cowboy' The word *stallion* presumably activated the concept underlying *horse*, thus causing direct and indirect object to exchange position. Levelt and Maassen (1981), however, found no effect of lexical difficulty upon constituent order. The lexical items they employed in their experiments were names of geometrical shapes (*circle, square, star,* etc.). The shapes were shown on a screen in various reaction time tasks. One task was non-verbal and involved shape identification: the subjects watched each presented shape and pressed a yes-button when recognizing a target shape. The no-button was pressed in response to any other shape. In a second task the subject had to pronounce the name of each figure presented (naming latency). Lexical difficulty was defined as the difference between mean naming latency and mean identification latency. The third task involved the description of moving figures in the form of short sentences, e.g., 'The square and the circle are going up'. Levelt and Maassen found no effect of lexical difficulty upon order of mention. There was no tendency for easy lexical items to occupy earlier positions in the description than more difficult items.

The contradiction between the observations by Bock and by Levelt and Maassen can be resolved on the assumption that identification latency in the

description experiments corresponds to L1-lexicalization time, and lexical difficulty to L2-lexicalization time. Since L2-lexicalization takes place *after* syntactic tree formation it cannot exert any influence on sentence form (at least in normal circumstances). Bock's operationalization of lexical accessibility includes L1-lexicalization as well, so that effects upon syntactic tree formation, including the ordering of constituents, are possible.

To conclude, we have argued for dividing the lexicalization process into an L1-stage which accesses a dictionary of abstract, pre-phonological (but syntactically specified) lexical items, and an L2-stage retrieving concrete phonological shapes for abstract items. In order to establish contact with experimental data such as utterance initiation latencies we introduced several additional assumptions, in particular the four operating characteristics listed at the beginning of this Section and the assignment of typicality and set size effects to stages L1 and L2 respectively. We feel that this set of ideas, which are in need of further experimental testing, provides a useful and attractive framework for designing more detailed process models of word finding in human speakers.

References

Bock, J.K. (1982) Toward a cognitive psychology of syntax: information processing contributions to sentence formulation. *Psychol. Rev., 89*, 1–47.

Brown, R., and McNeill, D. (1966) The 'tip of the tongue' phenomenon. *J. verb. Learn. verb. Behav., 5*, 325–337.

Butterworth, B. (1980) Some constraints on models of language production. In B. Butterworth (Ed.), *Language Production (Vol. 1 Speech and Talk)*. New York, Academic Press.

Clark, H.H., and Clark, E.V. (1977) *Psychology and Language*. New York, Harcourt Brace Jovanovich.

Fay, D., and Cutler, A. (1977) Malapropisms and the structure of the mental lexicon. *Ling. Inq., 8*, 505–520.

Garrett, M.F. (1975) The analysis of sentence production. In G. Bower (Ed.), *The Psychology of Learning and Motivation, Vol. 9*. New York, Academic Press.

Garrett, M.F. (1980) Levels of processing in sentence production. In B. Butterworth (Ed.), *Language Production (Vol. 1 Speech and Talk)*. New York, Academic Press.

Hudson, R.A. (1976) Lexical insertion in a transformational grammar. *Found. Lang., 14*, 89–107.

Kempen, G. (1977) Building a psychologically plausible sentence generator. In P. Seuren (Ed.), *Grammarij 9, Symposium on Semantic Theory*. Nijmegen, Katholieke Universiteit.

Kempen, G. (1977) Man's sentence generator: aspects of its control structure. *Com. Cog., 1*, 157–164.

Kempen, G. (1978) Sentence construction by a psychologically plausible formulator. In R.N. Campbell and P.T. Smith (Eds.), *Recent Advances in the Psychology of Language: Formal and Experimental Approaches*. New York, Plenum Press.

Kempen, G., and Hoenkamp, E. (forthcoming) An incremental procedural grammar for sentence formulation. *Cog. Sci.*

Levelt, W.J.M., and Maassen B. (1981) Lexical search and order of mention in sentence production. In W. Klein and W.J.M. Levelt (Eds.), *Crossing the Boundaries in Linguistics: Studies presented to Manfred Bierwisch*. Dordrecht, Reidel.

Lindsley, J.R. (1975) Producing simple utterances: how far ahead do we plan? *Cog. Psychol.*, 7, 1–19.

Lindsley, J.R. (1976) Producing simple utterances: details of the planning process. *J. Psycholing. Res.*, 5, 331–354.

Morton, J. (1970) A functional model for memory. In D.A. Norman (Ed.), *Models of Human Memory*. New York, Academic Press.

Sanders, A.F. (1980) Stage analysis of reaction processes. In G. Stelmach and J. Requin (Eds.), *Tutorials in Motor Behaviour*. Amsterdam, North-Holland Publishing Company.

Seymour, P.H.K. (1979) *Human Visual Cognition*. London, Collier MacMillan

Theios, J. (1975) The components of response latency in simple human information processing tasks. In P.M.A. Rabbit and S. Dornic (Eds.), *Attention and Performance V*. New York, Academic Press.

Van Wijk, C., and Kempen, G. (1982) Kost zinsbouw echt tijd? In Handelingen van het 37ste Nederlands Filologencongres.

Winer, B.J. (1971) *Statistical Principles in Experimental Design* (2nd ed.). Tokyo, McGraw-Hill Kogakusha.

Résumé

Au cours d'une série d'expériences les sujets décrivent des scènes visuelles simples soit avec des phrases soit avec des mots. Des données appuient les positions suivantes sur les processus de lexicalisation (recherche de mots); 1) les mots utilisés pour dénomer dans des phrases sont selectionnés suivant deux processus séquentiels, le premier travaille sur les items prephonologiques abstraits (items L1), le second ajoute la forme phonologique (items L2). 2) La sélection des items L1 dans un énoncé de plusieurs mots peut être simultanée. 3) Un dispositif de contrôle (moniteur) vérifie à la sortie de la lexicalisation L1 l'accord avec les contraintes sur le format de l'énoncé. 4) La recherche de l'item L2 correspondant à un item L1 donné ne commence qu'après la vérification de L1 par le moniteur et qu'après que tous les L1 nécessaires à la construction de l'énoncé soient disponibles. Une image cohérente des processus de lexicalisation commence à emerger lorsqu'on réunit ces points et les résultats expérimentaux obtenus avec les travaux sur la dénomination et le production de phrases, e.g., temps de réaction à la dénomination d'images (Seymour, 1979), erreurs (Garrett, 1980) ou préférences sur l'ordre des mots (Bock, 1982).