

Surface Form and Memory in Question Answering

WILLEM J. M. LEVELT AND STEPHANIE KELTER

Max-Planck-Institut für Psycholinguistik, Nijmegen

Speakers tend to repeat materials from previous talk. This tendency is experimentally established and manipulated in various question-answering situations. It is shown that a question's surface form can affect the format of the answer given, even if this form has little semantic or conversational consequence, as in the pair Q: *(At) what time do you close,* A: *"(At) five o'clock."* Answerers tend to match the utterance to the prepositional (nonprepositional) form of the question. This "correspondence effect" may diminish or disappear when, following the question, additional verbal material is presented to the answerer. The experiments show that neither the articulatory buffer nor long-term memory is normally involved in this retention of recent speech. Retaining recent speech in working memory may fulfill a variety of functions for speaker and listener, among them the correct production and interpretation of surface anaphora. Reusing recent materials may, moreover, be more economical than regenerating speech anew from a semantic base, and thus contribute to fluency. But the realization of this strategy requires a production system in which linguistic formulation can take place relatively independent of, and parallel to, conceptual planning.

It is known from studies of conversational interaction that speakers often repeat earlier material. Current speech frequently uses resources from speech produced before by either the same speaker or the interlocutor. Schenkein (1980) presents a variety of examples of such repeats. They can range from copying single words, as in (1), to copying clauses, as in (2), or repeating the structural format of a complex sentence or paragraph, as in (3).

- (1) A. Well, they join, they all say that you should stay here.
B. Well, the difference is they don't feel like (etc.).
- (2) A. But you can go to sleep tonight.
B. How am I going to sleep tonight?
- (3) A. Cor, the noise downstairs, you've got to hear and witness it to realize how bad it is.

Stephanie Kelter is now at the Institut für Psychologie, Technische Universität Berlin. We wish to thank Kitty Neilen, who has assisted in the design and execution of several of the experiments, and in some of the data analyses. We also want to thank Herbert Clark, Gerard Kempen, and Jan Koster for helpful discussions at different stages of the project. Bob Jarvella and Graham Richardson, last but not least, deserve our gratitude for giving our original nonnative and colloquial English the final appearance of being native and technical. Address reprint requests to W. J. M. Levelt, Max-Planck-Institut für Psycholinguistik, Berg en Dalseweg 79, NL-6522 BC Nijmegen, The Netherlands.

- B. You've got to experience exactly the same position as me, mate, to understand how I feel (etc).

It is as if previous talk sets up a more or less abstract frame in the mind of an interlocutor, which is then used in the formulation of the next turn. Observations like (3) are not entirely new in psycholinguistics. That such frames can have a certain preexistence in the process of speaking has already been observed by Karl Bühler (1934). Reporting on experiments where he used his method of systematic introspection for the analysis of sentence production activities, Bühler summarizes these introspections as follows: "And time and again it was described, that this or that wholly or partly empty syntactic schema preceded the actual formulation of an answer and in some way steered the effective speaking" (p. 253).¹ Similar phenomena in the formulation of utterances have been described by Selz (1922) and more recently and extensively by Kempen (1977) and Kempen and Hoenkamp (1981).

Still, the origin of these phenomena—or rather the origins, since it is not self-evident that cases (1) to (3) above have the same genesis—is far from clear. Repeating a word from a previous turn may be a different process from using a similar syntactic frame or copying an earlier argument structure. Even the repetition of a word may be due to a multitude of determinants, for example, the next speaker may want to deal with the same notion and as a consequence come up with the same wording (this may be so in case (2) above). Or, it may rather be a rhetorical reciprocation, as in (1). The repeat may also have a less functional reason, namely, that the word may have become activated by its previous use in the discourse and thus have a higher chance of appearing again.

What is lacking in the psycholinguistic literature is an experimental analysis of such phenomena. The present paper tries to make a preliminary contribution to the experimental study of what we will call the "correspondence effect." It will be limited to cases of word repeats, such as those in (1). We will first try to establish the effect experimentally (Experiment 1) in a question-answering situation. Next, we will study ways in which memory is involved in the genesis of the effect (Experiments 2, 3, 4, 5). Finally, we will concern ourselves with the question of whether a certain degree of "correspondence" is experienced as natural in conversational turns and how this depends on the "lexical support" the preposition receives from the verb (Experiment 6). The General Discussion will relate the experimental and theoretical findings to wider issues concerning

¹ "Und immer wieder wurde dann beschrieben, dass dies oder jenes ganz oder teilweise leere syntaktische Schema der eigentlichen Formulierung einer Antwort vorherging und das faktische Sprechen irgendwie erkennbar steuerte."

the production of surface anaphora and the relative autonomy of the process of linguistic formulation.

EXPERIMENT 1. THE CORRESPONDENCE EFFECT

This experiment was designed to establish a relatively simple case of correspondence in a two-turn interaction. As was mentioned, there may be a multitude of reasons why a correspondence effect arises. A respondent in a question-answering situation may repeat items for reasons of politeness, rhetoric, or simply thematic continuity. Is it possible to establish a case of correspondence where such functional reasons are not very apparent? Will respondents have a tendency to repeat items which have no particular pragmatic function? If this can be shown, one will have established the correspondence effect under quite minimal conditions. Such a "baseline case" can then be used in further experiments.

The case constructed is one in which a question is asked about a picture. The question can be phrased in either of two equivalent forms, one containing a preposition and the other not. The subject can freely answer the question, either matching the prepositional format of the question or not. For example, one picture depicts a situation in which a boy by the name of *Paul* shows his violin to a girl named *Toos*. The question asked (in Dutch) is either (4) or (5):

- (4) Aan wie laat Paul zijn viool zien?
(*To whom lets Paul his violin see?*)
- (5) Wie laat Paul zijn viool zien?
(*Whom lets Paul his violin see?*)

Given the picture, there is no noticeable semantic or pragmatic difference between questions (4) and (5). Both, moreover, can be answered by either (6) or (7):

- (6) Aan Toos.
(*To Toos.*)
- (7) Toos.
(*Toos.*)

or by longer forms including a prepositional phrase as in (6), or a mere NP as in (7), e.g., (*Aan*) *Toos laat hij zijn viool zien* ((*To*) *Toos lets he see his violin*). The issue is whether subjects will show a bias toward answering with a corresponding form, i.e., answer (6) to (4), and (7) to (5).

Method

Materials. The experimental material consisted of 16 question-picture pairs. Twenty-four additional question-picture pairs were used as fillers. All questions concerned two stick figures, *Paul*, a male, and *Toos*, a female, who were depicted in the pictures.

For experimental items each question had two versions: a prepositional version and a nonprepositional one. The information requested by the two versions was the same and unambiguous in the given picture context. Four different prepositions were involved in these questions: *aan* (to), *van* (of), *naar* (for), and *op* (on). (The English translations are somewhat arbitrary, and only valid in the present picture/question context.) We will call these *preposition types*; there were, among the 16 questions, 4 of each preposition type. Examples of the four types, each in its two versions (+ or – prepositional) are given in (8)–(11). Interlinear English translations are added.

- (8a) *Aan wie laat Paul zijn stok zien?*
(To whom lets Paul his cane see?)
- (8b) *Wie laat Paul zijn stok zien?*
(Whom lets Paul his cane see?)
- (9a) *Van wie is deze pet?*
(Of whom is this cap?)
- (9b) *Wiens pet is dit?*
(Whose cap is this?)
- (10a) *Waarnaar zoekt Paul?*
(What for searches Paul?)
- (10b) *Wat zoekt Paul?*
(What searches Paul?)
- (11a) *Op welk instrument speelt Paul?*
(On which instrument plays Paul?)
- (11b) *Welk instrument bespeelt Paul?*
(Which instrument plays Paul?)

Within a preposition type the four questions differed with respect to the person (*Paul* or *Toos*) and object (*stok* (cane), *tas* (bag), *pet* (cap), *hoed* (hat), *viool* (violin), or *piano* (piano)) involved. The questions used in the 24 filler items were of various sorts, but never asked for a location or anything else that would elicit a prepositional form in the answer given.

All questions were tape-recorded by a female Dutch native speaker. The 40 pictures corresponding to the 16 experimental and 24 filler questions were simple line drawings, presented on an Afigraf graphic display.

Subjects. Thirty-six persons recruited by advertising served as subjects. Twelve subjects were students, and the rest had various occupations. Their mean age was 25.7 years. All were native speakers of Dutch, and each was paid Dfl. 7,– (about \$3.50) for participation in the 30-min experiment.

Design and procedure. Each subject was presented all 40 items. For each subject, eight experimental questions (two within each preposition type) were given in the + preposition version, the other eight were given in the version without a preposition. The frequency with which the four items of a given preposition type were presented with or without preposition, was counterbalanced over the 36 subjects. The sequence of the 16 experimental and 24 filler items was randomly determined for each subject, with the restriction that no two experimental items occurred in immediate succession.

Subjects were tested individually in a soundproof and dimly illuminated room and sat at a distance of about 1 m from the graphic display. The experimenter was seated in an adjacent room, and could talk with the subject via an intercom system. The subject was told that he or she would be presented with pictures of simple events involving two stick figures, *Toos* and *Paul*. For each picture a question would be asked, which had to be answered as quickly as possible. The instructions were illustrated with seven practice trials. The experimenter would initiate a trial by pressing a button which caused the picture for that item to appear on

the screen. Two seconds after the picture appeared, the subject heard the corresponding question via earphones. The picture remained on the screen until the subject had answered. All test sessions were tape-recorded.

Results and Discussion

Table 1 gives percentages of answers containing prepositions and not containing prepositions. The mean number of subjects for the four presentation types in Table 1 ranged from 33 to 35 because of occasional missing data (8 of 576 responses) and consequent exclusion of a subject from the relevant comparison in each of these cases.

For all preposition types, there were more prepositional answers when the question contained a preposition than when it did not. Each difference is significant ($p < .05$ for *van* and $p < .001$ for the other preposition types); in total 73% of the answers are in corresponding format (with 50% as chance level). This finding demonstrates the existence of what we have called the "correspondence effect" between question and answer: it arises even if the preposition in the question carries little semantic or pragmatic weight in the given context. Except for preposition type *naar*, where nonprepositional questions never elicit a prepositional answer, in all other cases at least some subjects spontaneously produced noncorresponding forms, testifying to the nonobligatory character of the correspondence effect. It should be further observed that 68% of the answers contain the preposition. It would, however, be wrong to infer a general tendency in subjects to answer in a prepositional format. Only preposition types *aan* and *van* show this bias, *naar* and *op* are quite neutral in this respect.

TABLE 1
Percentage of Answers Containing the Preposition, and of Answers Not
Containing the Preposition

Question		Answer		
Type	Version	With preposition	Without preposition	Total
<i>aan</i>	With preposition	98	2	100
	Without preposition	64	36	100
<i>van</i>	With preposition	98	2	100
	Without preposition	86	14	100
<i>naar</i>	With preposition	81	19	100
	Without preposition	0	100	100
<i>op</i>	With preposition	89	11	100
	Without preposition	36	64	100

TABLE 2
Proportions of Main Verb Usage in Prepositional and Nonprepositional Answers,
Given Prepositional or Nonprepositional Questions

Answer	Question		All questions
	With preposition	Without preposition	
With preposition	.65	.68	.66
Without preposition	.73	.73	.73
All answers	.66	.71	.69

Nevertheless, one might wonder whether the correspondence effect is the result of a more holistic strategy on the part of the answerer, namely, to match the question in degree of elaboration. A question of the longer prepositional format would induce the respondent to give a longer answer than one of the shorter nonprepositional format. Such a holistic matching strategy would not only increase the tendency to use the longer prepositional format in the answer, but also to be less elliptical on the whole. In order to test this we checked the answers listed in Table 1 to determine whether or not they contain a main verb, i.e., whether they are more or less elliptical in that respect. The results of this analysis are presented in Table 2. It is immediately clear from this table that prepositional questions do not induce more main verb usage in answers than do nonprepositional questions. Nor is it the case that a prepositional answer is more likely to contain a main verb than a nonprepositional answer. So far, there is no indication that the correspondence effect results from a respondent's general tendency to match the degree of elaboration of the question. We will, however, check this result further in the subsequent experiments.

EXPERIMENT 2. THE MEMORY EFFECT

The first experiment established the existence of the correspondence effect, that is, the answerer's tendency to match the surface form of the question as far as the prepositional structure is concerned. However, there was nothing in the data so far to suggest the idea that the effect is caused by a general tendency on the part of the respondent, due to politeness or otherwise, to match the degree of elaboration of the question. Even so, matching the prepositional or nonprepositional format of the question may be the result of politeness on the part of the answerer, since by doing so, one may demonstrate that one has listened attentively and indicate to the questioner which part or aspect of the question one is reacting to (see Clark & Schunk, 1980 for a similar hypothesis). In this way the establishment of correspondence would fulfill a conversational function. It can be argued that this would become especially pertinent in

the conversation if the questioner asks more than one question at the same time, or otherwise adds confusing or distracting information to his question. The present experiment was designed to study the possible effects of such distracting information on question–answer correspondence. We wanted to test whether respondents show a greater tendency to match the prepositional form of the question when the questioner produces distracting information. But we should also consider the other possible outcome, namely, that distracting information causes a *decrease* in question–answer correspondence. An obvious explanation for such a result would be that, due to the distracting information, the answerer loses the relevant trace of the surface form from memory, so that matching it would become a matter of chance. An adequate way for him or her to express attentiveness under such circumstances would be to answer as far as possible with the more elaborate prepositional form, whatever the precise wording of the question: in other words, be maximally explicit in the event of distraction.

The experiment compares what happens to the correspondence effect under different levels of interference. More precisely, the experiment compares what is assumed to be slight proactive interference with stronger retroactive interference. This comparison was carried out by having the subject listen not to one but two questions which were presented in immediate succession. Only one of them had to be answered. Which one it was became apparent from the picture that followed the two questions. If the picture concerned the second question, question and picture were adjacent, and only some proactive interference might arise from the earlier question. If, however, the picture concerned the first question, the two would not be adjacent, and retroactive interference from the second question would occur. The object was to compare the correspondence effect in the nonadjacent case to the one observed in the other adjacent case.

The experiment also examined whether answerers adopt a strategy of adding the preposition in case of doubt (the “be explicit” strategy), and whether interference has an effect on the elaboration or ellipticity of the answer.

Method

Materials. The materials were the same 16 experimental question–picture pairs as in Experiment 1 plus 32 new question–picture pairs used as distractors. These distractors, to be used as interference material, concerned simple geometric figures (circle, triangle, square, star). Each picture showed four figures (though not necessarily different ones) in various sizes and positions. The questions used with these pictures concerned size and/or relative position, e.g., “Wat staat er boven de ster?” (“What is above the star?”). While the question did contain prepositions, the answers required did not. This feature was included to offset a potential set to answer with prepositional forms.

An additional variation—which turned out to be of little effect—was the length of distractor questions. There were 16 “short” questions which each had 6 words and mentioned no more than one geometric figure (as in the just-given example). Sixteen “long” questions each had 11 words and usually mentioned two figures each modified by adjectives, e.g., “Wat staat er tussen de grote driehoek en de kleine cirkel?” (“What is between the big triangle and the small circle?”). All distractors answered with a single word, or noun phrase, i.e., the name of a shape (“een driehoek”: “a triangle”). The correct answer was equally often the circle, triangle, square, or star, and these were also mentioned in the questions an equal number of times.

For both the experimental items, and the distractors, the pictures used were black-and-white line drawings, presented on slides using a Kodak Ektagraphic Carousel projector AF-2K. The image of a drawing on the screen was approximately 70×50 cm.

In contrast to Experiment 1, the questions were asked by the experimenter during the experimental session itself, i.e., they were not recorded beforehand.

Subjects. Thirty-two different persons from the same subject pool used in Experiment 1 took part in the study. All subjects were native speakers of Dutch. The 32 subjects were randomly assigned to two groups of equal size, one of which received the “short” distractors, the other received the “long” distractors. In each group there were 10 students and 6 others with various occupations. Mean age was 23.6 and 24.0 years for the two groups, respectively. Subjects were paid Dfl 7,— for their participation, for which about 30 min was needed.

Design and procedure. As mentioned the length of the distractor questions was varied between subjects. There were three within-subject factors: (i) type of question, with four levels corresponding to the four different preposition types involved (*aan*, *van*, *naar*, *op*); (ii) the version of the question presented—with preposition or without preposition; (iii) position of the distraction; the distracting question on experimental trials either preceded or followed the experimental question. Each subject was tested once in each of these $4 \times 2 \times 2$ conditions, which were counterbalanced in the obvious way. Additionally, there were 16 filler trials per subject, in which the question about geometric figures had to be answered. They were randomly interspersed with the experimental trials.

Both experimental and filler trials consisted of the following events: (i) two questions (one experimental and one distraction question) read by the experimenter one after the other and separated by a short pause (≤ 1 sec); (ii) immediately after the end of the second question a picture, corresponding to one of the two questions; (iii) the subject's answer to the corresponding question. In an experimental trial, the picture shown was always of the stick figures *Toos* and *Paul*.

Subjects were tested individually and sat in a dimly illuminated room, at a distance of ca. 1.20 m from the screen on which the slides were displayed. Questions were presented via earphones, spoken by the experimenter, who sat in the same room, but was invisible to the subject. The subject was told he would have to answer questions about pictures to be projected on the screen. He or she was told that on each trial two questions would be presented one after the other, but that only one of them should be answered, namely, the one which corresponded to the picture shown. Two illustrative examples and three practice trials were given. As previously, subjects' answers were tape-recorded.

Results and Discussion

The correspondence effect obtained for the two subject groups (those with short and with long distraction questions) did not differ significantly in any condition (in all cases $p > .10$ by χ^2 and Mann–Whitney U tests). The two groups were therefore combined in all further analyses.

Table 3 gives the percentage of answers with and without preposition to the experimental question. In the low-interference condition, i.e., when these questions were asked second, just before the picture, there was a clear correspondence effect: question and answer corresponded in 62% of the cases ($p < .001$, sign test). This replicates the main trend found in Experiment 1, though the correspondence was 73% there. The difference is most likely due to the proactive interference in the present experiment. A significant correspondence effect (58.5%, $p < .001$) was also found for the high-interference condition, i.e., when the experimental question came first, and the distractor question intervened before the picture.

The main issue in this experiment was whether the correspondence effect would be affected in one way or another by interim memory interference. Though the correspondence effect was highly significant for both conditions, there is also a significant ($p < .05$) difference in correspondence between the conditions. The effect is less pronounced in the high-interference condition when the subject has to answer the first of the two questions. This finding agrees with the theory that distracting information causes the answerer to loose the trace of the relevant aspect of the question. There is no evidence that he tries to increase correspondence in order to express attentiveness. It should be noted, however, that such an effort is doomed to failure once the relevant information for accomplishing it is no longer retraceable. A further check was therefore made as to whether the answers show evidence of attentiveness insofar as they are more explicit in the case of high interference. A glance at Table 3 makes it clear that no effect of this kind is apparent in the data. The number of prepositional answers to prepositional questions *decreases* significantly ($p < .01$) with higher interference, whereas the rate of prepositional answers to nonprepositional questions remains basically the same ($p > .20$).

TABLE 3
Percentage of Prepositional and Nonprepositional Answers to Questions Preceded and Followed by a Distractor Question

Answer	Question	
	With preposition	Without preposition
Low interference (distractor precedes exp. question)		
With preposition	54	30
Without preposition	46	70
Total	100	100
High interference (distractor follows exp. question)		
With preposition	44	27
Without preposition	56	73
Total	100	100

So far, therefore, we have no evidence that the correspondence effect established in these experiments results from a conversational strategy on the part of the answerer which is intended to express attentiveness to the questioner. It is not, of course, claimed that such a strategy is never used, for it should be kept in mind that we explicitly tried to establish correspondence under rather minimal conditions, i.e., for items which have little semantic or pragmatic function. This "baseline" correspondence so far does not seem to serve a conversational function. It is more like an autonomic process of copying on the part of the answerer, a process which can be interfered with by adding distracting information to the question.

In Experiment 1 we checked whether the correspondence effect could be explained by the answerer's more general matching of the degree of elaboration of the question, but no evidence could be found for such an explanation. For the present experiment, Table 4 gives the relevant data. For each combination of question and interference condition it lists the proportion of answers containing a main verb. It is evident from the table that a prepositional question does not induce a fuller answer than a nonprepositional question (the proportions were .36 and .36 for the low-interference condition and .39 and .38 for the high-interference condition). So such a holistic matching strategy again cannot explain the correspondence effect. On the other hand, the table does show a tendency toward "global elaboration": prepositional answers contain a main verb in a higher proportion of cases than nonprepositional answers (see values in "all questions" column). The differences, however, are not significant (sign test) because many subjects never gave prepositional answers, and

TABLE 4
Proportions of Main Verb Usage in Prepositional and Nonprepositional Answers for
Different Question and Interference Conditions

Answer	Question		
	With preposition	Without preposition	All questions
Low interference (distractor precedes exp. question)			
With preposition	.55	.39	.50
Without preposition	.14	.35	.27
All answers	.36	.36	.36
High interference (Distractor follows exp. question)			
With preposition	.61	.68	.64
Without preposition	.24	.27	.25
All answers	.39	.38	.38

therefore their results cannot contribute to a comparison of main verb usage in prepositional and nonprepositional answers. It happens that the same subjects almost never use a main verb in their (nonprepositional) answers. It is as if giving short answers is an answering style of certain subjects, encompassing both the prepositions and the main verb. This global strategy of shortness in answering, however, is independent of the processes underlying the correspondence effect. The latter relates to the prepositional form of the question, while the former does not. It is, in fact, unclear what induces the answerer to give a short answer. It is neither the form of the question, nor the type of interference: in the low-interference condition 36% of the answers contain a main verb, and in the high-interference condition the number is 38%. It should, however, be observed that although the experimental question/picture pairs were identical to those used in Experiment 1, the rate of answers containing a main verb dropped from 69 to 37%, and the rate of prepositional answers dropped from 64 to 39%. It seems, therefore, that the main difference between the experiments, namely, the presence of distracting questions, induced certain subjects to be more elliptical and generally shorter in their answers. This may, at the same time, have been the reason that these subjects did not bother too much about expressing attentiveness in their answers, as was noted above. It was therefore decided to try and replicate under more natural conditions the memory effect, i.e., the effect of distracting information on question/answer correspondence.

EXPERIMENT 3. CORRESPONDENCE AND INTERFERENCE EFFECTS IN TELEPHONE CONVERSATION EXCHANGES

In order to study the correspondence and interference effects under more everyday circumstances, where the answerer would "normally" be motivated to be polite and attentive, we turned to Clark's (1979) telephone technique. Clark obtained shopkeepers' responses to indirect requests by calling them for information. On the telephone, the merchants reacted to the direct and indirect meaning of a request in different degrees, depending on the particular form of the question asked. Since Clark found this method to be quite sensitive to this kind of difference, we hoped the technique would also allow us to establish *in vivo* a correspondence effect such as demonstrated in Experiments 1 and 2, and to study it for different degrees of interference.

In Nijmegen there exists great uncertainty about the closing time of shops on Saturday afternoons; this ranges from 1:30 to 9:00, and it is therefore quite normal to call a shop and to ask how long they are open. In Dutch, there are two perfectly normal phrases for doing so, namely, (12) and (13):

- (12) Hoe laat gaat uw winkel dicht?
(What time does your shop close?)
 (13) Om hoe laat gaat uw winkel dicht?
(At what time does your shop close?)

As is clear, the questions only differ in the presence of the preposition "om." When asked such a question, a merchant has good reason to answer correctly and attentively. The correspondence effect can show up in the answers given, which may or may not include the preposition (e.g., "Om vijf uur" (*at five o'clock*) vs "vijf uur" (*five o'clock*)).

Similarly, one can study the memory effect, by adding additional information at the end of the question. This was done here by appending an explanation and a tag question, as in (14) and (15).

- (14) Hoe laat gaat uw winkel dicht, want ik moet er speciaal voor naar de stad komen, ziet u?
(What time does your shop close, since I have to come into the town especially there for, you see?)
 (15) Om hoe laat gaat uw winkel dicht, want ik moet er speciaal voor naar de stad komen, ziet u?

The explanation that is added is identical across the two versions. A shopkeeper who follows it attentively may have trouble keeping track of the wording of the original question. Thus we can ask whether this factor leads to less preposition correspondence as in the former experiment, or to more correspondence, expressing special attentiveness. Alternatively, attentiveness may lead to the giving of fuller answers in the distraction conditions.

Method. A total of 228 shops in Nijmegen were called on four consecutive Saturdays, during which the four different questions (12)–(15) were asked on a rotating basis. The questions were put by the senior author, and the answers were registered separately by him and a Dutch assistant immediately after each call.² If the two transcriptions disagreed, or the answer did not contain the information asked for, the datum was dropped and replaced in the next call made. By the end of the fourth Saturday, 57 answers to each of the four versions of question had been collected.

Results and Discussion

Percentage of answers made with and without prepositions to the four question forms are shown in Table 5. For the simple questions (questions (12) and (13)) there is a significant correspondence effect ($\chi^2 = 4.25, p < .05$): answers correspond in preposition to the questions in 60.5% of the cases (chance level is 50%). For the questions plus additional information (questions (14) and (15)), there is no such effect (47% correspondence, $\chi^2 = 0.14, p > .70$). The difference in correspondence effect going from the

² It is illegal in the Netherlands to tape-record telephone conversations.

TABLE 5
Percentage of Prepositional and Nonprepositional Answers in Telephone
Conversations to Questions with or without Distracting Additional Clause

Answer	Question	
	With preposition	Without preposition
	Without additional clause	
With preposition	61	40
Without preposition	39	60
Total	100	100
	With additional clause	
With preposition	54	60
Without preposition	46	40
Total	100	100

simple to the more complex questions is significant ($\chi^2 = 3.97, p < .05$). Thus, the main findings from Experiments 1 and 2—correspondence and interference—are here replicated in a quite natural situation, and for a different preposition. The only difference here is that correspondence fully disappeared in the case of distraction, whereas some correspondence was left under interference in the former experiment. The difference may be due to the type of preposition used, as will be discussed under Experiment 6.

If the distracting information interfered with the memory for the prepositional form of the question, the shopkeeper would be unable to express attentiveness by matching the caller's question format even if he might want to do so. Is there evidence that attentiveness is expressed differently in these cases, namely, by giving prepositional answers in case of doubt, i.e., under conditions of distraction? Table 5 does show a slight increase of prepositional answers for complex questions (from 50 to 57%). This difference, however, is too small to reach significance. Moreover, as can be seen, if such an effect exists at all, it holds only for nonprepositional questions. Under this condition, going from simple to complex questions, the rate of prepositional answers increases from 40 to 60% ($\chi^2 = 3.51, p < .10$). For prepositional questions there is instead a decrease (from 61 to 54%), so that there is no basis for assuming that the shopkeepers generally used the preposition for such conversational purposes. Preposition correspondence can apparently arise without obvious semantic or conversational reason. Remembering the prepositional format of the question is in itself enough to induce the formulation of a prepositional answer, and one wonders what sort of mechanism is responsible for this. In the first two experiments we analyzed whether the mechanism might be something like a general matching of the length or shortness of the question, but there

TABLE 6
Proportions of Main Verb Usage in Prepositional and Nonprepositional Answers to
Different Types of Questions in Telephone Conversations

Answer	Question		
	With preposition	Without preposition	All questions
		Without additional clause	
With preposition	.11	.05	.09
Without preposition	.05	.00	.01
All answers	.09	.02	.05
		With additional clause	
With preposition	.22	.42	.32
Without preposition	.00	.18	.08
All answers	.12	.25	.22

was no evidence to support this. Before moving to other explanations, we will also check such matching in the case of the present results.

Table 6 shows the use of a main verb in the answers to the four types of question. The first thing to be observed is that elliptical answers abound, no less than 87% of all responses having this kind of form. Do prepositional questions release more answers containing a main verb? The table shows a slight and nonsignificant opposite tendency in the complex question (with additional clause) condition. This is the third experiment, therefore, in which no evidence could be found to suggest that subjects answer in more elaborate form to prepositional questions. There is, however, a clear replication in the data of the finding of the previous experiment that prepositional answers more often contain a main verb than nonprepositional answers ($\chi^2 = 11.57, p < .001$). This tendency toward "global shortness" on the part of the answerer is apparently not induced by the prepositional form of the question since it is orthogonal to the correspondence effect. What is new in the present data is the significantly higher proportion of main verb usage in answers to complex questions (.22 versus .05; the difference is significant $p < .001$). Though, as we saw, shopkeepers do not express attentiveness by the use of prepositions in their answers, they may be doing so by more frequently using a full sentence in answering complex questions.

EXPERIMENT 4. THE ARTICULATORY BUFFER

To create a corresponding answer the answerer might proceed in several ways. A very simple hypothesis is to assume the answerer stores the preposition he hears in an articulatory buffer, which is then released at an appropriate time when the requested information has been found. Such a

mechanism would be in agreement with Morton's (1970) hypothesis that the normal function of an articulatory loop is to form an output buffer in normal speech production: parts of utterances not yet spoken are temporarily stored while other parts are still being "worked on."

The present experiment was designed to examine this memory store as a possible source of the correspondence effect. During the last decade, effective techniques have been developed to eliminate the effects of an articulatory buffer from performance measures in memory research (Baddeley & Hitch, 1974; Baddeley, 1976). It seems appropriate to apply these techniques in our question-answer situations and ask if, when the articulatory buffer is filled extraneously, the correspondence effect will disappear or diminish. In the following experiment, subjects were asked to rehearse a six-digit number while a question was presented, thus preventing the storage of any part of the question in an articulatory buffer or "rehearsal loop."

Method

Materials. The experimental materials were 40 question-picture pairs and 20 six-digit numbers. The question-picture pairs were those from Experiment 1. Again, 16 pairs were experimental items, and the other 24 pairs served as filled items. The pictures and number series, each consisting of six different digits, were presented as black-and-white slides. The questions were spoken live (by the male experimenter) as the experiment progressed.

Subjects. Thirty-two students in various fields served as subjects. Their mean age was 22.5 years. All subjects were native speakers of Dutch, and were paid Dfl 7,- for their participation in the 40-min experiment.

Design and procedure. Trials were divided into two blocks, one "with interference" and one "without interference." The order of blocks was counterbalanced over subjects: Half first received 20 trials (8 experimental and 12 filler) during which they overtly rehearsed a six-digit string while being presented with the question, and then did the same number of trials without such interferences. The other half of the subjects received the noninterference trials first. There were three within-subject factors: (i) preposition type (*aan, van, naar, op*); (ii) question version (with or without preposition); and (iii) interference condition (already discussed). Each subject was tested once in each of these $4 \times 2 \times 2$ conditions, always on a different (experimental) question-answer pair. Subjects were assigned particular items using a 4×4 Latin square (questions per preposition type \times version of question \times interference condition). The sequence of the 8 experimental and the 12 filler trials within a block was always random with a restriction against immediate adjacency for two experimental items. A trial "with interference" took the following form: (i) a slide with a six-digit string was presented, and the subject began rehearsing the digits in a whispering voice and in order, (ii) after 1.5 sec of rehearsing the slide was removed and the experimenter read the question (the subject continued to rehearse), (iii) immediately after the question, a slide with an asterisk was presented signaling the subject to end rehearsal, (iv) 0.5 sec later, the slide was replaced again, this time by the picture of the two stick figures, corresponding to the question, (v) the subject gave his response. On trials "without interference" a black slide was initially shown instead of the digit string, and accordingly, the subject was not asked to rehearse. The procedure was otherwise the same, but with the asterisk now used as a "warning signal."

Subjects were tested individually. The apparatus used was the same as that in Experiment

2. The subject was informed that he or she would be asked questions about a pair of stick figures, and that the questions should be answered quickly and in reference to the picture following each question. On beginning the "with interference" block of items, the subject was told, that a picture with six digits would precede each question and that they should rehearse the digits over and over until an asterisk would appear on the screen after the question. For all subjects, at the beginning of each block, there were two examples and three practice trials. Subjects were informed that their answers and reaction times would be recorded. If the subject made errors in digit rehearsal on three consecutive trials, he was asked to be more careful in this respect. However, the correctness of rehearsals was not of concern in the data analysis.

Results and Discussion

By Mann-Whitney *U* tests blocks order did not affect frequency of question-answer correspondence ($p < .10$). The data of the two subject groups were pooled for further analysis. Table 7 shows the percentages of answers given with and without preposition in the two interference conditions. In good agreement with our previous findings, Table 7 reveals a strong correspondence effect in the no-interference condition. (There are 75% corresponding answers, $p < .001$, by sign test.) However, correspondence of about equal strength (73%, $p < .001$) now appears with interference as well. The nonsignificant difference ($p < .20$) between the two conditions is in contrast to our earlier experiments. As in Experiment 1, analogous analyses, performed for each of the four preposition types separately, give the same result. Clearly, rehearsing digits here while listening to the question and preparing the answer had no effect.

It is always problematic to draw conclusions from a null result. Still, it is safe to state that the present result makes it quite unlikely that the articulatory buffer hypothesis outlined above is correct. First, the technique used has always been effective in the Baddeley and Hitch experi-

TABLE 7
Percentage of Prepositional and Nonprepositional Answers to Prepositional and Nonprepositional Questions in Cases of Noninterference and of Digit Rehearsal

Answer	Question	
	With preposition	Without preposition
Without rehearsal of digits		
With preposition	91	41
Without preposition	09	59
Total	100	100
With rehearsal of digits		
With preposition	87	41
Without preposition	13	59
Total	100	100

ments, and second, we have already established in Experiment 2 that interfering information does affect the degree of correspondence for just these picture/question pairs. Apparently, the type of interference makes the difference. Moreover, the next experiment will also make it possible to rule out the articulatory buffer as the "site" for the correspondence effect.

EXPERIMENT 5. SHORT-TERM AND LONG-TERM MEMORY

How else might the correspondence effect be caused? One possibility is that a verbatim representation of the preposition is kept active in working memory while semantic search and other processes necessary for answering the question are performed. Alternatively, it may be that in the process of comprehending the question, a representation of the preposition's information is deposited in long-term memory and only a retrieval cue in working memory is preserved. In either case, we would expect verbal material which intervenes between the question and answer to cause the loss of this information from working memory, thereby leading to a diminished correspondence effect, as observed in Experiments 2 and 3. The following experiment was designed to determine whether the information retained about the question in long-term memory is sufficient to explain the correspondence effect.

The issue is prompted by a somewhat unexpected finding in Experiment 2. It was found there that some correspondence effect persisted even if another, irrelevant question intervened between the relevant question and the subject's answer. One may ask, whether, under such circumstances information from the relevant question, needed to produce this effect, could have remained active in working memory. In order to evaluate a more long-term memory type of explanation for this effect, in the present experiment the interfering material used was made even more extensive, so that continued short-term storage of the relevant question would be quite unlikely. This was done in a running memory task: subjects in the interference condition were required to answer not the last but the next to last question they had heard. In this way a previous picture, question, and answer intervened between a given question and its answer. In addition, long-term storage was examined more directly, by asking the subjects to recall a question itself at some later moment. In this way, the correspondence effect in question answering and long-term recall of the same question could be more directly compared.

Method

Materials. Sixty-five question-picture pairs were used, of the same type as that in Experiments 2 and 5. However, the present experiment was simpler in design, with three rather than four preposition types. Consequently, and in view of a strong bias for subjects in the earlier experiments to use *van* no matter what the question format was, we decided to

drop this preposition type. In its place, additional items with *aan*, *naar*, and *op* were constructed by introducing a third stick figure, the child *Eric*, and some additional objects (car and flute), so that the final set contained 32 experimental items instead of 16; there were 12 items for *aan*, 12 for *naar*, and 8 for *op* (see below). In addition, there were 33 filler items: these were the 24 filler items from Experiment 4 and 9 additional items constructed in the same way and involving the new stick figure *Eric*.

Subjects. Forty-eight new persons from our paid subject pool served in the experiment. They were randomly assigned to an "Interference Group" and a "Noninterference Group" ($N = 24$ each). Mean age in the two groups was 24.7 and 26.2 years, respectively. The majority of the subjects were students (21 and 19 of those in the two groups, respectively), the rest having various occupations. All were native speakers of Dutch, and paid Dfl. 7,- for the 45-min experiment.

Design. All subjects were presented with 55 question-picture pairs; 22 of which were experimental items, the remainder filler items. For the Noninterference group the question and picture always occurred in immediate succession as in Experiment 1, and subjects were instructed to answer as quickly as possible. The Interference Group differed from this in that after hearing each question, subjects received a picture related to the previous question. Thus each question had to be stored while the previous one was being answered and the following one presented. The procedure for the two groups is schematically rendered in Table 8.

The 55-item sequence was divided into four blocks. Among the last 4 items in a block there was only one experimental item, either type *aan* or type *naar*. Following each block, the subject being tested was cued to recall the question from this item, on the basis of the question's verb given as a prompt. Block size varied between 11 and 18 items so that the end of a block could not be anticipated. Each subject had to recall 4 questions in all, 2 of which were of type *aan*, and the other 2 of type *naar*. Half of the subjects in each group received a to-be-recalled question with a preposition present, the other half without it present. Across blocks, each subject was cued to one trial question heard, one next-to-trial question, one three back, and one four back from the end ("intervals" 1, 2, 3, and 4 respectively). Across-subjects within-groups assignment of "intervals" to blocks was made using a 4×4 Latin square. Preposition type and recall interval within subjects was partly yoked. Half of

TABLE 8
Sequence of Events in Experiment 5 for the Interference Group
and the Noninterference Group

Noninterference group	Interference group
Q_1 - Picture 1 - A_1	Q_1
Q_2 - Picture 2 - A_2	Q_2 - Picture 1 - A_1
Q_3 - Picture 3 - A_3	Q_3 - Picture 2 - A_2
⋮	⋮
Q_{n-1} - Picture $n-1$ - A_{n-1}	Q_{n-1} - Picture $n-2$ - A_{n-2}
Q_n - Picture n - A_n	Q_n - Picture $n-1$ - A_{n-1}
Probe word for Q_{n-k} - Recall of Q_{n-k}	Probe word for Q_{n-k} - Recall of Q_{n-k}
$(0 \leq k \leq 3)$	$(0 \leq k \leq 3)$

Note. Q_i = Question No. i ; Picture i = Picture corresponding to Q_i ; A_i = Answer corresponding to Q_i .

the subjects within each group received type *aan* at intervals 1 and 3 and type *naar* at intervals 2 and 4, and vice versa for the other half.

Procedure. Subjects were tested individually. The apparatus used was the same as that in Experiments 2 and 5. The subject was told that he would be asked questions about three stick figures and that the answer to each question would become known from a picture to be shown, and should be given as soon as possible. Subjects in the Noninterference Group were told that the pictures would be shown right after the question. Those of the Interference Group were told that the picture would be shown only after the following question was presented, so that his answer would always be made to the next-to-last question. The latter subjects were also shown a sketch of their task as in the right-hand part of Table 8. After a pair of examples subjects in both groups were further told that the experimenter would occasionally interrupt the procedure and ask them to recall a recent question. They were told the question would be cued by giving its verb, and that it would always be one of the last four heard. Four practice trials, ending with a recall test for the question in the second trial, were then given, before testing proper began. Except for the first trial in a block (where the interference group heard two questions), the procedure always followed the form: (i) the experimenter reads a question, (ii) a picture is then shown on the screen, (iii) the subject gives an answer related to the picture.

Results and Discussion

The results for preposition correspondence prior to the recall section of test blocks are summarized in Table 9. The pattern displayed resembles that in Experiment 2: correspondence between the form of the answer and the form of the question is significant for both the Noninterference Group and the Interference Group ($p < .001$ for each, sign test). In further agreement with Experiment 2, the correspondence effect is significantly ($p < .05$, sign test) greater in the Noninterference Group (70%) than in the Interference Group (59.5%). Thus, interleaving successive trials did result in lower probability of a corresponding answer, but even with this mas-

TABLE 9
Percentage of Prepositional and Nonprepositional Answers to Questions Containing or Not Containing the Preposition for the Noninterference Group and the Interference Group

Answer	Question	
	With preposition	Without preposition
Noninterference group		
With preposition	65	25
Without preposition	35	75
Total	100	100
Interference group		
With preposition	58	39
Without preposition	42	61
Total	100	100

sive interference the probability did not drop to chance level (50%). Thus, it appears that answerers even tend to give corresponding answers when the question's form has probably been erased from working memory. Surely, an articulatory buffer could not have been involved either.

The involvement of longer-term storage in this effect can be further assessed from the recall data obtained. It was predicted that if a subject in the Interference Group gave a corresponding answer to a question, that question's format would later be correctly recalled. So, for each question prompted at the recall trials we checked (a) whether the prepositional format had been correctly recalled, and (b) whether that question had earlier received a corresponding answer by the same subject. For each recall trial there was thus a four-way possible outcome: correct recall and corresponding answer, correct recall and noncorresponding answer, and similarly for incorrect recall. For each of the recall intervals 4, 3, and 2 we registered the outcome for the 24 subjects who had all been tested once for that interval. The resulting 2×2 contingency tables were analyzed by Fisher tests (two-tailed). Moreover, the degree of contingency was expressed in a phi coefficient. These coefficients and the results of the significance tests are given in Table 10, which also contains these data for the Noninterference Group. Additionally it gives the latter group's data for interval 1, the last interval; these will be discussed shortly. Table 10 clearly confirms the prediction for the Interference Group. For all three intervals recall of the question's format is significantly related to the way the question was answered. For these subjects the correspondence effect thus involves some form of long-term storage. We further tested for these subjects whether their *format-matching* in giving an answer was any better than their *format recall*. For each subject we compared his total recall performance over the three intervals, and his total answering performance for these three items. A sign test over the 24 subjects showed no significant difference ($p < .50$): they had performed equally well on giving a corresponding answer and recalling the preposition format of the question. There is thus no evidence that the subject's memory trace of the

TABLE 10
Association between Recall of Question Version and Question-Answer Correspondence.
Phi Coefficients for Intervals 1 (Noninterference Group Only), 2, 3, and 4,
and Their Significances

Recall interval	Noninterference group		Interference group	
1 (last)	.64	$p < .005$	—	—
2	-.06	n.s.	.58	$p < .01$
3	.07	n.s.	.57	$p < .01$
4	.21	n.s.	.48	$p < .05$

question's format was any better at the moment of answering it, than at the moment of recalling it. In other words, no additional short-term storage seems to have been involved at the moment of answering the question.

For the Noninterference Group, the results are quite different. As was discussed earlier, it is an empirical issue whether or not long-term storage is involved in the production of question-answer correspondence when no intervening material is presented. Table 10 gives a clear answer: for intervals 2, 3, and 4 there is no noticeable relation between question-answer correspondence and recall for the Noninterference subjects. The fact of a question's receiving a corresponding answer is not predictive of the question's long-term recall. Further evidence that the correspondence effect for these subjects is due to short-term recall comes from comparing the subject's level of performance for format matching in answering and the level of their recall performance. These were compared in the same way as was done above for the Interference Group subjects. The sign test (two-tail) showed a significant ($p < .02$) difference: the Noninterference subjects performed much better on question-answer correspondence than on recall. In this respect they differed significantly ($p < .05$, χ^2) from the Interference Group subjects who, as we have seen, performed alike on correspondence and recall. The Noninterference subjects must have worked from a short-term trace in producing their answers. A final piece of evidence for this is their recall performance on interval 1. Here, recall immediately followed the question-answer pair, so that recall is only short term. In this case a strong and significant relation appears between answering correspondence and recall of the question's prepositional format.

Thus the general conclusion from this experiment would be that in normal question-answering, as was the case for the Noninterference Group, long-term storage is not involved in the genesis of the correspondence effect but that there is a short-term trace of the question, which is immediately used in the formulation of the answer and which is subsequently forgotten. The recall instruction apparently does not induce subjects to try to memorize the prepositional form of the question. The subjects in the running memory condition, however, still produced some question/answer correspondence, and the pattern of their recall data makes it likely that this involved long-term storage of the question format. This strategy of long-term storage is probably not induced by the recall instruction, since that was the same for the Noninterference Group. Rather, the Interference Group subjects developed their long-term storage strategy simply in order to be able to answer the question, just as in Experiment 2 which had no recall condition. In a certain number of cases, that storage apparently involved the prepositional format of the questions.

Why that might be so will be further considered in the next and final experiment.

EXPERIMENT 6. NATURALNESS OF CORRESPONDENCE AND TYPE OF PREPOSITION

The prepositions used in the previous experiments, *naar*, *op*, *aan*, *van*, and *om* showed significant degrees of question/answer correspondence. Still, the size of the effect differed rather markedly for these prepositions. Table 11 presents the proportion of corresponding answers for these five preposition types under all conditions of the previous experiments. Size of question/answer correspondence generally decreases from preposition type *naar* (*for*), via *op* (*on*), and *aan* (*to*) to *van* (*of*) and *om* (*at*).

One could ask whether it would be more natural for a subject to give a corresponding answer if the preposition is *naar* than if it is *om* or *van*. There may indeed be good reasons for this to be so. The preposition *naar* (*for*) is part of the lexical entry of the verb *zoeken naar* (*search for*), but the same is not true for *om* (*at*): there is no verb *dicht gaan om* (*close at*). Also *van* (*of*) is unrelated to the verb: *the hat of Paul* and *Paul's hat* are just two different realizations of the genitive which have nothing to do with the main verb of the sentence. If it is natural for the answerer to use the question's main verb in the formulation of the answer, the corre-

TABLE 11
Proportion of Corresponding Answers^a for the Five Preposition Types
in the Conditions of Experiments 1-5

Condition	Type of preposition				
	<i>naar</i>	<i>op</i>	<i>aan</i>	<i>van</i>	<i>om</i>
Expt 1	.91	.77	.67	.56	—
Expt 2					
Low interference	.72	.64	.56	.56	—
High interference	.64	.55	.61	.55	—
Expt 3					
Simple question	—	—	—	—	.61
Complex question	—	—	—	—	.47
Expt 4					
No interference	.89	.88	.72	.53	—
Digit rehearsal	.91	.81	.66	.53	—
Expt 5					
Noninterference group	.77	.71	.61	—	—
Interference group	.61	.63	.54	—	—

^a Corresponding answers are prepositional answers to prepositional questions and non-prepositional answers to nonprepositional questions.

sponding preposition would automatically reappear in the case of *naar*, but not in the cases of *van* or *om*. It should, however, be remembered that we also found significant correspondence effects in the case of elliptical answers without a main verb. One would thus have to suppose that it is not necessary for the answerer to overtly express the main verb, but only to have it available in some form while formulating the answer. We will return to this issue in the General Discussion. Still, we also observed significant correspondence effects for *van* (*of*) and *om* (*at*), which are clearly not parts of the lexical verb. But there may be further reasons why it would be natural for an answerer to give a corresponding answer. The preposition may have particular *lexical support* in the question. If the supporting lexical items are remembered by the answerer, and if it is natural for him to use them in the construction of the answer, the corresponding preposition is likely to appear as well. The strongest lexical support for the preposition occurs if it is part of the verb, as just discussed. But there are also weaker forms of binding between preposition and verb. Compare, for instance, *op* (*on*) and *aan* (*to*). It is harder to move the prepositional phrase with *op* away from the verb *spelen* (*play*) than the one with *aan* from the verb *laten zien* (*let see*):

- (16) ? Ik denk dat Paul op de piano tijdens de pauze speelde.
 (*I think that Paul on the piano during the pause played.*)
 (17) Ik denk dat Paul aan Toos tijdens de pauze zijn hoed liet zien.
 (*I think that Paul to Toos during the pause his hat let see.*)

Sentence (17) is more acceptable in Dutch than (16). Comparable constructions with *zoeken naar* (*search for*) are even less acceptable:

- (18) ?? Ik denk dat Paul naar Toos tijdens de pauze zoekt.
 (*I think that Paul for Toos during the pause searches.*)

The preposition *om* (*at*) on the other hand gives no problem at all:

- (19) Ik denk dat de winkel om acht uur tijdens de pauze dicht gaat.
 (*I think that the shop at eight o'clock during the pause closes.*)

The degree of lexical support for the preposition from the verb is thus strongest for *naar* (*for*), slightly less for *op* (*on*), and again less for *aan* (*to*) and *om* (*at*). *Van* (*of*) has, as mentioned, no relation to the verb whatsoever.

The present experiment tests how natural it is to give a corresponding answer in the case of these five prepositions. Is naturalness related to the above-mentioned degree of lexical support, and can it be predictive of the size of the correspondence effect? In addition, it will be established whether naturalness of correspondence is greater when the answer contains the question's main verb (for prepositions *naar*, *op*, *aan*, *van*), and

whether for *om* (*at*) the addition of a clause (as in the telephone conversation experiment) enhances the naturalness of a noncorresponding answer.

Method

Materials. There were 80 question–answer pairs, half of which were “experimental” items, the other half being filler items. Of the 40 experimental question–answer pairs 32 were made up of the same questions plus (correct) answers that were used as experimental items in the preceding experiments. Four factors were varied: (i) preposition type (*aan*, *van*, *naar*, *op*); the content of the questions within preposition type was varied in order to have different questions for the different answer versions (factors (iii) and (iv) below); (ii) question with versus without preposition; (iii) answer with versus without preposition; (iv) answer in full versus in elliptical form.

The remaining eight experimental question–answer pairs were the four questions used in the telephone experiment (with or without preposition, and with or without the additional clause), each combined with an answer containing the preposition “om” and one not containing the preposition. The time mentioned in the answer was randomly varied, but always possible for a shop one might inquire about. These eight answers were always elliptical.

The 40 filler items also concerned the characters Toos and Paul, but used other verbs than the experimental ones (e.g., *What does Paul bring to Toos?*, or *Who is lying in bed?*). Experimental and filler items were tape-recorded in a randomly determined sequence, in which, however, no two experimental items of the same type were allowed to follow each other in immediate succession. Additionally, seven extra “warm-up” pairs were put at the beginning of the list. All questions were spoken by an adult male, and answered by an adult female. Both speakers were asked to use as much as possible natural speech and prosody.

Subjects. Twenty-nine new subjects from our paid subject pool participated in the experiment. All were native speakers of Dutch and paid Dfl. 7,— for 30 min of their time.

Procedure. Subjects were tested in two groups, 16 and 13 each in size. Each subject was given a nine-page booklet, with rating scales for 10 items vertically distributed over the page. Each rating scale consisted of a serial number and a string of five squares, the left end of which labeled “onnatuurlijk” (“unnatural”), and the right end labeled “natuurlijk” (“natural”). The numbers 1 to 5 were written in the five squares. The rating scales were numbered from 1 to 87. Subjects were told they would hear a series of question–answer pairs via tape recording, and that they were to judge for each pair, how “natural” the given answer sounded to that particular question. No information was provided about what aspect of an answer might be considered deviant, it was stressed that the answer content should be considered correct. Use of the rating scales was illustrated by means of two examples. Before each item, the experimenter announced the particular serial number of that item. Each item was presented only once. After each item, the experimenter waited until all subjects had marked one of the squares for that question–answer pair before going on.

Results and Discussion

The mean naturalness ratings for the various question–answer pairs are summarized in Table 12, which compares the corresponding and the noncorresponding pairs for the different prepositions. Let us first consider the top half of the table, relating to answers without a main verb. For each of the prepositions *naar*, *op*, *aan*, and *van*, corresponding pairs are significantly more natural than noncorresponding pairs (all $p < .00005$, sign test). It is, moreover, the case that the naturalness difference between

TABLE 12
Mean Naturalness Ratings for Corresponding and Noncorresponding
Question–Answer Pairs

	Type of preposition					
	<i>naar</i>	<i>op</i>	<i>aan</i>	<i>van</i>	<i>om</i> (no additional clause)	<i>om</i> (additional clause)
Answers without main verb						
Corresponding	4.48	4.52	4.41	4.12	4.29	4.07
Noncorresponding	2.05	2.50	2.52	2.47	4.53	4.16
Difference	2.43	2.02	1.89	1.65	-.24	-.09
Answers with main verb						
Corresponding	4.76	4.42	3.50	4.07	—	—
Noncorresponding	3.74	3.17	3.74	4.23	—	—
Difference	1.02	1.25	-.24	-.16	—	—

Note. Maximum naturalness = 5, minimum = 1.

corresponding and noncorresponding answers decreases gradually in the order *naar*, *op*, *aan*, *van*, *om* (some of these steps are statistically significant: for *naar* to *op*, $p = .010$; for *naar* to *aan*, $p = .16$; for *aan* to *van*, $p = .054$; and for *van* to *om*, $p < .00003$, sign test). This accords not only with the order of lexical support for the prepositions discussed above, but also with the order of size of the correspondence effect for the different prepositions, as presented in Table 11. It would, however, be wrong to conclude from these findings that a corresponding answer is given only when it is felt to be more natural than a noncorresponding answer. The results for *om* (*at*) show that corresponding and noncorresponding answers are both highly natural. Still, we obtained a significant correspondence effect for *om* in the telephone conversation experiment. Naturalness is not a *conditio sine qua non* for question–answer correspondence. It is not surprising, given these findings for *om*, that the addition of an extra clause to the question does not affect the difference in naturalness between corresponding and noncorresponding questions. In fact, naturalness is not significantly affected at all by this manipulation. We did find, however, that in the case of an additional clause, answers were judged to be more natural ($p < .04$, sign test) if they contained the preposition *om* than if they did not. This may have to do with the tendency observed in Experiment 3 for subjects to give fuller answers in case of complex questions.

The final observations to be made about naturalness concern the bottom part of Table 12. If the answer contains the question's main verb, naturalness generally increases ($p < .01$, sign test); the increase is also significant for the individual cases of *naar* and *van* ($p < .01$). The increase in naturalness is especially noticeable for noncorresponding question–answer pairs. Verb correspondence apparently compensates for the un-

naturalness which is due to preposition noncorrespondence. In other words, common elements, be they prepositions or main verbs, contribute to the perceived naturalness of a question-answer sequence. There is no evidence in the findings that naturalness due to preposition correspondence is enhanced if the answer contains the question's main verb. This may, however, be due to a ceiling effect: the naturalness of corresponding pairs is quite high already when there is no main verb in the answer.

GENERAL DISCUSSION

The present experiments have confirmed informal observations in the literature that speakers tend to repeat materials from previous talk, their own or their interlocutor's. The experiments established a "baseline" case: the repeat of a single word (preposition) in the answer to a question. In at least some of the experimental cases the word carried very little semantic or pragmatic function and repeating the question's preposition in the answer fulfilled no evident conversational function. The size of the question-answer correspondence effect could be manipulated by presenting the answerer with interfering information of various sorts. Systematic manipulation of task and interfering materials made it likely that, whatever the precise form in which the question's preposition was memorized, the relevant information is kept in working memory during the preparation of the answer. It was shown to be unlikely that the so-called "articulatory buffer" is involved in the generation of the correspondence effect. Also, there is no evidence that in the case of normal question-answer turns long-term memory is involved in the creation of question-answer correspondence. But when subjects work under great memory load, they are still able to produce some correspondence in their answers. We could show that this was due to a special long-term memory strategy which is not normally used. In a final experiment it appeared that question-answer turns are perceived as more natural when they agreed in prepositional form. Also, naturalness of correspondence increased with the amount of "lexical support" a preposition received from the question's main verb (e.g., *search for* versus *close at*). This, in turn, correlated with the size of the correspondence effect observed in the previous experiments. Still, neither naturalness nor "lexical support" can fully account for the established cases of correspondence. It rather seems to be the case that a previous element of speech which is available in the speaker's working memory can, by its mere presence, affect the formulation process and reproduce itself during the speaker's turn.

This state of affairs raises two important theoretical questions. The first one is: why are we equipped with a mechanism for preserving recent speech of our own or others? The second is: why would a preserved

element of speech be used in the generation of the next stretch of discourse?

As far as the first question is concerned, it is not necessary to review the extensive evidence from Jarvella's publications (1970, 1979) and subsequent studies for surface memory of the most recent clause. From the point of view of speech perception, such surface memory does seem to make sense for at least two reasons: (i) Though there is increasing experimental evidence that the listener interprets incoming speech "on-line" as deeply as possible with respect to meaning and reference (see, for instance, Marslen-Wilson & Tyler, 1980), the listener may occasionally fail to come up with the intended result. He or she may have misinterpreted some information, or simply have had an attentional lapse. In that case the availability of recent surface information may become a last resort for recomputation. Experiment 5 above shows that answerers who experience great difficulty in the on-line interpretation of a question manage to store the surface information in a more long-term fashion so that delayed (re)computation becomes possible. Garnham, Oakhill, and Johnson-Laird (1981) similarly found relatively good surface memory for texts that were hard to interpret (as compared to texts that were easy to understand). (ii) Even if recent speech is immediately and correctly interpreted, its surface form may critically determine the interpretation of ongoing talk. This is so in cases which Hankamer and Sag (1976) have called "surface anaphora." One example (which I owe to Philip Johnson-Laird) suffices to show this:

- (20) The Romans were sold the Britons as slaves, and the Gauls were too.
- (21) The Britons were sold to the Romans as slaves, and the Gauls were too.

If a listener only remembers the gist of the first clause of (20) and (21), there would be no way to interpret the second clause unambiguously. Though the gist of the initial clauses and the surface form of the following clauses are the same for (20) and (21), the interpretation of the second clauses should be radically different, and this hinges on the surface form of the first clauses. It is important to observe that surface anaphora can create the same problem for the speaker. If the speaker intends to formulate the state of affairs expressed in (20) and produces the first clause, he can only safely formulate the second elliptical clause if he has recourse to the surface form of the first clause. Verbatim retention of recent speech can thus be a requirement for the felicitous production of surface anaphora and ellipsis, which is quite normally observed in spontaneous discourse.

One would, in fact, expect similar interference effects as those ob-

served in the experiments above when a speaker produces coordinated constructions which allow for ellipsis. One example suffices to show this. It is quite natural for a speaker to use gapping where the two clauses of the sentence are adjacent as in:

(22) Mary visited the Rijksmuseum, and John the zoo.

It is less natural to do so if additional ‘‘interfering’’ clauses are inserted, as in:

(23) Mary visited the Rijksmuseum, you know I told you about this beautiful exhibition on expressionist art which is running there till the end of May, and John the zoo.

In the latter case a speaker is more likely to insert ‘‘visited’’ in the last clause as well. Correct production of the zero anaphor requires memory of the verb’s surface form, which may have been lost during production of the intervening clauses. The speaker will then generate the last clause exclusively from a semantic base. The resulting full form will be to the advantage of the listener as well, who might not have kept track of the main clause’s verb either.

The second theoretical issue is a more complicated one. If it is normal to keep a trace of one’s own or the interlocutor’s talk, why would one tend to reuse certain elements or turns of phrase in the newly produced speech? It has been suggested in the introduction that there may be a variety of semantic and pragmatic reasons to repeat elements of previous speech. The fact, however, that repetition can even be observed when such reasons are largely or wholly absent makes it likely that reusing previous discourse elements has the additional function of facilitating the fluency of the formulation process itself. It may require less effort to reuse available surface materials wherever possible than to generate speech every time anew from a semantic base. But if this is so, it presupposes the existence of a speech production mechanism in which the ‘‘formulator’’ which produces the surface form can function in relative independence of the speaker’s conceptual or pragmatic intentions. This view agrees well with other recent findings in the literature (cf. Kempen, 1977; Kempen & Hoenkamp, 1981; Garrett, in press; Levelt & Maassen, 1981).

REFERENCES

- Baddeley, A. D. *The psychology of memory*. New York: Harper & Row, 1976.
 Baddeley, A. D., & Hitch, G. Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation*. New York: Academic Press, 1974. Vol. 8.
 Bühler, K. *Sprachtheorie*. Stuttgart: Fischer, 1934, 1965.
 Clark, H. H. Responding to indirect speech acts. *Cognitive Psychology*, 1979, 4, 430–477.
 Clark, H. H., & Schunk, D. H. Polite response to polite requests. *Cognition*, 1980, 8, 111–143.

- Garnham, A., Oakhill, J., & Johnson-Laird, P. N. Referential continuity and the coherence of discourse. *Cognition* 1981, 10, in press.
- Garrett, M. Production of speech: Observations from normal and pathological language use. In A. W. Ellis (Ed.), *Normality and pathology in cognitive functions*. New York/London: Academic Press, in press.
- Hankamer, J., & Sag, I. Deep and surface anaphora. *Linguistic Inquiry*, 1976, 7, 391-426.
- Jarvella, R. J. Effects of syntax on running memory span for connected discourse. *Psychonomic Science*, 1970, 19, 235-236.
- Jarvella, R. J. Immediate memory and discourse processing. In G. H. Bower (Ed.), *The psychology of learning and motivation*. New York: Academic Press, 1979. Vol. 13.
- Kempen, G. Conceptualizing and formulating in sentence production. In S. Rosenberg, *Sentence production: Developments in research and theory*. Hillsdale, N.J.: John Wiley, 1977.
- Kempen, G., & Hoenkamp, E. A procedural grammar for sentence production. Report No. 81 FU 03/81 SO 07, Department of Psychology, Nijmegen University, 1981.
- Levelt, W. J. M., & Maassen, B. Lexical search and order of mention in sentence production. In W. Klein & W. J. M. Levelt (Eds.), *Crossing the boundaries in linguistics. Studies presented to Manfred Bierwisch*. Dordrecht: Reidel, 1981.
- Marslen-Wilson, W., & Tyler, L. The temporal structure of spoken language understanding. *Cognition*, 1980, 8, 1-71.
- Morton, J. A functional model for memory. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.
- Schenkein, J. A taxonomy of repeating action sequences in natural conversation. In B. Butterworth, *Language production*. Vol. 1. *Speech and talk*. New York/London: Academic Press, 1980.
- Selz, O. *Über die Gesetze des geordneten Denkverlaufs*. Bonn: Cohen, 1922. Vol. 2.

(Accepted July 7, 1981)