The man lifts up the BAG, the man lifts up the BOX

Instituut voor Perceptie Onderzoek
Postbus 513, 5600 MB Eindhoven

16.11.1992

Rapport no. 877

The man lifts up the BAG
The man lifts up the BOX

J.F.J. Fransen

The man lifts up the ***BAG***
The man lifts up the ***BOX***

On the Influence of Intonational Accent
on the Verification of Spoken Information.

Doctoraalscriptie voor . . . . . Fonetiek

Aan . . . . . de Rijksuniversiteit te Utrecht

Door . . . . . Jeroen F.J. Fransen
Student Nr. . . . . . (5)8637334
Datum . . . . . Augustus 1992

Scriptiebegeleiders . . . . . Dr. Hugo Quené (Utrecht)
Dr. Jacques M.B. Terken (Eindhoven)
Prof. Dr. René Collier (Eindhoven)

Deze scriptie werd geschreven naar aanleiding van een onderzoek dat gedeeltelijk plaatsvond en werd voorbereid in het kader van een externe stage bij het Instituut voor Perceptieonderzoek (IPO) te Eindhoven.

'This thesis is the result of research that was partly conducted and prepared during an external traineeship at the Institute for Perception Research (IPO), Eindhoven, the Netherlands.'

(Did you spot the difference between the 'men' in the title on the front page?)

≈M.A. Thesis for . . . . . Phonetics
At . . . . . the University of Utrecht

By . . . . . Jeroen F.J. Fransen
Student Nr. . . . . . (5)8637334
Date . . . . . August 1992

Supervisors . . . . . Dr. Hugo Quené (Utrecht)
. . . Dr. Jacques M.B. Terken (Eindhoven)
. . . . . Prof. Dr. René Collier (Eindhoven)

Mottos

"Nothing comes from dreamers, but dreams,
Nothing comes from talkers, but sound,
The world still goes around and around."

Prince, 'Round and Round'

"Man soll nie versuchen ein Gleichspiel zu machen. Wir können das überhaupt nicht."

Berti Vogts, coach of the German football team, interviewed on the eve of Holland-Germany in Sweden 1992 (3-1)

"Take care of the sense and the sounds will take care of themselves"

Lewis Carroll

"'Music carries me instantly and directly into the state of consciousness that was experienced by its composer. My soul merges with his, and together with him I'm transported from one state of consciousness into another; yet why this should be, I've no idea.'"

Taken from 'The Kreutzer Sonata', Leo Tolstoy

Preface and Acknowledgements

The present thesis is the result of a seven-month period during which I was forced to say no to most at one stage of time. I could not say no to everyone, which is my weak and, according to some, my strong side at the same time. I am grateful to all who could bear with me during this period of self test.

Caro I want to thank in particular. She was always behind and beside me, and to her I didn't want to say no. Without her ...

Many thanks go to my supervisors, who were there when I needed them most. Thank you Hugo, thank you Jacques and René. Thanks to Bert for useful comments. Special thanks go to Wilma and Yvette.

Many thanks also to Camiel, who was literally at my side most of the time, to Deirdre for living in the same house, to Maarten for being a friend with good ideas, to 'Mr President' Oliver, and to 'good food' Inge and 'actually' Astrid.

I am deeply grateful to my parents, who supported me in many ways. The same goes for Wibe. Thanks to Annemarie for being that very special kind of teacher, good luck!

Two people I owe a lot to are Theo Veenker from Utrecht, who could always find some spare time for me, and Marjolein de Vette from Eindhoven, who supported me in the creation of the pictures and actually did most of the drawing from photographs.

Thanks to Leo Vogten, Cees van Deemter, Ercan, and Jan-Roelof for general assistance, and thanks to everyone I forgot to thank here.

Table of Contents

Introduction

Chapter 1

This phonetically inspired paper about a psycholinguistic research on the influence of Accent will commence with a short introduction, consisting of the following elements. Section 1.1 aims to introduce the general body of ideas and questions. This discussion will focus in on the Given-new contract in 1.1.1 and the Given-new strategy in section 1.1.2. In section 1.1.3 the possible influence of given/new outside the actual Given-new strategy will be discussed, leading to the key question in section 1.1.4. The actual hypotheses are then proposed in section 1.2, and the chapter will finish with a preface to the hypothesis-testing experiment in 1.3.

1.1 Given and New Information

To be effective in communication, a speaker - or a writer for that matter - has to pay close attention to his audience. He has to keep track of what the audience knows and does not know, and he has to adapt the contents and form of his message to what he thinks the audience knows. He has to refer to things the audience knows, given

information, in order to make them understand him, and tell them things they do not know, new information, or they will not stay with him.

Communication takes two sides cooperating in the process. Not only do speakers have to think about the audience and their mutual environment, the audience itself has to be cooperative as well. Listeners are aware that the speaker has some idea about what the audience already knows, and prepare themselves for new information.

The message produced is thus tailored to the mutual knowledge and external points of reference of both participants in the communication process, so that its interpretation does not cause any problems.

Definitions of what may be considered as actually new and what as actually given information have been proposed by, for instance, Halliday (1967). Given information is defined by him as information *'offered as recoverable anaphorically or situationally'*. That is, given information is supposed to refer to one of two different kinds of information: either to things mentioned earlier in the current discourse ('anaphoric'), or to things that can be found in the environment surrounding speaker and listener ('situational'). In the definition of given information, 'recoverable' will be understood as *'already present in the consciousness'* (Chafe, 1974)[1]. New information, on the other hand, is defined by Halliday as *'... non-derivable information, either cumulative to or contrastive with what has preceded'*. In other words, new information cannot be extracted from already familiar knowledge, and either adds to or contrasts with information known to the listener.

1.1.1 THE GIVEN-NEW CONTRACT

A speaker is not only supposed to include given as well as new information in his utterances but, essential in the current research, he also has to make clear to the listener which is which, by marking it out. This implicit arrangement between

---

1 Given (already present) information will be interpreted in this research as mentioned in the previous sentence.

speaker and listener is defined more sharply in what is called the ***Given-new Contract***. According to this contract about assertions[2] the speaker agrees

a) to mark information as GIVEN if it refers to information he thinks the listener can uniquely identify from what he already knows and

b) to mark information as NEW if it refers to knowledge he believes is not already known to the listener (Clark and Clark, 1977).

In speech the GIVEN/NEW status of a word (which indicates if a word is marked either GIVEN or NEW) is usually expressed by means of a specific *accent pattern* (Hornby, 1971), either with or without a special *grammatical structure*. Take the following question: 'Who washed the dishes?'. Two possible types of answers to this question are exemplified in sentences 1 and 2 (fat capitals indicate +Accent).

1. Your **NEIGH**bour washed the dishes.

2. It was your **NEIGH**bour who washed the dishes.

Due to the preceding question both answers have 'dishwashing' as given information. In sentence 1 information is marked as NEW by realising it with ***+Accent***. In 2, the word 'neighbour' is additionally marked NEW by the specific grammatical construction used: 'It was ... who ...'. This cleft sentence construction points to 'neighbour' (the new information) as the one who performed the action (which is given information). In 2 the same information is conveyed as in 1: the neighbour did it. The difference is found in the form.

Baart, in his framework for Accent placement in Dutch, in which he combines ideas from the previous literature with his own, gives a more precise definition of what happens to given and new information in Dutch speech (Baart, 1987). According to him, new information in a sentence is placed *'in focus'*, while given

---

[2] This excludes yes/no questions, WH-questions, and requests from the further discussion.

information is placed *'out of focus'*. The focus distribution contained in the sentence is translated to an accent pattern on the basis of the syntactic structure of the sentence; in general, the +Focus *head* of a syntactic constituent receives +Accent on the appropriate syllable. This +Accent is, in turn, realised as a prominence-lending pitch movement on the syllable (see also 't Hart & Cohen, 1973; 't Hart & Collier, 1975).

Thus, in general, new information is marked NEW by realising it with +Accent, although words that are not realised with +Accent may also contain new information. This happens when a word falls under the *integrative accent* of another word in the same constituent, i.e. the entire (new) constituent is focused by means of a +Accent on a single word. In the rest of this paper, when a word is said to be realised with +Accent, this will stand for a word in a constituent which is in focus, and of which a syllable is realised with a prominence-lending pitch movement.

1.1.2 THE GIVEN-NEW STRATEGY

The conditions set by the *Given-new Contract* allow listeners to structurally process an incoming sentence. This is done by applying the three steps constituting the ***Given-new Strategy*** (Clark and Haviland, 1974).

a) Divide the sentence into given and new information.

b) Search memory for a unique antecedent that matches the given information.

c) Attach the new information at that point in memory.

In other words, when a listener is processing a sentence he first divides it into given and new parts on the basis of the GIVEN/NEW markers that are placed according to the Given-new contract, uses the given information as the basis of the interpretation

of the message being conveyed, and integrates the new information in his memory at that point.

1.1.3 VERIFICATION AND GIVEN/NEW

Hornby (1974) found that actually new information, which was also marked grammatically and intonationally as NEW, could be verified more reliably than given information, which was marked as GIVEN. He carried out the following experiment on Given/new using speech material. He read subjects a sentence like 3 and then showed them a picture for 0.05 second, which they had to judge as true or false.

3. It is the **BOY** who is petting the cat.

He found that listeners made more mistakes in the verification when the *cat* was misrepresented in the picture, than when this was done with the *boy*. In other words, he demonstrated that listeners performed picture verification on the basis of actually new information, which was also marked as NEW, whereas given information, marked as GIVEN, was taken for granted more often.

The question is, then, how Hornby's listeners determined which information was new. Two possible explanations will be given here.

Firstly, it is possible that listeners decide what is *new* on the basis of the GIVEN/NEW **markers**: the **form** aspects of the sentence. In other words, they use the markers agreed upon in the Given-new contract (the Accent pattern) to determine which information will have to be verified. This would imply that listeners are more or less conditioned by the Given-new strategy to use information *marked as* GIVEN as the basis of their interpretation, and to take it for granted.

Another possible explanation is that listeners rely on the actual lexical **contents** of the sentence to determine which information is *new* and consequently is to be verified.

This leads us to the key question of this research.

1.1.4 THE KEY QUESTION

The key question to be answered by the present research is whether listeners arrive at their verification decision on the basis of the ***form*** or the ***contents*** of the sentence, i.e. to what extent listeners, performing a verification task, depend on the GIVEN/NEW markers set by speakers. This very general question is translated to the experimentally answerable question whether the accent pattern (form) or the words (contents) determine which information is to be verified and which is taken for granted.

In order to try to find an answer to this question the Given-new contract will be 'violated': situations will be compared in which *actually given* information is realised either with *-Accent* ('normal' situation, indicates GIVEN) or with *+Accent* ('abnormal' situation, indicates NEW).

Another way of breaking the Given-new contract would have been to realise actually new information with -Accent. This was excluded from the design for two reasons. Firstly, as we saw at the end of section 1.1.1, unaccented new information can be considered as falling under the integrative accent of another part of the same syntactic constituent, so that a word realised with -Accent does not necessarily have to be given information. The second reason results from the fact that the sentences to be used consisted of a subject, a verb and an object, where only subject and object allowed +Accent (see Chapter 2 for details). If the contract were broken by realising new information with -Accent, the resulting situation would consist either of a sentence in which none of the words is realised with +Accent, or, in the case of a double violation of the contract, of a sentence in which actually given information is realised with +Accent and actually new information with -Accent. In these highly unnatural situations, it would be clear right from the start that form (accents) conflicted with contents, which would not allow to generalise conclusions to 'normal' situations.

1.2 Hypotheses

In order to answer the key question raised in the previous section the following hypotheses are proposed.

Firstly, in order to be able to investigate the independent influence of the GIVEN/NEW markers, a verification advantage for the use of actually new information marked as NEW has to be found to replicate Hornby's results discussed above. Therefore, as a first hypothesis it is proposed:

If a word is new it can be verified more easily than if it is given.

In addition to this, it is predicted that people rely on the GIVEN/NEW markers, the Accent pattern, to decide what is to be verified and what is not. More explicitly, the main hypothesis can be stated as follows.

If a given word is realised with +Accent it can be verified more easily than if it is not realised with +Accent.

The experiment that was used to test these hypotheses will be introduced in the next section.

1.3 Introduction to the Experiment

Although the detailed contents of the actual experiment will be accounted for in the next chapter, its general outlines are introduced here.

In the experiment a situation was to be created that would allow testing the hypotheses of the present research. Considering this purpose, it was clear that an environment was needed in which intonation and actual givenness could be varied independently.

The following experiment was conceived to test the hypotheses. Subjects would have to check spoken sentences for their correctness. A design was made

where words could be accented or unaccented, where they could either be actually given or new, and where they could be correct or incorrect with respect to an external point of reference.

These ingredients, + or -Accent, given/new, and correct/incorrect, could be obtained and varied along the following lines. The ± Accent status of a word could be controlled by manipulation of the pitch contour of the speech signal. Actual givenness could be created by presenting the same word in a preceding sentence (prime). Thus, a context was created, which made critical words in the target sentence either new or given. In the same fashion, correctness of a word can only be determined in relation to something else. This prompted the introduction of another reference, this time governing the true/false state of a word. Taking Hornby's experiment as an example, pictures were used against which the target sentence had to be verified (Hornby, 1974).

In addition to the ± Accent and given/new conditions, called for by the main question, one more factor was taken into account. Results could be expected to differ depending on whether the grammatical subject or the object in the sentence was given information, due to limitations within the experimental design. In order to keep the, already quite intricate, design relatively simple, it was decided to use only sentences of a specific configuration. Thus, throughout the entire experiment both context and target sentences contained only a **S**ubject (agens), a **V**erb, and an **O**bject (recipiens), in this order (details concerning sentences and word order will be explained in the next chapter). This specific sentence structure could cause a subject-object effect in two ways.

Firstly, the first constituent of a sentence is preferred as the place for given information (e.g. Kuno, 1971), and hence as the preferred place for -Accent. In the present experiment the grammatical subject would always be in this place.

In the second place, grammatical subjects are preferred as locations for given information over objects (Keenan, 1976). This factor may also play a role in the present research, as the subject will always be the first constituent in the test utterances, but will sometimes be used to express new information.

As a result of these two possible effects and their interactions, a comparison was to be made between subject and object for each effect.

Method

Chapter 2

*I*n the first section of this chapter the experimental method will be presented in broad outlines. In subsequent sections the technical details will be filled in, and an account will be given of the choices made.

2.1 Global Description of the Experiment

The basic idea was to test subjects on the accuracy and speed with which they could verify an utterance-picture combination. Target words could be given or new, by manipulating the (preceding) priming sentence. The intonation of the utterance was manipulated to obtain different accent patterns. In half of the cases the picture did not match the actual contents of the utterance.

Subjects were offered trials consisting of a target sentence followed by a very short presentation of a picture, and had to indicate as fast as possible whether or not the picture corresponded with the sentence.

In order to create a situation in which the intonation could be manipulated to coincide with, or differ from the actual given/new status of a word, target combinations of sentence and picture were preceded by similar combinations functioning as context. As a result of the context either the subject or the object of the target sentence was given. In the context the picture always corresponded exactly with the sentence, in the target only in 50% of the cases. One half of the actually given words was realised with -Accent, the other half with +Accent. New information was always realised with +Accent (see section 1.1.4).

Three tapes were made, each containing a complete test in a different randomised order. Each test (160 trials) for one subject consisted of two sessions with 80 trials each. Each session lasted approximately 13 minutes. The actual test was preceded by a training session (80 trials).

The actual Yes/No responses and their corresponding reaction times were stored in computer files.

2.2 Structure of the Experiment

2.2.1 Reference and Context

An external point of reference is an essential ingredient of a verification task. As the outcome of the present experiment depended highly on the influence of contextual information, it is essential to understand the reasons to include pictorial information in the role of this external reference. One point in favour of pictures is surely that, in everyday situations, linguistic information is constantly being processed simultaneously with and compared to pictorial information.

Furthermore, in this experiment two types of context were required, one type to be able to compare the influence of intonation together with, and separately from the actual new/given structure, the other type as a reference to be verified. One seemingly obvious way to introduce context was, of course, to simply present two utterances one after the other, the first one serving as context for the second. The subject's task would then consist of deciding whether or not the first utterance was

matched by the second. This task, however, would remove any context effect as both utterances would be, in a way, part of the target. The first utterance would then serve both as context to create the difference between given and new information and as a reference for comparison. Obviously, using one utterance as context would not do the job.

Using more than one sentence as context would not solve this context/reference dilemma. If another utterance were added a trial would be created that would consist of a context utterance, a target utterance, and one in between. Again, it would not be clear what purpose the context utterance should serve here, as the middle utterance (reference) may well keep the context from creating presupposedness in the target utterance (Chafe, 1974). This may, in turn, obstruct the intonation that comes with given information. In other words, it would not be clear whether a word, that has occurred two utterances before in the context, should or should not be accented. This obscuring effect only gains strength as more and more sentences are added. Clearly, using sentences as verification target would not work.

The present picture verification design[3] had clear advantages (see Table I on p. 12). As a picture-sentence combination preceded a sentence-picture combination, test subjects could simply be asked whether or not the second *picture* corresponded with the second *sentence*; in this situation context was separated from target information. The preceding acoustic-pictorial information together ($sentence_1$ and $picture_1$) served as context to the target utterance. In addition to that, the first picture gave the subject some idea as to what kind of picture he or she could expect for a target picture.

An additional practical advantage was that this design allowed uniform time and response measurements. In a sentence-sentence situation it would not be easy to decide from which point in the target sentence time measuring should start, and any choice would be an arbitrary one. The present design offers the advantage that time measurements can start from the moment the second picture has been presented.

---

3 A similar design, in which sentences had to be verified against pictures, has been used before with a similar hypothesis. One of the researches that was used as an example for the present experiment was that by Hornby (1974). He used picture-sentence constructions to show that listeners performed verification on the basis of given/new.

Although reaction time data will be treated only as secondary data, uniform time measurements were still preferred, if only for reaction time feedback.

TABLE I *Diagram of the time structure of a trial.*

Speech			sentence$_1$		sentence$_2$			
Screen		picture$_1$					picture$_2$	response recording and time measuring
Pause				<- 1 s ->		<- 0.05 s ->		
Time (s)->	*0*	*0.5*	*2*	*3*	*4.5*	*4.55*	*4.58*	

2.2.2 TIME STRUCTURE

It remains to explain the order of pictures and utterances within one trial, as exemplified in Table I. Subjects saw a picture, and after 0.5 second they heard the first sentence. At the end of the sentence the picture disappeared. After a short interval of 1 second they heard the second sentence, at the end of which the second picture flashed on for about 30 ms.

In the first combination of picture and sentence the picture is switched on 0.5 second before the sentence. It is not switched off during the utterance to provide ample time for subjects to process sentence$_1$ and picture$_1$ together. The first picture is presented even before the utterance begins because the sentences were rather short. They lasted only some 1.5 seconds, which was not enough time for the first picture to build a firm context for the following target.

The specific symmetrical *picture$_1$-sentence$_1$-sentence$_2$-picture$_2$* order was determined on two grounds. Firstly, this order guaranteed that sentence$_1$ could serve as context for sentence$_2$. If a picture had come between the two sentences, it would not have been clear whether an item mentioned in sentence$_1$ would still be given information in sentence$_2$ (Chafe, 1974). The second reason behind this was of a more practical nature. If the second picture were not the last part of a trial it would still be quite difficult to tell from which point time measurements should start, as was mentioned above. Thus, one advantage gained by using pictures would be lost.

2.2.3 *Presentation Time*

As was mentioned above, the presentation time of the target picture was only 32 ms. This duration was determined first on the basis of the description of Hornby's experiment (Hornby, 1974). Hornby presented simple black-and-white line drawings (similar to those used in the current experiment) on a tachistoscope. He found that a presentation time of 50 ms was just enough for subjects to interpret the image, but prevented them from actually studying the different elements in the image. This should force the subject to use given/new information to direct a verification strategy.

In order to check if the presentation time determined by Hornby was right for the present experiment a pilot test was conducted. In this pilot test[4] the presentation time was set at 48 ms (due to technical limitations of the video card and screen, presentation times could only amount to a whole multiple of 16 ms). In this test subjects gave almost 100% correct answers, which allowed no differentiating interpretations. Two factors were held responsible for these results. The first one was the virtually unlimited response time: from the presentation of picture$_2$ of the current trial until the presentation of picture$_1$ of the following trial, totalling some 4 seconds. The second factor, which was brought up by the pilot test subjects, lay in the fact that the picture could be analyzed quite thoroughly in the available presentation time of 48 ms. To prevent these flaws from influencing the actual experiment, the definitive presentation time was set at 32 ms, while a time-out message was added to elicit faster responses.

2.3 Summary of Conditions

Most matters concerning the external structure of the experiment have been accounted for in the previous sections. In the present section a summary will be

[4] Two subjects participated in this pilot test, which was the same as the real experiment, except that the presentation time was 48 ms, the learning session consisted of only 10 trials, and no feedback was given on reaction time.

given of all conditions that were tested in this experiment, illustrated by an overview in Table II.

TABLE II *Map of all conditions to be tested, with the number of presentations per subject. An index is in the upper left corner.*

		Subj Given Obj New	Obj Given Subj New
Given False New True	Given -Acc New +Acc	1 10	2 10
	Given +Acc New +Acc	3 10	4 10
New False Given True	Given -Acc New +Acc	5 10	6 10
	Given +Acc New +Acc	7 10	8 10
Given True New True	Given -Acc New +Acc	9 20	10 20
	Given +Acc New +Acc	11 20	12 20

Each trial could represent one out of twelve possible conditions as follows. The target sentence of a trial could contain either given information that was incorrectly pictured, or the new information was incorrect, or both given and new information completely matched the picture (in Table II, conditions 1, 2, 3, and 4 *vs* 5, 6, 7, and 8 *vs* 9, 10, 11, and 12). The given information could be realised either with or without +Accent (e.g. conditions 1 and 2 *vs* 3 and 4). The given constituent could be either the grammatical subject or the object in the target sentence (left column *vs* right column).

Taking all possibilities summed up above, the result was a 3 (New/given/both True) x 2 (given -Accent/+Accent) x 2 (subject/object given) = 12 conditions. For the sake of keeping an equal balance between true and untrue situations, a double number of trials were offered in conditions in which the total target sentence matched the picture (conditions 9, 10, 11 and 12).

2.4 Preparing the Ingredients

After the structural design of the experiment had been defined it had to be filled in concretely. In the following sections the ingredients and their preparation will be accounted for.

2.4.1 Stimulus Sets, Situations, and Sentences

In order to be able to offer ten replications of each of each condition, ten stimulus sets were created, each containing all twelve conditions. Each stimulus set started from a basic situation: the context sentence. From the context sentence the target sentences were derived: one target sentence in which the grammatical subject was given and one in which the object was given.

Combined with the right intonation contour and the appropriate picture, every condition could be created from these sentences.

The 10 basic sentences and their related target sentences were established which could easily be translated in pictures. As the different stimulus sets would be treated as replications of one and the same sentence, they also had to have similar characteristics. For this purpose the following criteria were adopted.

The choice of sentences was limited by using only utterances of the Subject Verb Object (SVO) type. This was done to secure grammatically and semantically equivalent situations which would be relatively easy to interpret.

The choice of verbs, already restricted by the previous decision to use only SVO sentences, was further narrowed down by the fact that the action expressed in it had to be common, clear, concrete, and easily recognisable at first sight. One more strong restriction was put on the choice of verbs. This restriction lay in the fact that, for each subject and each object used in the context sentence, an alternative of equal probability had to be available which could play the same role in the target sentence. These two requirements limited the number of verbs seriously.

The number of possible subjects and objects that were to play a role in the situation was again extremely limited by all previous decisions. Some verbs allowed

only two or three subjects and objects. From the small set of equally probable candidates that would go with a selected verb, the ones that were visually easily distinguishable were picked. Preference was given to minimal pairs, such as 'man/woman'. During the entire selection process, good use was made of intuition.

A comprehensive list of the sentences used in this experiment can be found in Appendix A.

2.4.2 SPEECH MATERIALS

2.4.2.1 SPEECH INPUT

The context and target sentences, 30 in all, were read aloud in the soundproof studio at IPO by an approved speaker. This speaker's voice had been pre-tested for the specific techniques that were to be used in the further speech manipulation process.

The sentences were pronounced at a distance of 30 cm from the microphone and recorded with a maximum sound level of -3 dB on a semi-professional DAT recorder. Subsequently, the analog signal from the DAT recorder was lowpass filtered at 5 kHz and used as input for 12-bit sampling at 10 kHz by a VaxLab.

The **P**itch-**S**ynchronous **O**ver**L**ap and **A**dd technique (PSOLA) was used for the actual manipulation of intonation contours (Hamon et al., 1989). PSOLA is a technique allowing manipulation of intonation and duration in the waveform, with high-quality re-synthesis results. The manipulation process is as easy as LPC-based manipulation, but the results are far more natural and, in favourable conditions, indistinguishable from original speech.

This technique uses the voiced/unvoiced (VUV) and F_0 parameters determined by means of LPC analysis. So, after cutting the signal into sentences the digital signal of each sentence was LPC analyzed on the basis of a source-filter model. LPC analysis was performed by determining the best 10-factor fit to be made within a time window of 25 ms, updating parameters every 10 ms. The original intonation contour was then determined on the basis of the LPC file and the sampled data file.

The software used to automatically determine the intonation contour did not change the original VUV decisions made in the LPC analysis. Each file was then checked for its VUV decisions by comparing it visually and auditorily to the original sampled signal. Where necessary, the VUV decisions were manually modified.

The PSOLA technique is based on pitch synchronous manipulation, where all desired changes are performed period by period of the speech signal. This technique works on the basis of pitch markers for the original sampled-data file. The next step thus consisted of performing pitch mark analysis on the files.

Artificial intonation (F_0) contours were made. Following the rules for Dutch intonation (RNK), so-called 'pointed hat' LPC F_0 contours were produced (see also 't Hart & Cohen, 1973; 't Hart & Collier, 1975). '+Accent' was defined as an excursion of 6 semitones from a declination line of 5 semitones/second. '-Accent' was defined as an excursion of 0 semitones from the declination line. In order to determine the positions where +Accents were to be placed, the vowel onsets of the syllable that was to carry the Accent were determined by eye and ear. Context sentences received two +Accents, one for the grammatical subject and one for the object. Of each target sentence two versions of the intonation contour were made: one where given information was realised with +Accent and one where it received -Accent.

The new intonation contour was merged with the rest of the speech signal using the PSOLA technique.

2.4.3 Stimulus Tapes

For each context sentence two target sentences had been recorded originally, one in which the grammatical subject constituent was given by the context, and one in which the grammatical object was given. By manipulating the intonation, as explained in the previous section, from each target sentence two versions had been created, one with +Accent for the actually given constituent and one with -Accent. Thus, by combining one context sentence with each of the possible target sentences four situations could be created.

Each of these four situations could, in turn, be combined with one of three pictures: the picture could either be completely correct, or be partly incorrect in that the new information or the given information was pictured incorrectly.

In this way trials were created with which every condition (see Table II) could be realised for every stimulus set. Each trial, combined with appropriate pauses was merged into one file. Special care was taken of the time alignment in view of the synchronisation of the experiment.

Three lists were made with 160 (10 situations x 16 conditions) trials in randomised order. These lists were split into two equal parts of 80 trials, which would be the two sessions making up a complete test.

Using these lists, 3 tapes were recorded, each containing a complete test in a different randomised order. Between each session a silence was inserted on the tape. Recordings were made directly[5] from the VaxLab DA-converter to the semi-professional DAT recorder, at a maximum sound level of -3 dB.

2.4.4 PICTURES

Starting from the list of situations defined by the set of sentences (see Appendix A on p. 47), the pictures were made according to the procedure described in 2.4.4.1. In the present section first a few more abstract matters will be dealt with.

The specific type of picture used in this experiment can be described roughly as a black-and-white detailed silhouette. This type of picture was chosen after consulting prof. Bouwhuis of the visual department at IPO. He was presented with the same picture in three formats: as a line drawing, as a silhouette and as a detailed silhouette. According to him, a line drawing is known to have the disadvantage of being an abstraction itself. When interpreting a line drawing one has to build up the picture in the mind. The pure silhouette, on the other hand, though a relatively natural way of depicting a situation, does not contain enough details to allow

---

5 As opposed to recordings made through the local stereo cable FM channel, used for real time speech manipulation at IPO. At first, this seemed the only possibility to make stereo recordings, which was necessary in my case. However, speech output on the FM channel had a low SNR. Just in time, I was allowed to use the first direct stereo output from the VaxLab. Thanks to Ercan!

situations other than pictograms to be interpreted fast. Thus, the detailed silhouette would be used which, stripped of unnecessary details, would still contain enough information to make it easy to produce uniquely recognisable pictures.

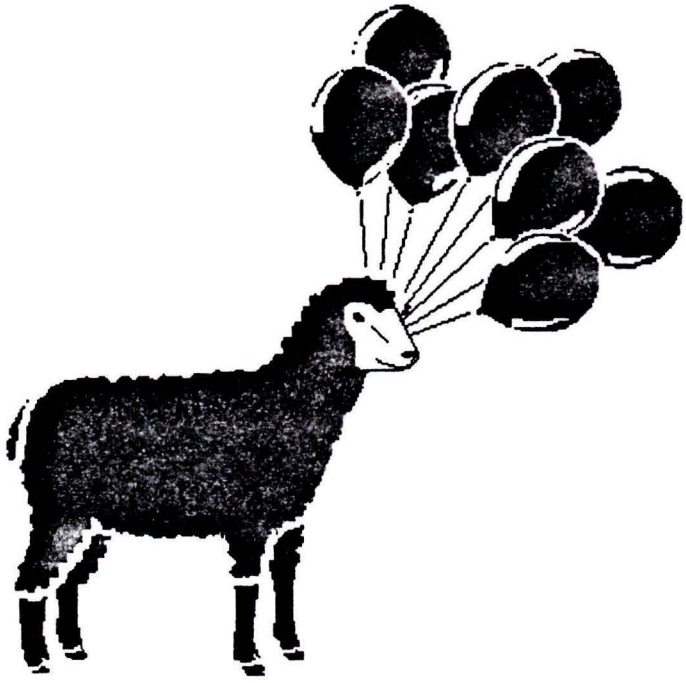

FIG. 1 *Picture of a sheep holding balloons. Example of the type of picture used in the experiment.*

The pictures were monochrome in order to keep the recognition process based on form, not on colour. In this way, the number of variables influencing the image interpretation was reduced, and drawings could be kept simple. Monochrome pictures also had the technical advantage of making up small files. As the pictures had to be put on screen in a very short time, large colour files might have caused a problem.

In the present experiment, in the context picture the 'subject' was on the left side of the screen, and the 'object' on the right side. In the target picture this was mirrored: the subject was on the right side and the object on the left. The purpose of this was that the same picture could be used both as context and, mirrored, as target. In this way, it was possible to avoid introducing a new character in the target picture, that is, a character that had not occurred earlier in the trial. A new character would have caused an unpredictable given/new situation.

2.4.4.1 MAKING THE PICTURES

In most cases the process of picture making started with taking a photograph of the situation to be portrayed in the picture, e.g. a man petting a dog. Where it was impossible to take a photograph of a situation, like a kangaroo holding an umbrella, combinations of existing pictures were used with essentially photographic qualities as a starting point.

The next step consisted of putting the photograph or other image under transfer paper, and tracing its main features with a black pen. Where this was necessary, features were added to the drawing, like the crown of a king.

This enhanced line drawing, with an average size of 10 by 10 cm, was then scanned on a black-and-white Apple flatbed scanner with a resolution of 300 dots/inch. The resulting file was checked and if a line had become discontinuous by scanning it was manually corrected. Thus, a pixel-format (TIFF) file was obtained which could be further manipulated.

The Apple TIFF file was converted to MS-DOS PCX-format, in order to use it as input for CorelTrace. This tracing software converted pixel-images to vector-images, which was necessary to transform each image to its required size, without losing the high resolution.

The traced images were thus sized, combined, mirrored, and otherwise modified in order to mould them into the pictures used eventually. The final files were saved in the appropriate size and density and converted to the IMA format used in the software controlling the experiment (see section 2.5)[6]. After this, they just had to be called by the right name to pop up on the screen.

---

[6] Many thanks to Jan Metselaar at IEC Progamma, Groningen, for providing the 'PACO' (CUT→IMA) conversion program so fast.

2.5 Making it Work: Controlling the Experiment

As has been explained above, a complete trial roughly consisted of first a picture, presented simultaneously with a spoken sentence, succeeded by a short break, and again succeeded by a sentence and a picture. The experiment was very sensitive to timing, as the images were to emerge instantly at a point in time relative to the beginning or end of a sentence, in particular the second (target) picture of each trial. As the spoken sentences were not of equal temporal length, and taking into consideration that the speech signal itself came from tape, it follows that the entire process had to be controlled through the speech signal from tape.

In order to achieve the required precise synchronisation between speech and images all recordings were made in stereo. One channel was used to record synchronisation ('sync') pulses, while the other channel contained the actual speech. The synchronisation signals were used by the controlling software for time measurements and picture presentation. In order to obtain exact time-alignment of speech and sync pulse, for each speech file (sentence) a twin file was produced of the same length, but with zero amplitude. In the essential timing positions sync signals were added on the synchronisation channel. These consisted of a 2000-Hz tone, recorded at 0 dB, and lasting 50 ms, which was paralleled by a 50-ms pause on the speech channel. The complete structure of a trial typically looked like Table III.

TABLE III *Schematic view of the stereo design of a trial.*

Channel	***Action for each channel***						
Sync			1		2		
Speech	silence	$sentence_1$	silence	$sentence_2$	silence	silence	
Screen	$picture_1$		empty	empty	empty	$picture_2$	response recording and time measuring
Pause			<- 1 s ->		<- 0.05 s ->		
Time (s) ->	*0* *0.5*	*2*	*3*	*4.5*	*4.55*	*4.85*	

The exact time structure of a trial thus looks as follows. Triggered by a signal at the end of the preceding trial, $picture_1$ is switched on. After 0.5 s, $sentence_1$ starts. At the end of this sentence, sync pulse 1 indicates that $picture_1$ can be switched off. A 1-second break follows, after which $sentence_2$ starts, at the end of which sync pulse 2 indicates $picture_2$ is to be presented for 32 ms. The response time and time-out period, together 2 seconds, start to count from there, after which a 2-second break follows. During this break a red cross appears on the screen as reaction time feedback when the reaction time exceeds 600 ms. The end of $sentence_2$ also starts a countdown of 4 seconds before $picture_1$ of the next trial is presented.

2.5.1 HARD- AND SOFTWARE

The controlling software (Flits[7]) ran on a PC. This computer, combined with a fast graphics card, was connected to a multi-sync colour screen. The software provided presentation and time measurement control over the entire experiment. It was able to detect the sync pulses and it measured the time (in *milliseconds*) elapsed between the presentation of the second picture and the subject's Yes/No reaction, the verification latency. Subjects gave their actual Yes/No response on a push-button device which was also connected to the computer. The presentation time of the second picture could be varied from 16 ms up, in steps of 16 ms each.

The experiment took place in a sound-proof booth of the Phonetics section in Utrecht. Subjects sat on a chair at a distance of approximately 60 cm from the screen. Speech was presented binaurally over high-quality mono headphones connected directly to a DAT recorder.

---

7 Thanks to Theo V.

2.6 Procedure

During the experiment a record was kept of the actual Yes/No responses and the corresponding reaction times. For this purpose a file was created for each subject, in which, together with each response, a sequential index number, a number from 1-12 indicating the condition, and a number 1-10 indicating the stimulus set was saved.

Subjects were told that, after a short warning beep, they would see a monochrome black and white picture on the monitor and shortly they would hear a sentence over the headphones that would match the picture. This combination would be followed by a short break, in which nothing would be seen or heard. Then another sentence would be presented on their headphones, immediately followed by a very short presentation of another picture.

They were asked to indicate whether the last sentence matched the last picture or not. They could give their response by means of pushing a button marked 'JA' (yes) if they thought they agreed, and 'NEE' (no) if they did not. They were also told that attempted corrections would cause the response to be excluded and that the same applied to pushing two buttons simultaneously. Subjects were instructed to give their response as fast as possible. If they did not react within the maximum time, a red diagonal cross would appear on their screen, together with the message 'TE LANGZAAM!' (too slow). Subjects were free to choose which button to hold left and which hand and finger to use in responding.

They were informed that the total test, plus breaks, lasted approximately 45 minutes, consisting of 3 sessions, The first session of three was always a training session, making subjects familiar with the required response speed and the short presentation period of the second picture. Between the sessions they could take a short break.

2.7 Subjects

From the students' population of Utrecht University 23 subjects were picked. Participants studied various subjects, but students of Phonetics and of Linguistics were excluded, as they might easily discover what the experimenter's goal was, and adapt their answering strategy accordingly. Of the 23 subjects, 8 listened to tape 1, 7 listened to tape 2, and 8 to tape 3.

Besides recording the Yes/No button responses and their corresponding reaction times, comments brought forward at any point during the experiment were also written down.

The observation made by almost every subject was that they had to give their answer too fast to be able to give a satisfactory answer ('If I had had one more second!'). This was confirmed by results from a preceding pilot experiment (see also footnote at page 13) conducted without reaction time feedback, but with otherwise identical circumstances and material. Here, subjects gave responses which were almost 100% correct, with reaction times exceeding 1.5 seconds. Thus, the time limit imposed by the real experiment was reported to be quite frustrating, as one became aware of judgment errors just after the button had been pressed.

Other observations included the difference in the experienced degree of complexity of the pictures.

Results

Chapter 3

A predicted effect of Accent was not found in the analyses on the data. Two other effects were found, as will be revealed in section 3.2, which is entirely dedicated to the analyses. First, however, in 3.1 a few words are said about the nature of the data and the missing data, followed by a short explanation of the arc sine transformation applied to the percentage data.

3.1 The Data

Two types of raw data were available in the analyses, one of which was used as primary data, the other as secondary data. The primary data were the mean percent correct scores for each condition for each of 23 subjects. Reaction times (verification latencies) for each response for each subject were counted only as secondary indications because there were too many factors that could influence them. For instance, the use of reaction time feedback imposed a ceiling on the reaction times. Another factor that might have influenced reaction times was the place of the target

in the sentence. When the target occurred in the early part of the sentence, subjects probably had ample time to process it before the picture appeared, whereas they were probably still busy processing the target when it had occurred late in the sentence, just before the picture appeared. In the results presented here it can be seen that there was indeed little RT variation, but the decision not to use RT as primary data was taken a priori, for the reasons given above.

Within this category of secondary data no distinction was made between RT for correct and RT for incorrect responses, as both correct and incorrect responses were expected and a part of the same design.

3.1.1 MISSING DATA

The data from four subjects who had scored below chance level (i.e. 50% in this research) were omitted from the analyses. Of the remaining 3040 (160 x 19) data, 52 (1.7%) were not recorded for various reasons: either the corresponding reaction time exceeded 2 seconds, or the subject pushed two buttons simultaneously, or the subject tried to correct his first choice.

The mean reaction time over all subjects and conditions was 403 ms, with a standard deviation of 85 ms. Each response with a reaction time that fell outside 3 standard deviations (255 ms) was excluded from the analyses, as it may have been based on reprocessing (> 650 ms, double take) or reacting before any picture had been presented (< 150 ms, premature response). By this selection another 184 responses were excluded (6.2%).

As the aim of the analyses consisted of finding out which factor had an influence on the accuracy of verification, and this can only be determined for cases in which errors can be made, the analyses of each effect were made only for the situations in which the target picture did not completely match the target sentence; either the given constituent or the new constituent was false. If the conditions had been included in which both were true, the influence of a factor on the decision process proper would be determined, instead of the influence of a factor on the accuracy of verification. Thus, although the figures for the condition 'Given True,

New True' (i.e. complete match existed between sentence and picture) can be found in the tables and charts below, they were not included in the actual analyses. This, again, reduced the total number of data to 1383. Under these circumstances, if subjects had simply pushed the 'NO' button, without paying any attention to the actual stimuli, they would have achieved 100% scores. The scores of each subject were checked for a preference of this kind, and admitted to the analyses only after showing no 'NO' bias over **all** data, where a 'NO' bias was defined as a 75% preference for the 'NO' button. This, actually, did not exclude any test subject.

3.1.2 THE ARC SINE TRANSFORMATION

When using an analysis of variance it is assumed that the variances of observations are equal across groups: the *assumption of homogeneity of variance*. With some types of data this assumption is not satisfied. This is the case with percentage data, the primary data of this research. Therefore, a simple transformation had to be applied (Ferguson, 1989, p. 264-267). This transformation, the arc sine transformation, replaces each of the original observations by an angle whose sine is the square root of the original observation. The actual formula to be used was:

$$\textit{arcsin}\ \sqrt{(x/100)},$$

where x stands for a percentage of correct responses.

3.2 Analysis

Univariate analyses of variance were calculated, separately for percentages and reaction times (RT), on the three fixed factors ('Subject Given' *vs* 'Object Given', 'Given False' *vs* 'New False', and 'Given +Accent' *vs* 'Given -Accent'). Two different percentages were calculated for each condition, collapsing either over stimulus sets or over test subjects. The analyses on RT also had to be made collapsing either over

subjects or over stimulus sets; if both subject and stimulus set were included the number of degrees of freedom would be zero. This allowed two separate analyses on all data, including either test subjects or stimulus sets as a random factor. The results of the analyses are given below, starting on the next page, where each effect will be discussed first in terms of primary data, then of secondary data.

TABLE IV *Percentage of correct responses and reaction times for ± Accent, separately for the given/new categories. Each cell is averaged over all subjects and stimulus sets.*

	Correct Responses (%)			Reaction Time (ms)		
	Given -Acc	Given +Acc	*Overall*	Given -Acc	Given +Acc	*Overall*
Given False New True	62.2	59.3	***60.7***	408	403	***405***
Given True New False	66.7	67.6	***67.2***	421	416	***418***
Given True New True	78.6	78.5	***78.6***	368	373	***371***
Overall	***69.0***	***68.4***	***68.8***	***397***	***398***	***398***

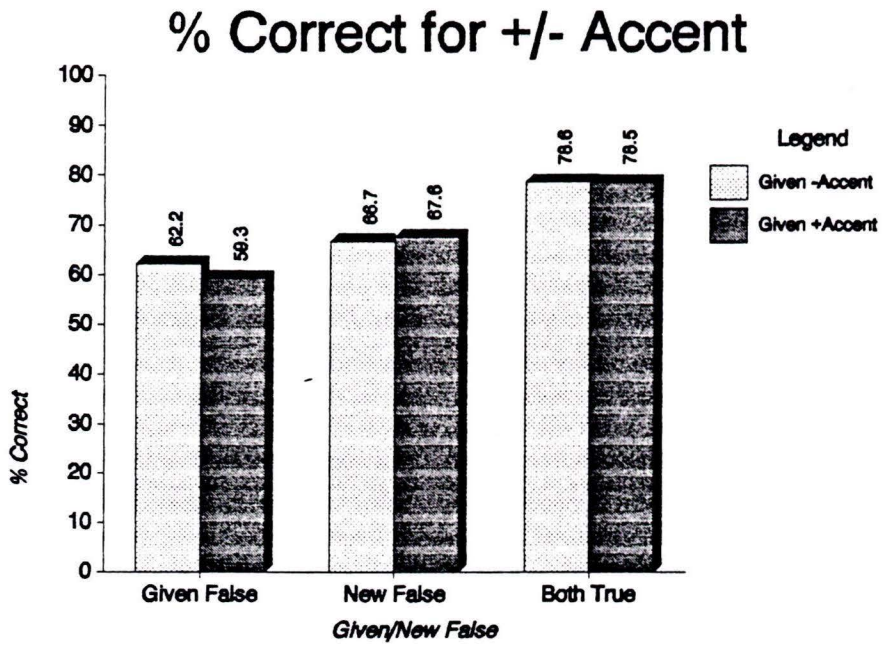

FIG. 2 *Chart of percentage of correct responses for ± Accent, separately for each given/new category. All figures are averaged over subjects and stimulus sets.*

3.2.1 +/- ACCENT

The main effect predicted by the hypothesis, of +Accent inducing higher percentages of correct responses, was not significant for primary (percentage) data: with test

subject as factor, F (1,18) = 0.01, n.s.; with stimulus set, F (1,9) = 0.27, n.s. The percentages are presented on the left side of Table IV and in Fig. 2.

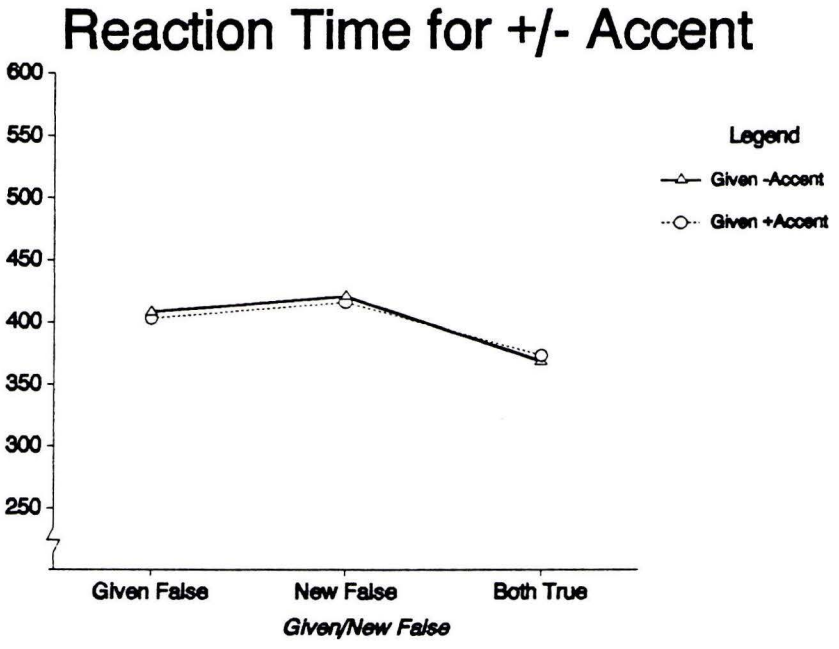

FIG. 3 *Chart of mean reaction times for factor ± Accent, separately for each given/new category. Figures are averaged over all subjects and stimulus sets. Lines have been included to indicate the direction of the effect.*

The same pattern of results was found for the secondary (RT) data: with subject as factor, F (1,18) = 0.74, n.s.; with stimulus set, F (1,9) = 0.35, n.s. This is illustrated in the right part of Table IV and Fig. 3.

3.2.2 GIVEN/NEW

A significant main effect that did play a role was 'Given/new false', as was predicted by the first hypothesis. The percentage correct was higher when the false information to be verified was new than when it was given (with subjects as factor, F (1,18) = 8.02, $p < 0.05$); although with stimulus set as factor the effect is less convincing, F (1,9) = 3.95, $p < 0.1$). The effect is illustrated in Table V and Fig. 4.

Again, the same pattern of results was found for the secondary data. With subjects as factor, F (1,18) = 5.62, $p < 0.05$); with stimulus set, F (1,9) = 3.66, $p < 0.1$). The effect is illustrated in Table V and Fig. 5.

TABLE V *Percentage of correct responses and reaction times for the Given/new effect, separately for subject and object as given constituent. Each cell is averaged over all subjects and stimulus sets.*

	Correct Responses (%)			Reaction Time (ms)		
	Subj given Obj new	Obj given Subj new	*Overall*	Subj given Obj new	Obj given Subj new	*Overall*
Given false New true	63.3	58.0	***60.6***	408	403	***405***
Given true New false	61.0	73.2	***67.1***	418	418	***418***
Given true New true	82.8	74.2	***78.5***	367	374	***370***
Overall	***69.0***	***68.4***	***68.8***	***398***	***399***	***397***

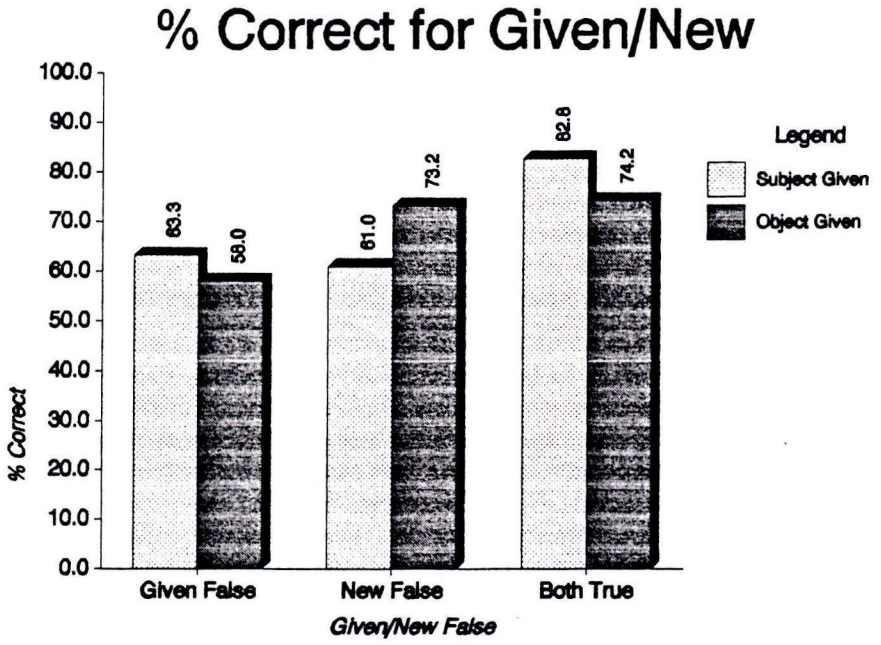

FIG. 4 *Percentage of correct responses for 'Given/new false' effect, specified for Subject and Object as given constituent. Percentages have been averaged over all subjects and stimulus sets.*

3.2.3 SUBJECT/OBJECT

No significant effect could be found for Object or Subject as the constituent containing the given information in the target sentence: with test subjects as factor,

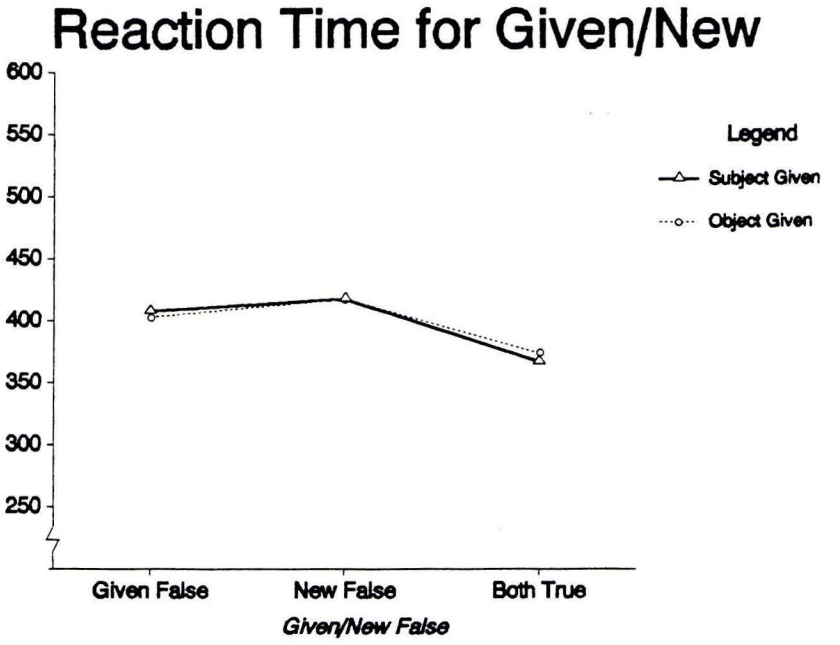

FIG. 5 *Mean reaction times for 'Given/new' effect, specified for Subject and Object as given constituent. Times have been averaged over all subjects and stimulus sets. Lines are drawn between points to indicate direction of the effect.*

F (1,18) = 2.17, $p > 0.1$; with stimulus set, F (1,9) = 1.4, $p > 0.1$). In Table V we see that if the percentages are averaged over the 'given false/new false' categories there is no effect of subject or object as given constituent, although the percentages differ for each individual cell.

The same absence of a main effect by this factor was found for secondary data: with subjects as factor, F (1,18) = 0.13, n.s.; with stimulus set, F (1,9) = 0.0, n.s.

3.2.4 INTERACTION BETWEEN SUBJECT/OBJECT AND GIVEN/NEW

As we have seen, 'Subject/Object Given' had no significant effect, while the only significant effect that was found so far was due to the Given/new factor. These two factors together produced the only significant interaction effect: for percentage, with test subjects as factor, F (1,18) = 7.45, $p < 0.05$, with stimulus set, F (1,9) = 3.18, $p < 0.1$). This effect is illustrated by Table V and Fig. 4.

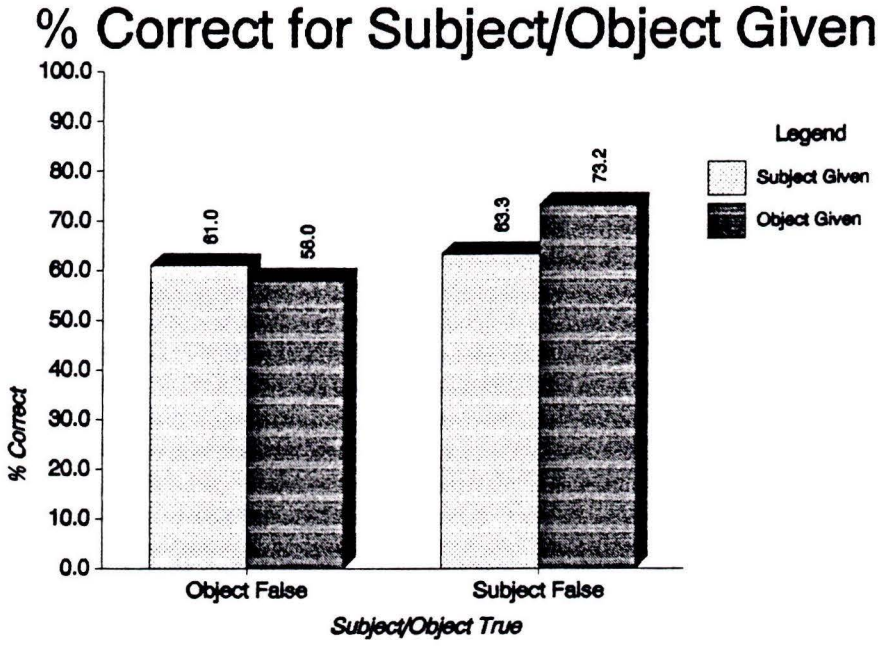

FIG. 6 *The same numbers as in Fig. 4, classified according to different categories. Percentages are averages over all subjects and stimulus sets.*

However, as can be seen on the right side of Table V and in Fig. 5, the interaction was not significant for secondary data: with test subjects included, F (1,18) = 0.52, n.s.; with stimulus sets, F (1,9) = 0.24, n.s.

Fig. 6, which is actually just a different way of graphing the numbers of Fig. 4, leads to a possible interpretation of the interaction between 'given/new false' and 'subject/object given'. According to this interpretation, if anything was pictured falsely, test subjects made less errors when the false part was the Subject (in Fig. 6: first pair of bars *vs* second pair of bars). Within these conditions we can see that verification is easier if it is the actual new constituent in the target sentence that is pictured falsely (Fig. 6: of the first two pairs of bars, the outer *vs* inner bars).

No further interactions were found, which indicates that the only significant forces at work were the ones mentioned above. It also shows that the effect of Accent, not found as a main effect, was not hidden in one of these interactions. The interactions, for which the analyses are given below, will not be included in the further discussion:

- the interaction between 'Subject/Object given' and 'Given ± Accent' (primary data: with test subjects as factor, $F(1,18) = 0.35$, n.s.; with stimulus set $F(1,9) = 0.77$, n.s.), (secondary data: with test subjects, $F(1,18) = 1.20$, $p > 0.1$; with stimulus set $F(1,9) = 1.81$, $p > 0.1$),
- the interaction between 'Given/new false' and 'Given ± Accent' (primary data: with test subjects, $F(1,18) = 0.05$, n.s.; with stimulus set, $F(1,9) = 0.67$, n.s.), (secondary data: with test subjects, $F(1,18) = 0.0$, n.s.; with stimulus set, $F(1,9) = 0.0$, n.s.), and
- the three-way interaction between all factors 'Subject/Object Given', 'Given/new false', and 'Given ± Accent' (primary data: with test subjects $F(1,18) = 0.01$, n.s.; with stimulus set $F(1,9) = 0.0$, n.s.), (secondary data: with test subjects $F(1,18) = 1.01$, $p > 0.1$; with stimulus set $F(1,9) = 0.57$, n.s.).

Summarising, it was found that Accent had no effect at all. It was also found that Given/new did have its predicted effect. Finally, it was found that an interaction existed between the factors 'Subject/Object given' and 'Given/new false', which was interpreted as an indication that false information is better verified if it is contained in the subject.

3.3 Survey of Data

In Appendix B, starting on p. 49, a few additional tables and figures can be found of the mean percentage of correct responses and reaction times.

(a) For each subject over all conditions and stimulus sets: Table VI on p. 49, Fig. 7 on p. 51, and Fig. 8 on p. 51.

(b) For each stimulus set over all subjects and conditions: Table VII on p. 49, Fig. 9 on p. 52, and Fig. 10 on p. 52.

(c) For each condition over all subjects and stimulus sets (see Table VIII on p. 50, Fig. 11 on p. 53, and Fig. 12 on p. 53 for surveys).

The tables under (a) and (b) indicate substantial differences between results for subjects and small differences between stimulus sets. This gives an explanation of why results differed when the data were collapsed over subjects (considerable differences were averaged) from when they were collapsed over stimulus sets (averaging over small differences). Collapsing over stimulus set is consequently more reliable which is reflected in the results of analyses presented in the preceding sections.

Finally, under (c) a table and 2 charts can be found which give a survey of the figures for each condition used in the analyses.

Discussion

Chapter 4

The main hypothesis on the influence of Accent was not confirmed by the data. The consequences of this result will be discussed in 4.2. Before that, in the sections of 4.1 a few reservations are made for possible flaws, both within the experiment (in 4.1.1) and concerning the experiment itself (in 4.1.2). First, in the current section the effects will be discussed that did render significant results.

The hypotheses proposed in Chapter 1 consisted of two parts: a baseline, inspired by results from earlier experiments, and the main hypothesis. The baseline expectation was that, in any case, in the verification of given information much more errors would be made than for new information (Hornby, 1974):

If a word is new it can be verified more easily than if it is given.

In addition to this duplicating hypothesis the main hypothesis was formulated:

If a given word is realised with +Accent it can be verified more easily than if it is not realised with +Accent.

The baseline hypothesis is confirmed in the results on primary data - percentage of correct responses - for given and for new information. Subjects made less errors for new information which did not match with the picture than for given information (see Fig. 4 on p. 31). In the analysis of the secondary data - reaction times - the effect found to be significant was a small increase of RT when the new information was pictured incorrectly (see Fig. 5 on p. 32). Since this concerns secondary data it is not of great interest. Thus, (part of) the results found in the literature were confirmed, in particular the observation that the 'actually new' status of a word causes it to be considered more carefully (e.g. Hornby, 1974).

However, the main effect predicted to come with the Given/new effect was not found. Neither higher percentages of correct responses (see Fig. 2 on p. 29), nor different reaction times (see Fig. 3 on p. 30) were found in the conditions in which the given information was realised with +Accent, as compared to the conditions in which given information was realised with the 'normal' -Accent. The implications of these findings will be discussed in 4.2.

One more effect was found in the analyses on the primary results: the interaction between 'Subject/Object false' and 'Given/new false'. If this interaction is interpreted along the lines set out in Chapter 3, namely as a subject/object **false** main effect, its effect can be explained. The data are, then, interpreted as showing higher percentages of correct responses if the false information was contained in the grammatical subject than if it was in the object (see also Fig. 6 on p. 33). This can be explained by a time advantage, due to place in the sentence. The subject was always the first constituent in the sentence in the present experiment. Consequently, when hearing the sentence, test subjects first heard the subject and had ample time to process the information it contained before the picture was presented. The object, on the other hand, would not appear until the end of the sentence, after which the picture immediately followed. The (hypothetical) claim made is that, if subject and object had alternately taken the first and last place of the sentence, the interaction would not have occurred.

4.1 Drawbacks

4.1.1 Within the Present Experiment

In this section I would like to draw attention to some more or less serious drawbacks of elements within the present experimental design, which may be responsible for part of the results. Suggestions will be given for possible improvements of the design.

The first problem is that, at no point during the preparation, a calibration of the materials took place, neither for pictures, nor for accents, nor for sentences. For the *pictures*, this would have implied a comparison of the relative degree of difficulty of each component of the picture. As several test subjects indicated, there was a clear difference between the 'difficulty' of the figures represented as well as of their combinations. The difference between the combinations is related to the difference in difficulty of interpreting the situations, or *sentences*. A subjective calibration of both pictures and sentence contents might have improved the strength of the design.

As for *accents*, no subjective calibration took place of the perceived prominence of accents between sentences, nor specified for accent on grammatical subject and object within each sentence, because the acknowledged Dutch Intonation Grammar was used to produce the patterns. Thus, even though I personally perceived a difference between +Accent for subject and +Accent for object, no modifications were applied. If more time had been available calibration would have started with measuring the relative subjective strength; possible modifications would be to give stronger accents to grammatical subjects, to reverse the observed trend, or to apply individual adaptations to each sentence. Test subjects did not spontaneously report differences in accent, but then, they were not consciously focusing on such differences.

In addition to the absence of calibration of the individual components of the pictures, the exact effect of mirroring pictures in the target situations is undetermined as yet. It may well be that mirroring prevented subjects from using a directive

strategy on the basis of Accent. Pictures were mirrored to prevent a subject-object effect in the pictures, in that the same picture was used as context and - mirrored - as target. Subjects reported the mirroring as very confusing.

A problem that is more linked to the statistical analysis, is that the factor 'stimulus set' cannot be unambiguously classified as either fixed or random. The sentences used in this experiment are in principle constructed from open class elements. However, due to the manifold semantic and pragmatic restrictions applied in selecting the eventually used sentences, the number of alternative sentences was very low. This makes the set of sentences something in between fixed and random. Test subjects could expect only a small number of different situations and elements. This may, in turn, have caused subjects to use an 'abnormal' verification strategy adapted to this limited situation. A choice for a less restricted type of sentence than SVO might have solved this problem, although it would have caused other problems (what other type, how many elements on the *given* side and how many on the *new*, etc.). Another solution to this problem might have been to use more different 'actors', although that would have prevented the use of dichotomies, like cat and dog. Under the present circumstances, results may not reflect effects that can be generalised to all types of sentence material.

Using the, relatively new, PSOLA technique for manipulation of the intonation involved a certain risk, especially under time pressure. Although the speaker used for the production of the speech material had been pre-tested, the eventual results after manipulation were not overall satisfactory. As PSOLA results are known to depend highly on the speaker's voice, all material was spoken for a second time by the same speaker, and all the required manipulations were performed on the new material as well. As results did not improve, however, the old material was eventually used in the experiment. The only way to possibly improve speech quality would have involved manual placing of pitch markers, which would have required another two to three weeks that would not fit into the research planning.

A final remark concerning the present experimental design goes to the answering device used: the Yes/No push button box. Most subjects indicated that it was easier to push one specific button, either left or right, depending on personal preference. They also reported perceiving a distinct time lag between the decided

reaction in their head and the translation to the push on the button, which never became an automatic process. A possible alternative means of answer recording would have been a Voice Key, which allows people to give oral answers like 'Yes' or 'No'. This device reacts to every sound above a certain threshold. Combined with the actual response, which would be recorded on tape, one would get the same kind of answers, but from a much more natural source of answering (for most people). A serious drawback of the Voice Key is, of course, the insolvable problem of calibrating the threshold amplitude to which the device reacts.

4.1.2 The Experimental Design

Drawbacks of elements within the present experiment have been accounted for in the previous section. At this point, I would like to propose a few objections to the actual experimental design as an instrument to test the hypothesis. This may be seen more or less as a brainstorming afterthought to what I have done, just as the remarks made in the previous section, but it is mainly intended as a suggestion for improving subsequent research in this area. The weakness of the present paradigm in testing the hypothesis is pointed out and suggestions are given for possibly better designs.

An on-line verification experiment on pictures presented only some 30 ms may have been the wrong choice of paradigm to test the present hypothesis: it focuses the test subject too much on checking what is correct and what is not, which does not reflect an everyday situation; during the experiment a test subject may dedicate an unnatural amount of attention to verification. This effect, reinforced by the very short picture presentation time, may have caused subjects to use a strategy tailored to cope with this special situation, which does not necessarily involve ± Accent as a cue. The reasoning underlying the choice of a short presentation in the first place was that subjects would then have to use all available cues, including ± Accent.

A plausible alternative for the present experiment could have been a Question-Answering test, where a picture is presented and the subjects are asked what they saw. One version could include a preceding sentence containing a

+Accented word, the other version the same word without +Accent. I would expect answers to more reliably reflect the use of Accent in such a 'normal' situation.

Another option could have consisted of a fake test of a different, non-verification, nature combined with post-test correctness evaluation, for instance a phoneme monitoring test combined with post-test scoring (see Cutler and Fodor, 1979). This might have rendered better results as, in retrospect, verification may not be the ideal primary task if 'abnormal' scoring strategies are to be prevented. The disadvantage of a fake test serving as cover for the real test is that, in particular in the case of phoneme monitoring, post test scoring indeed allows no more than a post-hoc analysis. It is hard to imagine an on-line correctness test combined with another test, where attention is not primarily focused on the correctness verification.

The scope of the conclusions, on the basis of the present results, is in any case limited by the very short presentation time and the requested immediate response by test subjects. The results allow only conclusions on the effects of the on-line use of Accent in verification. This does not mean that the conclusions drawn from the results from this experiment are incorrect in themselves.

4.2 Conclusion on the Present Results

In this section the implications of the experimental findings will be discussed. In the previous section a few reservations were made concerning the validity of conclusions drawn on the basis of the present results. In this respect it was pointed out that, in any case, the conclusions to be drawn can only concern the on-line use of Accent in verification, and some reservations have to be made about generalising results from the sentence material used here to sentences in general.

In spite of the reservations made, there are still good reasons to believe in the validity of the experiment. The fact that the new/given effect itself was found convincingly may prove that the experiment was apt to test the hypothesis proposed in Chapter 1. Thus, the conclusion still holds that the main hypothesis is contradicted by the results: *contents* are more important than *form*.

In the rest of this section the present absence of an Accent effect is assumed to reflect the fact that there is no effect of Accent in the process of on-line verification in general. On the other hand, the observed given/new effect indicates that given/new does influence this process. The implication of these findings is that, at the level at which verification decisions are made, given/new apparently is a relevant distinction, while ± Accent is not. Consequently, given/new and ± Accent are to be considered distinctive categories at different levels in speech perception. In order to give a possible explanation of this observation, a few findings from previous research have to be taken into account.

Investigating on the word level, Terken & Nooteboom (1987) found that, measuring reaction times in a sentence verification experiment, given information was recognised faster if it was realised with - Accent, while new information was recognised faster if it was realised with +Accent.

On the phoneme level, Cutler and Fodor (1979), conducting a phoneme monitoring experiment, found that reaction times to a phoneme target in a particular word were faster when that word was accented than when it was not.

Also on the phoneme level, Van Donselaar (1990) used a mispronunciation detection task to record reaction times and detection miss rates depending on the position of the accented syllable. She found that new information produced faster reactions if the information was accented, whereas it made no difference whether or not given information was accented or not.

Taking these findings together, an interpretation can be used that was also proposed by Van Donselaar. According to her hypothesis the search for sentence accent and the search for (higher order) semantic information are two different strategies that work in parallel to achieve the same goal. When the two sources are conflicting, the higher order information is stronger than accent information. Assuming this hypothesis to be true, the present results could be explained as indicating that, as higher order (lexical) information is readily available, conflicting accent information does not have an influence on verification. Alternatively, the present and previous results could be taken to signify that accent has influence only up to a certain level: the word recognition level. Beyond that level, *only* higher order information is available. This explanation differs from the first one only in strictly

limiting the domain of influence of accent; the 'parallel' explanation interprets results by saying that accent information is *weaker* than higher order information at the verification level; the other explains it by saying that accent information is *absent* at that level. Neither explanation can be ruled out by the present observations, either apart from or in combination with observation from previous experiments. This is mainly the result of the widely varying materials and paradigms, adopted to test hypotheses from equally varying perspectives in order to be able to develop a credible theory on accent and given/new. A major goal of future research will have to be to either provide a comprehensive series of experiments covering all levels using similar materials and operationalisations, or to use experiments of which results can be compared directly to those produced from previous research.

4.3 Summary

In conclusion, I can say that given/new is not the same as ± Accent in verification; the main hypothesis has proved unable to correctly predict results. Although part of the explanation for the results may be found in drawbacks of the experiment, the larger part is due to real effects. While accent apparently does not play a role in the short term verification of a sentence it is likely to play a role at the recognition level. The experimental design, though perhaps flawed in some respects, was a powerful testing tool that may have needed some fine-tuning. Future research on given/new and ± accent should seek to link up with previous research in order to allow an eventual comprehensive analysis.

References

Baart, J.L.G. (1987). *Focus, Syntax, and Accent Placement*, Doctoral Dissertation, University of Leiden.

Chafe, W.L. (1974). Language and Consciousness, *Language*, 50, Nr. 1, 111-133

Clark, H.H. and Clark, E.V. (1977). *Psychology and language*, Harcourt Brace Jovanovich, N.Y.

Clark, H.H. and Haviland, S.E. (1974). Psychological processes as linguistic explanation, in: D. Cohen (Ed.), *Explaining linguistic phenomena*, Washington D.C., Hemisphere Publishing, 91-124.

Cutler, A. and Fodor, J. (1979). Semantic focus and sentence comprehension, *Cognition*, 7, 49-59.

Donselaar, W. van (1990). The use of prosodic cues in speech recognition, *OTS Yearbook 1990*, University of Utrecht, 1-12.

Ferguson, G.A. and Takane, Y. (1989). *Statistical analysis in psychology and education*, 6th edition, McGraw-Hill, N.Y., 264-267.

Halliday, M.A.K. (1967). Notes on transitivity and theme in English, *Journal of Language*, 3, 199-244.

Hamon, C., Moulines, E., and Charpentier, F. (1989). A diphone synthesis system based on time-domain prosodic modifications of speech, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. ICASSP-89*, 238-241.

Hart, J. 't and Cohen, A. (1973). Intonation by rule, a perceptual quest. *Journal of Phonetics*, 1, 309-327.

Hart, J. 't and Collier, R. (1975). Integrating different levels of intonation analysis, *Journal of Phonetics*, 3, 235-255.

Hornby, P.A. (1971). Surface structure and topic-comment distinction: a developmental study, *Child Development*, 42, 1975-1988.

Hornby, P.A. (1974). Surface structure and presupposition, *J. of Verbal Learning and Behavior*, 13, 530-538.

Keenan, E.L. (1976). Towards a universal definition of "subject", in: C.N. Li (Ed.), *Subject and Topic*, Academic Press, N.Y., 303-333.

Kuno, S. (1971). The position of locatives in existential sentences, *Linguistic Inquiry*, 2, 333-378.

Terken, J.M.B. and Nooteboom, S.G. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for Given and New information, *Language and Cognitive Processes*(2), 3/4, 145-163.

Appendix A
Sentence Material

Sentence material used in the experiment. The first sentence of each stimulus set is the context sentence, the other two were the possible target sentences.

Stimulus set	*Dutch sentence*	*English translation*
1	De man aait de hond	('The man is petting the dog')
	De man aait de kat	('The man is petting the cat')
	De vrouw aait de hond	('The woman is petting the dog')
2	Het meisje schopt het boek	('The girl kicks the book')
	Het meisje schopt de bal	('The girl kicks the ball')
	De jongen schopt het boek	('The boy kicks the book')
3	De agent stopt de fiets	('The constable stops the bicycle')
	De agent stopt de auto	('The constable stops the car')
	De soldaat stopt de fiets	('The soldier stops the bicycle')
4	De draak bedreigt de ridder	('The dragon endangers the knight')
	De draak bedreigt de jonkvrouw	('The dragon endangers the lady')
	De leeuw bedreigt de ridder	('The lion endangers the knight')
5	De koning gooit de speer	('The king throws the spear')
	De koning gooit de stoel	('The king throws the chair')
	De kok gooit de speer	('The cook throws the spear')

6 De motor raakt de molen .... ('The motorcycle hits the windmill')
De motor raakt het huis ..... ('The motorcycle hits the house')
De tank raakt de molen ..... ('The tank hits the windmill')

7 De koe opent het hek ..... ('The cow opens the gate')
De koe opent de deur ..... ('The cow opens the door')
Het paard opent het hek ..... ('The horse opens the gate')

8 Het meisje pakt de doos ..... ('The girl picks up the box')
Het meisje pakt de tas ..... ('The girl picks up the bag')
De man pakt de doos ..... ('The man picks up the box')

9 De vogel eet de maïskolf ..... ('The bird is eating the corn')
De vogel eet het brood ..... ('The bird is eating the bread')
Het varken eet de maïskolf ..... ('The pig is eating the corn')

10 Het schaap heeft de ballonnen ..... ('The sheep has the balloons')
Het schaap heeft een paraplu ..... ('The sheep has an umbrella')
De kangaroe heeft de ballonnen .. ('The kangaroo has the balloons')

Appendix B
Additional Tables and Figures

Surveys of percentage of correct responses and mean reaction times

(a) for each subject over all conditions and stimulus sets: Table VI on p. 49, Fig. 7 on p. 51, and Fig. 8 on p. 51,

(b) for each stimulus set over all subjects and conditions: Table VII on p. 49, Fig. 9 on p. 52, and Fig. 10 on p. 52, and

(c) for each condition over all subjects and stimulus sets (see Table VIII on p. 50, Fig. 11 on p. 53, and Fig. 12 on p. 53 for surveys).

TABLE VI *Percentage of correct responses and reaction time (ms) for each subject, averaged over all stimulus sets and conditions.*

Subject	% Correct	RT
1	62.7	380
2	87.9	498
3	66.3	413
4	67.3	369
5	61.4	351
6	64.3	346
7	85.7	394
8	65.8	469
9	59.3	367
10	77.4	413
11	53.0	268
12	71.3	399
13	71.6	445
14	53.4	432
15	59.3	343
16	84.6	426
17	64.2	378
18	73.7	429
19	77.5	446
Overall	*68.8*	*397*

TABLE VII *Percentage of correct responses and reaction time (ms) for each stimulus set, averaged over all subjects and conditions.*

Stimulus	% Correct	RT
1	59.2	391
2	75.5	394
3	61.1	396
4	64.7	410
5	70.2	399
6	66.2	412
7	77.5	403
8	63.8	389
9	77.5	390
10	72.8	382
Overall	*68.8*	*397*

TABLE VIII *Percentage of correct responses and reaction time for each condition, averaged over subjects and stimulus sets, specified for role. Index is in the upper left corner.*

		Correct Responses (%)		Reaction Time (ms)	
		Subj Given Obj New	Obj Given Subj New	Subj Given Obj New	Obj Given Subj New
Given False New True	Given -Acc New +Acc	1 63.2	2 61.1	1 411	2 405
	Given +Acc New +Acc	3 63.5	4 54.9	3 406	4 401
New False Given True	Given -Acc New +Acc	5 59.6	6 73.6	5 426	6 416
	Given +Acc New +Acc	7 62.4	8 72.8	7 411	8 421
Given True New True	Given -Acc New +Acc	9 84.4	10 72.8	9 361	10 375
	Given +Acc New +Acc	11 81.3	12 75.7	11 374	12 373

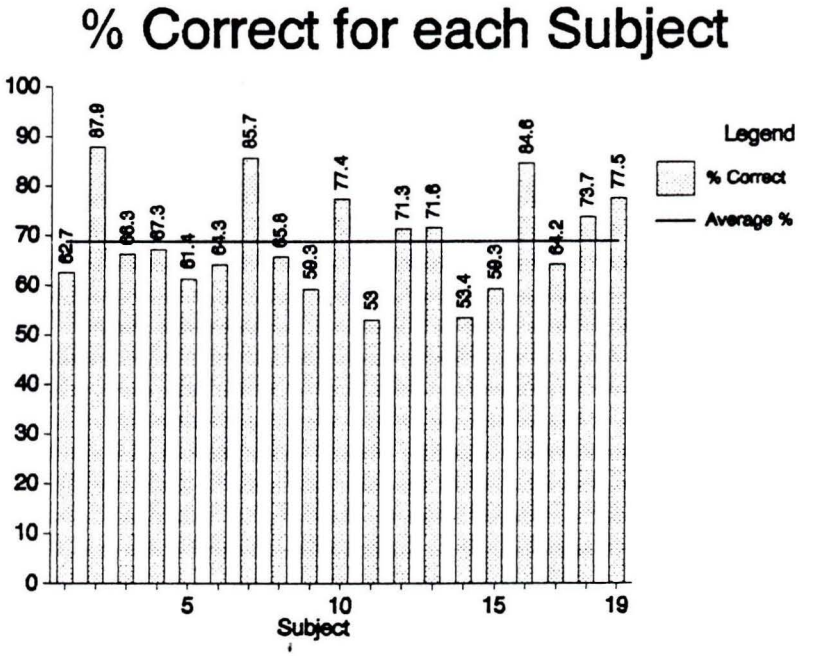

FIG. 7 *Percentage of correct responses (%) for each subject averaged over conditions and stimulus sets.*

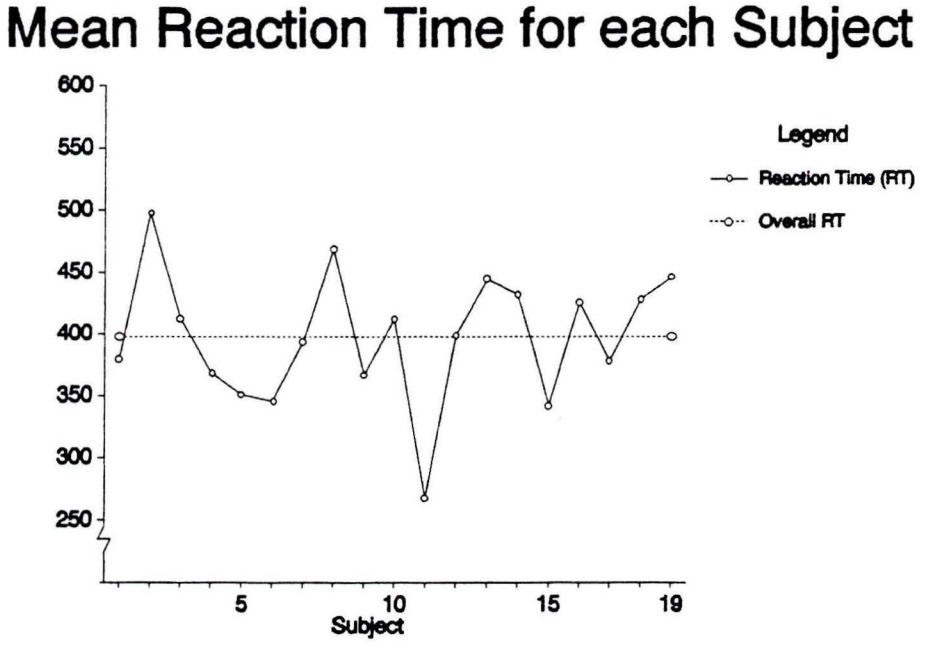

FIG. 8 *Reaction time (RT) in ms for each subject, averaged over conditions and stimulus sets.*

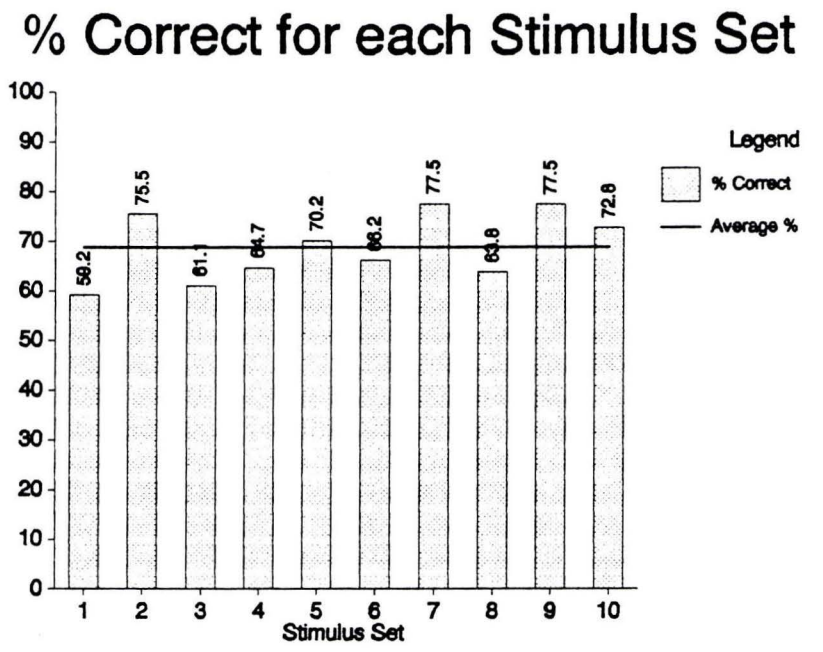

FIG. 9 *Percentage of correct responses (%) for each stimulus set, averaged over subjects and conditions.*

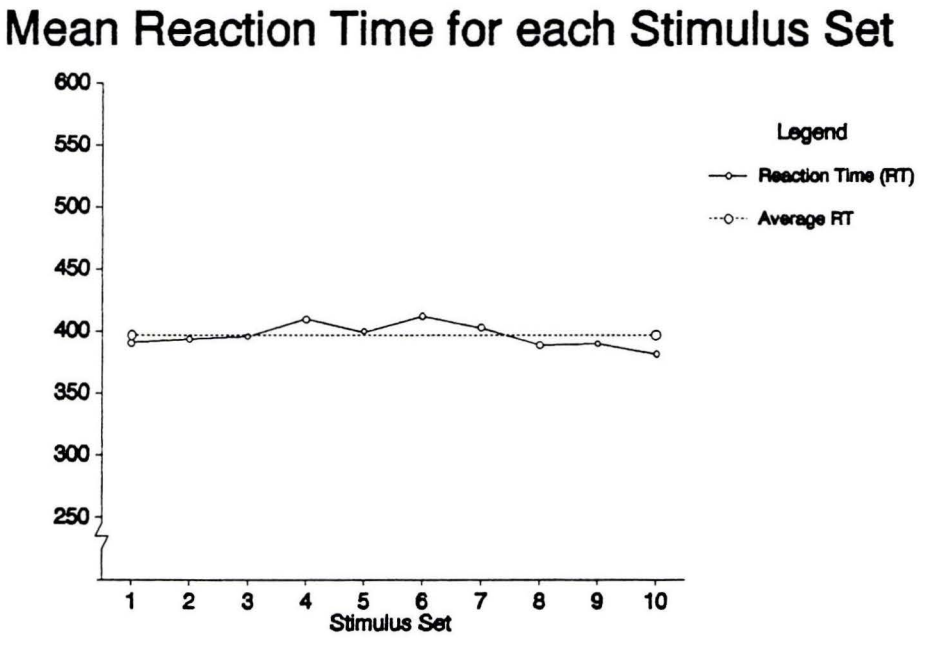

FIG. 10 *Reaction time (RT) in ms for each stimulus set, averaged over subjects and conditions.*

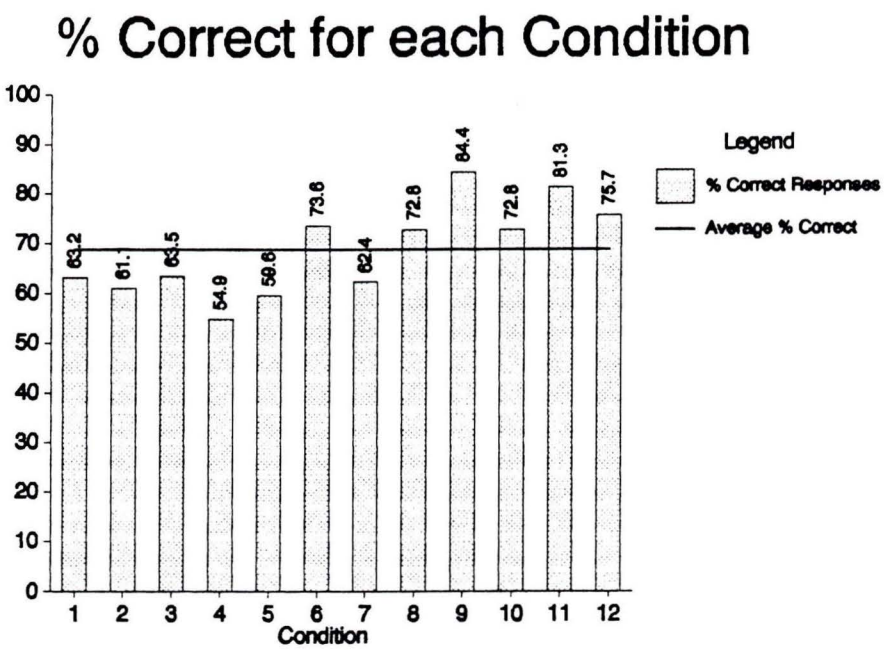

FIG. 11 *Percentage of correct responses (%) for each condition, averaged over subjects and stimulus sets.*

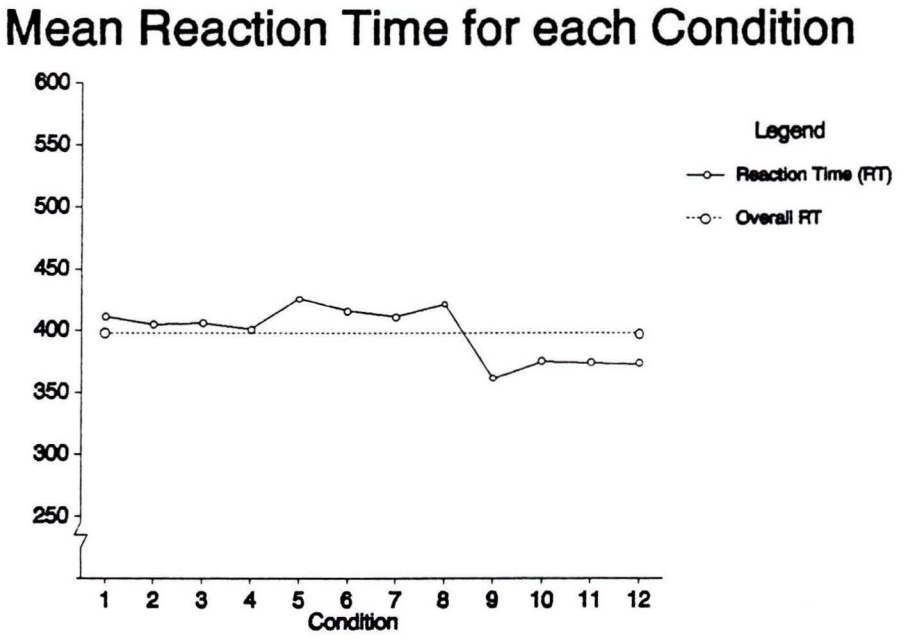

FIG. 12 *Reaction time in ms for each condition, averaged over subjects and stimulus sets.*

Mailing list

Ten Bosch
Collier
Custers
Dirksen
Fransen, L. Nieuwstraat 21 bis, 3512 PB Utrecht (2 ex.)
Geluykens
Hermes
Meerding
Mozziconacci
Nooteboom
Odé
De Pijper
Rump
Sanderman
Sanders
Spaai
Swerts
Terken
Vroomen

Dr. H. Quené, Fonetisch Instituut RUU, Trans 10, 3512 JK Utrecht

Bouma
A.C. der Kinderen, Bibl. en Inf. Nat.Lab., WY 1.36 (8 ex.)
Dr.ir. W. Strijland, Octr. en Merken, WAH

8 archief