Implicit and incidental second language learning: experiments in the processing of natural and partly artificial input
Hulstijn, J.H.

Published in:
Interlingual processes

Citation for published version (APA):
Implicit and Incidental Second Language Learning: Experiments in the Processing of Natural and Partly Artificial Input

Jan Hulstijn

Introduction

Young children are said colloquially to “pick up” their mother tongue (L1). What does this expression mean? In this paper, picking up a language is considered an implicit and incidental process. First, L1 acquisition takes place implicitly. The child’s caretakers do not teach the rules of the grammar underlying the language, nor does the child consciously try to induce these rules on his own. Second, L1 acquisition takes place incidentally. Young children do not have an intention to learn the language. It is true that they may intend to understand what caretakers and peers say to them, and that they may have an intention to express their feelings, thoughts, wishes, etc. verbally. But they lack the intention of trying to remember what they hear.

It is not only first language acquisition that takes place implicitly and incidentally. Second language (L2) acquisition may also have these characteristics. Many adults, living in a foreign community, may “pick up” a fair amount of an L2 without following language courses, or even consulting dictionaries (implicit learning). They may not even have the intention to learn the other language, and yet gain some proficiency in it, just from hearing it being used in their environment (incidental learning).

Implicit learning, although not involving reflection on grammar rules, still requires that the learner pays attention to the formal features of the language. This may seem fairly obvious for the acquisition of content words: The learner needs to pay attention to both the meaning and the phonological form of a word, separately and combined. But it may be less obvious (especially to those who advocate a “natural” approach to L2 learning and teaching) that attention to form is also required for the acquisition and skillful application of function words and morphological and syntactic rules (Hulstijn & Hulstijn, 1984; Hulstijn, 1986).

An example may illustrate this. Assume that a learner has learned the content words flower, girl, pick and red, but has not learned any function words or any syntactic rules. This learner could easily guess the meaning of the sentence The flower the girl picked was red (Clark & Clark, 1977, p. 73). That is, this sentence can be considered to contain “comprehensible input” (Krashen & Terrell, 1983), in that the learner can successfully apply a semantic, conceptually driven, top-down strategy in processing the sentence. The question remains, however: Why would learners bother to pay attention to formal properties (function words, syntax) of such a sentence, if they can get away with a semantic strategy? Perhaps input sentences need not be too
comprehensible! Perhaps they must instead offer the learner a small problem to solve, just to make sure that the formal L2 properties will receive the attention, minimally required for their implicit acquisition to take place. This is the central issue underlying the research to be reported here: For the implicit and incidental acquisition of formal properties of a language to take place, is it sufficient for learners to pay attention to the meaning of the linguistic materials they hear or read, or is it necessary to pay attention to their form as well?

It is suggested that this research question should be investigated within the following framework, an expanded and adapted version of Jenkins' (1979) tetrahedral model of memory experiments. (Jenkins’ model has been borrowed by psychologists, such as Bransford (1979), Eysenck (1982), and especially Brown (1982); see also Brown, Bransford, Ferrara, & Campione (1983).) The adapted framework comprises five factors:

1. Processing mode: the way in which the linguistic input is processed by the learner,
2. the learner's current L2 knowledge,
3. linguistic characteristics of the grammatical features to be learned: target structures,
4. number and frequency with which the target structures appear in the input,
5. compatibility between learning and retention tasks.

It is important to note here the cognitive nature of this model. It categorizes the potential interrelations among variables pertaining to L2 learning as a cognitive phenomenon, leaving aside, for methodological reasons, L2 learning as an emotional and social event (Brown, Bransford, Ferrara, & Campione, 1983, pp. 147-150).

A Cognitive Framework for the Study of Second Language Acquisition

For a better understanding of the framework and the design of the present experiments, take, as an example, intermediate L2 learners who have acquired a fair number of words, and perhaps also a few basic rules of morphology and syntax — as in the “flower picking” example mentioned earlier. These learners are presented with L2 sentences containing instances of some grammar rules, which they haven’t acquired as yet. These grammar rules will be here referred to as target rules, and the sentences in which they appear, as target sentences. The five factors of the framework will be clarified in succession.

1. Processing Mode

According to Craik and Lockhart (1972), what is critical to learning, are the mental encoding activities during study (processing), rather than rehearsal or the intention to learn. Since 1972, many hypotheses have been proposed and tested concerning the retention effect of the various ways in which
Implicit and Incidental SLA

words, sentences and texts can be processed. (For an overview, see Zechmeister & Nyberg, 1982; Eysenck, 1982; Franks, Bransford, & Auble, 1982.) Notions such as depth of processing, elaboration, congruity of encoding, transfer-appropriate processing, cognitive effort, distinctiveness of encoding, etc. have been proposed in rapid succession during the last fifteen years. There are two characteristics that are shared by these various approaches: (1) They emphasize the qualitative (type of processing) rather than the quantitative (duration, rehearsal, etc.) aspects of learning; and (2) they consequently concentrate on incidental rather than on intentional learning, since “memory performance is determined far more by the nature of the processing activities engaged in by the learner than it is by the intention to learn per se” (Eysenck, 1982, p. 203).

2. The Learner’s Current L2 Knowledge

The most relevant learner variable in a cognitive study of implicit and incidental grammar learning, is obviously prior L2 knowledge. The extent to which new grammar rules can be acquired depends to a large extent on the number and type of words and rules already acquired, on the routes by which this knowledge base can be reached, as well as on the ease with which it can be reached.

3. Linguistic Characteristics of the Targets

One of the most important factors in implicit and incidental grammar rule learning is the linguistic complexity and saliency of the target structures to be acquired. In English, for instance, some form distinctions constitute the markers of relatively clear and easily accessible concepts, for example singular versus plural suffixes, present versus past tense markers, and declarative versus interrogative word order. Other form distinctions, however, express fairly complicated concepts, such as definite versus indefinite articles, and simple past versus present perfective tense markers. Some other distinctions express subtle differences in the presentation of information, for example cleft versus non-cleft word order, and passive versus active mode. Finally, there is a great number of morphological and syntactic rules, the functions of which are mainly formal and hardly ever semantic. Examples of this last category are: that as a formal marker of clause subordination, there as a formal marker of existential sentences, basic word order patterns such as SOV and DET ADJ N, and (e.g., in German) suffixes of articles and adjectives reflecting the grammatical gender of the head noun. One may hypothesize that grammatical structures (targets) which differ in linguistic complexity along these lines, may require different amounts and types of attention in order to be acquired. Hence, interactions are to be expected between the targets’ linguistic characteristics and the modes in which the targets can be processed.

4. Target Number and Frequency

One may expect that the number of targets being presented to the learner as well as the frequency with which the target sentences are being presented,
will affect learning rate. Possibly, they interact with variables from the other framework clusters.

5. Compatibility Between Learning and Retention Tasks

Tasks to be performed after a learning phase (retention or criterial tasks), may be compatible, incompatible, or neutral to the processing mode of the previous learning task. In connection with this phenomenon of (in)compatibility between learning and retention task, Bransford, Franks, Morris, and Stein (1979) introduced the notion of transfer appropriateness. Bransford and his associates (Morris, Bransford, & Franks, 1977) found an interaction between encoding processes (semantic and nonsemantic learning tasks) and the product of retrieval processes (semantic and nonsemantic retention tasks). Subjects who had been administered compatible learning and retention tasks (semantic-semantic, or nonsemantic-nonsemantic) achieved higher retention scores than subjects who were given incompatible learning and retention tasks (semantic-nonsemantic, or nonsemantic-semantic). Apart from the theoretical relevance of this and similar experiments, there is also a methodological lesson to be learned here, namely that an accurate assessment of information processing mode requires a joint consideration of learning task and retention task (Eysenck, 1982, p. 225). Tulving (1979) therefore advocated the inclusion of at least two different retention tasks in all learning experiments.

Research Questions and Design

The experiments reported here constitute the first steps in an attempt to systematically study the mutual influence of the five factors mentioned in the previous section, on the implicit and incidental learning of second language grammar rules, with special focus on the first factor. For an easy understanding of the present experiments, the operationalization of the factors in the present experiments will be first described in general terms, focusing on their relevance to the research questions and design. The method section provides more detailed information concerning their operationalization.

Processing Mode

Target processing mode was manipulated in the following way: Instructions to subjects (Ss) were aimed at orienting their attention inadvertently either to the formal properties of the target sentences, or to their meaning, or to both form and meaning. Thus, along with this dimension of target processing mode, there were three experimental conditions: Form, Meaning, and Form & Meaning (details follow below).

The introduction section contained the assumption that, for implicit and incidental learning to take place, it is necessary to pay attention to form. On the basis of this assumption, the hypothesis in these experiments is that
learning is more successful in the Form and Form & Meaning conditions, than in the Meaning condition, disregarding for the moment possible interactions with variables from the remaining four framework clusters.

**Current L2 Knowledge**

Prior L2 knowledge was not an independent variable in the present study, but an attempt was made to control for prior knowledge, by selecting Ss who could be assumed to possess enough vocabulary to grasp the meaning of the target sentences, without having, however, a command of the target rules themselves (as in the “flower picking” example). Ss were therefore pretested by means of a special sentence copying test (details below). Data of Ss with floor and ceiling scores on this pretest were discarded from the analyses.

Conducting an experiment with “real” L2 learners, implies experimenting with a number of individuals, of whom not even two may have exactly the same prior L2 knowledge, even though they may well have exactly the same pretest scores. To accommodate this methodological problem, the first experiment (which was conducted with “real” learners of Dutch as a second language and with Dutch target sentences containing Dutch target rules) was replicated by a second experiment, in which native speakers of Dutch were presented with target sentences containing an artificial target structure. The content words in this second experiment remained the same natural Dutch content words, conveying the same meaning as in the first experiment, but some function words were replaced by artificial morphemes, and the word order was changed into an artificial order. Thus, in the second experiment, almost absolute control of prior knowledge was achieved, in that all Ss were familiar with all of the content words, but no S was familiar with any feature of the (artificial) target structure (as in the “flower picking” example). This study thus followed a twofold research strategy: The first experiment, with natural stimuli, was replicated with a second experiment, using partially artificial stimuli. In this way a somewhat less valid but more reliable procedure (Experiment 2 with partly artificial targets) was added to the more valid but less reliable procedure (Experiment 1 with natural targets). Perhaps such a twofold strategy will prove useful for future L2 acquisition research.

**Grammatical Characteristics of the Target Structures**

For the present experiments, targets were selected with a predominantly formal function. In later experiments, the design will be expanded, such as to include both targets that do and targets that do not reflect semantic notions.

**Number and Frequency of Target Structures in Input**

In the present experiments, number and frequency of targets was held constant.

**Retention Tasks**

The present experiments contained two retention tests: (1) a cued recall test, in which Ss’ recollection of all target sentences that had been presented
during the learning task was tested, and (2) a sentence copying test, consisting of target sentences. Some of these sentences were identical to the ones presented in the learning task, but most of them were new. The new target sentences contained the same target structures (function words and word order), but they contained other content words, and thus conveyed other meanings. The cued recall test was assumed to be neutral (unbiased) to the processing mode of the previous learning phase. This test was therefore deemed particularly appropriate to prominently exhibit differences between experimental groups in what they had learned implicitly and what they had learned incidentally. It was hypothesized that learners who had served in the Form condition of the learning task would be best at recalling the formal features of the target sentences (function words and word order), whereas learners who had served in the Meaning condition of the learning task would be best at recalling the semantic features of the target sentences (content words). The second retention test, however, was assumed to offer a bias in favour of the Form Group Ss. They were expected to perform better than the Meaning Group Ss, because they could successfully transfer their formal knowledge, acquired in the learning task (compatibility). The Meaning Group Ss, on the other hand, would have focused most of their attention on the meaning (content words) of the sentences in the learning task. Their memory for these semantic features could hardly be successfully applied in the second retention test, since most target sentences in this test contained other content words (incompatibility).

The experiments adopted the incidental learning paradigm; that is, Ss were not informed about the research questions until after completion of the last test, and, while carrying out a current task, did not know whether a subsequent test would follow. The following hypotheses were tested:

Hypothesis 1: On the cued recall test (an unbiased procedure to ascertain recollection of the nominal target sentences previously shown), Form Group Ss will perform better than Meaning Group Ss on the formal aspects of the target sentences (function words and word order). Meaning Group Ss, however, will perform better than Form Group Ss on the semantic aspects of the target sentences.

Hypothesis 2: On the sentence copying test (the second retention test), Form Group Ss will perform better than Meaning Group Ss, since retaining a long sentence in memory requires encoding its grammatical structure.

The third experimental group, the Form & Meaning Group, was added to establish a control for joint attention to form and meaning.

Both experiments consisted of the administration of four tasks in the following order (all tasks coming unexpectedly to Ss):

1. Pretest: sentence copying test 1,
2. learning task: processing target sentences while paying attention to Form, Meaning, or Form & Meaning (three independent subject groups),
3. retention test 1: cued recall of the learning task sentences,
4. retention test 2: sentence copying test 2, containing mostly new target sentences.

In the second experiment, a baseline Control Group was added (fourth subject group), whose Ss did not participate in the learning task or in the cued recall test. These Ss were only administered the pretest (sentence copying test 1) and the second retention test (sentence copying test 2).

Experiment 1

Subjects

The first experiment was conducted with 145 adult intermediate learners of Dutch as an L2. Almost all Ss were taking Dutch classes in order to prepare for various forms of higher education in the Netherlands. Most of them had non-Indo-Germanic native languages.

Target Structure

A fairly complicated target structure was selected, as is illustrated in the top half of Table 1. The subordinate clause in this structure contains features which pose difficulties, even to advanced learners. Subclause word order is generally found difficult, but the present target structure is even further complicated by the presence of a modal auxiliary, the passive auxiliary worden, and the function word er. This word has no meaning of its own; it accompanies an indefinite existential subject NP. However, it does not immediately precede, or follow the subject NP to which it belongs, but it is placed at clause initial position, after the conjunction dat. The reason for having chosen this target structure was that intermediate learners of Dutch were unlikely to be familiar with all its grammatical properties.

Copying Tests 1 and 2 (Pretest and Retention Test 2)

These tests consisted of a visual dictation task. Ss were shown single sentences, which they had to write down. All target sentences contained the target structure mentioned above. Sentences were projected with a slide projector. Each sentence was exposed for seven seconds. This period is just long enough to carefully read the sentence only once. It must then be held and rehearsed in short term memory for a few more seconds, in order to be written down. Sentences of the length and complexity as the target sentences used here, can only be held in memory if subjects are familiar with both their content words and their structural characteristics. Performance on such a copying test is considered to reflect grammatical competence (Albertini & Forman, 1983; Spitze & Fischer, 1981). Thus, by means of this procedure, the pretest was used to assess the extent to which Ss had already acquired the target structure. Retention test 2, containing mostly new target sentences, was used as a posttest, testing whether performance differed according to experimental condition in the learning task.
Table 1
Target structures and sample target sentences

<table>
<thead>
<tr>
<th>Experiment 1*</th>
<th>Experiment 2**</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vindt u</strong> +</td>
<td><strong>Vinden u</strong> +</td>
</tr>
<tr>
<td><strong>dat</strong> +</td>
<td><strong>tek</strong> +</td>
</tr>
<tr>
<td><strong>er</strong> +</td>
<td><strong>ADV. LOCAL/TEMPORAL PP</strong> +</td>
</tr>
<tr>
<td><strong>ADV. LOCAL/TEMPORAL PP</strong> +</td>
<td><strong>MODAL AUX</strong> +</td>
</tr>
<tr>
<td><strong>INDEF. SUBJECT NP</strong> +</td>
<td><strong>PAST PART. MAIN V</strong> +</td>
</tr>
<tr>
<td><strong>PAST PART. MAIN V</strong> +</td>
<td><strong>INDEF. SUBJECT NP</strong>?</td>
</tr>
<tr>
<td><strong>MODAL AUX</strong> +</td>
<td><strong>barden</strong> +</td>
</tr>
<tr>
<td><strong>worden?</strong> +</td>
<td><strong>ra</strong> +</td>
</tr>
</tbody>
</table>

1. **Vindt u** **dat** **er** **op alle scholen**
   Find you that there 
   Would you say that dancing classes should be taught in all schools?
   **dansles** **gegeven** **moet** **worden?**

2. **Vindt u** **dat** **er** **in trams en bussen**
   Find you that there 
   Would you say that the consumption of French fries should be allowed in street cars and busses?
   **patates frites** **gegeten** **mag** **worden?**

3. **Vindt u** **dat** **er** **in elke grote stad**
   Find you that there 
   Would you say that one should allow the construction of a subway in every big town?
   **een metro** **aangelegd** **mag** **worden?**

1. **Vinden u** **tek** **op alle scholen**
   Find you in all schools
   **moeten** **elgeven** **barden** **ra** **dansles** ?

2. **Vinden u** **tek** **in trams en bussen**
   Find you in trams and busses
   **mogen** **elenet** **barden** **ra** **dancing lesson** ?

3. **Vinden u** **tek** **in elke grote stad**
   Find you in every big town
   **mogen** **elaanleggen** **barden** **ra** **een metro** ?

---

* Experiment 1: Dutch target structures with English translations.
** Experiment 2: Artificial target structures with English translations (artificial elements printed in italics).
Copying test 1 consisted of twelve items: two practice sentences, followed by ten test sentences — six nontargets (distractors) and four targets (items 3, 5, 8, and 10). Only targets were scored. Copying test 2 consisted of two practice items, followed by nine test sentences: two nontargets and seven targets (items 1, 2, 3, 4, 6, 7, and 9). Targets 3 and 7 had also formed part of the stimulus set in the learning task. Thus, in copying test 2, five out of seven targets were new.

Assignment to Conditions

Originally, Experiment 1 was designed with two conditions only — the Form Group and the Meaning Group. This part of the study was carried out during the spring of 1984. Ss were pretested in groups varying from eight to 20 Ss, and afterwards, randomly assigned to either condition. Testing in the remaining three tasks (learning task and retention tests) also took place in small groups. Not more than ten Ss were tested simultaneously. The results of this part of the study indicated that the experiment needed to be extended with two more experimental conditions. Therefore, in the fall of 1984, more learners were tested, again in small groups. After pretesting these groups were assigned to either the second Meaning condition (see below), or the Form & Meaning condition. Together, the spring and fall studies constitute one experiment with four experimental groups. With this combining procedure, the principle of random assignment could be essentially, although not entirely adhered to.

Learning Task

During the learning task, all groups were presented with twelve sentences: three practice items not containing the target structure, followed by nine target sentences. Three such sentences are shown in the top half of Table 1. These sentences were shown one by one on a screen by means of a slide projector.

The Form Group was presented with the target sentences not only on the projection screen but also in a test booklet. However, in this booklet the sentences were divided into eight fragments, which were randomly ordered and printed vertically. Ss were instructed to write a number before each fragment, corresponding to its place in the correct sentence, while consulting the correct sentence on the screen. Thus, the Form Group had to perform what can be called an anagram task, which directed Ss’ attention to word order and did not require the encoding of the sentence meaning.

The Meaning Group Ss were also presented with the target sentences one by one on the screen, but they were asked to think about the targets’ meanings (all sentences being opinion questions), and then write their opinion in short statements, such as yes, yes of course, no, not at all, perhaps, I don’t know, and so on.

The experiment was originally designed in such a way that the only difference between experimental groups was to be the Ss’ focus of attention. Everything else was to be held constant, including exposure time to the target. In a pilot study, it was found that 30 seconds was sufficient time
for Ss in the Form condition to perform their anagram task. Therefore 30 seconds was originally chosen for the Meaning Group as well. However, it turned out that many Ss completed their task in about 15 seconds or so. During the latter period they may well have engaged in something other than semantic processing of the target sentences. In order to determine the differential impact of a purely semantic task on the one hand and the effect of 30 seconds exposure without any particular activity to perform on the other hand, two more experimental groups were added to the study: the second Meaning Group, and the Form & Meaning Group.

The second Meaning Group performed exactly the same learning task as the first, the only difference was that, for the second Meaning Group, the sentences were visible for only 10 seconds. The total response time remained 30 seconds, however.

The fourth experimental group was simply instructed “to pay attention to both form and meaning of the sentences,” but no concrete task was assigned. Thus, no particular type of processing was enforced, and Ss were free to choose their own encoding strategies. To this extent, the Form & Meaning Group can be conceived of as a control group.

**Cued Recall Test**

The learning task was followed by an unexpected cued recall test. Ss were asked to recall all nine target sentences that had been presented in the learning task (in the same order). For each of the nine sentences to be recalled, a cue was given, consisting of the adverbial local/temporal phrase (see Table 1). Each cue was printed on a separate page in the test booklet. For each sentence 60 seconds response time was given. Ss were not allowed to turn the pages backward, and turned the pages forward only when the experimenter instructed them to do so. The instruction read as follows: “Do you still remember the questions that you’ve just read? Try to write them down as precisely as you can. We’ll help you a little by giving you a few words with every question.” (The original instruction was in Dutch.)

**Scoring**

Responses in copying test 1, cued recall test, and copying test 2, were scored for the number of constituents correctly reproduced. Each of the eight constituents making up a target sentence (see Table 1) was credited with one point if reproduced in the right place. Spelling errors and minor form errors were ignored, the criterion being whether the morphological and syntactic structure was intact. It was felt that this scoring procedure was a valid measure of Ss’ sensitivity to the target structure.

The responses in the cued recall test were not only scored for grammatical structure, but also for content. Each response could be credited with five points according to whether the five most important information units of the target sentence had been reproduced comprehensively. These units were: (1) opinion solicitation, that is, the fact that the sentences asked for the reader’s opinion, (2) modality, that is, obligation or necessity (cf. the modal auxiliaries in the sample sentence of Table 1), (3) the meaning of the subclause
subject NP, (4) the meaning of the subclause main V, and (5) the meaning of the local/temporal phrase. (This last unit, time or place, was already given with the cue. Thus, one point was easily earned. Yet, in quite a few responses, this cue was left unused, and not incorporated into the response.) Grammatical and spelling errors were ignored, and so were deviations from the target's grammatical structure. For example, a response containing no subclause, no passive voice, nor the word *er* (as in the original target sentence), could still receive the maximum content score.

Results

Not included in the analyses were the data of Ss with pretest scores of more than 91% correct, or less than 28% correct, to avoid floor and ceiling effects. Nor were the data included of those few Ss who found the recall task too difficult and had given up on this test. Of the 145 Ss tested, there were 123 Ss with complete data within these boundaries.

The results are presented in Tables 2 and 3. From Table 2, which shows the results of the cued recall test, it can be seen that the responses of the two Meaning Groups were better than those of the Form Group, when scored for content, but that they were poorer when scored for grammatical structure. This can be taken as evidence for Hypothesis 1. The Form & Meaning Group responses were best of all, both ways. Note that Ss in this group had been able to look at the targets during an uninterrupted period of 30 seconds, since they did not have to perform any writing task within this period. From Table 3, it can be inferred that performance on the pretest fortunately was not significantly different across groups, although the principle of random assignment of Ss to conditions had not been completely adhered to. At first sight, the pretest scores appear to be extremely high, but one has to bear in mind that, in a visual dictation, it is fairly easy to remember at least the first four of all eight constituents, which is good for a 50% score. Thus, this 50% is in fact a bottom score, indicating that the target structure as such has not been encoded. Since pretest scores are about 20% higher than this bottom score, it can be suspected that most Ss had been familiar with some, although not all grammatical features of the target structure.

There is an average 12% performance increase from pretest to retention test 2 (i.e., almost one sentence fragment), indicating that Ss had considerably improved their performance. However, there are no significant differences between groups. Thus, although the learning task had a differential impact on the recall test, this differential effect was not transferred to retention test 2, consisting of another task and containing other target sentences.

Since pretest scores were much higher than the assumed bottom scores, and since high pretest scores may cause ceiling effects, a second comparison between copying test 1 and 2 was made. This time, an analysis was carried out on data of Ss with pretest scores not exceeding 70%. (This analysis was carried out after Experiment 2 had been conducted, showing pretest scores not exceeding 70%, i.e., 5.6 out of 8 fragments.) However, the pattern of
### Table 2

Experiment 1: Correct reproduction of grammatical structure and content (in %) in cued recall test.

<table>
<thead>
<tr>
<th></th>
<th>Form</th>
<th>Meaning 1</th>
<th>Meaning 2</th>
<th>Form &amp; Meaning</th>
<th>One-Way ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=34</td>
<td>N=38</td>
<td>N=32</td>
<td>N=19</td>
<td>F (3,119)</td>
</tr>
<tr>
<td>Grammatical Structure¹</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>N=34</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Grammar</td>
<td>66.7</td>
<td>16.1</td>
<td>50.5</td>
<td>20.4</td>
<td>76.2</td>
</tr>
<tr>
<td>Structure</td>
<td>58.8</td>
<td>19.4</td>
<td>50.5</td>
<td>20.4</td>
<td>76.2</td>
</tr>
<tr>
<td>Content²</td>
<td>84.0</td>
<td>15.9</td>
<td>84.6</td>
<td>15.7</td>
<td></td>
</tr>
</tbody>
</table>

1 100% = 8 sentence fragments  
2 100% = 5 content units  
3 Meaning Group 2 significantly lower (p < .01) than both Form Group and Form & Meaning Group; Meaning Group 1 significantly lower (p < .01) than Form & Meaning Group (Newman-Keuls).  
4 Form Group significantly lower (p < .01) than all other groups (Newman-Keuls).
Table 3

Experiment 1: Correct reproduction of grammatical structure (in %) in sentence copying tests 1 and 2.$^1$

<table>
<thead>
<tr>
<th></th>
<th>Form</th>
<th>Meaning 1</th>
<th>Meaning 2</th>
<th>Form &amp; Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=34</td>
<td>30 sec. exposure</td>
<td>N=38</td>
<td>10 sec. exposure</td>
</tr>
<tr>
<td>M SD</td>
<td>M SD</td>
<td>M SD</td>
<td>M SD</td>
<td>M SD</td>
</tr>
<tr>
<td>Copying Test 1</td>
<td>71.1 15.5</td>
<td>70.3 17.6</td>
<td>62.0 15.5</td>
<td>70.0 16.5</td>
</tr>
<tr>
<td>Increase</td>
<td>11.0 11.6</td>
<td>10.5 11.0</td>
<td>13.2 10.3</td>
<td>14.1 10.3</td>
</tr>
<tr>
<td>Copying Test 2</td>
<td>82.1 12.9</td>
<td>80.8 13.8</td>
<td>75.2 15.8</td>
<td>84.1 14.8</td>
</tr>
</tbody>
</table>

$^1$ 100% = 8 sentence fragments
results from this second analysis remained essentially identical to the pattern obtained from the first analysis: All group means in the second analysis were obviously lower than those in the first analysis, but the differences between groups did not change. It can therefore be concluded that Experiment 1 provided evidence in favour of Hypothesis 1, but failed to support Hypothesis 2.

At this point in the study, it seemed likely that in Experiment 1, it had not been possible to adequately control for prior L2 knowledge, and this might have caused the absence of differences between experimental groups on the second retention test. Therefore, in order to improve on the generalizability of the experiment, Experiment 2 was conducted with artificial target features, of which simply no subject could possibly have had prior knowledge.

Experiment 2

Subjects
In Experiment 2, 80 high school students (ages 15-19) participated on a voluntary basis. All Ss, who were paid for their participation, were native speakers of Dutch. They had signed up for an additional hour after a normal school day. As in the first experiment, the whole testing session (pretest, learning task, retention tests 1 and 2) lasted 45 to 50 minutes. Tests were administered in groups; each group consisting of 20 Ss. At the beginning of the session, Ss were told that their participation was asked for in order to investigate reading and writing processes. They were debriefed after all groups had been tested.

Target Structure
The target sentences were very similar to the ones used in Experiment 1, their meaning and grammatical deep structure being the same, but their surface structure being different. Table 1 shows the differences. Finite verb forms were changed into infinitives; the conjunction dat into artificial tek; the passive auxiliary worden into artificial harden; the function word er into artificial ra; past participle prefix ge- into artificial el-; and finite verb forms vindt, moet, and mag into infinite vinden, moeten, and mogen, respectively. Furthermore, subclause word order was changed into an artificial order. Thus, although all sentences had undergone considerable changes in their syntactic and morphological surface structure, they contained the same natural Dutch content words, conveying the same meaning as in Experiment 1. Following only a semantic encoding strategy, as in the flower picking example, a native speaker of Dutch would not have much difficulty in comprehending the meaning of the partly artificial target sentences. (Neither would an English speaker have much difficulty with a sentence such as “Would you say tek in every big town may elconstruct harden ra a subway?”.)

Copying Test 1 and 2, and Cued Recall Test
Tasks and procedures for these tests were identical to those in Experiment 1.
Implicit and Incidental SLA

Experimental Treatment in the Learning Task

There were four groups: Form, Meaning, Form & Meaning, and Control. The Form Group performed the same anagram task as the Form Group in Experiment 1. The Meaning Group performed the same task as the second Meaning Group in Experiment 1, that is, target sentences were visible for ten seconds, followed by 20 seconds provided to write down an answer. The Form & Meaning Group also performed the same task as the Form & Meaning Group in Experiment 1. Added to the design was a Control Group, which was administered the pretest and the second retention test only (sentence copying tests 1 and 2). During the twenty minutes needed by the other groups to complete the learning task and the cued recall test, the Control Group was given a reading comprehension test (texts and multiple choice questions, in normal, nonartificial Dutch). Ss were led to believe that this test belonged to the experiment.

Results and Discussion

Scoring procedures were the same as in Experiment 1. Table 4 shows the results of the cued recall test. The pattern of results is similar, but not identical to that of the cued recall test in the first experiment (compare Table 4 with 2). As hypothesized, the responses of the Meaning Group were better than those of the Form Group, when scored for content, and poorer when scored for grammatical structure. Such a result was also obtained in Experiment 1, but in Experiment 2 the Meaning Group performance is more pronounced, compared with the Meaning Group performance in Experiment 1 (higher content performance, and lower grammatical performance). Supposedly, this is due to the higher reliability in Experiment 2. Comparing the performance of the Form & Meaning Group with that of the Form Group, it can be observed that, while it is still higher (as in Experiment 1), and even significantly higher when scored for grammatical structure, the magnitude of the difference is small in comparison to that in Experiment 1. Thus, one could not characterize the Form & Meaning Group performance as truly superior, as was the case in Experiment 1.

From Table 5 which presents the results of the two sentence copying tests, it can be seen first of all, that the Form & Meaning Group performed somewhat lower than the other groups, on the pretest. This difference, however, should be regarded as incidental and irrelevant to the question under investigation, since it does not reach significance, and since Ss had been assigned to conditions on a completely random basis.

Secondly, it should be noted that average pretest performance is in the neighbourhood of 50% (i.e., four out of eight fragments correctly reproduced), the supposed bottom level, indicating that the grammatical structure of the target sentences had not been encoded. To that extent, the conditions of the flower picking example appear to have been successfully simulated under experimental conditions.

Turning to the second retention test, and to its differences with the pretest, it can be observed firstly, that all groups considerably increased their performance, and secondly, that the magnitude of this increase differed
Table 4
Experiment 2: Correct reproduction of grammatical structure and content (in %) in cued recall test.

<table>
<thead>
<tr>
<th></th>
<th>Form N = 20</th>
<th>Meaning N = 20</th>
<th>Form &amp; Meaning N = 20</th>
<th>One-Way ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>F (2,57)</td>
</tr>
<tr>
<td>Grammatical Structure¹</td>
<td>63.3</td>
<td>11.3</td>
<td>68.0</td>
<td>17.310</td>
</tr>
<tr>
<td>Content²</td>
<td>60.3</td>
<td>12.3</td>
<td>69.7</td>
<td>19.840</td>
</tr>
</tbody>
</table>

1 100% = 8 sentence fragments
2 100% = 5 content units
3 Meaning Group significantly lower (p < .01) than both other groups (Newman-Keuls).
4 Meaning Group significantly higher (p < .01) than both other groups; Form & Meaning Group significantly higher (p < .05) than Form Group (Newman-Keuls).
Table 5

Experiment 2: Correct reproduction of grammatical structure (in %) in sentence copying tests 1 and 2.¹

<table>
<thead>
<tr>
<th></th>
<th>Form N = 20</th>
<th>Meaning N = 20</th>
<th>Form &amp; Meaning N = 20</th>
<th>Control N = 20</th>
<th>One-Way ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>F (3,76)</td>
</tr>
<tr>
<td>Copying Test 1</td>
<td>52.6</td>
<td>9.1</td>
<td>52.2</td>
<td>10.9</td>
<td>1.304</td>
</tr>
<tr>
<td>Increase</td>
<td>31.8</td>
<td>8.5</td>
<td>24.4</td>
<td>9.9</td>
<td>8.321</td>
</tr>
<tr>
<td>Copying Test 2</td>
<td>84.4</td>
<td>11.4</td>
<td>76.6</td>
<td>14.2</td>
<td>4.011</td>
</tr>
</tbody>
</table>

¹ 100% = 8 sentence fragments

2 Control Group significantly lower (p < .01) than both Form Group and Form & Meaning Group, and significantly lower (p < .05) than Meaning Group; Form Group significantly higher (p < .01) than Meaning Group (Newman-Keuls).

3 Control Group significantly lower (p < .01) than Form Group (Newman-Keuls).
significantly across conditions, as predicted by Hypothesis 2. The Control Group, which had not been presented any target sentences in between, also increased its performance, the increase being the slightest of all groups, showing a bottom line test-retest effect. This is to be taken into account when interpreting the progression made by the three experimental groups. (Incidentally, this illustrates the importance of a control group in an experimental design.) The performance increase from the first to the second sentence copying test took place jumpwise (i.e., between the last item of the copying test 1 and the first item of the copying test 2). This is demonstrated in Figure 1 (page 67), showing mean group performance for every individual item in both tests. There is only one within test jump, namely from the first target item to the second, in the pretest. This is a natural phenomenon, since the very first target sentence, which displays all artificial characteristics of the surface structure, struck Ss not only as new, but also as somewhat bizarre. Importantly, however, performance within the second copying test remained constant over all items, showing only slight, nonsignificant differences between items. This was held true for all four groups.

While Table 5 shows the results of three separate one-way ANOVAs, two analyses of covariance, using the first copying test scores as covariates, were also run. One analysis of covariance included, the other excluded the Control Group. Both ANCOVAs yielded essentially the same results as the One-way ANOVA on the difference scores. This is no surprise, since the experiment had been conducted with a sample of homogeneous Ss, who had been randomly assigned to conditions. The ANCOVA, including the Control Group, yielded a highly significant group effect, \( F(3,75) = 8.153, p < .001 \). Newman Keuls post-hoc comparisons showed the Control Group as being significantly lower than the Form Group (\( p < .01 \)), the Form & Meaning Group (\( p < .01 \)), and the Meaning Group (\( p < .05 \); the comparison between the Form Group and the Meaning Group almost reached significance at the .05 level. The ANCOVA, excluding the Control Group, yielded a significant main effect \( F(2,56) = 3.4, p = .04 \). Newman Keuls post-hoc comparisons showed the Form Group as significantly higher than the Meaning Group (\( p < .05 \)).

Individual differences within groups did not change substantially from the first to the second copying test. This is illustrated in Table 6, giving correlations amongst the scores on the first and the second copying test, as well as the difference scores between the two. For all four groups, first and second copying tests correlate substantially, whereas small and insignificant correlations were obtained between first copying test and difference scores. No ceiling or bottom effects existed, and therefore the individual error in the ANOVA on the difference scores can be assumed to be slight.

It can thus be concluded that performance increase from first to second copying test was brought about by a combination of a test-retest effect and a treatment effect, caused by the manipulation of encoding processes in the learning task. As predicted, Form Group Ss progressed significantly more than Meaning Group Ss. Such a treatment effect had not manifested itself in Experiment 1. However, the magnitude of the difference is small: Form Group Ss increased their performance with only 7.4% more (somewhat
Figure 1: Mean scores for individual target items of sentence copying tests 1 and 2.
Table 6

Experiment 2: Within group intercorrelations (Pearson) between sentence copying tests 1 and 2, and their difference.

<table>
<thead>
<tr>
<th></th>
<th>Form Copying Increase</th>
<th>Meanin Copying Increase</th>
<th>Form &amp; Meaning Copying Increase</th>
<th>Control Copying Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copying Test 1</td>
<td>0.68</td>
<td>0.72</td>
<td>0.61</td>
<td>0.75</td>
</tr>
<tr>
<td>Increase</td>
<td>-0.17</td>
<td>-0.07</td>
<td>-0.11</td>
<td>-0.07</td>
</tr>
<tr>
<td>Increase</td>
<td>0.61</td>
<td>0.64</td>
<td>0.72</td>
<td>0.60</td>
</tr>
</tbody>
</table>
more than half a constituent on a maximum of eight) than the Meaning Group Ss.

The analyses presented and discussed so far, were based on scores which represented the total number of sentence fragments which had been reproduced in the right place. In these scores, artificial fragments, such as *tek*, *barden*, etc. had been summed with natural fragments, such as the local/temporal adverbial phrase and the subclause subject NP. In order to discover whether the slight superiority of the Form Group Ss over the Meaning Group Ss would be visible more clearly in the reproduction of the artificial characteristics of the target sentences, separate analyses were carried out on all eight sentence fragments as well as on word order violations, in the second copying test. The results are shown in Table 7.

A comparison of the results of the overall analysis (shown on the bottom line of Table 5) with the breakdown for individual fragments (Table 7) shows that the general pattern remains the same. Group differences are significant for the following three fragments only.

Firstly, the Form & Meaning Group performed significantly lower than the Form Group on the fourth fragment, the modal auxiliary. (For some unexplained reason, three Ss in this group did not include the modal auxiliary in any response.)

Secondly, the Control Group performed significantly lower than each of the other groups on the last fragment, the subclause subject NP. It is uncertain whether this might be due to general memory constraints. On the basis of evidence from the literature (e.g., Meara, 1980; Spitze & Fischer, 1981), one would expect the burden of memory to be most detrimental in the middle of a string of elements which must be remembered, especially if this string consists of unrelated items. This pattern is also present in the data of the sentence copying tests. Reproduction performance of all groups is best in the first two fragments, then shows a decline, and finally rises a little towards the end. This is illustrated in Table 7. Thus, it may seem that Ss in all groups, while dealing with the copying task, adopted an approach that is more characteristic of processing strings of unrelated elements, rather than processing normal sentences.

Thirdly, performance on the seventh fragment is higher for the Form Group and Form & Meaning Group than for the Meaning Group and Control Group. The overall effect is significant, but due to the substantial variance in the scores of the latter two groups, Newman Keuls comparisons do not show any pair of group means as significantly different from one another. Yet, it is important to note that it is precisely the most difficult artificial word *ra* (corresponding to Dutch *er*, which has no meaning of its own, and functions in sentences with an indefinite, existential subject NP), which showed Form Ss to perform better than Meaning and Control Ss.

Finally, all instances of interference errors in Ss’ responses were counted, that is, when Ss wrote Dutch words instead of the artificial target words. Surprisingly, Ss’ responses in both the recall test and the sentence copying test, contained very few cases of interference, with the exception of Meaning Group responses in the cued recall test. Replacements were counted of the three artificial unbound morphemes *tek*, *barden*, and *ra* by Dutch *dat*, wor-
Table 7

Experiment 2: Correct reproduction of sentence fragments, and word order deviations, in sentence copying test 2.

<table>
<thead>
<tr>
<th></th>
<th>Vinden u</th>
<th>tek</th>
<th>ADV. LOCAL/ TEMPORAL PP</th>
<th>MODAL AUX</th>
<th>PAST PART. MAIN V</th>
<th>barden</th>
<th>ra</th>
<th>SUBJECT NP</th>
<th>word order deviations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form</td>
<td>6.9</td>
<td>6.7</td>
<td>6.75</td>
<td>5.75</td>
<td>4.65</td>
<td>4.9</td>
<td>5.95</td>
<td>6.7</td>
<td>1.9</td>
</tr>
<tr>
<td>Meaning</td>
<td>6.75</td>
<td>6.2</td>
<td>6.9</td>
<td>4.9</td>
<td>4.45</td>
<td>4.25</td>
<td>4.85</td>
<td>6.65</td>
<td>3.15</td>
</tr>
<tr>
<td>Form &amp; Meaning</td>
<td>6.75</td>
<td>6.45</td>
<td>6.5</td>
<td>3.85</td>
<td>4.6</td>
<td>4.7</td>
<td>5.75</td>
<td>6.45</td>
<td>2.7</td>
</tr>
<tr>
<td>Control</td>
<td>6.55</td>
<td>5.63</td>
<td>6.6</td>
<td>4.8</td>
<td>3.3</td>
<td>3.6</td>
<td>4.4</td>
<td>5.75</td>
<td>3.6</td>
</tr>
</tbody>
</table>

One-Way ANOVAs (df 3,76)

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>p</th>
<th></th>
<th>F</th>
<th>p</th>
<th></th>
<th>F</th>
<th>p</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.110</td>
<td>.35</td>
<td>1.753</td>
<td>1.126</td>
<td>.152</td>
<td>2.988</td>
<td>.036</td>
<td>3.066</td>
<td>.033^2</td>
</tr>
<tr>
<td></td>
<td>1.485</td>
<td>.225</td>
<td>1.812</td>
<td>.344</td>
<td>.016^3</td>
<td>2.7</td>
<td>3.679</td>
<td>.260</td>
<td></td>
</tr>
</tbody>
</table>

1 Form & Meaning Group significantly lower (p < .05) than Form Group (Newman-Keuls).
2 No significant pairwise differences (Newman-Keuls).
3 Control Group significantly lower (p < .05) than Form Group (Newman-Keuls).
den, and er respectively. The cued recall test offered 540 possible cases of interference per group (nine target sentences, three target words in every target sentence, 20 Ss in every group). For the Form Group, 10 cases of interference were found, for the Meaning Group 90 cases, and for the Form and Meaning Group 6 cases. The second sentence copying test offered 420 possible interference cases (seven sentences, three target words, and 20 Ss). In the Form Group responses no interference errors were found, the Meaning Group made 1, the Form & Meaning Group 8, and the Control Group 7 interference errors. These low figures demonstrate that Ss, generally, had not been aware of the fact that the artificial function words corresponded to Dutch function words. The relatively high incidence of interference errors in the Meaning Group responses of the cued recall test, provides further evidence that, during the learning task, Ss had not had much time to focus attention on the grammatical features of the target sentences.

Conclusion

Reviewing recent research on L2 learners’ perception and processing (intake) of L2 input, Chaudron (1985) concludes that “second language research has inadequately explored the process of intake,” and he points out that “both theory and the research derived from it need to develop consistent, well-defined models and methods for investigating language learners’ processing of input” (p. 1).

The present study aimed at contributing precisely to these two goals: theory and method development. A five factor framework was presented as a useful pretheoretical tool to help conceptualize the most relevant variables in the cognitive study of second language acquisition. The experiments were designed within this framework, and based on theoretical assumptions concerning encoding processes necessary for grammar learning to take place. Both experiments attempted to simulate a situation in which learners were presented with sentences consisting of familiar content words but unfamiliar structural features (function words, word order). By experimentally manipulating Ss’ encoding of target sentences, orienting Ss’ attention to form, meaning, or simultaneously to form and meaning, the differential impact of these processing modes on the acquisition of the structural features could be investigated. It was hypothesized that the Form Group Ss would demonstrate a better recollection of the structural features than the Meaning Group Ss, and that the latter Ss would exhibit a better recollection of the contents of the target sentences. These two expectations were borne out in both experiments.

Cognitive psychologists have repeatedly pointed out that for the study of learning a joint consideration of learning (encoding) and retention (retrieval) is crucial. A retention test may be compatible, incompatible, or neutral to encoding strategies in a previous learning task. The cued recall test in the present experiments was assumed to be neutral in this respect, thus providing each experimental group with the opportunity to exhibit a recollection of precisely those features, to which their attention had been
directed in the previous incidental learning task. However, the sentence copying test (the second retention test) was assumed to provide a bias in favour of the Form Group Ss. Of the two retention tests, the latter was the most important one. A test in which the acquired knowledge has to be transferred and applied to new sentences, is obviously of more relevance than a test requiring a reproduction of the learning task's nominal instances. Sentences of the same length and complexity as the target sentences used in these experiments, can only be kept in memory if subjects are familiar with both its content words and its structural characteristics. The assumption was that, in the second copying test, these two conditions would be fulfilled only for the Form Group Ss. Hence it was hypothesized that they would perform better than the Meaning Group Ss. The latter Ss were expected to either leave out the less familiar function words, or to copy only the first few sentence fragments.

These expectations were not supported by the results of the first experiment: Form Group Ss did neither substantially nor significantly outperform Meaning Group Ss. It was reasoned that this lack of supporting evidence could have been due to a lack of adequate control of Ss' prior knowledge. Many of them might well have been familiar with some of the structural features, at the outset. That is why the second experiment was designed, allowing for maximal control of prior knowledge. In the second experiment, allocations were made so that, at the outset, all Ss were familiar with all content words, but that no S could be familiar with any (artificial) function word, nor with the sentences' (artificial) word order. In this experiment, the results obtained by the second sentence copying test do indeed show the predicted difference. However, the results show another phenomenon as well: The Control Group Ss were able to reproduce a substantial number of artificial elements, although they had not served in the learning task. The writing down of partly artificial sentences from a screen thus appeared to be a task allowing for successful processing of at least some artificial elements. Hence, one may question the appropriateness of the adopted copying procedure.

Returning to the question underlying the project, of which these two experiments constitute the first two parts, it can be concluded that this study provides supporting evidence for the claim that, for implicit and incidental learning of structural language elements to take place, attention to form at input encoding is a sufficient condition. However, only modest (and therefore inconclusive) evidence was obtained in support of the claim that exclusive attention to meaning inhibits the acquisition of structural language elements. Further research is therefore necessary to determine the extent to which top-down, conceptually driven processing of second language sentences, allows for the acquisition of syntactic and morphological features. Presumably, meaning is always the learner's first priority. Form comes into play as a fail-safe, backup procedure, in case meaning alone is not sufficient to arrive at an adequate input interpretation (e.g., when the input comprises an ambiguous sentence).

Future experiments are planned, in which the third framework factor, linguistic target characteristics, will be experimentally manipulated (formal
versus semantic function words; ambiguous versus unambiguous input sentences) and its interaction with the first factor, target processing (focus on form or on meaning) will be investigated to test the hypothesis mentioned above: "meaning first, and form only as a fail-safe procedure." The sentence copying procedure will be replaced by a technique calling for production rather than for reproduction.

References


