A Psycholinguistic Analysis of
Word Guessing Using Cloze Procedures

Nobuya ITAGAKI
Miyagi University of Education

Kunihiro SATOMI
Eikoh Incorporation

and

Masumi TAKAHASHI
Marumori Junior High School

This paper is concerned with three psychological aspects of word guessing: (a) "contextual constraint" of texts, (b) "decontextualization" of contextual cues and (c) "problem-solving strategies" for word guessing. An experiment was carried out in which subjects were asked to infer words in cloze procedures, and they were also asked to report on how they came to infer the words. An analysis of these self-reports suggests (a) word guessing should be viewed as a complex interplay among the above three elements and English proficiency of learners; (b) heuristic and procedural strategies should be recognized as characteristic of low and high proficiency learners; and (c) word guessing should be related to the question of how deeply people understand textual contents as well as the English ability of learners. These findings will be discussed in conjunctions with several issues of word guessing and word acquisition.

1. Introduction

This paper is concerned with the question of how Japanese Learners of English (JLEs) infer words and their meanings in cloze tests or procedures. A number of studies have been reported on the usefulness and validity of cloze tests as measures of various language skills, such as reading and writing (e.g., Mochizuki, 1992; Naito, 1992). A simple question still remains to be addressed, however. That is, how do JLEs guess missing words and their meanings? In other words, what kind of psychological processes and strategies might be involved in word
guessing? The present paper aims to report a psychological analysis of word guessing by JLEs, providing some insights into the psychological processes of word guessing and word acquisition.

Two psychological aspects of word guessing are the foci of this study. The first is how contextual elements are related to word inferences; and the second is what kind of “problemsolving” strategies are manifested in word guessing. Among plausible contextual elements are the notions of “contextual constraint” and “decontextualization.” According to Schwanenflugel and Shoben (1985) and Schwanenflugel (1991), contexts in cloze procedures differ in their predictability of acceptable target words, and under the “high constraint” context, target words should be psychologically easier to infer, decide, and recognize than those under the “low constraint” context (also see Dubin & Olshtain, 1993; Kellas, Paul, Martin & Simpson, 1991). A few examples (cited in Schwanenflugel, 1991, p. 23) may help understand the notion of contextual constraint. Given “John kept his gym clothes in the ___,” we would fill in the blank with either “locker,” “closet” or “hamper.” The point is, however, that the first word, “locker,” is more likely to be chosen than the others. It is in this sense that the sentence context can be judged to be a “high constraint” context. On the other hand, a context sentence, “In the valley, there were three small ___,” can be considered to be “low constraint” because any particular word could be chosen. It can, thus, be assumed that high and low contextual constraints might lead to psychologically distinct processes of word guessing in cloze procedures. It must also be made clear that we are not concerned with whether words provided are acceptable or not, but how people infer target words regardless of whether they are acceptable or not. In other words, what concerns us here is not how many people are right in guessing target words, but what psychological processes may be involved in word guessing.

A second contextual element is that of “decontextualization.” Decontextualization processes have been discussed as a major cognitive process or component of word meaning acquisition (e.g., Dallen-Kapteijns & Elshout-Mohr, 1981; Elshout-Mohr & Dallen-Kapteijns, 1987; Just & Carpenter, 1987). The question is what use of contextual information is made by JLEs as cues for word guessing in cloze procedures. Our tentative answer is that some contextual information might be psychologically more decontextualized than others. Let us see how decontextualization processes work in learning the meaning of an unknown word using an example from Elshout-Mohr and Daalen-Kapteijns (1987, p. 62). Suppose that people are to infer the meaning of a coinage "kolper" in a context sentence - “During a heat wave a lot of people all of a sudden want to have kolpers, so the sales of sunblinds then reach a peak.” The contextual information about kolper could be, for example, decontextualized in terms of the following three levels:

Level 0: Kolpers are much asked for during a heat wave;
Level 1: Kolpers in some respects resemble sun-blinds;
Level 2: Kolpers have a cooling effect.

Note that the “having a cooling effect” and “resembling sun-blinds” of the levels 2 and 1 can be claimed to be “inferred” information, which is not directly specified in the above context.
sentence. In other words, the content of the context sentence is more or less decontextualized in the levels 1 and 2, while almost the same content as “wanting to have kolpers” is stated in the level 0. Furthermore, it is on the basis of the “general” knowledge about “a heat wave” and “summer” that we can infer “having a cooling effect” of the level 2. The point is that some information might be more decontextualized than others, and some people might make greater use of the decontextualized information than others.

It is thus possible to assume that decontextualization processes might be a major cognitive component of word guessing in cloze procedures. To be specific, contextual information might be decontextualized at several levels of abstraction (in the sense of Elshout-Mohr & Daalen-Kapteijns (1987)), and somewhat decontextualized information might be considered to be more useful contextual cues for word guessing.

This paper proposes the following three levels of decontextualization as psychological processes of word guessing: “Direct Translation,” “Inferred Information” and “Outside Knowledge.” Their operational definitions are as follows:

Direct Translation (DT) : JLEs simply translate contextual information in English into Japanese, making use of it as contextual cues.

Inferred Information (II) : JLEs infer some information from given contextual information, making use of it as inferred cues.

Outside Knowledge (OK) : JLEs interpret some contextual information in relation to their outside knowledge, making use of it as world-related cues.

For the sake of convenience, we will hereafter refer to these three as “DT,” “II” and “OK” cues, respectively.

Let us turn to the “problem-solving” aspect of word guessing in cloze procedures. In psychological literature, several strategies have been proposed and investigated as problem-solving strategies, for example, for “chess play,” “arithmetic problems,” “anagrams” and so on (for more details, see Bourne, Dominowski, Loftus & Healy, 1986; Glass, Holyoak & Santa, 1979). Our question is which strategies are used by JLEs to infer target words in cloze procedures. We claim that Siegler and Shragler’s (1984) triparty, “heuristic-retrieval,” “elaboration” and “algorithm,” of problem-solving strategies should be related to word guessing under consideration here.

Following Siegler and Shragler (1984), let us suppose that we are to spell an English word “vocabulary.” Some people would choose to spell it simply by retrieving their own orthographic information, whether correct or not, from the “mental lexicon”; for example, “bocabulary,” “vokaburary” or “vocabulary” would be among possible ones. Some people might visualize or try a few possible spellings to make reasonably sure that his or her solution is correct before making the final decision. Others would consult their dictionaries at hand to find the correct spelling. These three could be considered as “heuristic-retrieval,” “elaboration” and “algorithm” strategies, respectively.
What concerns us here is whether or not the heuristic-retrieval and elaboration strategies could be claimed to be psychologically plausible for word guessing in cloze procedures, and how the elaboration strategies can be interpreted in the context of word guessing phenomena. Among possible elaboration strategies there may be a “hypothesis – testing” strategy. That is, we propose that a primary distinction should be made between heuristic – retrieval and hypothesis – testing strategies with the former adapted by default and the latter as a kind of elaboration strategy. In this paper, these two strategies are operationally defined as follows:

Heuristic - Retrieval Strategy: JLEs provide target words simply by retrieving a certain word, whether correct or not, from their mental vocabulary without any specific attempt to formulate hypotheses about the target words on the basis of contextual information.

Hypothesis - Testing Strategy: JLEs provide target words by following such a procedure that they formulate and then test some specific hypothesis, whether correct or not, about the target words on the basis of contextual information.

These two will hereafter be referred to as “HR” and “HT” strategies, respectively.

With the above theoretical framework, we address the following research questions: (1) What effects do high and low contextual constraints exert on word guessing by JLEs in cloze procedures? (2) How do JLEs decontextualize contextual information as cues for word guessing in cloze procedures? (3) What kind of psychological strategies do JLEs use for word guessing in cloze procedures? (4) What relationship can be found among the proficiency level of JLEs, contextual constraints, guessing strategies and decontextualization processes? This paper aims to investigate these research questions from a psycholinguistic perspective, discussing several issues related to word guessing and word acquisition.

2. Method

2.1 Subjects

Twenty-eight students volunteered to participate in this experiment during a regular class. All of the subjects were native speakers of Japanese, and have been studying English as a foreign language for more than 6 years. They were randomly selected from two English proficiency groups, namely “high” and “low.” These two groups were divided on the basis of their scores in a simple English test which was performed three months before this experiment. Their mean scores (out of 100) and standard deviations are these: $\bar{X}=64.4$, $SD=7.0$ and $\bar{X}=42.1$, $SD=6.1$ for high and low proficiency groups. The mean difference is also significant ($t(26)=8.6$, $p<0.01$). These two groups will hereafter be referred to as the “high JLEs” and “low JLEs.”

2.2 Materials

Three short passages, independent of each other, were chosen from Hill (1988), and they were
slightly modified for the sake of the experiment (see Appendix). The three passages were all composed of three or four sentences only. Each of the three passages had three blanks which were supposed to be filled in by the same English noun word. These tests are, of course, not a conventional cloze test. Our concern is not with how well subjects could perform conventional cloze tests, but with what psychological processes may be involved in inferring words, regardless of whether the words are acceptable or not. This is the reason why we decided to use this cloze-like procedure as an experimental method. On the basis of Itagaki (1991), the three passages were also classified into the "high," "mid" and "low" contextual constraint ones. That is, the passages were judged to be high, mid and low constraint passages in the sense that the most popular guess, whether acceptable or not, was provided nearly 70%, around 50% and below 30% of the time, respectively. This criterion for the contextual constraint is adapted basically following Dubin and Olshtain (1993) and Schwanenflugel (1991).

2.3 Procedures
Subjects were provided with a four-page booklet with the experimental instructions on the first and the three cloze tests (i.e., passages) on the next three pages. They were first asked to fill the three blanks with the same word; immediately after filling the blanks, they were asked to describe whatever they thought about during word guessing as much and precisely as they could. That is, subjects' "self-reports" (i.e., "introspective reports") on word guessing were analyzed whereby certain aspects of mental processes of word guessing might be examined. We must note, however, that this method is always subject to criticism, but we must also admit that this method has always a certain methodological validity (e.g., Huckin & Bloch, 1993). This issue is far beyond the scope of the present paper.

3. Results and Discussion

3.1 Contextual Constraints
As expected, the most expected target words turned out to be the most popular words provided. That is, the most likely words, "company," "park" and "computer," were provided as the most popular ones 68%, 54% and 21% of the time, respectively (see Table 1). The notion of contextual constraint or textual support can, therefore, be viewed as a psychological variable in the sense that high constraint texts helped JELs more to provide the most likely target words in cloze tests than low constraint ones. We can suggest that texts themselves should be examined in terms of their degree of contextual constraints on word guessing, and word guessing should be related to the inherent textual constraints as well as the English ability of JELs.

A simple yet important question can be raised here: Why do some texts have inherently higher constraints on word guessing than others? One plausible answer is suggested by Dubin and Olshtain (1993, p.183). According to them, it is not only by the language command of readers, but
TABLE 1. FREQUENCIES AND PERCENTAGES OF WORDS PROVIDED FOR THREE PASSAGES

<table>
<thead>
<tr>
<th></th>
<th>High Constraint Passage</th>
<th>Mid Constraint Passage</th>
<th>Low Constraint Passage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>word</td>
<td>n.</td>
<td>%</td>
</tr>
<tr>
<td>company</td>
<td>19</td>
<td>(68)</td>
<td></td>
</tr>
<tr>
<td>business</td>
<td>3</td>
<td>(11)</td>
<td></td>
</tr>
<tr>
<td>office</td>
<td>1</td>
<td>( 4)</td>
<td></td>
</tr>
<tr>
<td>dream</td>
<td>1</td>
<td>( 4)</td>
<td></td>
</tr>
<tr>
<td>others</td>
<td>4</td>
<td>(14)</td>
<td></td>
</tr>
</tbody>
</table>

also textual constraints, which they call “textual support,” that the ease or difficulty of word guessing might be determined. They also propose that the textual support should be composed of the five linguistic and psychological elements: (a) Extratextual knowledge: the reader’s general knowledge extending beyond the text; (b) Thematic knowledge: the reader’s overall grasp of the content of this particular text; (c) Semantic 1: information extending over large discourse units in the text beyond the paragraph level; (d) Semantic 2: information available locally at the sentence or paragraph level; and (e) Syntactic: relationship within the immediate sentence or paragraph. If their conception of textual support is true, we can address two important questions. First, how much contribution does each of the five components make to the ease or difficulty of word guessing? Second, what use of each component do JLEs make in guessing words and their meanings? Further research is needed to investigate these empirical questions, which can, no doubt, provide insights into the psychological processes of word acquisition and word processing as well as word guessing.

3.2 Decontextualization of Cues

We analyzed all pieces of information which were explicitly stated in the self-reports as cues for word guessing. As a result, 176 cues were identified and classified into the above three types: “DT cues,” “II cues” and “OK cues.” This classification was done with the criterion of the agreement among the three authors. The frequencies and percentages of each type are presented in Table 2.

TABLE 2. FREQUENCIES AND PERCENTAGES OF THREE TYPES OF CUES FOR HIGH AND LOW PROFICIENCY GROUPS

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>Type of Cues ***</th>
<th>DT Cues</th>
<th>II Cues</th>
<th>OK Cues</th>
</tr>
</thead>
<tbody>
<tr>
<td>High JLEs</td>
<td>DT Cues</td>
<td>39 (46%)</td>
<td>41 (49%)</td>
<td>4 (5%)</td>
</tr>
<tr>
<td>Low JLEs</td>
<td>DT Cues</td>
<td>64 (70%)</td>
<td>24 (26%)</td>
<td>4 (4%)</td>
</tr>
</tbody>
</table>

*Note: DT - “Direct Translation” II- “Inferred Information” OK- “Outside Knowledge”
***p < 0.01 when evaluated with chi-square
Table 2 clearly indicates that there is a significant association between the two proficiency levels and the type of cues (χ² = 10.6, p < 0.01, φ = 0.2). That is, high JLEs are inclined to decontextualize textual information as cues for word guessing, while low JLEs tend to rely on DT cues in inferring target words and their meanings. These results suggest two points. First, the decontextualization can be claimed to comprise a major psychological process of word guessing as well as word meaning acquisition (Daalen-Kapteijins & Elshout-Mohr, 1981; Elshout-Mohr & Daalen-Kapteijins, 1987; Just & Carpenter, 1987). Second, high JLEs have more access to the decontextualization processes than low JLEs.

Two important questions can be brought up here: (1) Why do high JLEs make more use of decontextualization processes than low JLEs? (2) What effects does the greater use of decontextualization processes cause on the strategic aspects of word guessing? (This second question will be discussed in the next section.) A possible answer to the first may have to do with the question of "level of discourse representation" (in the sense of Johnson-Laird (1983) and Singer (1990)). Some people, probably including low JLEs, may result in a simple, surface propositional meaning of textual contents; others, probably including high JLEs, may come up with a deeper, more elaborated understanding of textual contents in the sense that it integrates "situational" information and "general" knowledge. It should be the decontextualization information that comprises a major part of such "cognitive," "situational" or "mental" representations of textual contents. This account is, of course, a speculative one, but the question here is whether or not high JLEs might have a better chance to establish a more schematic, coherent understanding of textual contents than low JLEs. Further research is needed with regard to this empirical question.

3.3 Strategies for Word Guessing

The 84 self-reports (i.e., 3 passages times 28 subjects) were also analyzed to determine which of the Heuristic-Retrieval (HR) or Hypothesis-Testing (HT) strategies was manifested. As a result, either the HR or HT strategy was identified in 76 out of the 84 self-reports, with no specific strategies revealed in others. The identification was determined with the criterion of the

<table>
<thead>
<tr>
<th>TABLE 3. FREQUENCIES AND PERCENTAGES OF STRATEGIES FOR THREE PASSAGES AND TWO PROFICIENCY GROUPS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strategy</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>HT Strategy</td>
</tr>
<tr>
<td>HR Strategy</td>
</tr>
</tbody>
</table>

*Note. HT- "Hypothesis-Testing"  HR- "Heuristic-Retrieval"
***p < 0.01 when evaluated with chi-square
agreement among three experimenters. Table 3 shows the frequencies and percentages of the strategies.

Table 3 clearly suggests two points. First, our speculative strategies can be claimed to be of psychological plausibility in that they manifested themselves 94% of the time. That is, a primary, superordinate distinction should be drawn between the heuristic type (e.g., the HR strategy) and the procedural type (e.g., the HT strategy) with regard to psychological strategies for word guessing. In this sense, Siegler and Shrager (1984) are right in suggesting that problem solving strategies should fall into the basic three types of “heuristic,” “elaborative” and “algorithmic” ones. Second, the choice between the HR and HT strategies was found to be largely dependent on the proficiency levels of JLEs and the contextual constraints of texts. That is, a significant interaction was verified among these three factors. For the high JLEs, an association was found to be highly significant among the choice of strategies and the passage constraints \( \chi^2 = 11.2, p<0.01, \gamma^2=0.41 \). In other words, high JLEs are significantly more likely to use the HT strategy than low JLEs, which is also more salient for high constraint passages than for low constraint passages. A similar association was seen for the low proficiency group, but it did not turn out to be significant.

There are two important questions: (1) Why is the HT strategy used by high JLEs more often than low JLEs? (2) Why do the high constraint passages elicit the HT strategy more often than low constraint passages? We speculate that the availability of the HT strategy may be related to the utilization of the decontextualization process discussed above. High JLEs are more likely to make use of decontextualization processes than low JLEs (see Table 2) with the result that the former can formulate a more schematic, situational representation of passage contents. This more elaborated understanding of passages might lead high JLEs to a more elaborated, procedural strategy (e.g., the HT strategy). This hypothetical reasoning may account for the significant interaction among the choice of strategies, English proficiency levels and contextual constraints. This account, thus, has to be investigated as an empirical one.

4. Conclusions

Several conclusions and implications regarding English teaching can be now drawn from the results above. First, word guessing should not be accounted for only in terms of the English proficiency of JLEs, but also in terms of the contextual constraint of texts. Teachers, therefore, have to be concerned with both of these two aspects of word guessing in preparing cloze tests and word guessing exercises. Second, decontextualization processes can be viewed as a crucial part of word guessing as well as word acquisition. We speculate that the decontextualization processes might assist JLEs in relating textual contents to their world (or schematic) knowledge with the result of a more situational, schematic representation of the textual contents. Teachers should, thus, encourage their JLEs to decontextualize textual information in such a way as to
deepen and elaborate their understanding of textual contents. Third, low JLEs tend to use more heuristic-retrieval strategies for word guessing, that is, a “less ideal” problem solver. Our future research should be focussed on why low JLEs fail to become more procedure-oriented solvers of problems. Teachers have to emphasize procedural, rather than heuristic, aspects of and processes underlying the four basic skills of speaking, listening, writing and reading.

References


Mochizuki, A. “Effectiveness of a Multiple-Choice (M-C) Cloze Test and the Number of Words in its Text.” *Annual Review of English Language Education in Japan*, 3, 1992, 33-42.


**APPENDIX**

Three (High, Mid and Low Constraint) Passages Used in the Experiment (The Most Likely Words Are “company,” “park” and “computer,” Respectively.)

**High Constraint Passage**

“Mary Johnson, a saleswoman, had just begun to work for a big (  ) that employed a lot of salesmen and saleswomen. They had to go to various parts of the country, visiting factories, offices and shops and trying to sell them the things which (  ) made. They worked very hard for the (  ) from Monday to Friday, but they had Saturday and Sunday off.”

**Mid Constraint Passage**

“Greenville, a small town in England, was full of houses and shops, but in the middle of the town there was a small (  ), which had a playground for small children. The (  ) had always been closed at six very evening, and the playground was closed at the time too. But now the town concil is discussing whether, in the summer, the (  ) should be opened till later.”

**Low Constraint Passage**

“When Joan was a college student, she majored in science, and she was especially interested in (  ). When she finished her college, she decided to work with (  ). She very much enjoyed the work in the office she joined, and soon she was able to do unusually clever things with (  ).”