Testing the Quality of Word Knowledge in a Second Language by Means of Word Associations: Types of Distractors and Types of Associations

TINE GREIDANUS
Vrije Universiteit Amsterdam/IFOTT
The Netherlands
Email: dt.greidanus@let.vu.nl

LYDIUS NIENHUIS
Universiteit Utrecht/OTS
The Netherlands
Email: nienhuis@let.uu.nl

The quality of word knowledge in a second language (L2) and the assessment of it is a growing field of interest. Read (1993) presented a test format that assesses the quality of word knowledge by means of word associations. The present study examined the development of the word knowledge of 2 groups of advanced learners of French as a L2 by means of a slightly revised version of this test format. Three aspects in particular were studied. First, there was the type of distractor most suited to the participants: distractors semantically related to the stimulus word versus semantically nonrelated distractors. Second, the role of the 3 types of associations was distinguished: paradigmatic, syntagmatic, and analytic. Third, the study linked qualitative aspects of word knowledge to quantitative aspects, by distinguishing 5 frequency ranges. The semantically related distractors were found to be more appropriate to assess the quality of word knowledge of advanced learners than the semantically nonrelated distractors. The participants showed a preference for paradigmatic responses, as expected. There was a relation between frequency and quality of knowledge: The more frequent a word, the better the knowledge of the tested aspects.

ACCORDING TO NATION (1990), KNOWLEDGE of a word can be divided into knowledge concerning its form, its position, its function, and its meaning. Thus, it is not the case that a word is either known or unknown. A word can be known in all sorts of degrees: from knowing that a given form is an existing word to knowledge including all four aspects mentioned above. These degrees of word knowledge apply to native speakers as well as to second language (L2) learners. Their vocabularies expand in breadth (the number of known words grows) but also in depth (the knowledge concerning the words already known increases, in other words, the quality of what the learner knows increases).

The majority of vocabulary tests concern breadth of word knowledge.1 There are as yet few test formats that test the quality of word knowledge in an efficient way. Read (1993) and Schoonen and Verhallen (1998) made a start by developing deep-word knowledge tests based on word associations. Both tests assess receptive word knowledge and address, in particular, knowledge about the meaning of a word, the words with which it is associated, and the collocations in which it occurs. Read looked at university students of English and tested their knowledge of “academic” words, whereas Schoonen and Verhallen tested the base word knowledge of 9- and 11-year-old Dutch children, natives as well as nonnatives. An item on Read’s test consists of a target word followed by eight other words, four of which are semantically related to the target word, and four of which are not. The L2 learner has to indicate which are the related words.

An item has the following structure:
It is obvious that the association words are film, publishing, revise, and text. In the Schoonen and Verhallen test, six words are placed round a target word. Three of them are related to the target word; the other three have a looser relationship to the target word. The pupils have to draw lines between the three related words and the target word.

The following item is an example from the test:

vrucht ape
lekker banaan glijden
(nice) (banana) (glide)
schil krom
(in) (curved)

The association words (fruit, skin, and curved) have a more inherent relationship to the target word than the distractors (ape, nice, and glide). Fruit is a superordinate, a skin is a part of a banana, and curved is a distinguishing feature of it, whereas ape, nice, and glide can be associated with banana, but in a more contextualized way.

The results of the test have clearly shown that the test format is a promising one. The two tests of 50 items each, developed by Read (1993), turned out to have high internal reliability (K-R 20 = .92 for both tests). They also showed a rather high correlation with a word test in which five words have to be matched with their meaning definition (.76 and .81). The psychometric data for the Schoonen and Verhallen test are satisfying too: The reliability for two versions (.83 and .82 at grade 3; .75 and .80 at grade 5) and the first validity checks appear to be good. As an indication of "known group validity," the test discriminates well between the two age groups (η² = .22 and between first language (L1) and L2 learners (η² = .12). The concurrent validity appears to be satisfactory as well. The correlation (corrected for attenuation) with another type of vocabulary test (a test measuring definition skill) is reasonable (.80 at grade 3, and .84 at grade 5). Moreover, this type of test is efficient in use, compared with methods such as those used by Anglin (1985), who studied the quality of word knowledge in structured interviews. So, we entirely agree with Read (1993) when he states that the test format has "potential as a tool in research on the organization of the mental lexicon and the processes by which word knowledge is acquired" (p. 359).

Apart from the nature of the associations, the test formats of Read (1993) and Schoonen and Verhallen (1998) differ on one important point. Whereas the distractors in the Read test are not semantically related to the target word (arithmetic, pole, risk, and surprise in the edit item), they are semantically much closer to the target word in the Schoonen and Verhallen test (ape, glide, and nice in the banana item). Because it was our aim to create a test to assess aspects of the deep-word knowledge of university students of French as a foreign language, the question that presented itself was: which type of distractor to choose, a distractor not related at all to the target word or a distractor semantically close to the target word? This question was the starting point of the present study. The second aspect we investigated was that of the association words. The association words in both tests were of three different types. Read and Schoonen and Verhallen did not work out this issue of the types of associations. We attempted to shed more light on those two points.

THE STUDY

The test we developed assesses certain semantic and syntagmatic aspects of the deep-word knowledge of two groups of Dutch university students of French as a foreign language. The format is comparable to the formats of Read (1993) and Schoonen and Verhallen (1998). Our study especially examined a point on which both test formats differ, namely the nature of the distractors. The choice of one of both types, that is, distractors semantically close to the target word or semantically unrelated to it, should in our view be studied more closely than is the case in the two studies mentioned above. In addition, we considered the types of word associations, a point not well developed by Read and Schoonen and Verhallen. We will give a detailed account of what is mentioned about this aspect in the literature and discuss what can be concluded from our results in this respect.

Semantically Related Distractors versus Semantically Nonrelated Distractors

Read (1993) remarked about his distractors that they "were chosen on the basis that they had no semantic link with the stimulus word. No attempt was made to create 'distraction' by using words that had some orthographic or morphological relation to the stimulus" (p. 360). He did not present the reason for this choice, however.

Schoonen and Verhallen (1998) made a different choice, which they explained as follows:

We have chosen to make the differences between correct and incorrect answers gradual ones. In the
The question of the semantic distance of the distractors from the target word rarely arises in the literature. When it does, scholars consider items in which the distractors are semantically unrelated to the target word to be easier than items in which the distractors are semantically closer to the target word (see, e.g., Bahrick & Bahrick, 1964; Dale & Baddeley, 1962; Joe, 1995; Joe, Nation, & Newton, 1998; Nagy, Herman, & Anderson, 1985).

It is indeed likely that items whose distractors are semantically related to the target word are more difficult than items whose distractors are semantically more remote from the target word. However, one may wonder whether, in addition, items with more difficult distractors do not give a more refined image of the lexical knowledge of a learner, in particular with respect to the ability to abstract from context. A test with easy, or too easy, distractors gives the learner fewer opportunities to demonstrate his or her lexical knowledge than a test with difficult distractors.

Read (1993) and Schoonen and Verhallen (1998) calculated the correlations between the deep-word knowledge tests they developed and vocabulary tests in which words must be matched with their meaning definition (Read), or in which a meaning definition must be given (Schoonen & Verhallen). As a similar, reliable vocabulary test was not available for the level we had in view, we had to find another solution. Because our aim was to develop a vocabulary test that provides an adequate idea of the mastered lexicon, in breadth as well as in depth, we decided to take as a criterion the mean of the marks obtained for a number of language tests in which vocabulary plays an important role. These tests are part of the regular curriculum of French studies at Dutch universities and are administrated after courses of 1 or 2 semesters and graded by the teachers of these courses.

The first-year students took three tests (oral comprehension and production, written comprehension and production, translation from the mother tongue into French); the third- and fourth-year students took nine tests (three oral comprehension and production tests, three written comprehension and production tests, three translation tests in both directions). The mean of these marks offered an overall idea of the language proficiency of our participants. We refer to this mean as the “average mark for language proficiency.” Our hypothesis was that a test with difficult distractors would correlate higher with this average mark than a test with easy distractors.

Types of Associations

Read (1993) distinguished three types of associations, on the basis of “preliminary drafting of items”: (a) paradigmatic (“The two words are synonyms or at least similar in meaning, perhaps with one being more general than the other”); (b) syntagmatic (“The two words are collocates that often occur together in a sentence”); (c) analytic (“The associate represents one aspect, or component, of the meaning of the stimulus word and is likely to form part of its dictionary definition” [p. 359]).

For the purposes of our study, we distinguished the same types of associations as did Read (1993). 2 The paradigmatic-syntagmatic distinction is also a generally accepted distinction in studies of word associations. The four most important types of links between a stimulus word and a response listed by Aitchison (1994)—co-ordination, collocation, superordination, and synonymy—are either syntagmatic or paradigmatic. The types of associations chosen furthermore correspond to the characteristics of words spontaneously produced by Anglin’s (1985) participants. Students are familiar with the three above-mentioned forms of relationship between words, because every monolingual dictionary, whatever its lexicographical design, presents a definition, plus paradigmatic and syntagmatic information.

We will now discuss what is known from studies of word associations of native speakers and L2 learners. These studies examine free word associations, in contrast to our study and those of Read and Schoonen and Verhallen, which asked participants to choose from eight or six possibilities offered, respectively.

One of the most striking results of word association studies was summarized by Read (1993) as follows:

One of the basic findings is that native speakers have remarkably stable patterns of word association, which can be taken to reflect the sophisticated lexical and
semantic networks that they have developed through their acquisition of the language. On the other hand, second language learners produce associations that are much more diverse and unstable; often their responses are based on purely phonological, rather than semantic, links with the stimulus words. (p. 358)

The L2 data are the reason Read opted for a more controlled form of word association.

In free word association experiments in the mother tongue, adults usually give responses from the same word class (80% for nouns, 50% for verbs and adjectives, according to Aitchison, 1994, who probably took her data from Deese, 1965). We checked whether our L2 learners also preferred to select paradigmatic responses, if they had a choice.

According to Clark (1970), syntagmatic responses are influenced to a considerable extent by the left-to-right production of sentences. Thus, stimulus words that are transitive verbs give responses that are objects, rather than subjects. We checked whether this effect could also be observed in our study.

METHOD

An experiment was carried out to examine the above mentioned aspects of deep-word knowledge. In the following sections, we will describe the design of the experiment.

Participants

Our participants were 30 first-year and 24 third- and fourth-year university students at the French departments of the two universities in Amsterdam, the Free University and the University of Amsterdam. The students had a secondary education with an average of 2 or 3 hours of French per week over a period of 6 years. They were thus advanced (first-year university students) and very advanced (third- and fourth-year university students) learners of French.

With these two groups, we were also able to study the development of the quality of word knowledge.

Tests

All the words used in the two versions of the test came from the frequency list of Juillard, Brodin, and Davidovitch (1970), which contains 5,083 words. These words are arranged in three sublists, according to usage, frequency, and dispersion. We used the frequency sublist. The Juillard list is based on a corpus of written prose texts of 500,000 words (plays, novels and short stories, essays, texts from periodicals, and scientific and technical texts). By selecting our stimulus words from a large frequency list, we were able to assess the quantity, as well as the quality, of the word knowledge of our participants. Each of our two tests contained 350 words (50 stimulus words and 300 association words and distractors). The 50 stimulus words represented 1% of the Juillard list, and the 300 association words 6% of the list. We thus assessed, albeit in different ways, the knowledge of 7% of the words in the Juillard list.

Ten content words were chosen from each range of 1,000 words, including concrete words as well as abstract words, nouns, verbs, and adjectives. It is well known (see, e.g., Juillard et al., 1970; Sciarone, 1979) that the proportion between the different word classes changes as the frequency falls. The first 1,000 words contain many grammatical words and adverbs. In the subsequent frequency ranges, the proportion of nouns grows rapidly, compared to that of verbs and adverbs. On the basis of Sciarone’s (1979) work, we selected 5 nouns, 3 verbs, and 2 adjectives from the first three ranges of 1,000 words and 6 nouns, 2 verbs, and 2 adjectives from ranges four and five.

The association words belonged to the same frequency range as the stimulus word concerned, or were more frequent. There were three types of associations: (a) paradigmatic (synonyms, sub- and superordinates, and antonyms); (b) syntagmatic (collocations and selection restrictions); and (c) analytic (defining characteristics, such as those used in dictionary definitions).

For each stimulus word, two test items were constructed. Both test items contained the same three association words (one of each type: paradigmatic, syntagmatic, and analytic). One item had three distractors that were semantically related to the stimulus word and the other item had three different distractors that were semantically nonrelated to the stimulus word. The association words and the distractors were ordered alphabetically. The participants received the instruction to choose three of the six words given for each item. The following is an example of two items for the stimulus word rive (bank).

\[
\begin{align*}
\text{rive} & \quad \text{artificiel (artificial)} & \quad \text{rive} & \quad \text{bord (bank)} \\
\text{bord (bank)} & \quad \text{côté (side)} & \quad \text{fleuve (river)} & \quad \text{gauche (left)} \\
\text{fleuve (river)} & \quad \text{vague (wave)} & \quad \text{paquet (parcel)} & \quad \text{prudent (careful)} \\
\text{gauche (left)} & \quad \text{tombe (grave)} & & \\
\end{align*}
\]
The association words are bord (paradigmatic), gauche (syntagmatic), and fleuve (analytic). The three semantically related distractors are artificiel, côté, and vague; the three semantically nonrelated distractors are paquet, prudent, and tombe.

There were two versions of the test: In test A, item 1 had semantically related distractors, while in test B the distractors of item 1 were semantically nonrelated; with item 2, the opposite was the case, and so on. Test A and test B thus both had 25 items with semantically related distractors and 25 items with semantically nonrelated distractors. The stimulus word and the association words were always the same in both tests. The items were presented in order of frequency of the stimulus word: Item 1 came from frequency range 1 (i.e., the 1,000 most frequent words), item 2 from frequency range 2, and so on up to and including item 5; item 6 came again from frequency range 1, item 7 from frequency range 2, and so forth. There were also two order variants of both tests: One half of the participants responded to the test items in the order from 1 to 50; the other half responded to them in the reverse order. This procedure was used to distribute any effect attributable to fatigue.

A preliminary version of the test was tried out with 4 university teachers of French, 2 of whom were native speakers. The items to which they responded incorrectly were modified. The test was preceded by a rather extensive instruction (three-quarters of a page) treating the different types of relations between stimulus word and association words, followed by an example item with comments. The participants were to choose three words for each item, even if they did not know the word.

The test was administered in class (first-year students) or individually (third- and fourth-year students). No time limit was imposed. The participants needed 10 to 40 minutes to complete the test. The test forms of 3 individuals who did not always mark three responses per item or who did not answer all the items were eliminated. As for scoring, one point was awarded for each correctly recognized relation.

**Hypotheses and Research Question**

Our hypotheses were as follows:

1. Third- and fourth-year students give more correct answers than first-year students.
2. Items with semantically nonrelated distractors are easier than items with semantically related distractors.
3. Items with semantically related distractors correlate more strongly with the average mark for language proficiency than items with semantically nonrelated distractors.
4. There is a relation between frequency and quality of knowledge; that is, the more frequent the word, the better one’s knowledge of different aspects of the word.

We further addressed the question of whether the results concerning the type of association related to what is known from L1 association studies.

**RESULTS**

*Results per Test Version and per Student Level*

Because we had two levels of students and two parallel versions, A and B, of the test (see section TESTS), we calculated first the mean scores and the standard deviations for the two student levels, 1 and 4, separately for the two versions, A and B, of the test, as well as for the groups of scores totaled over the two test versions. These data appear in Table 1.

We observed a small difference between the two levels for each of the test versions A and B, as well as for the total group of participants. With a maximum score of 150, the mean scores were relatively high, and the standard deviations were rather low. The sample groups were, therefore, rather homogeneous. The coefficient reliabilities of versions A and B were .76 and .63 respectively.

In order to test our first hypothesis, we conducted an ANOVA to know whether the mean scores of the students of Levels 1 and 4 differed significantly. Because we wanted to determine also whether the use of the two test versions A and B resulted in different scores, we decided to carry out a two-way ANOVA. The results of this analysis appear in Table 2.

The first finding is that the mean scores for the
two levels were significantly different at the $p < .001$ level of confidence. Therefore, one may conclude that the test had a good discriminating power at these advanced levels. Furthermore, there was no significant difference between the two versions, A and B, of the test, nor was there any evidence of an interaction between test version and student level; in other words, the version of the test did not appear to influence the relationship between student level and test performance. In the following sections, we therefore take the two versions of the test together.

**Results per Type of Distractor**

In order to test our second hypothesis, we made two versions of each test item: one with semantically related distractors and one with semantically nonrelated distractors. Each of the two versions of our test had 25 items with semantically related distractors and 25 items with semantically nonrelated distractors; these two kinds of items were distributed on an alternating basis over the two versions of the test. Table 3 shows the mean scores and the standard deviations for all the items with semantically related and semantically nonrelated distractors, for each student level group and for the total group of participants.

As expected, for both levels of students as well as for the total group of students, the items with semantically nonrelated distractors were easier than the items with semantically related distractors; the differences between the mean scores were rather large and, according to the results of the $t$-tests we conducted, these differences were significant at the $p < .001$ level of confidence. The mean scores for the items with nonrelated distractors approached, notably for Level 4, the maximum score of 75; this finding may mean that this type of distractor is less well suited for tests for the higher levels of learners.

**Correlations with an Average Mark for Language Proficiency**

In the section **Results per Test Version and per Student Level** we concluded that the test had a good discriminating power between groups of students of different levels. To determine to what extent our test is more or less suited for discriminative use within groups of students of approximately the same level, we investigated the degree of relationship between the test scores of the students of each group on the one hand and a general indication of the proficiency of these students on the other hand. To establish this proficiency, we used an average mark for language proficiency, as explained above in the section on the tests. We calculated the correlations between the mean scores of our test and the average mark for language proficiency. We also calculated correlations separately between the scores

---

**TABLE 2**

Results of Two-Way ANOVA of Mean Test Scores

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
<td>476.181</td>
<td>1</td>
<td>476.181</td>
<td>15.369</td>
<td>.000</td>
</tr>
<tr>
<td>Test</td>
<td>61.165</td>
<td>1</td>
<td>61.165</td>
<td>1.974</td>
<td>.166</td>
</tr>
<tr>
<td>Level × Test</td>
<td>17.882</td>
<td>1</td>
<td>17.882</td>
<td>.577</td>
<td>.451</td>
</tr>
<tr>
<td>Residual</td>
<td>1549.211</td>
<td>50</td>
<td>30.984</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 3**

Mean Scores per Type of Distractor and per Student Level Group (Max = 75)

<table>
<thead>
<tr>
<th>Level</th>
<th>Distractor SR</th>
<th>Distractor SNR</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Level 1</td>
<td>56.1</td>
<td>4.0</td>
<td>68.2</td>
</tr>
<tr>
<td>N = 30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 4</td>
<td>59.9</td>
<td>3.8</td>
<td>70.3</td>
</tr>
<tr>
<td>N = 24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>57.8</td>
<td>4.3</td>
<td>69.1</td>
</tr>
<tr>
<td>N = 54</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. SR = semantically related; SNR = semantically nonrelated.
for the two item types, semantically related and semantically nonrelated distractors, and the language proficiency mark, so as to answer the question in a detailed manner. Table 4 shows the correlations between the scores for the whole test and for the two item types, semantically related and semantically nonrelated distractors, as well as for the average mark for language proficiency.

At Level 1, the proficiency marks did not correlate significantly with the test scores (semantically related distractors, semantically nonrelated distractors, and total test), and at Level 4, they correlated only modestly with the semantically related distractors and the total test score. The absence of correlation at Level 1 might be attributable to the fact that the language proficiency marks within this group were rather homogeneous: a lot of 6s and 7s on a 10-point scale. At Level 4, the picture was clearer; the scores for the items with semantically related distractors correlated more highly with the language proficiency marks and, thus, seemed to be better suited for this level than for the lower level. We see, in fact, further confirmation of what was said above with respect to Table 3: The items with nonrelated distractors lead to scores that are rather high; in other words, these items are too easy for this level.

Results per Frequency Range

The test we devised was for use in assessing vocabulary knowledge in depth and in breadth. We wanted to know whether there was a relation between quality and quantity, that is, whether deep knowledge about frequent words was better established than the same kind of knowledge about less frequent words. To answer this question, that is, to test Hypothesis 4, we distinguished five levels of frequency on the basis of the frequency ranges of Juillard et al. (1970). For each frequency level represented by 10 test items, we calculated the scores of the students of both level groups. Table 5 shows the mean scores and standard deviations per student level, as well as for the total group, for the five frequency ranges.

The mean scores diminished with the decline in frequency, from frequency range 1 to frequency range 5, in line with our expectations. This finding held across the board, with one exception, but the differences were not always very important.

### Table 4

<table>
<thead>
<tr>
<th>Level</th>
<th>SR — SR</th>
<th>SNR — SNR</th>
<th>TEST — TEST</th>
<th>MARK — MARK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>.46 **</td>
<td>.79 **</td>
<td>.91 **</td>
<td>.20 *</td>
</tr>
<tr>
<td>N = 30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 4</td>
<td>.37</td>
<td></td>
<td>.89 **</td>
<td>.63 *</td>
</tr>
<tr>
<td>N = 24</td>
<td></td>
<td></td>
<td>.76 **</td>
<td>.18</td>
</tr>
</tbody>
</table>

Note. SR = semantically related; SNR = semantically nonrelated; TEST = Whole Test; MARK = average mark for language proficiency.
* p = .05, ** p = .01, *** p = .001.

### Table 5

Mean Scores of Students of the Two Levels 1 and 4 and of the Total Group of These Students for Groups of 10 Test Items, Ordered by Frequency

<table>
<thead>
<tr>
<th>Level/N</th>
<th>FR1 M</th>
<th>SD</th>
<th>FR2 M</th>
<th>SD</th>
<th>FR3 M</th>
<th>SD</th>
<th>FR4 M</th>
<th>SD</th>
<th>FR5 M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>26.9</td>
<td>1.3</td>
<td>25.3</td>
<td>1.8</td>
<td>25.4</td>
<td>2.0</td>
<td>23.7</td>
<td>1.8</td>
<td>23.0</td>
<td>2.5</td>
</tr>
<tr>
<td>N = 30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 4</td>
<td>27.3</td>
<td>1.8</td>
<td>26.8</td>
<td>1.7</td>
<td>26.2</td>
<td>1.9</td>
<td>25.4</td>
<td>1.9</td>
<td>24.5</td>
<td>1.8</td>
</tr>
<tr>
<td>N = 24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>27.1</td>
<td>1.5</td>
<td>26.0</td>
<td>1.9</td>
<td>25.7</td>
<td>2.0</td>
<td>24.4</td>
<td>2.0</td>
<td>23.7</td>
<td>2.3</td>
</tr>
<tr>
<td>N = 54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. FR = frequency range.
For Hypothesis 4, we conducted a MANOVA. We wanted to know whether the scores for the items of any frequency group from frequency group 2 through frequency group 5 were significantly lower than the scores for the items of the more frequent groups. Table 6 shows the pairs of mean scores for the items composed of words from successive frequency ranges with mean scores for the items of the next, less frequent range that differed significantly.

The results of this analysis revealed an almost regular trend for the mean scores of the words of the more frequent ranges to exceed significantly the mean scores of the less frequent ranges. Thus, there appeared to be a rather obvious relation between the frequency of the item words and their difficulty. In other words, and generally speaking, the more frequent a word, the better the knowledge of the tested aspects, at least in the present study. The difference between the mean scores of Level 4 students was not significant for the item groups of frequency range 1 and frequency range 2; a possible explanation for this might be that among learners at this near-native level the words in these two frequency ranges were equally well known.

### Results per Type of Association

As noted above, the association words used in the test represented different types of associations: paradigmatic, syntagmatic, and analytic. Each test item contained one paradigmatic association, one syntagmatic association, and one analytic association. Generalizing from what is known about mother tongue free associations and assuming that for adults paradigmatic associations are more frequently given than syntagmatic ones, and, consequently, are activated more easily, one would expect that the scores on the paradigmatic associations would be higher than those on the syntagmatic ones. For the analytic associations there were no assumptions, because this category has been absent in mother tongue association studies. The data concerning the types of associations appear in Table 7.

The means for the paradigmatic associations were, in every case, higher than those for the syntagmatic ones; the results of the tests showed that these differences were always significant. Our assumption was, therefore, confirmed for Levels 1 and 4, as well as for the group as a whole. The mean scores for the analytic associations were also higher than those for the syntagmatic ones; these differences were also significant in all cases. It appeared that the more defining associations were activated more easily or were better known than the syntagmatic associations.

In addition to the degree of difficulty or, in other terms, the availability of the types of associations, a question that can be asked is: To what extent would scores for the different types of associations correlate with each other? In other words, to what extent would an appeal be made to the same or to different types of knowledge? In this respect, it was also interesting to determine to what extent the scores for the different

---

### Table 6

Results of MANOVA Analysis of Contrasts for Mean Scores of the Items with Words from the Five Different Frequency Ranges

<table>
<thead>
<tr>
<th></th>
<th>Level 1</th>
<th>Level 4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>FR1 v FR2</td>
<td>***</td>
<td>NS</td>
<td>***</td>
</tr>
<tr>
<td>FR12 v FR3</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>FR123 v FR4</td>
<td>***</td>
<td>**</td>
<td>***</td>
</tr>
<tr>
<td>FR1234 v FR5</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
</tbody>
</table>

Note. FR = frequency range.
* * p = .05. ** p = .01. *** p = .001.

---

### Table 7

Mean Scores, Standard Deviations, and Significances of the Differences between the Means for the Three Different Types of Associations: Paradigmatic, Syntagmatic, Analytic

<table>
<thead>
<tr>
<th>LEVEL/N</th>
<th>PAR M</th>
<th>PAR SD</th>
<th>SYN M</th>
<th>SYN SD</th>
<th>ANA M</th>
<th>ANA SD</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 50</td>
<td>Level 1</td>
<td>43.6</td>
<td>2.4</td>
<td>36.7</td>
<td>3.0</td>
<td>44.1</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>Level 4</td>
<td>46.2</td>
<td>1.9</td>
<td>40.0</td>
<td>3.2</td>
<td>44.1</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>44.8</td>
<td>2.6</td>
<td>38.2</td>
<td>3.5</td>
<td>44.1</td>
<td>2.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N = 24</th>
<th>N = 54</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>43.6</td>
</tr>
<tr>
<td>Level 4</td>
<td>46.2</td>
</tr>
<tr>
<td>Total</td>
<td>44.8</td>
</tr>
</tbody>
</table>

Note. PAR = paradigmatic; SYN = syntagmatic; ANA = analytic. *** p = .001.
types of associations correlated with the average mark for language proficiency. To this end, we calculated the correlations between the scores for the three different types of associations and the average mark for language proficiency (see Table 8).

The correlations were rather small and not significant; consequently, we could only conclude that there was no significant relationship among the three types of associations, nor between these types and the average mark for language proficiency. It may be that our type of test appealed to three types of relatively dissimilar knowledge, and none of the three correlated significantly with the average mark.

At the end of the section treating the type of associations, we raised the question of the possible influence of the order of stimulus word and syntagmatic association in discourse, given that this effect exists in mother tongue association studies. For 18 of our 50 test items, we found a left-to-right order between stimulus word and syntagmatic association (e.g., main [hand] → droite [right]); in 29 cases, there was a right-to-left order (e.g., foudre [lightning] ← coup [knock, blow]); in 3 cases, finally, both possibilities were open (e.g., guérir [to heal] and malade [ill]: Malade can be both subject and object of the intransitive and transitive verb guérir). We examined whether the left-to-right order yielded more correct answers than the right-to-left order, as is the case with free associations, according to Clark (1970). Table 9 shows that this was not the case.

These results might be explained by the fact that in a word association test the response is spontaneous, whereas in our test the participants could ponder their answers at leisure.

DISCUSSION AND CONCLUSIONS

We formulated four hypotheses, none of which was falsified. As to Hypothesis 1, we found that the third- and fourth-year students had more correct answers than the first-year students. The differences were, however, less clear-cut than might have been expected, owing to the fact that the test was rather easy for Level 1 and too easy for Level 4. A more difficult test might have shown more differences between the two levels. It is evidently not hard to devise a more demanding test, but it will be impossible to study the breadth and depth of vocabulary knowledge concurrently, because there is no general frequency list of French words that goes beyond the 5,000-word limit. Moreover, according to Nation (1990), “The 5,000-word level is on the boundary of high- and low-frequency vocabulary” (p. 79). The knowledge of individual learners rapidly diverges beyond this level. Assessments as to the vocabulary size of learners vary enormously, but it is clear from our results that the receptive vocabulary size

<table>
<thead>
<tr>
<th>Level</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>.74</td>
<td>.11</td>
<td>.72</td>
<td>.08</td>
<td>.30 (ns)</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>.80</td>
<td>.09</td>
<td>.79</td>
<td>.08</td>
<td>.54 (ns)</td>
</tr>
</tbody>
</table>
of our first-year and advanced university students was much larger than 5,000 words.

The results pertaining to Hypothesis 2 were, not surprisingly, the clearest: The items with semantically nonrelated distractors were markedly easier than the items with semantically related distractors. The latter type of distractor, thus, appeared to be the appropriate one for our purpose, the assessment of the quality of word knowledge of advanced learners of a L2. Semantically related distractors may have another advantage. Read (1993) gathered verbal report data from 8 participants to gain insight into the way they proceeded in taking the test. The more proficient learners tried to guess the meaning of an unknown stimulus word and were often quite successful at selecting the correct responses. The reason was that “the associates not only were linked individually to the stimulus word but often also had semantic connections among themselves” (p. 366). A test with semantically related distractors makes guessing harder, because there are semantic connections between associates and distractors and not just between associates.

Hypothesis 3 stated that the items with semantically related distractors would correlate better with the average mark for language proficiency than the items with semantically nonrelated distractors. The results concerning this hypothesis were influenced by the homogeneity of the Level 1 group and the ceiling effect of the test. A more demanding test and a more heterogeneous Level 1 group could lead to more clear-cut evidence about the differences between the two types of test.

There exists a clear relationship between the frequency of a word and the quality of the knowledge of it (Hypothesis 4). Apparently, vocabulary grows in breadth as well as in depth, with the knowledge of the frequent, earlier acquired words being qualitatively better than that of the less frequent, more recently acquired words. These results could also be taken to suggest that a frequency list, in this case the Juilland list, correlates to a psycholinguistic reality. The Juilland list did not play a role in the teaching the students had received. Furthermore, the Juillard corpus on which the list was based is small by present-day standards (500,000 words). Thus the critical remark Wesche and Paribakht (1996) made about “the appropriateness of the word frequency list approach to vocabulary sampling for measurement of vocabulary size” (p. 16) does not seem justified in this case.

We formulated a general research question in addition to the four hypotheses. This research question addressed the relationship between results from L1 association studies and our results. Our assumption concerning the preference of our learners for paradigmatic responses was confirmed, but we found no evidence to support a left-to-right preference in our learners. As to our selection of the three types of associations in line with Read (1993), we can conclude that this choice proved to be a happy one: The three types appear to be different in terms of lexical knowledge. One wonders why Read (1994) reduced the number of associations to two types: paradigmatic and syntagmatic. Considering our results, this modification was neither necessary nor justified.

The word association test format as used in this study indeed turned out to be a promising one for assessing the quality of word knowledge of rather advanced L2 learners. We hope our study has shed more light on the development over time of deep-word knowledge by linking depth to breadth and by examining the different kinds of knowledge that it is possible to have about words. The issue of the development of the quality of word knowledge of less frequent words deserves further empirical study.

ACKNOWLEDGMENTS

We wish to thank Vivien Hoekstra, who composed the test under the supervision of the first author and who took care of the administration of the test. We are also very grateful to Tom de Wolf of the University of Amsterdam, who allowed us to administer the test to his students. Finally, we would like to thank three anonymous reviewers for their useful comments on earlier versions of this article.

NOTES

1 For an overview of different testing procedures, see Wesche and Paribakht (1996) and Joe et al. (1998).
2 Schoonen and Verhallen (1998) distinguished the following three types: paradigmatic relations (superordination, subordination, synonymy), partonomic relations (constituents), and decontextualized syntagmatic relations (defining perceptual characteristics, inherent characteristics, and/or means-end relations).
3 As regards the meaning of these terms, we adopt the definition as given by Deese (1965):
Paradigmatic associations are those in which the stimulus and response fit a common grammatical paradigm and syntagmatic associations are, in general, sequential elements or at least elements which usually occupy different positions within phrases or sentences.
4 Compare to Meara (1980) and Meara (1982).
This finding can be caused by two factors: (a) There are often fewer paradigmatic and analytic associations than syntagmatic associations for a given word; and (b) the frequency of a collocation, considered as a lexical unity, is evidently lower than that of the two composing words separately. Knowing the words deuil ([mourning], frequency range 3292, or joindre [to join], frequency range 3009), does not imply knowing collocations with those words, even with easy words such as prendre (to take) in prendre le deuil (to go into mourning) or efforts (efforts) in joindre les efforts (to join forces).

Anv, however, this greater difficulty of syntagmatic associations is in keeping with the experience of teachers that even for advanced learners of a L2 the knowledge and use of collocations and common combinations remains one of the greatest problems.

To this category now belong synonyms of the target word and words that express one aspect of its meaning. Categories 1 and 3 in Read (1993) are thus combined, with the old category 1 being reduced to synonyms and the old category 3 being reduced to words of the same part of speech as the target word.

**REFERENCES**


---

**Introducing Perspectives, a New MLJ Regular Feature**

Beginning with Volume 86, 2, 2002, the *MLJ* will inaugurate a new regular feature, *Perspectives*. Shaped and edited by Heidi Byrnes, *Perspectives* seeks to foreground the “discussion” portion of the *Journal’s* editorial mission by addressing issues beyond the specific classroom or the specific research study that, nevertheless, affect teaching, learning, and research.

*Perspectives* will appear twice a year, in the summer and winter issues. Each topic will be addressed through solicited, reviewed contributions, a primary paper, and a series of commentaries. The first two columns will address enrollment issues at the college and the K–12 levels respectively. Readers are encouraged to join the discussion through their own contributions to the Readers’ Forum in subsequent issues.