What Makes Comprehension of Derivatives Easier?: Pedagogical Implications for the Order of Vocabulary Introduction

Naoki SAKATA
Graduate School, Kumamoto Gakuen University

Abstract

To increase learners’ vocabulary size, using morphological knowledge is useful. It has been suggested to lead to a 1.6-2 increase in a vocabulary size. There are three kinds of morphological knowledge, abbreviated RK, SK and DK. RK is relational knowledge, referring to the knowledge of the relationship between morphologically related words (e.g., happy and happiness). SK (syntactic knowledge) refers to the explicit knowledge of syntactic functions of suffixes (e.g., -ness). With DK (distributional knowledge), learners become aware that to which part of speech a certain suffix attaches. A previous study (Sakata, 2006) has shown that Japanese EFL learners experience developmental stages through RK→SK and that those learners have rather sufficient RK. This present study investigated how four independent variables (the presence of contexts, the semantic relatedness of the base words and derivatives, the difficulty of suffixes and the difference in frequency between base words and derivatives) affect the RK ratio as a dependent variable. The results showed that contextual help and utilizing semantic relatedness needed a minimal vocabulary size. Suffix difficulty didn’t affect the RK. Frequency affected all vocabulary size groups’ RK.

1. Introduction

EFL learners usually face a great number of words to learn. Native speakers of English are said to have a receptive vocabulary of around 20,000 word families. Coping with this great number of words is normally demanding for learners of English. Laufer (1992) suggested that 3,000 word families are a minimal vocabulary for academic reading. Even so, even this goal is not so easy to reach.

There are many ways to increase vocabulary. For example, we can memorize words directly through using word cards or word lists. We can build up our vocabulary through inferring the meaning of the novel words in texts. Besides them, there is another way to increase the learners’ vocabulary, using morphological knowledge. For example, when learners face a novel word happiness, they can figure out the meaning by compounding the meaning of the base word happy and the meaning of the derivational suffix –ness. In the case of speaking or writing, they have to know every part of the word. For listening and reading comprehension, however, they don’t have to memorize every part (Melka, 1997). Therefore,
if they can separate morphemes in derivatives, they can guess the meaning. If so, what amount of word knowledge can be built up through the use of morphological knowledge?

Laufer (1992) claimed that 1,000 word families equal 1,600–2,000 word items. This suggests that the ability to infer the meaning of derivatives through form relatedness to the base words leads to a 1.6 increase in a learners’ vocabulary size. Nagy and Anderson (1984) suggested that 12.8% (roughly one eighth) of the different types in the American Heritage corpus (Carroll, Davies and Richman, 1971) were derivatives. These studies pointed out that guessing the meaning of derivatives is quite effective in facilitating vocabulary increase.

Derivatives include those compounded with prefixes (e.g., unhappy) and those compounded with suffixes (e.g., happiness). This present study only deals with the latter derivatives and their derivational suffixes.

We now know the importance of the ability to infer the meaning of derivatives. Next, we have to classify the morphological knowledge related to the ability. Tyler and Nagy (1989) showed that native speakers of English experience a developmental sequence of morphological knowledge from Phase 1 (relational knowledge, RK) through Phase 2 (syntactic knowledge, SK) to Phase 3 (distributional knowledge, DK). Let us explore the details of these three phases. First, RK enables speakers to extract the base form of a certain derivative when facing derivatives. For example, when speakers having RK face happiness, they can figure out the meaning by extracting the base form happy from the derivative. This phase doesn’t require speakers to have the explicit knowledge of suffixes (e.g., -ness). Suffixes usually provide the syntactic functions of derivatives and those functions don’t always have to be known when reading or listening is taking place since that information has been already provided by the contexts surrounding the derivatives.

SK refers to the knowledge related to the understanding of the syntactic information certain suffixes contain. In the example of happiness, SK refers to the knowledge of the function of the suffix -ness, which changes words into nouns. This knowledge enables us to infer the part of speech of novel words and can facilitate the comprehension of the whole word. Finally, with DK, learners become aware that to which part of speech a certain suffix attaches. For example, this knowledge allows us to know that the suffix -ness only attaches to adjectives, not to nouns or verbs. Tyler and Nagy (1989) showed that native speakers’ acquisition order was RK→SK→DK.

In the case of Japanese EFL learners, to what extent do they have morphological knowledge? Mochizuki and Aizawa (2000) conducted an experimental research with respect to SK through questions that asked Japanese high school and university students into which part of speech a certain suffix changes words. The rate of correct answers was 67%. This result showed that Japanese EFL learners are in the developmental stage in terms of SK. Sakata (2006) investigated the RK and SK of the Japanese EFL learners of university students. In the result, although the rate of SK was 58.8%, which showed that their SK was in the developmental stage, they showed somewhat sufficient RK. When they knew the meaning of base words, they could answer their derivatives at the rate of 80% or above of their
vocabulary levels. For example, learners having a 3,000 word-family vocabulary could answer more than 80% of the derivatives of the 3,000 word-family level when they knew the base words. These results revealed that university level EFL learners have relatively sufficient RK, and suggested that EFL learners also have a developmental order from RK to SK.

This previous study also demonstrated that the difference of the rate of RK due to whether they had the explicit suffix knowledge of a certain derivative or not was around 6%. In the example of *happiness*, the knowledge that the suffix *-ness* turns words into nouns doesn’t necessarily affect the comprehension of the derivative *happiness* very much.

In summary, Sakata (2006) revealed that university level EFL learners have a rather sufficient level of RK. However, what influences RK has not been explored. RK can be considered important and sufficient knowledge for comprehension skills since the syntactic information (SK) contained in suffixes isn’t always needed for comprehending derivatives. This present study will reanalyze the data of the results of Sakata (2006) and investigate the cause of facilitation of RK. The results of this analysis will hopefully benefit the syllabus design in terms of when certain derivatives are to be introduced and in which order they are to be instructed. The presence of contexts, the semantic relatedness of the base words and derivatives, the difficulty of the suffixes attached to the derivatives and the difference in frequency between base words and derivatives are to be researched as causes affecting RK.

2. Method

2.1 Participants

Seventy-four low-intermediate level EFL students in a Japanese university took part in this experiment. The subjects were separated by the results of the Vocabulary Levels Test (VLT, Schmitt, Schmitt and Clapham, 2001) into three groups called 1,000 word-family holders, 2,000 word-family holders and 3,000 word-family holders (hereafter, 1,000 holders, 2,000 holders and 3,000 holders). The numbers of the subjects for each group were 17, 27 and 30, respectively.

2.2 Test materials

Three tests (the suffix test, a vocabulary test and the derivative test) asking explicit suffix knowledge and knowledge of base words and derivatives were conducted. The suffix test asked the subjects to indicate into which part of speech a certain suffix changes words. The test form was similar to Mochizuki and Aizawa (2000). VLT was used to measure subjects’ base word knowledge. The derivative test was a self-made multiple-choice test with 6 answers for each question that asked the meaning of the 16 derivatives of 2,000 and 3,000 word-family levels whose base words had been asked in the VLT.
2.3 Analysis

The purpose of this experiment is to investigate how the four factors (the presence of contexts, the semantic relatedness of the base words and derivatives, the difficulty of the suffixes attached to the derivatives and the difference in frequency between base words and derivatives) influence RK. To start the analysis, how to measure RK is a problem. In this study, an RK ratio was calculated by dividing the number of correctly answered derivatives by the number of whole derivatives of which base words had been answered correctly in VLT. For example, if the base words elect and admire were answered correctly but introduce was not answered correctly, and also if the derivatives electorate and introduction were answered correctly but admiration was not answered correctly, then the ratio would be 1/2 = 0.5 (The item introduction was not incorporated in the calculation since a base word, introduce, had not been answered correctly). We will explore how the four factors above as independent variables influence RK ratio as a dependent variable through a series of 3 x 2 ANOVA tests including the investigation of the interaction of the vocabulary size to each variable. The influence of each independent variable was analyzed separately. Let us here explore what the four independent variables are.

2.3.1 The presence of contexts

In the derivative test (cf. 2-2), every derivative was tested with and without contexts. It can be claimed that when they have base word knowledge, it is easier to comprehend the meaning of derivatives with contextual help. When contexts are provided, we can analyze the sentence containing the target derivatives to figure out the function of derivatives. In the example of happiness, when we know the meaning of the base word happy but do not know the function of -ness, we can guess that happiness is related to happy because the forms of the two words are similar. However, we could not find out the syntactic function of happiness because we don’t know the function of -ness. On the other hand, when the target word happiness is surrounded by a one-sentence context, it can be considered that they can figure out the syntactic function of the derivative by analyzing the structure of the sentence. How this kind of contextual help affects the derivative comprehension will be investigated. Contexts were made within high frequency words (within 2,000 word-family vocabulary) in order to ensure that the context would be understandable.

2.3.2 The semantic relatedness of the base words and derivatives

Nagy and Anderson (1984) measured this variable by using a judge and 6 grades. In line with their study, this present study used three raters to evaluate by 6 grades how deeply a certain derivative and the base word were related. Using the results, 16 derivatives were divided into two groups, one of which had deeper relationships with the base words (11 derivatives: admiration, noisy, manufacturer, importable, administer, arrangement, introduction, development, awareness, stability and durable) and the other of which had shallower relationships (5 derivatives: electorate, preference, nervous, taxpayer and pursuit). We will explore how this difference in the semantic relatedness affects the RK
2.3.3 The difficulty of the suffixes attached to the derivatives

The difficulty to answer the suffixes of the twelve derivatives out of 16 was investigated by the suffix test (cf. 2-2). The number of suffixes was 9 since some suffixes were attached to more than one derivative. There seemed to be a large difference between the scores of six suffixes scoring under 64% and those of three suffixes scoring over 81%. Therefore, seven derivatives with the difficult suffixes were assigned as difficult-suffix words and five derivatives with the easy suffixes were assigned as easy-suffix words.

2.3.4 The difference in frequency between base words and derivatives

It can be assumed that if derivatives are less frequent than the base words, learners have more difficulty comprehending the derivatives due to a relative lack of opportunity to encounter them. In this experiment, the frequencies were investigated using Kilgarriff's 1997 data given by analyzing BNC (British National Corpus). The ratio of the difference of the frequency was calculated as frequency of derivative / frequency of base word. For example, the frequency per 100 million words of the derivative electorate is 898. The frequency of the base word elect is 4329. Then, the ratio is 898 / 4329 = 0.207. If the ratio is 1, the frequency of the derivative and the base word is identical. There seemed to be a cut-off point between the 10 derivatives with the ratio under 0.6 and the 6 derivatives with the ratio over 0.6. Therefore, the former derivatives will be called low frequency derivatives and the latter high frequency derivatives.

3. Hypotheses

This study will test the following 4 hypotheses:

1. When contexts are provided, it is easier to comprehend derivatives through the use of base word knowledge. This tendency is stronger for low vocabulary holders because they complement the lack of vocabulary knowledge with contextual help.

2. Semantically transparent derivatives are easier to comprehend. This tendency is stronger for low vocabulary holders because high vocabulary holders have suitable number of words so that even if derivatives are semantically opaque, they can guess the meaning with the use of their vocabulary knowledge.

3. Easy suffix derivatives are easier to comprehend. Suffixes are important components of derivatives. Therefore, they will affect derivative comprehension greatly. This tendency is stronger for low vocabulary holders because they don’t have suitable vocabulary knowledge, so the lack of suffix knowledge damages comprehension of derivatives more greatly than for high vocabulary holders.

4. High frequency derivatives are easier to comprehend. Low vocabulary holders will be affected more
greatly than high vocabulary holders because they might have seldom encountered low frequency words.

4. Results and Discussion

4.1 The presence of contexts

There was no main effect of contexts presence ($F=2.111$, $p>.05$). However, the interaction (vocabulary size $\times$ presence of contexts) was observed ($F=3.223$, $p<.05$). The graph below showed that the subjects having larger vocabulary (3,000 holders and 2,000 holders) were assisted with contextual help, while the subjects holding less vocabulary (1,000 holders) weren’t, in contrast to hypothesis 1. Actually, high vocabulary holders can make use of contextual help.

<table>
<thead>
<tr>
<th>Vocabulary size</th>
<th>N</th>
<th>Mean with contexts (S.D.)</th>
<th>Mean without contexts (S.D.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3,000 holders</td>
<td>30</td>
<td>0.93 (0.09)</td>
<td>0.80 (0.14)</td>
</tr>
<tr>
<td>2,000 holders</td>
<td>27</td>
<td>0.81 (0.19)</td>
<td>0.73 (0.21)</td>
</tr>
<tr>
<td>1,000 holders</td>
<td>17</td>
<td>0.65 (0.32)</td>
<td>0.72 (0.19)</td>
</tr>
</tbody>
</table>

This result indicates that to make use of contexts, learners need a minimal vocabulary size. The contexts actually helped learners, but only more advanced learners. The contexts were made to facilitate the syntactic analysis of derivatives. The results suggest that this facilitation actually happened with larger vocabulary holders, but not with smaller vocabulary holders. What does this mean?

First, the contexts might have been too difficult for smaller vocabulary holders. As described in the Method section, contexts were made within high frequency (easy) words in order to be understandable. However, 1,000 holders have less than 2,000 word-family vocabulary. In effect, 2,000 word-family level words in the contexts might have prevented them from utilizing the contexts. It might be one cause.

Second, smaller vocabulary holders might have been less proficient in analyzing syntactic structures. We can guess that grammatical competence and lexical competence correlate to some extent.
Therefore, 1,000 holders might have had less grammatical knowledge so that they had difficulty in analyzing the structures around the tested derivatives. That could have led to poor results in comparison with high vocabulary holders.

Anyway, whatever the reason, it can be said that utilizing contexts needs some minimal vocabulary. It was found in contrast to hypothesis 1 that contexts don’t complement a lack of vocabulary knowledge. The threshold to utilize contextual help might be between 1,000 and 2,000 word families. However, we have to interpret the results cautiously. The contexts were made with easier words in this study. In authentic materials, learners might have to have more vocabulary to utilize the surrounding contexts.

4.2 The semantic relatedness between base words and derivatives

There was no main effect of the semantic relatedness ($F=0.751, p>.05$). However, the interaction (vocabulary size $\times$ semantic relatedness) was observed ($F=3.465, p<.05$). The graph below showed that larger vocabulary size group (3,000 holders) could make use of the semantic relationship between morphologically related words, while lower vocabulary holders (2,000 holders and 1,000 holders) couldn’t. This was in contrast to hypothesis 2. In reality, high vocabulary holders can effectively use semantic relatedness.

<table>
<thead>
<tr>
<th>Vocabulary size</th>
<th>N</th>
<th>Mean for deeper (S.D.)</th>
<th>Mean for shallower (S.D.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3,000 holders</td>
<td>30</td>
<td>0.88 (0.09)</td>
<td>0.81 (0.14)</td>
</tr>
<tr>
<td>2,000 holders</td>
<td>27</td>
<td>0.75 (0.19)</td>
<td>0.82 (0.17)</td>
</tr>
<tr>
<td>1,000 holders</td>
<td>17</td>
<td>0.65 (0.25)</td>
<td>0.74 (0.25)</td>
</tr>
</tbody>
</table>

How can we interpret this result? The lexical network might be concerned with this result. Qian and Schedl (2004) showed that vocabulary breadth and depth correlate—in other words, when learners’ vocabulary grows, the network in learners’ lexicon gets denser. Words are more closely bound together in higher vocabulary holders’ lexicons.

In this present study’s result, words in a word family might be grouped together in 3,000 holders’
mental lexicons. For example, *happy, happier, happiest, happily and happiness* could be grouped together in the lexicon. This effect of grouping might have affected the results. Three thousand holders' words in the mental lexicon are grouped in a word-family manner so that they have access to the relationships within word families. However, in the mental lexicon of 2,000 and 1,000 holders, the relationships within word families are sparse so that they might have had difficulty utilizing the semantic relationships between derivatives and base words.

We predicted in hypothesis 2 that higher vocabulary holders would be able to figure out the meaning of derivatives that are semantically distant from base words due to their high vocabulary knowledge. We may now predict that those things will happen with more advanced learners. Learners with 5,000 word families or more vocabulary might bear different results. This should be investigated in future research.

### 4.3 The difficulty of suffixes

Neither main effect (*F*=.578, *p*>.05) nor interaction (*F*=1.505, *p*>.05) was observed. This contradicted hypothesis 3. The component we had considered important didn’t affect the RK ratio at all.

<table>
<thead>
<tr>
<th>Vocabulary size</th>
<th>N</th>
<th>Mean for easy suffix (S.D.)</th>
<th>Mean for difficult suffix (S.D.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3,000 holders</td>
<td>30</td>
<td>0.94 (0.13)</td>
<td>0.94 (0.09)</td>
</tr>
<tr>
<td>2,000 holders</td>
<td>27</td>
<td>0.73 (0.30)</td>
<td>0.84 (0.20)</td>
</tr>
<tr>
<td>1,000 holders</td>
<td>17</td>
<td>0.73 (0.25)</td>
<td>0.70 (0.31)</td>
</tr>
</tbody>
</table>

How can we interpret this result? This result supports Sakata’s (2006) result showing that lack of suffix knowledge doesn’t necessarily lead to lack of derivative comprehension. With respect to comprehension skills such as reading or listening, the results showed that learners don’t always need to have explicit suffix knowledge. Suffixes usually supply syntactic information. However, in a real communicative situation, this information is provided by the context. Therefore, we don’t always have to have proper knowledge of suffixes when we have base word knowledge because the semantic
meaning can be provided with base words similar in forms with derivatives. As far as relational knowledge (RK) is concerned, it can be said that suffix knowledge might not be necessary.

However, we need to be cautious to some extent. In this study, a simple one-sentence context was used so that it was not so difficult to analyze the sentences. When learners have to read more complex materials, suffix knowledge might affect the derivative comprehension. This should be investigated in the future.

4.4 The difference in frequency between base words and derivatives

Only the main effect ($F=20.352, p<.01$) was observed (for the interaction, $F=2.128, p>.05$). For every group, if the frequency of the derivative was relatively higher in comparison with that of the base word, the RK ratio was higher. Regardless of the vocabulary size of subjects, more frequent derivatives were easier to comprehend. This result partially disproved hypothesis 4. Frequency actually affects the RK ratio. However, the influence was maintained through all of the subject groups.

<table>
<thead>
<tr>
<th>Vocabulary size</th>
<th>N</th>
<th>Mean for more frequent (S.D.)</th>
<th>Mean for less frequent (S.D.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3,000 holders</td>
<td>30</td>
<td>0.95 (0.10)</td>
<td>0.81 (0.11)</td>
</tr>
<tr>
<td>2,000 holders</td>
<td>27</td>
<td>0.80 (0.24)</td>
<td>0.75 (0.15)</td>
</tr>
<tr>
<td>1,000 holders</td>
<td>17</td>
<td>0.82 (0.18)</td>
<td>0.62 (0.22)</td>
</tr>
</tbody>
</table>

How can we interpret this result? First, we can assume that learners have to face derivatives at least once to comprehend them. If they don't encounter a derivative at all, it would be difficult to figure out the meaning. In another study (Sakata, forthcoming), it was found that novel derivatives were comprehended at the rate of 50% if they have base word knowledge. The rate was below that of normal derivatives (about 80%, novel and familiar derivatives are mixed in this condition). Summing up the results of this present and other studies, it can be suggested that novel (unfamiliar) derivatives are more difficult to comprehend. Preferably, they should have faced a derivative once to comprehend it.

Second, it was found that frequency effect is maintained across all learner groups. Larger
vocabulary holders cannot complement the frequency effect by their vocabulary size. We can assume that derivatives of relatively low frequency are rarely encountered even for larger vocabulary holders, and the frequency effect was therefore maintained. If subjects having higher levels vocabulary (5,000 – 10,000 word-family holders) had been included, they might bear different results. The frequency effect might diminish. This should be investigated in the future.

5. Conclusion and Pedagogical implications

Four things were found in this study. First, to utilize contextual information, learners have to have a minimal vocabulary size. Second, some lexical network in learners’ mind is important in comprehending derivatives. Third, suffix difficulty doesn’t affect derivative comprehension. Fourth, frequency effect influences all of the learners’ groups’ derivative comprehension.

When introducing derivatives to learners, we have to be careful of the order of the introduction. What we found is that regardless of the suffix difficulty, semantically more related and more frequent derivatives are easier to comprehend. In the syllabus design, we have to keep in mind that this order should be kept so as to facilitate learners’ relational knowledge development. If we ignore this order, learners’ development of morphological knowledge might be delayed. Semantically distant and less frequent derivatives are more difficult to comprehend, so these derivatives should be introduced later. This sort of treatment might be compatible with a natural acquisition order.

Selected references


