

Journal of Universal Language 18-2, September 2017, 89-104
DOI 10.22425/jul.2017.18.2.89
eISSN 2508-5344 / pISSN 1598-6381

Principles of Developing Lexicon in Artificial Languages: Based on an Empirical Study

Jaeyoung Kim

Seonam University, Korea

Abstract

Even though the importance of the lexicon has not always been focus of mainstream language acquisition research, it is widely accepted that the lexicon is an integral element of language. Therefore, it was argued in this paper that effective strategies to build up the lexicon should be carefully considered when artificial languages are developed. To support this argument, an experiment of two groups (i.e., an experimental and control group) of

Jaeyoung Kim

Department of Liberal Arts, Seonam University
7-111 Pyeongchon-gil, Songak-myeon, Asan-si, Chungnam, Korea 31556
Email: jykim@seonam.ac.kr

Received 10 August, 2017; Revised 2 September, 2017; Accepted 13 September, 2017

Copyright © 2017 Language Research Institute, Sejong University

Journal of Universal Language is an Open Access Journal. All articles are distributed online under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

university students was carried out and it was found that vocabulary significantly affected their reading comprehension. Even though in the preliminary vocabulary test, no statistically significant differences were detected between the two groups, the experimental group obtained statistically higher scores in reading comprehension than the control group in English official test.

Considering the results of the present research, it was proposed that the lexicon may play a crucial role in improving reading abilities in artificial languages. Hence, when developing artificial languages, it is better to create a set of vocabulary that has commonality in many natural languages and is unmarked phonologically, morphologically, and syntactically. With this carefully chosen set of vocabulary, it is expected that users of the artificial languages would have better reading comprehension.

Keywords: vocabulary, reading comprehension, artificial languages, word order, syllable types

1. Introduction

A number of scholars have been advocating the importance of vocabulary in language education, and Wilkins (1972: 111) is one of them who argued that; vocabulary is more important than grammar in conveying meaning. With the rise of the communication approaches in the 1970s, the role of words that had long been ignored started to be highlighted (Thornbury 2002: 13-14). Words play an important role in communication, but its importance is often neglected when opening an English course in academic setting (Davies & Pearse 2000: 59-60).

In the same argument, words may also have equally heavy influence on reading comprehension in artificial languages or constructed languages. This study was to find out whether artificial

language users can improve their communication skills by learning a carefully built vocabulary list. To do that, we had a comparative analysis of 3 artificial languages, Esperanto, Ido, and Unish, and thereby propose a guideline for creating lists of artificial languages.

This study aims to investigate the effectiveness of vocabulary-focused teaching in general English courses at universities, in comparison with the conventional teaching method which focuses on grammar and reading comprehension, and then apply it in building word lists for artificial languages. Research questions of this study are as follows: (i) How does the vocabulary-based teaching model, compared with the conventional problem-solving one, help students get higher test scores?; (ii) How will vocabulary-focused teaching affect reading comprehension?; (iii) How does vocabulary learning of artificial languages improve students' reading comprehension in general?

2. Theoretical Background of Research

2.1. Vocabulary

Most scholars argue that knowing vocabulary means knowing the meaning and use of the words. Davies & Pearse (2000: 60) maintained that when learners are acquiring a new word, they must know not only the meaning of the word, but also its usage in communication, pronunciation, spelling, and grammar. That is, knowing the meaning of a word is just a starting point and we should know how to use it in real communication. Thornbury (2002: 15-16) also said that learning vocabulary in any languages at the most basic level is knowing the form and meaning of words, and learning new

words is not simply to know their meaning but to know their pronunciation, structure, usage, and many other things in comprehensive manners.

2.2. Correlation between Vocabulary and Reading

Ha (2012) divided her class students into 2 groups; a listening-oriented group and a reading-oriented group. The reading-oriented group involved in vocabulary learning got higher scores in both listening and reading in post-class assessments and the composite scores of the reading-focused group turned out higher as a result. In the experiment of comparing the vocabulary-focused class with the grammar-focused class for reading (Lee 2011), both classes showed a significant improvement in the reading tests.

3. Research Methods

3.1. Research Subject

Participants in the experiment were students of a 4-year university located in the central region of South Korea. They were from different majors, and total number of students was 40. The experiment was carried out over 2 different student groups, the experimental and the control group.

3.2. Research Material

The Barron's Essential Words for the TOEIC (Lougheed 2014)

was used in the classroom experiment. The description of the book on the publisher's homepage reads that it combines 600 commonly used words and is designed to meet the most-up-to-date test trends.¹

3.3. Research procedure

The experimental group and the control group both used the same textbook. At the beginning of the semester, lectures on morphology such as prefixes, suffixes and etymology were given to the experimental group, and students were briefed how they would learn vocabulary for the semester.

In the meantime, the control group was also given, at the beginning, an introduction of how the class would proceed during the semester. The students were told to do the homework from textbook.

On the 3rd week of the semester, both groups took a preliminary vocabulary test. Throughout the semester both the experimental and the control group took achievement tests twice. A total of 60 questions were presented in each achievement test.

4. Data Analysis

4.1. Test Results

The empirical data was analyzed using SPSS statistical processing program with the significance level $p < .05$.

¹ Refer to <<http://barronseduc.com/1438077297.html>>.

4.1.1. Preliminary Vocabulary Test Results

The preliminary vocabulary test was conducted to find out the language proficiency level of both experimental and control group.

Table 1. Preliminary Vocabulary Test Results

	N	M	SD	<i>t</i>	<i>p</i>
Experimental group	30	5.67	4.47	-.953	.347
Control group	10	7.25	4.80		

Table 1 shows the results of the preliminary vocabulary test for the experimental group and the control group. The average test score of the experimental group was 5.67 and the standard deviation was 4.47. And the average score of the control group was 7.25 and the standard deviation was 4.80. The score gap was 1.85 and the average score of the control group was slightly higher. However, the significance probability was .347, higher than .05. Therefore, the average score gap of the two groups was not statistically significant. In other words, there was no significant difference between the experimental group and the control group. They can be regarded as a statistically homogeneous group.

4.1.2. Achievement Test Result

4.1.2.1. Total Scores of the Achievement Tests

The data was analyzed to compare the total score of the achievement test.

Table 2. Total Scores of the Two Achievement Tests

	N	M	SD	<i>t</i>	<i>p</i>
Experimental group	30	62.93	21.49	.380	.706
Control group	10	60.00	19.89		

Table 2 compares the composite scores of the experimental group and the control group. The average score of the experimental group in the first and second achievement tests was 62.93 and the standard deviation was 21.49. And that of the control group was 60.00 and the standard deviation was 19.89. The average test score gap between the experimental group and the control group was 2.93 and the average score of the experimental group was higher. However, there is no statistically significant difference between the average scores of the two groups since the significance probability of the both were .706.

4.1.2.2. Reading Comprehension Scores

Tests were carried out to evaluate the different reading comprehension scores between two groups.

Table 3. Reading Comprehension Scores

	N	M	SD	<i>t</i>	<i>p</i>
Experimental group	30	91.37	10.16	4.605	.001*
Control group	10	62.25	19.12		

* $p < .05$

As shown in Table 3, the average reading comprehension score of the experimental group was 91.37 and the standard deviation was 10.16. And that of the control group was 62.25 and the standard

deviation was 19.12. The score gap between the two groups was 29.12 points. The average score of the experimental group was higher, and this is statistically significant because the probability of the significance of the both was .001, smaller than .05. In other words, it was statistically significant that the reading comprehension scores of the experimental group were higher than those of the control group. What it implies is that the experimental group with vocabulary lessons gained higher scores of statistical significance in the reading test.

The score gaps of the achievement tests in this study were found to have no statistical significance between the two groups in contrast with the study by Ha (2012). However, in the field of reading comprehension, vocabulary-focused teaching brought a statistically significant increase in reading scores of the experimental group, just as demonstrated in the studies by Ha (2012) and Lee (2011).

4.2. Esperanto, Ido, and Unish Vocabulary

4.2.1. Developing Lexicon

This paper deals with three artificial languages: Esperanto, Ido, and Unish. Esperanto, which is considered to be the most commonly used artificial language, has mainly stemmed from European languages like Latin, Spanish, French, German, and English. Ido, which is communicable with Esperanto and thus considered dialectic of the same language, has vocabulary derived from six languages including English, French, Spanish, German, Russian, and Italian.² Contrary to the Esperanto and Ido that have been developed from natural languages only, Unish created by a Korean research team has

² <https://en.wikipedia.org/wiki/Comparison_between_Esperanto_and_Ido>

derived words in reference of Esperanto as well as 14 natural languages such as English, Spanish, Portuguese, Italian, French, German, Russian, Korean, Chinese, Japanese, and Arabic. Words are selected from these languages in accordance with seven criteria: commonality, simplicity, diversity, clarity, convenience in pronunciation, cultural connectivity, and complexity.³

In terms of the number of words, Esperanto has approximately 15,000 words.⁴ Ido, also known as an improved Esperanto, has about 14,000 words.⁵ Lastly, Unish consists of around 13,000 words.⁶

4.2.2. Linguistic Traits

4.2.2.1. Phonological Traits

Heo (2015) investigated the vowel system of natural languages ranging from 3 vowel system to 7 vowel system, and found out that 92 languages have a 5 vowel system accounting for 20.40%, highest out of the total 451 languages. All of the three artificial languages that this paper deals with have a 5 vowel system.

It is better that words of artificial languages are described in Roman alphabets on the basis of a one-to-one correspondence of spelling and pronunciation for convenience's sake (Chung 2001).

Especially in terms of the correspondence of spelling and pronunciation, Esperanto is considered to have almost a one-to-one correspondence⁷. Esperanto consists of 23 consonants and 5 vowels⁸.

³ <http://www.unish.net/introduce/e_unish_Summary.jsp>

⁴ <<https://www.esperanto-panorama.net/vortaro/eoen.htm>>

⁵ <<http://idolinguo.org.uk/idan.htm>>

⁶ <http://www.unish.net/search/e_unish_Dic_Search.jsp>

⁷ <https://en.wikipedia.org/wiki/Esperanto_orthography>

⁸ <<https://en.wikipedia.org/wiki/Esperanto>>

The double consonant /j/ and the half vowel /i̇/ are written in the same letter *j*, so its one-to-one correspondence is allegedly almost perfect. Ido⁹ consists of 21 consonants and 5 vowels, and also has 2 letters that have 2 different sounds. The letter *j* is pronounced /z/ and /d͡z/ and *x* is pronounced /ks/ and /gz/. Unish is composed of 25 consonants and 5 vowels. Here are five exceptions: the letter *q* is pronounced /kw/ and *x* is pronounced /ks/. Also, the pronunciation symbol /ʃ/ is expressed as *sh*, /ŋ/ is *ng*, and /t͡ʃ/ is expressed as *ch* (Chung 2004).

Davis (2002) examined the syllable structure of an artificial language. All natural languages have CV syllables, consisting of just one consonant preceding a vowel. As it is the syllable structure that appears in all languages, we may assume that CV would be one of the most preferred syllable types in artificial languages as well. However, if a word consists only of CV syllable structure, it could become 3 syllables (CVCVCV) or 4 syllables (CVCVCVCV) and be uncomfortably long. In the case of natural languages which have only CV syllables, they employ 2 ways to solve the problem of inconveniently long words. First, make words with long syllables and drop consonants from the first syllable leaving it a vowel-only syllable. Second, use intonation and make it a tone language. However, these two methods are not practical in artificial languages because it is not easy to learn words with long syllables and using pitch to distinguish words. They would make the language acquisition even more difficult.

Here are some examples of triple onsets of each artificial language. In Esperanto, *skiribi* ('to write') and *sklavo* ('slave')¹⁰ have triple onsets. Likewise, Ido also has *skiribar* ('to write') and *sklavo*

⁹ <https://en.wikipedia.org/wiki/Ido_language>

¹⁰ This data is from Oostendorp (1999: 63).

(‘slave’)¹¹ which have triple onsets. Unish also has triple onsets such as *skrambl* (‘scramble’) and *strait* (‘straight’)¹². As we can see, these three artificial languages have complex syllable structure.

4.3.2.2. Types of Word Order

Dryer (2005) made a table about word order, as presented in Table 4, especially in declarative sentences. There are six possible word orders of the three elements Subject (S), Object (O), and Verb (V). 1228 languages were considered in this box.

Table 4. Types of Word Order (Adapted from Dryer 2005: 330)

Type	Number of language	Percentage of language (%) ¹³
Subject + Object + Verb	497	40.47
Subject + Verb + Object	435	35.42
Verb + Subject + Object	85	6.92
Verb + Object + Subject	26	2.12
Object + Verb + Subject	9	0.73
Object + Subject + Verb	4	0.33
Lacking dominant word order	172	14.00
Total	1228	99.99

¹¹ <<http://idolinguo.org.uk/idan.htm>>

¹² <http://www.unish.net/search/e_unish_Dic_Search.jsp>

¹³ Percentage of language is adjusted from the original table. Round off the numbers to the nearest hundredths.

As shown in Table 4, the most common word order in natural languages goes as subject (S) + object (O) + verb (V). However, the three artificial languages have subject (S) + verb (V) + object (O) word order in declarative sentences. Given this, it may be a good idea to apply the subject (S) + object (O) + verb (V) word order which is the most common in natural languages into artificial languages.

5. Implications

The purpose of this study was to investigate the effect of vocabulary-focused learning on the scores of Achievement tests by carrying out a classroom experiment. The outcome of this experiment shows that competence of vocabulary affects reading comprehension. From this, we can say that artificial language speakers can develop their reading comprehension by learning carefully selected vocabulary. Three research questions of this paper are presented in the introduction. The first question was whether and how does the vocabulary-centered teaching model have effects of improving test scores, compared to the general problem-solving lessons. The preliminary vocabulary test on the experimental group and the control group turned out that the two groups were homogeneous showing not so much difference in their vocabulary knowledge. Following the course of 10 weeks or so, the experimental group showed an improvement in total composite scores, but it was of no statistically significance. This was different from what Ha (2012) had found in her study.

The second question was how the vocabulary-based lesson affects reading comprehension scores. In this experiment, there was a

statistically significant improvement in the reading comprehension score of the experimental group. The outcome is in line with the previous studies by Lee (2011) and Ha (2012) that had found a significant increase in reading scores likewise.

The third question was to find out how to improve the reading comprehension ability through vocabulary acquisition in artificial languages. As demonstrated in this experiment, vocabulary learning has helped students with their reading comprehension skills. This leads to a question whether there should be a standard or a guideline to select right kinds of vocabulary to help the learners improve their reading skills. While Esperanto and Ido have borrowed words from mostly European languages, Unish has derived words from a wide range of natural languages and an artificial language such as Esperanto. Vocabulary selection process of Unish was made in accordance with 7 criteria of commonality, simplicity, diversity, clarity, convenience in pronunciation, cultural connectivity, and complexity.

The three artificial languages tried to make one-to-one correspondence between spelling and pronunciation as much as possible, but none have achieved an absolute one-to-one correspondence. Therefore, when building a new list of vocabulary for an artificial language, it would be better to take into consideration one-to-one correspondence between spelling and pronunciation. In terms of phonological features, a 5 vowel system is the most common in natural languages, and all of the three artificial languages have a 5 vowel system as well. The CV syllable is the most common syllable structure in natural languages. However, it is not easy to learn words with long syllable or using tone to distinguish a phoneme. Because of the various problems mentioned above, though, some degree of complexity is required. The three artificial languages

mentioned in this paper have complex syllable structure.

Finally, as for the word order, the most common type in natural languages is in the order of subject (S), object (O) and verb (V). All three artificial languages surveyed in this paper, on the other hand, have employed subject (S) – verb (V) – object (O) word order. Therefore, in the future when adopting a word order for an artificial language, it is worth considering the order of subject (S), object (O), and verb (V) as in most of natural languages.

6. Conclusions

In this experiment, it was proposed that lexical knowledge is a very important factor in reading comprehension in natural languages. Likewise, the lexicon plays a crucial role in the artificial languages when developing reading comprehension. Therefore, artificial languages need to have carefully selected principles of developing the lexicon. As mentioned above, it is more desirable to develop vocabulary that shares common linguistic features in natural languages and is phonologically, morphologically, and syntactically unmarked.

In a future comparative study of natural languages and artificial languages, it would be meaningful to analyze more natural languages other than English. It would also be interesting to include words that are inflectional and with multiple meanings when building word lists for an artificial language.

References

- Chung, Y-H. 2001. Borrowing for a Universal Language. *Journal of Universal Language* 2, 24-33.
- Chung, Y-H. 2004. English, Unish, and an Ideal International Language: From a Perspective of Speech Sound and Writing System. *Journal of Universal Language* 5.2, 21-36.
- Davies, P & E. Pearse. 2000. *Success in English Teaching*. Oxford: OUP.
- Davis, S. 2002. Syllable Structure for an Artificial Language Based on Universal Principles. *Journal of Universal Language* 3.1,1-13.
- Dryer, M. 2005. 81 Order of Subject, Object, and Verb. In M. Haspelmaht et al. (eds.), *The World Atlas of Language Structures* 330-333. New York: OUP.
- Ha, M-A. 2012. Effects on the Improvement of TOEIC Scores of College English Learners Using Listening-focused and Reading-focused Teaching Methods. *Foreign Languages Education* 19.4, 323-348.
- Heo, Y. 2015. A Research on Sound Pattern in Relation to the Types of Vowel Systems-focusing on 3 to 7 vowel systems. *The Language & Culture* 11.2, 319-349.
- Lee, M-K. 2011. The Effects of Grammar-focused and Vocabulary-focused Teaching in TOEIC Reading Classes. *Modern Studies in English Language & Literature* 55.2, 155-177.
- Lougheed, L. 2014. *Essential Words for the TOEIC*. New York: Barron's.
- Oostendorp, M. 1999. Syllable Structure in Esperanto as an Instantiation of Universal Phonology. *Esperanto Studies* 1, 52-80.
- Thornbury, S. 2002. *How to Teach Vocabulary*. London: Person Longman.

Wilkins, D. 1972. *Linguistics in Language Teaching*. London: E. Arnold.