KOREAN JOURNAL OF ENGLISH LANGUAGE AND LINGUISTICS

The Study of "Loanwords" in Elementary School English Textbooks based on the 2015 Revised National Curriculum

Jiyong Lee (Seoul National University)

Jiyong Lee
Research Professor,
Seoul National University
Tel: 02-880-6020
jiyonglee0123@gmail.com

ABSTRACT

Lee, Jiyong. 2022. The study of "loanwords" in elementary school English textbooks based on the 2015 Revised National Curriculum. *Korean Journal of English Language and Linguistics*, 169-184.

As a preliminary attempt to define an authentic list of English source words for loanwords, this study investigated how frequently, widely, and evenly the 200 source words listed in the 2015 RNCE are used in elementary school English textbooks. For this purpose, a corpus was created based on 10 elementary school textbooks and their supplementary e-textbooks. Examination of the distributional properties of the 200 source words in the corpus revealed two findings. First, elementary school English textbooks are not homogeneous with respect to the frequency of these source words. Second, out of the 200 source words, 26 were identified as representative source words of elementary school English textbooks. The usage coefficient of these source words is equal to or greater than 96, their dispersion figure is greater than .60, and their corresponding loanwords are true cognates. This indicates that 26 source words can be useful resources for Korean elementary school students learning English.

KEYWORDS

2015 Revised National Curriculum of English, source word for loanword, elementary school English textbook, coefficient of usage

1. Introduction

Four years have passed since the 2015 Revised National Curriculum of English (henceforth, the 2015 RNCE) was first implemented in the Grade 3 classroom at elementary schools in Korea. Among the more prominent changes brought by this curriculum is the guideline about the utilization of source words for loanwords from English.[1] According to this guideline, as many as 50 out of the 200 source words in Table 1 can be used in each of the 20 grade-level textbooks, without being counted as new words to learn.

Table 1. The List of Source Words for Loanwords of the 2015 RNCE

alarm, album, alcohol, amateur, ambulance, apartment, arch, bacon, badminton, bag, banana, belt, bench, biscuit, bonus, box, bus, butter, cabinet, cake, calcium, camera, camp, campaign, campus, card, carol, carpet, catalogue, center, champion, channel, chart, cheese, chicken, chocolate, click, coat, coffee, comic, computer, course, court, crayon, cream, cup, data, diamond, diet, disc, doughnut, drama, dress, drill, drum, echo, elevator, elite, energy, essay, event, fashion, feminist, fence, festival, fiction, film, fork, gallery, game, gas, golf, gown, graph, guard, guitar, gum, hamburger, harmony, highlight, hint, hormone, hotel, image, interior, internet, interview, issue, jacket, jam, jazz, juice, jump, kangaroo, kiss, kiwi, laser, league, lemon, lobby, magic, manual, marathon, market, mask, medal, media, melon, member, menu, message, model, motor, mystery, news, notebook, okay, opera, orange, oven, page, panda, parade, partner, party, pen, percent, piano, pie, pilot, pipe, pizza, plastic, plug, program, project, quiz, radio, recreation, rehearsal, ribbon, robot, rocket, rugby, salad, sample, sandwich, sauce, scarf, scenario, schedule, section, seminar, service, set, shirt, skate, sketch, ski, snack, soup, spaghetti, sponsor, sport, spray, spy, staff, star, steak, stereo, studio, style, sweater, tank, taxi, team, technique, technology, television, tennis, tent, terror, ticket, toast, tomato, topic, towel, track, truck, vaccine, veil, video, villa, violin, virus, vision, waiter, website, wine, yacht

This RNCE guideline is based on two assumptions: 1) the source words listed in Table 1 correspond to English loanwords that are commonly used in everyday life, and 2) the relationship between these source words and their loanword counterparts is so transparent that they should not be counted as new vocabulary items to learn (NVILs) for Korean learners of English. In other words, Korean learners are already expected to have access to the concepts conveyed by the source words through their corresponding loanwords. These assumptions, however, have been challenged recently. Lee's (2020) corpus-based investigation showed that a considerable number of source words in Table 1 are ones whose Korean counterparts (loanwords) are not of high frequency and wide range. In fact, 41 loanwords appeared less than 50 times in the Sejong tagged corpora, comprising written texts of six genres (with 1,323,068 sentences and 34,446,300 tokens of morphemes) and spoken texts of three genres (with 216,718 sentences and 1,618,529 tokens of morphemes). Moreover, 26 loanwords appeared in the Sejong corpora in three genres at most. This finding directly contradicts the initial assumption that the source words in Table 1 are based on loanwords that are commonly used in everyday life. Lee (2020) also pointed out a considerable number of cases in which the relationship between a source word and its Korean equivalent lacks transparency, due to phonological/morphological changes, e.g., *seupocheu < sport* or *interieo < interior decorating*, or lack of a one-to-one correspondence in meaning or usage, e.g., *daieoteu < diet*. It has been generally acknowledged that loanwords whose derivation involved a drastic phonological, morphological, or semantic change are difficult for L2 learners to recognize and use as useful resources in their learning of the target language (Lado 1955; Masson 2013; Rubin 1987; Uchida 2001).

---

[1] According to Haspelmath (2009: 36-37), a loanword is "a word that at some point in the history of a language entered its lexicon as a result of borrowing", and "the word that served as a model for a loanword" is its source word.

Another premise behind defining the words in Table 1 as source words in the 2015 RNCE is that they should be used somewhat frequently in curriculum-based English textbooks. If there is a remote possibility that a word in Table 1 can be used as a source word in any textbook, there is no reason to define it as a source word in the national curriculum. Following this line of reasoning, in order to investigate the eligibility of, or lack thereof, the source words in Table 1, Lee (2021) investigated how frequently, widely, or evenly they are used across four subcorpora that comprise texts taken from 13 middle school English textbooks, 11 high school English textbooks, 9 high school English I textbooks, and 9 high school English II textbooks. It was found that 19 source words never appear in any textbook, and as many as 79 source words occur less than 10 times in the subcorpora; 24 source words appear only in one subcorpus. Such findings suggest that more than 110 source words in Table 1 may not be eligible.

Lee's (2020, 2021) work on the eligibility of the 200 source words in Table 1 and their application in English textbooks shows that it is necessary to define a much more constrained set of English source words if they can serve as useful resources for Korean learners of English. However, the selection of reliable English source words is not possible without knowing how frequently, widely, and evenly the source words in Table 1 occur in elementary school English textbooks. Assuming that the loanwords available to elementary school students differ from those known to secondary school students, this study aims to answer the following research questions:

    1) Are elementary school English textbooks comparable in terms of the frequency of source words listed in Table 1?
    2) What are the representative English source words in elementary school English textbooks?

In answering the first question, a preliminary to the second one, the present work is not concerned with the number of types of source words that are introduced, but the number of tokens. Since the 2015 RNCE allows every grade-level English textbook to use up to 50 out of the 200 source words in Table 1 as non-NVILs, it is expected that elementary school English textbooks use a similar number of types of source words, but how many tokens of these source words are used depends on the amount of effort textbook writers put into helping students utilize readily available resources for themselves. The more tokens of source words are used in a textbook, the smaller the learning burden is for students. Regarding the second question, the distributional statistics of source words as well as their authenticity are dealt with. Source words can be considered genuine only when they occur frequently, widely, and evenly in the textbooks and are used to mean what they are originally intended to convey as such.

2. Loanwords and L2 Acquisition

When learners acquire new L2 words at the early stages of L2 learning, they map them to pre-existing concepts or L1 equivalents because their conceptual and semantic systems in the L1 have already been developed. (Schmitt 2000, Takač 2008). Put simply, L1 is utilized during the process of accessing the meaning of L2 words, during which the degree of similarity between the L1 and L2 affects the level of difficulty in acquiring new L2 words (Jiang 2002, Nation 2001, Takač 2008). Because greater similarity between the new L2 word and its equivalent is claimed to result in easier acquisition of the new L2 word, loanwords can be considered to be particularly facilitative of L2 vocabulary learning. This claim has been supported by numerous studies (Ard and Homburg

　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　171

Jiyong Lee
The study of "loanwords" in elementary school English textbooks
based on the 2015 Revised National Curriculum

1983, Banta 1981, Lee 1958, Palmberg 1987). In fact, Nation (2003) proposed that loanwords are a very effective strategy for vocabulary expansion in the learning of English.

Focusing on the similarities and differences between the L1 and L2, much research has attempted to investigate the reason why certain L2 features are more difficult to learn than others. Such contrastive analyses postulate that similarities between the two result in positive cross-linguistic transfer, and differences lead to negative transfer. Based on this assumption, loanwords have been considered to be a useful learning tool (Ard and Homburg 1983, Banta 1981, Daulton 1998, Lee 1958, Palmberg 1985). For instance, Lee (1958) conducted one of the first investigations of the effects of English loanwords on the learning of English as foreign language (EFL), and found that cognacy characteristics of English loanwords increase learners' motivation, thus having a positive effect on L2 learning. Similar findings were obtained regarding the beneficial effects of English loanwords for native speakers of Spanish (Rodigriguez 2001). Daulton (1998) also found that high-frequency English loanwords in Japanese assisted the learning of their corresponding English source words. From an opposite perspective, differences between the loanwords and their source words have been argued to interfere with L2 learning (Kent 1999, Martin 2004, Sheperd 1996, Simon-Maeda 1995). Such interference effects would most likely manifest when the loanword and its corresponding source word differ in orthography. Because of the globalization of English and the steady increase in the number of loanwords entering the language, it is likely that EFL learners will continue to face interference problems during L2 learning.

The scholarly debate on the learning burden of loanwords have given rise to numerous studies that investigated the factors that determine whether a loanword facilitates or hinders L2 learning (Uchida 2001, Daulton 2008, Masson 2013). Five types of loanwords and their source words have been proposed (Uchida 2001): 1) true cognates, 2) convergent cognates, 3) divergent cognates, 4) close false friends, and 5) distant false friends. True cognates, whose primary meanings are equivalent in the L1 and L2, constitute a very low learning burden, thus facilitating L2 learning. Convergent cognates and divergent cognates are distinguished in terms of whether the loanwords have a more restricted or extended meaning than their corresponding source words. Close false friends are words that are similar conceptually, but have clearly different meanings in the L1 and L2. Distant false friends are words that completely differ in meaning in the two languages. With the exception of true cognates, the degree of cognacy between the loanword and its source word has been claimed to determine whether the other four types of cognates may have a beneficial or negative effect on L2 learning. Besides semantic aspects, other factors that affect the cognacy of a loanword and its source word include the way the loanword was coined, its similarity with the source word in terms of pronunciation and usage, and the level of difficulty L2 learners perceive its relationship to its source word. It should be noted that even though cognates are easier to learn than non-cognates because of the relatively identifiable form-meaning connection, they may be more difficult to use in context (Rogers, Webb and Nakata 2015).

In an attempt to investigate the relationship between English loanwords and their source words and its potential effect on English learning, Lee (2020, 2021) conducted corpus-based analyses of 200 English source words in the 2015 RNCE that are claimed to be used in everyday life. Because these words are excluded from the list of NVILs, Lee (2020) claimed that a more rigorous justification is necessary for specifying the criteria on which the said source words are selected. In fact, findings of the study indicated that a considerable number of the source words out of the 200 in the 2015 RNCE had corresponding Korean loanwords that are not of high frequency and wide range. In other words, many of the loanwords are not in fact used in everyday life, as originally assumed. Moreover, many cases of loanwords and the source words that serve as their models were found to lack transparency due to phonological/morphological changes or a lack of a one-to-one correspondence in meaning or usage. As a follow-up study, Lee (2021) conducted a more scrutinized examination of the 200 source words in the RNCE, looking into

their frequencies and overall distributions based on several corpora of 13 middle school English textbooks and 29 high school English textbooks. Findings showed that only less than half of the words under investigation—87 to be exact—may satisfy the necessary conditions to be included as source words in the RNCE.

3. Method

The purpose of the study was to find out how the English source words listed in the 2015 RNCE are used in English textbooks for elementary school students. More specifically, the frequency and distribution of such source words were investigated. Accordingly, a corpus was built using all of the English textbooks available to elementary school students. Then the frequency, range, and usage coefficient of each English source word in the 2015 RNCE were examined. The following provides a schematic description of the procedural steps that the present study took, from the construction of the corpus to the corpus-based analytical methods of the study.

3.1 Construction of the Corpus

For the purpose of constructing a more comprehensive corpus, texts were taken from all of the English textbooks officially allowed by the Korean government to be used for elementary school students. According to the 2015 RNCE, these textbooks are divided into two groups: lower-level English textbooks for Grades 3 and 4 students and upper-level English textbooks for Grades 5 and 6 students. The former textbooks can include no more than 240 NVILs, and the latter textbooks, 260 NVILs. Because textbooks of the two levels are different in vocabulary size, it is natural to think that different source words are used in textbooks of different levels. Thus, data were drawn from the five lower-level textbooks and the five upper-level textbooks.

Each textbook of the study had an accompanying e-textbook. Therefore, in compiling the corpus for the present study, texts were collected from each of the 10 textbooks and their accompanying e-textbooks. This means that there were 10 compilations, each of which constituted a subcorpus. For instance, the subcorpora compiled from the lower-level materials were termed ESEL1, … ESEL5, and the subcorpora compiled from the upper-level materials, ESEU1, … ESEU5.

WordSmith version 7 was used to count the number of tokens and the occurrences of the source words used in the 10 subcorpora, whose results are presented in Table 2.

Table 2. Subcorpora Word Count

Level	Textbook	Type	Token	Source Word
Lower	ESEL1	241	23,275	35(814)*
	ESEL2	230	17,706	40(817)
	ESEL3	236	22,254	32(1,074)
	ESEL4	240	12,335	37(546)
	ESEL5	232	12,963	42(517)
Upper	ESEU1	467	43,574	53(886)
	ESEU2	488	43,442	66(958)
	ESEU3	477	24,395	59(666)
	ESEU4	465	39,817	59(848)
	ESEU5	471	29,200	58(610)

* The number outside ( ) is the number of types and that inside ( ), the number of tokens.

According to the 2015 RNCE, 240 is the maximum number of new English word types that can be introduced in a lower-level textbook, and 260 in an upper-level textbook. All the textbooks comply with this requirement. However, there appears to be a considerable variation in the number of tokens used in the textbooks, depending on the numbers of chants, songs and other supplementary materials included. In spite of this great variance, there seems to be little difference in the number of source words for loanwords, particularly with regard to the types of the source words. The number of the source words ranges from 32 to 42 for the lower-level textbooks, and 53 to 66 for the upper-level textbooks.

3.2 Analytical Framework

The majority of research on high frequency word lists are based on the frequency, range, and dispersion of the occurrence of such words (Leech, Rayson and Wilson 2001, Nation 2004, Nation and Waring 1997). It is worth noting that the frequency of a word is not the sole factor in determining whether a word should be included in a resource list of high frequency words. In fact, a word should occur frequently across a wide range of texts (range) and should be evenly distributed (dispersion). While the frequency of a word does not necessarily have to be the same across different texts, a word should appear in one way or another in *most* of the different texts.

First proposed by Juilland and Chang-Rodriguez (1964), the coefficient of usage, Juilland's *U*, is one of the most common ways to determine whether a word should be included in a high frequency word list or a frequency dictionary. This coefficient is calculated by multiplying a word's frequency in a corpus by a coefficient of dispersion, *D* (Davis and Gardner 2010). Also known as Juilland's *D*, this index measures the dispersion of a word's subfrequencies over *n* equally-sized subcorpora of a corpus. Based on the standard deviation of these subfrequencies, a value of 1 indicates that the occurrence of a word is evenly distributed over the subcorpora, and a value of 0 indicates that the occurrences of a word are found only in a single subcorpus. In the case in which the value of Juilliand's *D* is 1, *U* is equal to the total corpus frequency. The following formula is used to calculate Juilland's *D*:

Juilland's *D* = 1 - (coefficient of variation / SQRT(number of subcorpora - 1)

According to Juilland and Chang-Rodriguez (1964: LXVIl), Juilland's *U* assumes that "coefficients of word usage aim to predict or approximate word occurrence in an 'ideal,' that is, perfectly representative and perfectly unbiased sample". Based on this assumption, a word type that appears in a number of different texts of a corpus with an equal probability would be more representative of the basic vocabulary of a language than one that appears uniquely or erratically in one of these texts. When a word has an even distribution, the coefficient of usage should reflect the total frequency in the corpus. On the other hand, if a word occurs only in one subcorpus, the coefficient of usage should be scaled downwards from its total frequency because the total frequency might be an 'inflated' result from accidents in sampling of a given subcorpus.

A critical drawback of Juilland's *U* for the purpose of the present study lies in the assumption regarding subcorpora size. In other words, as formulated above, it is applicable only when the subcategories have equal sizes. Juilland and Chang-Rodriguez (1964: XXV) argued that category sizes should be equal since "too many difficulties arise in the weighting of genres," although there are many situations in which category sizes are unequal, either by necessity as in this work or by inevitable circumstances. Therefore, it is not possible to use Juilland's *U* in this work.

An alternative to Juilland's U was proposed by Carroll (1970). This alternative can be applied to cases with unequally-sized subcorpora, since the index of dispersion is not based on raw frequencies, but the probabilities of a word in the several subcorpora. The only condition is that the category sizes are large enough to yield reasonably good estimates of true probabilities. MSE, HSE, HSE1, and HSE2 are large enough to meet this condition.

The following is a step-wise calculation of Carroll's D_2 and U_m:

Let n = the number of categories (subcorpora);
    f_j = the frequency of a given word-type in category j (j = 1, 2, ..., n);
    F = the total frequency of the word in the corpus = $\sum_j f_j$;

    s_j = the number of tokens in category j;
    N = the total number of tokens in the corpus = $\sum_j s_j$;

    p_j = the proportion of tokens in category j that are instances of the given word-type = f_j/s_j;
    P = $\sum_j p_j$. (P may take any positive value; it is not in general equal to unity.)

Then:
    $H = \log P - \sum_j p_j * \log p_j/P$; ($p_j * \log p_j = 0$ for $p_j = 0$).

    $D_2 = H/\log n$ = the index of dispersion;
    $s_{pmax} = s_j$ for the category that has the largest value of p_j;
    $f_{min} = F * s_{pmax}/N$;
    $U_m = (1{,}000{,}000/N) * (F * D_2 + (1 - D_2) * f_{min})$

As seen in the last part of the calculation, the metric of U_m is scaled in terms of a standard corpus of one million words.

4. Result and Discussion

4.1 Testing for Homogeneity of the Frequency of Source Words for Loanwords

The first research question of this study is to see if elementary school English textbooks are homogeneous in terms of the frequency of source words for loanwords. To answer this question, the following null hypotheses were tested:

H_0 1: The five lower-level English textbooks, ESEL1~5, are homogeneous with regard to the frequency of source words for loanwords.
H_0 2: The five upper-level English textbooks, ESEU1~5, are homogeneous with regard to the frequency of source words for loanwords.

Chi-square tests for the two hypotheses are shown in Tables 3 and 4. Since $\chi(4)$, $p < .001$, H_0 1 is rejected, indicating that the lower-level textbooks significantly differ in the frequency of source words. The same result is obtained for the upper-level textbooks ($\chi(4)$, $p < .001$).

Table 3. Chi-square Tests for Elementary School Lower-Level English Textbooks

Textbook	Source Word	Others	Token	χ^2	p
ESEL1	814	22,461	23,275		
ESEL2	817	16,889	17,706		
ESEL3	1,074	21,180	22,254	59.363	.000**
ESEL4	546	11,789	12,335		
ESEL5	517	12,446	12,963		

*$p < .05$, **$p < .001$

Table 4. Chi-square Tests for Elementary School Upper-Level English Textbooks

Textbook	Source Word	Others	Token	χ^2	p
ESEU1	886	42688	43574		
ESEU2	958	42484	43442		
ESEU3	666	39151	39817	243.490	.000**
ESEU4	848	23547	24395		
ESEU5	610	28590	29200		

*$p < .05$, **$p < .001$

Given the statistically significant results above, Bonferroni post hoc tests were conducted. Table 5 presents the pairs of textbooks that show a significant difference based on Bonferroni corrections.

Table 5. Multiple Pair-wise Comparisons

Textbook	Textbook	χ^2	p	corrected p
	ESEL2	32.829	.000	.000
ESEL1	ESEL3	5.536	.000	.000
	ESEL4	18.947	.000	.000
ESEL3	ESEL5	13.330	.000	.000
ESEU1	ESEU3	14.818	.000	.000
	ESEU4	13.953	.000	.000
ESEU2	ESEU3	3.813	.000	.000
	ESEU4	97.369	.000	.000
ESEU3	ESEU4	213.709	.000	.000
	ESEU5	16.095	.000	.000
ESEU4	ESEU5	96.629	.000	.000

Regarding the rejection of H_0 1, four pairs, ESEL1-ESEL2, ESEL1-ESEL3, ESEL1-ESEL4 and ESEL3-ESEL5, show a significant difference, indicating that the proportion of source words for loanwords relative to the total number tokens is too small in ESEL1 and too large in ESEL3. With respect to the rejection of H_0 2, seven pairs, ESEU1-ESEU3, ESEU1-ESEU4, ESEU2-ESEU3, ESEU2-ESEU4, ESEU3-ESEU4, ESEU3-ESEU5 and ESEU4-

ESEU5, show a significant difference, indicating that the proportion of source words for loanwords is too small in ESEU3 and too large in ESEU4.

4.2 Representative Source Words for Loanwords in Elementary School English Textbooks

The second research question of the study dealt with which source words for loanwords are representative of elementary school English textbooks. First, it should be noted that a considerable number of the source words in the RNCE are not used in the textbooks. Based on a scrutinized investigation, the following 84 words never occur in any of the subcorpora (Table 6).

Table 6. Source Words Not Used in Any of the Elementary School English Textbooks

alarm, alcohol, amateur, ambulance, arch, bonus, calcium, campus, catalogue, champion, channel, chart, click, data, diamond, diet, disc, doughnut, drama, drill, echo, elite, essay, feminist, fiction, film, gallery, gas, golf, harmony, highlight, hint, hormone, image, interior, issue, jazz, kiss, laser, league, lobby, manual, marathon, medal, media, member, message, motor, mystery, oven, page, partner, percent, pipe, plug, radio, recreation, rehearsal, rugby, scenario, schedule, section, seminar, service, sketch, sponsor, spray, spy, staff, stereo, style, tank, technique, technology, terror, toast, topic, track, vaccine, veil, villa, vision, waiter, wine

With the exception of *ambulance*, *champion*, *doughnut*, *drama*, *film*, *marathon*, *medal*, *oven*, *radio*, *toast*, *vaccine*, *waiter*, the words in Table 6 have corresponding loanwords that are generally less commonly used by students from Grades 3 to 6. Therefore, it can be concluded that some of the words in Table 6 are accidentally missing in the elementary school English textbooks, and others are not suitable for these textbooks.

With regard to the remaining 116 English sources words listed in the 2015 RNCE that do appear in the subcorpora, Table 7 summarizes all of the information needed to rank these source words in terms of their frequency, range, and usage coefficient.

Table 7. Distributional Properties of Source Words in Elementary School English Textbooks

Rank	Source Word	Frequency	Range	D_2	U_m
1	OKAY	1388	10	0.90	4,697
2	BAG	366	10	0.89	1219
3	PIZZA	250	10	0.89	829
4	GAME	242	9	0.81	753
5	BOX	232	10	0.82	722
6	BADMINTON	221	9	0.84	703
7	ROBOT	201	10	0.91	684
8	JUMP	263	7	0.68	677
9	BUS	205	7	0.77	615
10	CUP	157	10	0.91	535
11	SALAD	168	8	0.76	484
12	ORANGE	151	9	0.69	402
13	FESTIVAL	161	5	0.61	391
14	PIANO	132	7	0.75	379
15	CAKE	117	10	0.79	352
16	JUICE	113	9	0.82	349
17	CHICKEN	141	7	0.61	332
18	CREAM	111	8	0.78	328
19	PIE	112	8	0.77	325
20	JACKET	112	8	0.72	309

21	PARTY	123	6	0.60	291
22	SKATE	123	6	0.57	279
23	SHIRT	100	8	0.73	278
24	MARKET	107	5	0.59	262
25	SPAGHETTI	93	8	0.73	258
26	PEN	110	5	0.60	254
27	GUITAR	72	6	0.73	207
28	SKI	83	6	0.61	195
29	CAMP	91	5	0.48	192
30	SCARF	82	7	0.60	190
31	CHEESE	85	8	0.56	183
32	PILOT	97	5	0.45	170
33	CHOCOLATE	59	7	0.75	169
34	VIOLIN	78	6	0.54	166
35	COAT	62	6	0.61	147
36	COURSE	67	4	0.49	134
37	DRESS	51	5	0.61	122
38	SPORT	59	4	0.48	119
39	FASHION	55	3	0.45	111
40	TOMATO	41	7	0.69	108
41	STAR	41	7	0.65	107
42	STEAK	61	4	0.42	101
43	COMPUTER	43	4	0.56	97
44	LEMON	43	6	0.57	96
45	CENTER	63	4	0.30	96
46	MELON	37	5	0.64	91
47	FORK	38	4	0.60	88
48	ENERGY	49	4	0.41	86
49	QUIZ	48	4	0.27	68
50	PROJECT	36	3	0.44	68
51	MAGIC	29	5	0.61	67
52	KANGAROO	28	5	0.58	62
53	TENNIS	42	5	0.35	60
54	CAMERA	38	2	0.30	58
55	PLASTIC	33	4	0.32	52
56	JAM	30	4	0.42	51
57	BANANA	28	3	0.43	49
58	MASK	29	3	0.36	45
59	COMIC	29	2	0.30	44
60	KIWI	22	4	0.45	40
61	SANDWICH	25	4	0.32	39
62	SOUP	14	5	0.68	37
63	GUM	27	2	0.27	33
64	TENT	20	3	0.39	33
65	NOTEBOOK	28	3	0.27	32
66	CRAYON	22	3	0.33	32
67	ROCKET	20	3	0.30	27
68	MODEL	15	3	0.39	27
69	TAXI	25	2	0.16	27
70	TEAM	20	3	0.20	24
71	SWEATER	28	3	0.14	22
72	BELT	38	1	0.00	21
73	TICKET	11	3	0.40	20

74	RIBBON	11	4	0.42	20
75	VIDEO	13	2	0.27	16
76	CARD	8	3	0.41	15
77	SNACK	7	3	0.42	13
78	CARPET	10	2	0.30	12
79	NEWS	5	4	0.59	12
80	MENU	6	3	0.41	11
81	PANDA	7	3	0.35	10
82	DRUM	6	3	0.32	8
83	TOWEL	19	1	0.00	8
84	SET	3	3	0.48	6
85	INTERNET	7	2	0.14	6
86	TRUCK	3	3	0.47	6
87	BENCH	4	2	0.28	6
88	EVENT	3	3	0.46	6
89	WEBSITE	5	2	0.19	4
90	GOWN	13	1	0.00	4
91	ELEVATOR	10	1	0.00	4
92	GRAPH	3	2	0.23	3
93	HAMBURGER	2	2	0.28	3
94	ALBUM	2	2	0.29	2
95	STUDIO	4	1	0.00	2
96	SAUCE	3	1	0.00	2
97	CAROL	3	1	0.00	2
98	GUARD	3	1	0.00	2
99	BUTTER	3	1	0.00	2
100	PARADE	4	1	0.00	1
101	TELEVISION	7	1	0.00	1
102	COFFEE	2	1	0.00	1
103	PROGRAM	2	1	0.00	1
104	OPERA	2	1	0.00	1
105	CAMPAIGN	2	1	0.00	1
106	CABINET	1	1	0.00	1
107	FENCE	1	1	0.00	1
108	BACON	1	1	0.00	1
109	HOTEL	1	1	0.00	1
110	APARTMENT	1	1	0.00	1
111	YACHT	3	1	0.00	1
112	BISCUIT	1	1	0.00	0
113	COURT	1	1	0.00	0
114	INTERVIEW	1	1	0.00	0
115	SAMPLE	1	1	0.00	0
116	VIRUS	1	1	0.00	0

The words in Table 7 appear at least once in the 10 textbooks. These source words are ranked and arranged in descending order of their usage coefficient figures, Um. The top-ranked source word is *okay*, whose U_m is 4,697. This figure indicates that one can probably encounter *okay* 4,697 times in a one-million-word text of the same type as the textbooks. Note that the average of word tokens of the textbooks is 26,896. Based on calculations, if a text is of an average length, *okay* would appear 126 times in that text. The second-ranked word is *bag*, but its U_m is 1,219, which is far lower than the figure for *okay*. The word *bag* is immediately followed by *pizza*, whose U_m is 828. There is a considerable difference in U_m between the second and the third source word, after which a gradual decrease is detected. Forty-five source words have a U_m figure equal to or greater than 96, and except for *fashion*,

they appear in at least four of the 10 textbooks. It is most likely that they should occur more than three times in a text of 26,896 words. Nine source words have a U_m figure that is between 58 and 91, and they are expected to appear twice in a text of the same length. Twenty source words have a U_m figure that is between 20 and 52, and it is very likely that they appear only once in a text of 26,896 words. The remaining 42 source words have a U_m figure that is lower than 20, which means that there is no possibility that they are ever used in an average elementary school English textbook for Korean learners. These words are listed in Table 8 below.

Table 8. Source Words Hard to Find in Elementary School English Textbooks

album, apartment, bacon, bench, biscuit, butter, cabinet, campaign, card, carol, carpet, coffee, court, drum, elevator, event, fence, gown, graph, guard, hamburger, hotel, internet, interview, menu, news, opera, panda, parade, program, sample, sauce, set, snack, studio, television, towel, truck, video, virus, website, yacht

4.3 Discussion

Results obtained from statistical analyses showed that although the same number of types of source words for loanwords were introduced in the elementary school English textbooks of the study (see Table 2), such words were not utilized in a consistent manner. As pointed out in 4.1, the proportion of these words relative to the total number of tokens is too small in ESEL1 and too large in ESEL3. On average, they account for 4.3% of the total number of tokens of a lower-level textbook. However, the proportion of these source words is 3.5% in ESEL1 and 4.8% in ESEL3. Even though it is not yet known what proportion is best for beginner learners of English, the ideal proportion is assumed to be around the average 4.3%. The average proportion of source words that occur in an upper-level textbook is 2.2%. This is natural, considering that students in Grades 5 and 6 tend to make less use of such words in their learning of English, compared to students in Grades 3 and 4. The fact that the proportion of source words is 1.7% in ESEU3 and 3.5% in ESEU4 suggests that there was no consistency in utilizing such words as useful resources in the upper-level textbooks.

As already noted in the previous section, the average number of word tokens of the 10 elementary school English textbooks is 26,896. Therefore, it seems natural to assume that representative source words for elementary school English textbooks should appear at least three times in a text of the same kind with 26,896 words. The top 45 source words in Table 7 (such as *okay*, *bag*, … *center*) meet this qualification. Note, however, that these words are chosen only based on the coefficient of usage of each source word. If a question is raised about whether they are legitimate source words of loanwords, it is highly likely that some of these source words should be excluded.

Source words for loanwords are said to facilitate L2 learning of English under the assumption that L2 learners easily find them to be true cognates of their corresponding loanwords. Therefore, 200 English words were defined as such in the 2015 RNCE. However, some cases were found in which they failed to do so. Words such as *okay*, *jump*, *chicken*, *skate*, *shirt*, *market*, *ski*, *camp*, *course*, *sport* and *star* are among the 45 source words in Table 7, but their occurrences in the elementary school English textbooks appear to disqualify them as source words.

For example, *market* cannot be a source word since its loanword version, *maket*, has not been accepted as standard language, according to *The Standard Korean Language Dictionary* 'Pyojungugeodaesajeon', published online by the National Institute of Korean Language.[2] Thus, *market* must be introduced as a basic word to learn in elementary school English textbooks. The following three words are also problematic: *chicken*, *shirt* and *sport*. To be true cognates, the relationship between loanwords and their source words must be linguistically transparent.

---

[2] It is interesting that *syupeomaket* 'supermarket' is a loanword accepted by the same dictionary.

As pointed out by Lee (2020, 2021), similar to the 51ᵗʰ-ranked *magic* in Table 7, *chicken* is morphologically opaque. While *maejik < Magic Marker* is a case of back clipping, *chikin < fried chicken* is a case of fore-clipping. Because Korean learners of English are not familiar with the morphological changes involved here, and *chicken* is used to mean a common domesticated fowl, not fried chicken, it must be introduced as a basic word rather than a source word in elementary school English textbooks. On the other hand, *shirt* and *sport* are cases of morphophonetic opaqueness. These words were incorrectly defined as source words in the 2015 RNCE that correspond to loanwords, *syeocheu* and *seupocheu*, respectively. These pairings allow Korean learners of English to dismiss the convention that the two English words are used only in the plural forms in Korean. Instead of *shirt* and *sport*, the plural *shirts* and *sports* must be defined as source words in the curriculum.

Furthermore, all loanwords in Korean are nouns, resulting in the 200 English source words in the 2015 RNCE being nouns as well. Therefore, it is surprising that the top-ranked *okay* is presented as a source word in every textbook. In fact, its loanword counterpart *okei* has a very specialized meaning, as discussed in Lee (2020, 2021). According to *The Standard Korean Language Dictionary*, it is meant to be a word that signals approval when proofreading or editing a manuscript. None of the 1,388 occurrences of *okay* in the textbooks, however, illustrates this specialized meaning. They are used as what they are supposed to mean in English, which indicates that *okay* must be introduced as a basic word to learn. Note further that most occurrences of *jump*, *skate* and *ski* in the textbooks are verbs, rather than nouns. This also means that they should not be introduced as source words, but as basic words to learn.

The remaining source words, *course*, *camp* and *star* are also spurious, at least for elementary school students learning English. All the occurrences of *course* in the elementary school English textbooks are simply part of the idiom *of course*. They have nothing to do with any of the three meanings of the loanword *koseu*: the route or path taken by something that moves, a division or part of a meal, and a series of lectures, medicines etc. Thus, *course* is not an appropriate source word. Likewise, there is a semantic mismatch in the cases of *camp* and *star*. The loanword counterparts of the words are *kaempeu* and *seuta*, respectively. The loanword *kaempeu* means a place equipped with facilities for recreation, sports, and sometimes academic instruction, usually for children during summer break, but this meaning is not the primary meaning of *camp*. The primary meaning is a place where an army or other group of persons is lodged in tents or other temporary shelters. The loanword *seuta* means either an artistic performer or athlete whose leading role or superior performance is acknowledged, or an officer of a rank senior to lieutenant general, especially one who commands a large military formation. Again, these meanings are not the primary meaning of *star*. Therefore, presenting *camp* and *star* as source words does more harm than good to Korean learners of English because it gives them a false sense of being aware of the target words that they are supposed to learn.

Finally, if a list of basic words is made based on dispersion, the typical cut-off point is .60 (Dang, Coxhead and Webb 2017) or .80 (Gardner and Davies 2014). If the former is applied to the source words in Table 7, *skate* (.57), *market* (.59), *camp* (.48), *cheese* (.56), *pilot* (.45), *violin* (.54), *course* (.49), *sport* (.48), *fashion* (.45), *steak* (.42), *computer* (.56), *lemon* (.57), and *center* (.30) must be excluded.

Based on the arguments presented above, 26 source words presented in Table 9 can be classified as representative of the elementary school English textbooks of the study.

Table 9. Representative Source Words in Elementary School English Textbooks

bag, pizza, game, box, badminton, robot, bus, cup, salad, orange, festival, piano, cake, juice, cream, pie, jacket, party, spaghetti, pen, guitar, scarf, chocolate, coat, dress, tomato

As long as elementary school English is concerned, the source words in Table 9 and their corresponding loanwords are true cognates, and hence these source words can be useful resources for Korean students learning English at elementary schools.

5. Conclusion

One of the preparations for a complete authentic list of English source words for loanwords is to choose the appropriate loanwords that are believed to facilitate L2 learning for elementary school students. For this purpose, loanwords must meet two conditions. First, only loanwords that are actually available to these students are helpful. No matter how simple a loanword is, it is not helpful if it is not available or known to the students. Second, only loanwords whose relationship with their source words are transparent to the students are helpful. That is, loanwords and their source words must be true cognates. Otherwise, there is no guarantee that they are helpful. Then how can one be sure what loanwords are available to the students? A survey could be conducted to gather responses from elementary school students, but the more efficient and effective way is to investigate how frequently, widely, and evenly the 200 source words listed in the 2015 RNCE are used in elementary school English textbooks.

For this purpose, a corpus was created that consists of 10 subcorpora. These subcorpora are based on texts taken from 10 textbooks and their accompanying e-textbooks, five of which are for the lower-level elementary school English, and the other five for the upper-level elementary school English. Examination of the distributional properties of the 200 source words across the subcorpora revealed two findings. First, the elementary school English textbooks of the study are not homogeneous with respect to the frequency of these source words. While they do present a similar number of types of source words, chi-square tests showed that the frequency of tokens of source words differ significantly. Since tokens of source words play a more important role than their types, the difference in tokens indicates that some textbooks are more helpful in allowing elementary school students to utilize source words as useful resources. Second, out of the 200 source words, 26 were selected as representative source words of the elementary school English textbooks (see Table 9). Their coefficient of usage is equal to or greater than 96, their dispersion figure is higher than .60, and their corresponding loanwords are true cognates. In other words, these words mean what they are supposed to mean as source words and should appear more than three times in the average elementary school English textbook.

It is important to make a final comment on the 26 representative source words of elementary school English. Using the same method adopted here, Lee (2021) identified 53 representative source words of secondary school English textbooks. Eighteen source words from Table 9 are among these 53 source words, but the following eight are not included: *badminton*, *box*, *coat*, *dress*, *party*, *pie*, *scarf*, and *tomato*. Note that these eight source words were ranked relatively low in Lee's (2020) list where the 200 English source words were ranked based on how frequently, widely, and evenly their Korean equivalents occur in the Sejong tagged corpora. None was ranked within the top 50 source words. The source word *party* was ranked 59th, followed by *box* (76th), *tomato* (82nd), and *coat* (94th). The other four were ranked below 100th. This indicates that when it comes to defining an authentic list of source words for the national curriculum of English, loanwords commonly used in everyday life must be chosen based on not only the distribution of loanwords in a large corpus, such as the Sejong corpus, but also their occurrences in a corpus compiled from texts specifically for young learners of English.

References

Ard, J. and T. Homburg. 1983. Verification of language transfer. In S. M. Gass and L. Selinker, eds., *Language Transfer in Language Learning*, 157-176. Rowley, MA: Newbury House.

Banta, F. G. 1981. Teaching German vocabulary: The use of English cognates and common loan words. *Modern Language Journal* 65(2), 129-136.

Carroll, J. B. 1970. An alternative to Juilland's usage coefficient for lexical frequencies. *ETS Research Bulletin Series*, i-15. https://doi.org/10.1002/j.2333-8504.1970.tb007 78.x

Choi. H., J. Seo, E. Moon, M. Lee, J. Yun, M, Park, et. al. 2018a. *Elementary School English 3.* Seoul: YBM.

Choi. H., J. Seo, E. Moon, M. Lee, J. Yun, M, Park, et. al. 2018b. *Elementary School English 4.* Seoul: YBM.

Choi. H., J. Seo, E. Moon, M. Lee, J. Yun, M, Park, et. al. 2019a. *Elementary School English 5.* Seoul: YBM.

Choi. H., J. Seo, E. Moon, M. Lee, J. Yun, M, Park, et. al. 2019b. *Elementary School English 6.* Seoul: YBM.

Dang, T. N. Y., A. Coxhead and S. Webb. 2017. The academic spoken word list. *Language Learning* 67(4) 959–997.

Daulton, F. E. 1998. Japanese loanword cognates and the acquisition of English vocabulary. *The Language Teacher* 22(1), 17-25.

Daulton, F. E. 2008. *Japan's Built-in Lexicon of English-based Loanwords*. Clevedon, England: Multilingual Matters.

Davis, M. and D. Gardner. 2010. *A Frequency Dictionary of Contemporary English: Word Sketches, Collocates and Thematic Lists*. New York: Routledge.

Gardner, D. and M. Davies. 2014. A new academic vocabulary list. *Applied Linguistics* 35(3), 305–327.

Ham, S., Y. Lee, H. Kim, S. Park, J. Park, S. An, et. al. 2018a. *Elementary School English 3.* Seoul: Chunjae.

Ham, S., Y. Lee, H. Kim, S. Park, J. Park, S. An, et. al. 2018b. *Elementary School English 4.* Seoul: Chunjae.

Ham, S., Y. Lee, H. Kim, S. Park, J. Park, S. An, et. al. 2019a. *Elementary School English 5.* Seoul: Chunjae.

Ham, S., Y. Lee, H. Kim, S. Park, J. Park, S. An, et. al. 2019b. *Elementary School English 6.* Seoul: Chunjae.

Haspelmath, M. 2009. Lexical borrowing: concepts and issues. In M. Haspelmath and U. Tadmor, eds., *Loanwords in the World's Language: A Comparative Handbook*, 35-54. Berlin: De Gruyter Mouton.

Jiang, N. 2002. Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition* 24, 617-637.

Juilland, A. G. and E. Chang-Rodriguez. 1964. *Frequency Dictionary of Spanish Words*. The Hague: Mouton & Co.

Kent, D. B. 1999. Speaking in tongues: Chinglish, Japlish and Konglish. In Korea TESOL, ed., *Proceedings of the Second Pan Asian Conference*, 197–209. Seoul: Korea TESOL.

Kim, H., C. Hwang, Y. Kang, H. Lim, J. Kyeong, T. Kim. et. al. 2018a. *Elementary School English 3.* Seoul: YBM.

Kim, H., C. Hwang, Y. Kang, H. Lim, J. Kyeong, T. Kim. et. al. 2018b. *Elementary School English 4.* Seoul: YBM.

Kim, H., C. Hwang, Y. Kang, H. Lim, J. Kyeong, T. Kim. et. al. 2018c. *Elementary School English 5.* Seoul: YBM.

Kim, H., C. Hwang, Y. Kang, H. Lim, J. Kyeong, T. Kim. et. al. 2018d. *Elementary School English 6.* Seoul: YBM.

Lado, R. 1955. Patterns of difficulty in Vocabulary. *Language Learning* 6(1-2), 23-41.

Lee, H. S. 1958. English-Korean cognates. *Language Learning* 8(3-4), 57-72.

Lee, J. 2020. On the Source Words for Loanwords of the 2015 Revised National Curriculum of English. *Korean Journal of English Language and Linguistics* 20, 851-880.

Lee, J. 2021. A corpus-based study of the source words for loanwords of the 2015 Revised National Curriculum of English. *Journal of Linguistic Studies* 26(1), 177-204.

Lee, J., J. Kim, K. Na, D. Lee, J. Yin, M, Kwon, et al. 2018a. *Elementary School English 3.* Seoul: Daekyo.

Lee, J., J. Kim, K. Na, D. Lee, J. Yin, M, Kwon, et al. 2018b. *Elementary School English 4.* Seoul: Daekyo.

Lee, J., J. Kim, K. Na, D. Lee, J. Yin, M, Kwon, et al. 2019a. *Elementary School English 5.* Seoul: Daekyo.

Lee, J., J. Kim, K. Na, D. Lee, J. Yin, M, Kwon, et al. 2019b. *Elementary School English 6.* Seoul: Daekyo.

Leech, G., P. Rayson and A. Wilson. 2001. *Word Frequencies in Written and Spoken English.* Harlow, UK: Longman.

Martin, A. 2004. The 'Katakana effect' and teaching English in Japan. *English Today* 77(20), 50-55.

Masson M.-E. 2013. How L1 loanwords can create a false sense of familiarity with L2 vocabulary meaning and usage. *Vocabulary Learning and Instruction* 2(1), 8-14.

Ministry of Education. 2015. *English Curriculum*. Seoul: Ministry of Education.

Nation, I. S. P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nation, I. S. P. 2003. The role of the first language in foreign language learning. *Asian EFL Journal* 5(2), 1–8.

Nation, I. S. P. 2004. A study of the most frequent word families in the British National Corpus. In P. Bogaards, P. and B. Laufer, eds., *Vocabulary in a Second Language: Selection, Acquisition, and Testing*, 3-13. Amsterdam: John Benjamins.

Nation, I. S. P. and R. Waring. 1997. Vocabulary size, text coverage and word lists. In N. Schmitt and M. McCarthy, eds., *Vocabulary: Description, Acquisition and Pedagogy*, 6-19. Cambridge: Cambridge University Press.

Palmberg, R. 1987. On lexical inferencing and the young foreign-language learner. *System* 15(1), 69-76.

Park, G., K. Ahn, J. Hong, E. Lim, G. Jung. et. al. 2018a. *Elementary School English 3*. Seoul: DongA.

Park, G., K. Ahn, J. Hong, E. Lim, G. Jung. et. al. 2018b. *Elementary School English 4*. Seoul: DongA.

Park, G., K. Ahn, J. Hong, E. Lim, G. Jung. et. al. 2018c. *Elementary School English 5*. Seoul: DongA.

Park, G., K. Ahn, J. Hong, E. Lim, G. Jung. et. al. 2018d. *Elementary School English 6*. Seoul: DongA.

Rodriguez, T. A. 2001. From the known to the unknown: Using cognates to teach English to Spanish-speaking literates. *The Reading Teacher* 54(8), 744-746.

Rogers, J., S. Webb and T. Nakata. 2015. Do the cognacy characteristics of loanwords make them more easily learned than noncognates? *Language Teaching Research* 19(1), 9−27.

Rubin, J. 1987. Learner strategies: Theoretical assumptions, research history and typology. In Wenden, A. and J. Rubin, eds., *Learner Strategies in Language Learning*, 15-30. New York: Prentice Hall.

Schmitt, N. 2000. *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.

Sheperd, J. 1996. Loanwords: A pitfall for all students. *The Internet TESL Journal* 2(2). Retrieved from iteslj.org/Articles/Shepherd-Loanwords.html

Simon-Maeda, A. 1995. Language awareness: Use/misuse of loan-words in the English language in Japan. *The Internet TESL Journal* 1(2). Retrieved from iteslj.org/Articles/Maeda-Loanwords.html

Takač, V. P. (2008). *Vocabulary Learning Strategies and Foreign Language Acquisition*. Clevedon: Multilingual Matters.

Uchida, E. 2001. *The Use of Cognate Inferencing Strategies by Japanese Learners of English*. Doctoral dissertation, University of Essex, Colchester, UK.

Examples in: English
Applicable Languages: English
Applicable Level: All