

A Reconsideration of the Construct of “A Vocabulary for Japanese Learners of English”: A Critical Comparison of the JACET Wordlists and New General Service Lists

Shin'ichiro Ishikawa
Kobe University

Abstract

By comparing the vocabularies included in the Japan Association of College English Teachers (JACET) wordlists (1993, 2003, and 2016 editions) and recently released New General Service Lists (Brezina & Galasova, 2013; Browne, 2013), we tried to identify the construct of “the vocabulary for Japanese learners” with which JACET researchers have been concerned. Our quantitative analysis has shown that it includes the vocabulary concerning (a) family and people, (b) houses and daily life, (c) foods and cooking, (d) clothes and fashions, (e) sports, (f) social meetings, (g) transportation, and (h) emotions and mental states, as well as (i) spoken English vocabulary and (j) non-basic forms. This finding will shed new light on the discussion of what kind of vocabulary should be included in pedagogical wordlists.

Keywords: vocabulary selection; corpus-based approach; pedagogical adjustment

Introduction

It is true that recent developments of mega-corpora and the increasing sophistication of corpus query techniques have made the compilation of wordlists much easier than before, but this may not apply to the creation of pedagogical wordlists, for there exists a certain amount of pedagogically indispensable vocabulary that cannot be extracted easily from corpora (Ishikawa, 2007, 2015). Unlike L1 children, L2 learners, especially in English as a Foreign Language (EFL) situations, usually learn the vocabulary not from natural input but from institutionalized education at schools, which is undoubtedly influenced by their L1, nationality, and culture. This suggests the need to consider learners' backgrounds in making a pedagogical wordlist.

The Japan Association of College English Teachers (JACET), which is one of the largest expert associations in the field of applied linguistics in Japan, has been engaged in the compilation of “a pedagogical vocabulary list for Japanese learners of English” (JACET 2003, p. 2). The first JACET list (J1) was published in 1981, and the most recent edition (J5) was published in 2016 (Fig. 1; Table 1). This means that the JACET has continued the challenging task of wordlist compilation for more than three decades.

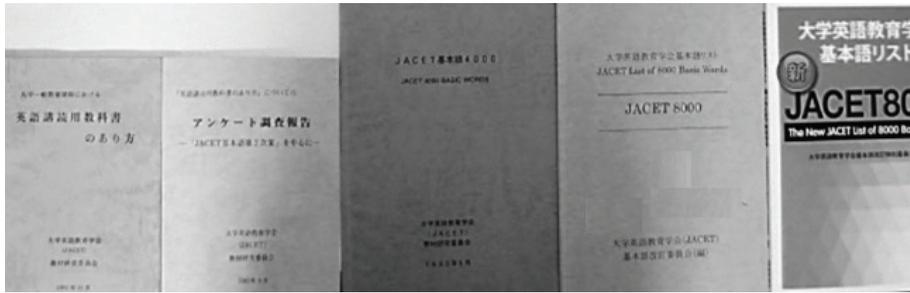


Figure 1. Five Editions of the JACET Wordlist.

Table 1. Outline of the JACET wordlists

Edition	Year	Headwords	Methods to choose the vocabulary
1st (J1)	1981	4064	
2nd (J2)	1983	3990	Combination of several existing wordlists + pedagogical adjustments
3rd (J3)	1993	3990	
4th (J4)	2003	8000 + 250	Corpus-based frequency analysis + pedagogical adjustments
5th (J5)	2016	8000	

Note: Concerning the details about each of the five editions, see Ishikawa (2018).

Although the JACET researchers have regarded word frequencies obtained indirectly from established wordlists (J1, J2, J3) or directly from corpora (J4, J5) as an important base of judgment in vocabulary selection, they have also been interested in rational adjustments so as to make purely frequency-based wordlists more suitable for Japanese learners of English.

However, how the JACET researchers operationalized the construct of “a vocabulary for Japanese learners of English,” what kinds of adjustments they really made, and whether their adjustments were consistent and truly justifiable still remain uncertain. Therefore, this article aims to probe the construct of “a vocabulary for Japanese learners of English” with which the JACET researchers have long been concerned.

Identification of the Vocabulary for Japanese Learners of English

Aim

The history of JACET wordlists illustrates the developers’ lasting belief that there should exist some special set of vocabulary for Japanese learners of English that presumably cannot be extracted directly from corpora. In other words, they believe that words for learners should be chosen not only based on their frequency and dispersion (range) in corpora but also based on some pedagogical considerations.

What kind of vocabulary then have the JACET researchers regarded as “the vocabulary for Japanese learners of English” in the long history of their wordlist developments? To what extent is it different from the ordinary high-frequency words obtained from corpora? We discuss three research questions here.

RQ1. What kind of vocabulary items are regarded as the core of the JACET wordlists and frequency-based wordlists for general purposes?

RQ2. What kind of vocabulary is missing from the JACET core vocabulary?

RQ3. What kind of vocabulary is included only in the JACET core vocabulary?

Data

As Japanese learners currently learn 3000 words at secondary schools, we pay attention to the roughly 3000 words placed at the top of three JACET wordlists: J3_3K (2864 words in Levels 1–4), J4_3K (2750 words and the “plus 250” words), and J5_3K (3000 words), as well as two recently compiled corpus-based wordlists for general purposes (Table 2): Browne (2013) (2801 words and 52 supplementary words; Browne hereafter) and Brezina and Gablasova (2013) (2494 entries; B & G hereafter). These two corpus-based lists are intended as New General Service Lists (NGSL), namely, modernized alternatives to the General Service List (GSL) (West, 1953).

Method

Principles for dealing with tokenization differ greatly among the five lists to be compared. Some include derivative forms (e.g., *ing*-nouns and *ly*-adverbs), proper nouns (e.g., *England* and *Japan*), numerals (e.g., *one* and *fourteen*), and phrases (e.g., *a lot of* and *according to*) but others do not. Therefore, we re-lemmatized all five wordlists with the lemma table used for the development of J5 and identified the number of word forms to be compared in the current analysis, where we do not distinguish different meanings and functions of a single word form.

After re-lemmatization, the number of word forms included in each of the five lists was found to be between 2201 and 3016 (Table 3). We can say that the

Table 2. Outline of the Two Corpus-based Wordlists for General Purposes

Wordlists	Source corpora	Entries
Browne	Subsection from the Cambridge English Corpus (273 million words)	2801 words + 52 supplementary words (e.g., numerals, and names of months and days)
Brezina and Gablasova (B & G)	The Lancaster–Oslo–Bergen Corpus (1 million words), BNC (100 million words), BE06 Corpus of British English (1 million words), and EnTenTen12 (11 billion words)	2494 entries including phrases. The word-forms of different POS are treated separately.

Table 3. The Number of Lemmas to be Compared in the Current Analysis

Lemmas	J3_3K	J4_3K	J5_3K	Browne	B & G	All
Each	2740	2865	3016	2849	2201	
Merged						3896

sizes of these lists are roughly equivalent to the size of the vocabulary needed by Japanese secondary students.

Concerning RQ1, we chose the words included in all three editions of the JACET wordlist, which will be called the JACET core vocabulary. Next, we chose the words included in both Brown's and B & G's lists, which will be called the NGSL core vocabulary. We then compared the JACET and NGSL core vocabularies and identified the words included only in the NGSL core (RQ2) and those included only in the JACET core (RQ3). The latter could be regarded as an embodiment of what JACET wordlists editors have regarded as the appropriate vocabulary for Japanese learners of English.

Results and Discussions

RQ1: JACET and NGSL Core Vocabulary

The number of words included in all of the J3_3K, J4_3K, and J5_3K was 2161. Table 4 shows the first 20 words in alphabetical order. We can say that this is core vocabulary of the JACET wordlists.

Meanwhile, the number of words included in both of the NGSL wordlists was 2100. Table 5 shows the top 20 words in alphabetical order. This can constitute the core vocabulary of corpus-based wordlists for general purposes.

It is of note that the sizes of the JACET core vocabulary and the NGSL core vocabulary were almost identical, which might suggest that the top 2000 words are relatively more stable than lower-ranked words.

RQ2: What Is Missing from the JACET Core Vocabulary

By comparing the JACET and NGSL core vocabularies, we identified 1729 words that were commonly included in both lists. Meanwhile, it was found that 373 words were included only in the NGSL and absent from the JACET core vocabulary. Table 6 shows the first 100 words in alphabetical order.

This suggests that JACET has regarded words primarily concerning business (agenda, asset, budget, buyer, colleague, consultant, consume, corporate, delivery) and higher-order mental activities (accuse, acknowledge, advise, analysis, analyze, anticipate, appreciate, approval, assess, assessment, awareness, complaint, comprise, confirm, consideration) as relatively less important for Japanese learners. This list seems to include many morphologically complex suffixed words

Table 4. A Part of the JACET Core Vocabulary

a, abandon, ability, able, about, above, abroad, absolutely, absorb, accept, accident, accompany, accord, account, accurate, achieve, acid, acquire, across, act

Table 5. A Part of the NGSL Core Vocabulary

a, abandon, ability, able, about, above, absence, absolute, absolutely, abuse, academic, accept, acceptable, access, accident, accommodation, accompany, accord, account, accurate

Table 6. Words Missing from the JACET Core Vocabulary

absence, absolute, abuse, academic, acceptable, access, accommodation, accuse, achievement, acknowledge, ad, addition, additional, adjustment, administration, advise, agenda, album, alcohol, alongside, alternative, analysis, analyze, angle, anticipate, anxiety, apparent, appoint, appreciate, approval, approximately, arrival, assess, assessment, asset, assist, assistance, awareness, barely, barrier, beer, bind, boost, brand, breast, brilliant, budget, buyer, cancer, category, chapter, charity, chart, chat, childhood, chip, chocolate, cite, clinical, colleague, commit, commitment, complaint, component, comprehensive, comprise, concentration, confident, confirm, consequence, consequently, considerable, consideration, consistent, constantly, construction, consultant, consume, context, conventional, core, corporate, county, coverage, creation, creative, criminal, criterion, crucial, cultural, currently, decline, defense, definitely, definition, delivery, dependent, depression, derive, deserve

Table 7. The Words Included only in the JACET Core Vocabulary

abroad, absorb, acid, admire, advertise, ah, airport, alarm, altogether, anger, angry, anxious, anybody, apple, arrest, aside, asleep, aspect, assembly, assignment, assistant, aunt, authority, autumn, awake, badly, banana, bare, bark, baseball, basket, bat, bath, bay, bean, beginning, behave, being, bell, belt, bench, bend, beneath, beside, besides, best, better, bicycle, birthday, bite, blanket, blind, bomb, born, bowl, brick, broadcast, brush, building, burst, bury, bush, butter, calm, cap, captain, carpet, castle, cattle, ceiling, ceremony, champion, chase, cheer, cheese, classroom, clay, clerk, clever, clue, coat, coin, command, communicate, compete, complicate, concert, conference, confuse, conscious, continent, cook, cooking, corn, cotton, counter, courage, cousin, cow, crack

(e.g., -able, -tion, -ment, -ate, and -ness), and the average length (number of letters) of these 373 words is 7.90 ($SD = 2.38$).

RQ3: What Is Included Only in the JACET Core Vocabulary

Our analysis found 435 words that were included only in the JACET core vocabulary, which we can conclude is the essence of the vocabulary for Japanese learners of English that JACET researchers have consistently prioritized. Table 7 presents the first 100 words in alphabetical order.

These words may be roughly classified into vocabulary concerning (a) family and people (*aunt, cousin, assistant, authority, captain, clerk, champion*), (b) houses and daily life (*bath, bench blanket, brick, brush, carpet, aisle, ceiling, counter*), (c) foods and cooking (*apple, banana, bean, bowl, butter, cheese, cook, cooking, corn*), (d) clothes and fashions (*belt, cap, coat, cotton*), (e) sports (*baseball, basket [ball], bat*), (f) social meetings (*assembly, concert, conference*), (g) transportation (*abroad, airport, bicycle*), (h) emotions and mental states (*anger, angry, anxious, asleep, awake, calm, cheer, clever, conscious*), and (i) spoken English (*ah, badly*), as well as (j) non-basic forms (*beginning, being, building, ceiling, cooking, best, better, born*).

This list includes many relatively simpler words in terms of not only semantics but also morphology. The average length (number of letters) of the 435 words is 5.83 ($SD = 1.96$), which is significantly shorter than the average length of the words missing from the JACET core vocabulary (Welch's two-sample *t*-test yielded: $t = 13.3643$; $df = 720.6922$; $p < 0.001$; Cohen's $d = 0.959$).

Summary

This article tried to identify the construct of “the vocabulary for Japanese learners” with which JACET wordlists editors have been concerned. Our data-based analysis has shown that it consists of the vocabulary concerning (a) family and people, (b) houses and daily life, (c) foods and cooking, (d) clothes and fashions (e) sports, (f) social meetings, (g) transportation, and (h) emotions and mental states, as well as (i) spoken English vocabulary and (j) non-basic forms.

Although it is not necessarily clear how the JACET wordlists will change in the future, one thing that could be done is to incorporate the results of the analysis of Japanese learners’ L2 outputs into their vocabulary selection. The recent development of learner corpora has made it possible for researchers to identify the problems that learners with a particular L1 typically have in the use of L2 English vocabulary (Ishikawa, 2013). If JACET continues to try to choose words exclusively for “Japanese” learners of English, it would be indispensable for them to examine how they actually use English vocabulary in their speech and writing.

References

- Brezina, V., & Gablasova, D. (2013). Is there a core general vocabulary? Introducing the new general service list. *Applied Linguistics*, 36(1), 1–22. doi: 10.1093/applin/amt018.
- Browne, C. (2013). The new general service list: Celebrating 60 years of vocabulary learning. *The Language Teacher*, 7(34), 13–16.
- Ishikawa, S. (2007). Eigo kyoiku no tame no kihongo o do erabuka: Kopasu gen-gogaku kara no shiten. *The English Teachers’ Magazine*, 55(13), 10–13. [How should we choose the basic English words to be taught at schools? A viewpoint of corpus linguistics].
- Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world*, 1 (pp. 91–118). Kobe, Japan: Kobe University.
- Ishikawa, S. (2015). A new corpus-based methodology for pedagogical vocabulary selection: Compilation of “HEV1800” for Japanese high school students. *Journal of the Chubu English Language Education Society*, 44, 41–48.
- Ishikawa, S. (2018). A critical survey of JACET English word lists: Reconsideration of the validity of the frequency integration method. *Journal of Corpus-based Lexicology Studies*, 1, 53–80.
- JACET Kyozaikai Kenkyu Iinkai (JACET, Committee for the Research of Teaching Materials). (1981). *Daigaku ippan kyoyo katei ni okeru eigo kodoku yo kyokasho no arikata*. Tokyo: JACET. [A report on how textbooks for English reading classes at college general education should be]
- JACET Kyozaikai Kenkyu Iinkai (JACET, Committee for the Research of Teaching Materials). (1983). “*Eigo kodoku yo kyokasho no arikata*” ni suite no anketo chosa hokoku: *JACET kihongo dai 2 jian o chushin ni*. Tokyo: JACET. [The results of the survey on *A report on how textbooks for English reading classes at college general education should be*: With a focus on the second version of JACET list of basic words].

- JACET Kyozaï Kenkyū Iinkai (JACET, Committee for the Research of Teaching Materials). (1993). *JACET 4000 basic words*. Tokyo: JACET.
- JACET Kihongo Kaitei Iinkai (JACET, Committee for Revision of the JACET Wordlist). (2003). *JACET List of 8000 Basic Words*. Tokyo: JACET.
- JACET Kihongo Kaitei Tokubetsu Iinkai (JACET, Special Committee for Revision of the JACET Wordlist) (2016). *The New JACET List of 8000 basic words*. Tokyo: Kiriara Shoten.
- West, M. (1953). *A General Service List of English Words: With Semantic Frequencies and a Supplementary Word-list for the Writing of Popular Science and Technology*. Longmans, Green.