Lexical types and tokens found in the classroom speech of native and non-native English language instructors in a Japanese high school

Noriaki Katagiri¹,* and Goh Kawai²

¹Hokkaido Sapporo Kaisei Senior High School
²Center for Language Learning, Hokkaido University

(Received 25 May 2012, Accepted for publication 12 October 2012)

Abstract: Within the context of English language taught solely using English language at Japan’s secondary schools, no research quantifies the differences between native instructors (first language English, may or may not speak Japanese) and non-native instructors (first language Japanese; second language English). We developed a video corpus of an English language classroom, and examined the speech of 3 native and 1 non-native instructors. The corpus contains 49 English lessons of 45 minutes each in a Japanese public high school with monolingual learners of English as a foreign language. The native and non-native instructors occasionally taught together. Almost all speech in the lessons was in English. We compared lexical tokens and types found in our transcriptions with a collection of typical classroom English dialogues, and a wordlist created from large bodies of written and spoken English. We obtained the distributions of words, and words preferred by either native or non-native instructors. Results suggest that (a) native and non-native instructors share a core vocabulary of classroom English, (b) native instructors teach vocabulary depth via open-ended conversations, (c) non-native instructors teach vocabulary breadth via textbook explanations, and (d) native and non-native instructors differ in teaching roles but not in language ability.

Keywords: Classroom corpus, Native vs. non-native language instructor speech, Frequency ratios, Keyness

PACS number: 43.71.Sy [doi:10.1250/ast.34.94]

1. INTRODUCTION

English language learners in Japan, their instructors, parents, and school administrators share a preconception that, in contrast with native instructors (NIs), non-native instructors (NNIs) are somewhat if not totally incapable of teaching English through English. (NIs are teachers whose first language (L1) is English, and who may or may not speak Japanese as their second language (L2). NNIs are teachers whose L1 is Japanese and whose L2 is English.) Because NNIs outnumber NIs by over 15 to 1 in Japan’s middle schools and high schools [1,2], parents and learners fear that they are statistically likely to be taught by inferior teachers. The roles and effectiveness of NIs and NNIs are poorly researched, however. Instructors, learners, parents, and school administrators deserve to know how NIs and NNIs differ, and how to leverage or correct those differences.

Let us consider how the general public perceives NIs. Many for-profit language schools stress in advertisements that their instructors are all NIs, implying that NNIs are less desirable. College students tend to judge NIs as more authentic and demanding. Everybody claims they would love to speak English “like a native.” What better way than to learn from a native speaker? But being a native speaker per se is no guarantee of teaching ability. Most Japanese people would be inadequate in teaching Japanese language to non-native learners. Yet learners of English language clamor for NIs. Most NIs are native speakers of English with a bachelor’s degree, but with no formal training in teaching any subject to any age group (examples of job announcements include [3,4]). Some capabilities of NIs are misunderstood and overvalued.

Next, let us consider the advantages of NNIs. While some NIs are monolingual, all NNIs in middle schools and high schools are formally trained in college or university, are certificated by MEXT (the ministry of education, culture, sports, science & technology in Japan), are at least somewhat bilingual, and tend to be more aware of linguistic and cultural differences between L1 and L2. When the NNIs and learners share the same L1 (as is the case with the vast majority of English language classrooms
in Japan), learning is facilitated by the NNIs’ own experience in learning L2, along by their training in applied linguistics specifically comparing L1 and L2. For instance, when a learner says “John’s attitude changed like turning over a palm,” an NNI would recognize that the learners wants to say “John abruptly changed his attitude,” but an NI would be baffled without knowledge of Japanese idioms or mismatches in lexical meaning between Japanese and English. An NNI might teach faster and more precisely than an NI. Some capabilities of NNIs are undervalued, perhaps foremost by the NNIs themselves.

How do NNIs teach? As recently as in 2007 in Japan, only 30% of middle school and 10% of high school English language classes were conducted in English [5]. Starting in 2013, all English language courses must be delivered in English [6,7], but some NNIs are reluctant. A Japanese governmental nation-wide survey estimated that less than half of NNIs have L2 skills sufficient to teach using only L2 [8]. In one case, 30 NNIs in Japan experienced difficulty incorporating communicative language teaching (CLT) in their English language courses [9]. (CLT is a language teaching technique that emphasizes the active usage of L2 with the primary purpose of achieving tasks and the secondary purpose of acquiring language accuracy. CLT discourages the use of L1.) In another case, NNIs outside Japan preferred mixing L1 and L2 to using solely English as both the medium and content of instruction. South Korean NNIs code-switched from L2 to L1 when explaining grammar, vocabulary, and background information [10]. (Code-switching refers to changing languages.) As cautioned [11], the current course of study by MEXT has not elevated the use of English language as a means of instruction and medium of communication as high as it expected. A considerable number of NNIs resist CLT [11–13].

In sum, these surveys indicate that (a) about half of Japan’s NNIs cannot teach using CLT, (b) many NNIs avoid CLT even if they can use it, and (c) when CLT is used, it is often in the form of co-teaching where the NNI depends on the NI. (Co-teaching means multiple instructors teaching the same class together, often by dividing tasks among them.) The unwillingness to use CLT is deemed by learners, parents, and school administrators that NNIs teach poorly.

But there is no evidence from the classroom that supports or refutes the claim that NIs are superior to NNIs who use CLT. Either way, evidence is quickly needed because in middle schools English classes are already being taught under the new course of study, and high schools will teach English through English starting in 2013. If there exists a difference in teaching ability between NNIs and NIs, then the educational community has the right to know what those differences are, and how those differences affect learning. If in fact NNIs are less skilled in CLT than NIs, then to succeed in CLT either NNIs need more training, or more NIs need to be hired. If NNIs and NIs are equally skilled, then NNIs and NIs can be positioned interchangeably. If NNIs and NIs have different roles, then in-service and pre-service instructors should either become explicitly aware of their complementing responsibilities so that they can co-teach, or receive cross-training to acquire skills in both roles. (Cross-training means being additionally trained in a colleague’s skill in order to broaden one’s skill set.)

Let us consider how to clarify the differences between NNIs and NIs by analyzing their spoken language production in the classroom. Could it be that the NNIs’ total vocabulary size or basic vocabulary size are substantially smaller than the NIs’? What can NNIs do to use English as the main or sole medium of instruction and content? If the vocabularies of NIs and NNIs are equally effective, then parents and school administrators can confidently ask NNIs to use CLT. (The stake-holders needing confidence the most may well be the instructors themselves.) When NNIs and NIs both use English language more, the government mandate might succeed.

In order to achieve CLT use by both NIs and NNIs, this study focuses on their classroom speech. Previous studies on the classroom speech of NIs and NNIs, and on NI vocabulary size provide scope for our study.

Speech in English language lessons consists of roughly two types: (1) *lesson proper*, involving teaching speech such as teacher-initiated one-way informatives and directives [14], and (2) *lesson peripheral*, involving classroom management speech, and including so-called classroom English (CE). In a third-year classroom at a lower secondary school in the Netherlands, 72% of all instructor utterances were lesson proper, and the rest were lesson peripheral [15].

Research on vocabulary size shows that by 18 years of age native speakers of English know approximately 18,000 words, of which 2,000 words constitute a basic vocabulary [16], beyond which the frequencies of occurrence drop sharply [17].

Based on the literature, the following questions deserve attention: (1) Do the classroom utterances of NIs and NNIs differ in terms of vocabulary size? (2) What vocabularies do NIs and NNIs share? (3) What characterizes their different vocabularies? (4) What can NNIs learn from NIs, and vice versa? (5) How can we encourage and justify the use of English by NNIs in classrooms of English as a foreign language where NIs participate only on limited occasions?

Answering these questions requires analyses of NI and NNI speech. This paper addresses that requirement by developing a corpus of classroom spoken English, and...
proposing methods to analyze lexical types and tokens found in such a corpus. (A lexical token is an instance of a word form. A lexical type is a headword that represents a group of inflected forms. For brevity, this paper occasionally refers to lexical types and tokens as types and tokens respectively.)

We hope this paper generates interest among the speech processing community to (a) engage at a larger scale in quantitative research of CLT classroom behavior by NIs and NNIs, and (b) quantify English production and reception by analyzing spoken interactions between instructors and learners in classrooms (which this paper has not done). We also hope that such research methods and findings from the speech processing community will aid language teacher training program administrators, language learning policy-makers, NIs, and NNIs in becoming cognizant of the instructors’ developed or latent talents.

The remainder of this paper consists of research materials (Sect. 2), methods, (Sect. 3), results and analyses (Sect. 4), and discussion and conclusion (Sect. 5).

2. MATERIALS

We compared NI and NNI speech using the following 3 materials: (1) transcriptions of English spoken by instructors in a language classroom, (2) a collection of CE phrases, and (3) a corpus created from large bodies of written and spoken English. Of these, (1) is our work, as little research exists on NI and NNI English [18]. Materials (2) and (3) were obtained from the literature as explained later in this section.

2.1. Classroom Video Corpus

In order to obtain NI and NNI speech data, we videotaped a Japanese public senior high school freshman English language class. This class was observed over a 6-month period. The video recordings were transcribed at the lexical token level.

There were 3 NIs, 1 NNI, 23 boy students, and 17 girl students. 40 students is a common class size for a Japanese public school.

The NIs had arrived at the school district after completing college in the USA, the UK, or New Zealand. None had prior teaching experience. The NIs were hired fulltime for a limited duration as assistant language teachers through a junior instructor exchange program in Japan [19]. The NNI was a permanent employee of the school district, with 20 years’ teaching experience.

The learners were typical Japanese high school freshmen, who, when videotaping began, had just finished 9 years of compulsory education (all of it in Japan), of which during the last 3 years they learned English language in 50-minute class periods meeting 3 times per week, for 35 weeks per school year. Their estimated vocabulary size was approximately 1,000 lexical types. We believe the learners’ English language proficiency was at the elementary level, perhaps S-1 on the ILR (FSI) proficiency scale [20] or between A1 and A2 on the Council of Europe scale [21].

The high school English class consisted of 45-minute class periods meeting 4 times a week. 1 NNI, who was always present, taught the class during the entire videotaped period. Once every 5 lessons on average, an NI (but never more than 1 at a time) co-taught with the NNI; that is, (a) NIs co-taught with the NNI every time they appeared in the lesson, (b) NIs never taught solo, but (c) the NNI did teach solo. When the speech data were being collected, the instructors were not explicitly aware of any differences in their teaching roles or lexical types. Differences might have arisen because (a) NIs always co-taught while NNIs co-taught and solo-taught, meaning that the NNI may have played the roles of the NIs had they been co-teaching, and (b) NNIs in general seem to depend upon NIs for communicative language teaching.

During class, an audio CD provided by the textbook publisher was used from time to time. CD speech appearing in the classroom video recordings was also transcribed in order to show the size of NI and NNI speech relative to CD recordings. Some readers of this paper may be familiar with classes where NNIs speak L1 most of the time, and an L2 recording is played on occasion. The CD speech in our corpora occurs at the same order of magnitude. This may give readers an idea of how much L2 was spoken by the instructors.

2.2. NNI Reliability

To show that results from our lone NNI are generalizable, we compared our NNI’s speech with that of other classes taught in English language by 5 NNIs [22]. The 5 NNIs had been video-recorded at 5 schools in as many regions across Japan by a government agency to serve as role models for NNIs throughout Japan. Each class had different learners and language material. In Sect. 4.2, we will show by using word token distributions that the classroom language was similar across all 6 NNIs, and therefore that results from our 1 NNI can be (with reservations) be generalized among NNIs who can teach using CLT.

We did not wait to collect more NNI data because we believe schools deserve to learn of our findings as quickly as possible. The pedagogical issues are pressing because the government new course of study for middle schools has already started, and its new course of study for high schools mandates English-only classes starting in April, 2013. If our findings are practically useful in schools, then teachers and learners are entitled to benefit from them now. We hope that our research will stimulate our colleagues to collect convincing amounts of NNI and NI data.
2.3. Classroom English

One narrow definition of CE consists only of questions and commands uttered to the students, for example, “Who hasn’t got a book?” “Sit down, please.” and “Take out your books.” [23]. We define CE as teacher-student interactions that are not teaching moves, that is, instructor utterances that are neither task-centered elicitations (questions using interrogatives), directives (commands using imperative forms), nor informatives (explanations using declaratives). Examples of CE conversations are given in the literature [23,24] that list classroom phrases intended as a guide for NNIs wishing to conduct classes in English. Among the NNI-student conversation examples given in [24], there were a total of 10,839 lexical tokens, of which 6,824 were spoken by the NNI, and 4,014 by the students.

2.4. Benchmark Wordlist

As the benchmark wordlist for comparison [25], we chose a wordlist that contains basic words [26], which lists the 8,000 most frequent lexical types in order from 1 (the most frequent) down to 8,000 (the least frequent). These 8,000 lexical types are ranked into 8 bands of 1,000 types each. The frequency rankings and bands were used to compare lexical types and tokens found in our transcriptions.

3. METHODS

After transcribing NI, NNI, and CD speech, utterances were separated into lexical tokens. Lexical tokens were grouped into lexical types using Perl scripts that lemmatized tokens (i.e., grouped lexical tokens belonging to the same lexical type) [26]. Lexical tokens that were derived forms were counted as different types.

We analyzed tokens appearing in NI and NNI speech in 3 ways: (1) we divided NI, NNI, and CE tokens and types into sets based on their commonality (Sect. 3.1); (2) we compared NI and NNI tokens and types against the benchmark wordlist with regards to frequency rank, and correlation coefficients. (Sect. 3.2); and (3) we compared NI and NNI types by first normalizing frequencies of NI and NNI tokens, and then identifying NI and NNI types with high keyness, i.e., tokens that occur with unusually high frequency (Sect. 3.3).

3.1. Classifying Tokens and Types

For convenience, we will refer to sets consisting of NI, NNI, and/or CE tokens as sets S1 through S7 as defined in Fig. 1. Such classification is of interest because, for instance, S7 might suggest a core vocabulary of language instructors [27], while the difference between S1 and S2 might suggest how NIs and NNIs differ from each other.

3.2. Comparing NI and NNI Tokens and Types against the Benchmark Wordlist

The sets NI and NNI were compared against the benchmark wordlist in the following ways:

(1) NI and NNI tokens were ordered in the order of frequency that their corresponding types appear in the benchmark wordlist.

(2) Correlation coefficients were obtained between NI and NNI vs. those in the benchmark wordlist. Raw frequencies of lexical types in each 1,000-type rank in [26] were used for this procedure, and yielded 8 correlation coefficients.

3.3. Finding Types with High Keyness

We identified NI and NNI types with high keyness using the following procedure:

(1) Normalize NI and NNI token frequencies by calculating frequency ratios of NI and NNI tokens with respect to their maximum frequency counts.

(2) Apply a chi-square test on NI and NNI type frequencies. This yields types with potentially high keyness.

(3) Remove context-dependent words, because they are specific to the lesson being taught. We allow CE tokens uttered by our NIs and NNI (S5, S6, and S7 in Fig. 1) because discourse is expected in most if not all classroom interactions [28].

4. RESULTS AND ANALYSES

This section focuses on NI and NNI speech, although learners were exposed to CD speech as well. First we show that our 4 NIs can be statistically treated as 1 group, and
that our NNI is statistically similar to 5 other NNIs who use CLT to teach different classes. Next we statistically describe the data. We end with results and analyses of methods that were explained in Sect. 3.

4.1. NI and NNI Inter-Instructor Equivalence

We examined the similarities among the 3 NIs used in this study, and the 5 NNIs taken from [22]. Figures 2 and 3 show distributions of spoken tokens in respective subjects across the 8 ranks of the benchmark corpus.

One-way analysis of variance (ANOVA) of the utterance token distributions of the 3 NIs across the 8 ranks of the benchmark corpus showed no statistical significance $F(2, 21) = 0.005$, $p < 0.001$. Therefore we combined the speech from the 3 NIs and treated them as a one set of data.

One-way ANOVA of the utterance token distribution of our NNI and the 5 NNIs from other classes across the 8 ranks of the benchmark corpus showed no statistical significance $F(5, 42) = 0.00011$, $p < 0.001$. Therefore we tentatively propose that our NNI might be representative of similar NNIs, i.e., the 10 percent of all NNIs in Japan who mostly or totally use English language during their classes.

We did not combine the speech of all 6 NNIs because they taught different English classes to different students.

4.2. Classifying Tokens and Types

Table 1 shows descriptive statistics of the number of lexical tokens per classroom lesson. The number of tokens per class period shows that on average NIs and the NNI produce comparable numbers of tokens.

Table 1 Summary of the number of tokens per lesson. The number of tokens per class period shows that on average NNIs and the NNI produce comparable numbers of tokens.

|       | NI  | NNI | CD   | Total |
|-------|-----|-----|------|-------|
| Sum   | 18,572 | 96,460 | 11,178 | 126,210 |
| Mean  | 2,064  | 1,969 | 373  | 2,576  |
| SD    | 769    | 901  | 208  | 936    |
| Min   | 1,171  | 222  | 73   | 496    |
| Max   | 3,374  | 3,456 | 997  | 4,533  |
| NC    | 9      | 49   | 30   | 49     |

Note. Means and SD were rounded at the first decimal place. NC is the number of classes the NIs and the NNI appeared in.

when the lesson context allows. The CD is used in 60% of all lessons and its number of tokens is roughly 20% compared to those of the NIs and the NNI.

The distribution of lexical types across $S_1$ and $S_7$ is shown in Fig. 4. We found that the sizes of $S_1$ and $S_7$ are relatively large. We also found that $S_2$ is substantially larger than $S_1$ and $S_7$.

Based on the lexical types found in each set, along with the utterances in which the lexical types appeared, and teacher behavior in general, we surmise the following.

$S_7$ was large because every language depends on a relatively small set of words to express any idea, of course, but in this study the largeness of $S_7$ is pedagogically meaningful because it indicates similarities between the NIs and NNI. The greater the similarity, the greater the equivalence, hence the greater interchangeability between NNIs and NNIs.
Further evidence of NI-NNI similarity was found in $S_4$ being larger than $S_5$ and $S_6$. Most instructors adapt to their learners without explicit guidelines of teaching conduct. Perhaps the instructors unconsciously adapted their vocabulary to their learners’ capabilities (i.e., aid comprehension by adjusting their vocabulary) even though the instructors did not plan on doing so.

$S_2$ was markedly larger than $S_1$. The reason may be that NI lexical types remained fairly constant across lessons because NIs and students engaged in oral exchanges that were less dependent on textbook content, while NNI types varied across lessons because the NNI was responsible for introducing and reviewing new vocabulary when they appeared in the course material. This division of labor probably arose spontaneously because (a) the NNI was always present in class, knew what lexical types were being reviewed and what were being introduced, and drafted the lesson plan, whereas (b) the NIs were not present in every class, were not all trained as teachers, and were more comfortable being conversation partners.

In sum, the largeness of $S_4$ and $S_7$ suggests that NIs and NNIs tailor lexical types to their learners. The difference between $S_1$ and $S_2$ suggests that NIs repeat the same lexical types while NNIs cover a wider range.

4.3. Comparing NI and NNI Tokens and Types against the Benchmark Wordlist

4.3.1. NI and NNI tokens and types ranked in the order of the benchmark wordlist

Out of the 126,210 tokens in our classroom corpus, 103,218 tokens (82%) were found in the benchmark wordlist. The breakdown was NI 15,765 (85%), NNI 77,765 (81%), and CD 9,688 (87%).

Figure 5 shows the cumulative percentages of NI and NNI tokens for each 1,000-type rank of the benchmark wordlist. As expected, coverage asymptotes quickly, and reaches 100% coverage only at the end, reflecting the long tail of lexical type distribution even at the upper-beginner to lower-intermediate level of course content.

Figure 6 shows the number of lexical types preferred by either NI or NNI per 1,000-type rank of the benchmark wordlist. We found that NIs should produce more types in the essential type group to become NNI-like, and the NNI should produce more types in the 2–1,001 rank to become NI-like.

4.3.2. Correlation coefficients of lexical types vs. the benchmark wordlist

Figure 7 shows correlation coefficients of lexical types appearing in our data sets vs. the benchmark wordlist. We found that the correlation between the union set $S_4 \cup S_7$ and the benchmark is high for the 1,000 most common types, but that the correlation coefficients drop as a rough function of benchmark wordlist frequency (i.e., the less frequent the type, the lower the correlation tends to be). The union sets $S_5 \cup S_7$ and $S_6 \cup S_7$ both highly correlate with the 1,000 most common types, but not with others. Beyond the 1,000 most frequent types we see low or negative correlation between NI and NNI vs. the benchmark wordlist. These findings are consistent with the interpretation that the 1,000 most frequent types would form the very core vocabulary of all instructors, and that textbook materials relate to specific topics that lack the breadth and depth of written and spoken English from which the benchmark wordlist is drawn.
4.4. Finding Types with High Keyness

4.4.1. Normalized token frequencies

The first step in identifying NI and NNI types with high keyness is comparing normalized NI and NNI token frequencies. We calculated frequency ratios of NI and NNI tokens with respect to their maximum frequency counts, and examined their distributions (Figs. 8 and 9). Although the frequency plots appear similar (Fig. 8), the distribution of types differs noticeably (Fig. 9), showing that types with high keyness may exist.

4.4.2. Chi-squared test on type frequencies

The second step in identifying NI and NNI types with high keyness is finding high-key candidates. Treating NI and NNI tokens as constituting mutually independent sets, we used a chi-square test to compare type frequencies with benchmark wordlist frequency rank.

![Benchmark wordlist frequency rank](image1)

**Fig. 6** Number of lexical types preferred by either NI or NNI per benchmark wordlist rank. NI or NNI token frequencies were normalized for each of the benchmark wordlist ranks according as the number of appearance in the video recording. The number of each NNI token was multiplied by the ratio of appearance in the classroom corpus i.e., the value of 9 (NNI’s number of classes) divided by 49 (NI’s number of classes) as shown in Table 1. The leftmost 2 columns include 137 essential lexical types that the benchmark wordlist deems equally worthy of instruction regardless of frequency in language (e.g., cardinal numbers, days of the week). The benchmark collectively treats these essential types as rank 1. The NNI uses such types more often than NIs (e.g., when referring to textbook page numbers). Types in the 2–1,001 rank are twice as likely to be preferred by NIs than the NNI. Types in the 1,002–2,001 rank appear at roughly identical frequencies for the NIs and the NNI.

![Correlation coefficients](image2)

**Fig. 7** Correlation coefficients of lexical types appearing in $S_4$, $S_5$, $S_6$, $S_7$ vs. the benchmark wordlist. The correlation between the union sets $S_k \cup S_7$, $S_k \cup S_7$, and $S_k \cup S_7$ vs. the benchmark wordlist is high for the 1,000 most common types.

![Normalized token frequencies](image3)

**Fig. 8** Normalized token frequencies of NIs and NNI ordered by their respective frequency ranks. NI and NNI raw frequency values were divided by their maximum values (rank = 1) and plotted on a logarithmic scale in the order of their respective ranks. Data points are connected with lines help visualize tendencies. Both NI and NNI plots steeply drop at around rank = 100. The NNI used more token types than the 3 NIs combined; hence the NNI’s plot extends further to the right.

![Normalized frequency](image4)

**Fig. 9** Normalized token frequencies of NIs and NNI ordered by NNI frequency rank. This figure is similar to Fig. 6 except that NNI normalized token frequencies are plotted in the order of NNI ranks where common tokens were found, and otherwise in the order of NI ranks. Data points are connected with lines to help visualize tendencies. The NI’s jagged plot shows how frequency counts for the same lexical type vary between NIs and NNI. The same tendency was found when this same graph was plotted in order of NI ranks.

We calculated frequency ratios of NI and NNI tokens with respect to their maximum frequency counts, and examined their distributions (Figs. 8 and 9). Although the frequency plots appear similar (Fig. 8), the distribution of types differs noticeably (Fig. 9), showing that types with high keyness may exist.

4.4.2. Chi-squared test on type frequencies

The second step in identifying NI and NNI types with high keyness is finding high-key candidates. Treating NI and NNI tokens as constituting mutually independent sets, we used a chi-square test to compare type frequencies with
Fig. 10 Histogram of significant NI vs. NNI ratio differences, sorted in the order of difference magnitude. The height of each bar represents the difference of NI and NNI frequency ratios for the same lexical type. NI and NNI frequencies are relative to their respective frequency values at rank = 1. Positive values indicate NI high-key candidate types \((n = 112)\), and negative values indicate NNI candidates \((n = 43)\).

Table 2 Top-5 NI high-key candidates. The top-5 NI high-key candidates \(i.e., \text{lexical types with the greatest NI vs. NNI ratio differences}\) are shown below. Not all candidates are true high-key types. For instance, “torch” is a content-dependent type that will be removed later.

| Candidate | \(f\) [dB] |
|-----------|-------------|
|           | NI         | NNI        | NI – NNI    |
| guy       | -21.46     | -36.51     | 11.05       |
| anyone    | -21.46     | -36.51     | 11.05       |
| wonder    | -21.46     | -36.51     | 11.05       |
| torch     | -21.46     | -36.51     | 11.05       |
| full      | -21.46     | -36.51     | 11.05       |

respect to the entire number of tokens. Figure 10 shows a histogram of significant NI vs. NNI ratio differences, sorted in the order of difference magnitude. The 155 types found are high-key candidates, of which samples are shown in Tables 2 and 3.

4.4.3. Removal of content-dependent types

The third step in identifying NI and NNI types with high keyness is removing context-dependent types. Because such types are specific to the lesson being taught, they interfere with observing the unbiased language use of NIs and NNI. As mentioned in Sect. 3.3, we did not remove CE tokens because discourse is expected in most if not all classroom interactions.

The context-dependent types we found were newly introduced types that appeared in the class textbook, e.g. “backpack,” “chimpanzees,” and “dome.” We removed them because they were obviously being taught through explanations or drills. Had these types not been the focus of instruction, it is unlikely they would have appeared in our data.

A total of 25 context-dependent types were removed, yielding 130 high-key types \((\text{NI 87, NNI 43})\).

Finally, we added an arbitrary constraint to reduce the likelihood of accidentally selecting a high-key type. We chose to retain types that were realized by the NIs or NNI as at least 2 tokens on average for each class period they chose to retain types that were realized by the NIs or NNI yielding 130 high-key types (NI 87, NNI 43).

Table 3 Top-5 NNI high-key candidates. The top-5 NNI high-key candidates \(i.e., \) lexical types with the greatest NI vs. NNI ratio differences are shown below.

| Candidate | \(f\) [dB] |
|-----------|-------------|
| underline | -29.24     | -12.21     | -17.03      |
| forty     | -29.24     | -14.15     | -15.09      |
| textbook  | -29.24     | -15.11     | -14.13      |
| CD        | -29.24     | -15.64     | -13.60      |
| page      | -21.46     | -8.26      | -13.20      |

Table 4 NI lexical types with high keyness. The NIs used 51 types significantly more frequently than the NNI \((p < 0.001)\) without regard to topics of the lessons. These might constitute NIs’ core vocabulary when conducting lessons. The least frequent type’s frequency was \(-16.69\)dB relative to the maximum value of 840.

| Type | \(f\) [dB] | Type | \(f\) [dB] | Type | \(f\) [dB] |
|------|------------|------|------------|------|------------|
| be   | 0.00       | about| -10.38     | just | -13.80     |
| you  | -2.54      | like | -10.49     | true | -14.06     |
| and  | -2.88      | but  | -10.55     | any  | -14.33     |
| so   | -4.63      | who  | -10.92     | could| -14.33     |
| I    | -4.76      | can  | -11.05     | his  | -14.33     |
| it   | -5.06      | would| -11.18     | little| -15.26     |
| of   | -6.76      | why  | -11.39     | talk  | -15.26     |
| that | -6.99      | know | -11.84     | most  | -15.26     |
| go   | -7.54      | because| -11.92    | everyone| -15.44    |
| not  | -7.78      | how  | -12.08     | much  | -15.63     |
| he   | -8.38      | then | -12.17     | maybe | -15.63     |
| yes  | -8.79      | my   | -12.34     | someone| -16.02     |
| say  | -8.99      | guy  | -12.52     | yourself| -16.02    |
| or   | -9.56      | group| -12.52     | him   | -16.02     |
| they | -9.70      | false| -12.52     | bit   | -16.23     |
| think| -10.16     | thank| -12.62     | great | -16.46     |
| people| -10.27   | she  | -12.81     | anyone| -16.69     |
There are types, as a whole, often used for oral interaction between instructors and students. Table 5 contains imperatives (e.g., “look,” “repeat,” “check,” “underline,” “begin,” “turn”), meta-instruction phrases (e.g., “question,” “right,” “textbook”), and meta-language phrases (e.g., “word,” “phrase,” “meaning”) that are often used for textbook-based drill activities. Over 90% of the types in Tables 4 and 5 fall within the 2,000 most common types (Fig. 5).

It is highly unlikely that the NIs are incapable of using types in Table 5, or that the NNI is ignorant of types in Table 4. What causes this skew?

As touched upon in Sect. 4.2, NIs were more likely to use communicative types to engage in open-ended conversations with students. The NNI was more likely to engage in teaching moves regarding textbook material. We believe that the lexical differences between these 2 tables reflect the differences of teaching roles between the NIs and the NNI, instead of differences in language ability.

5. DISCUSSION AND CONCLUSION

We state our findings (Sect. 5.1), implications to language pedagogy (Sect. 5.2), limitations (Sect. 5.3), and suggestions for further research (Sect. 5.4).

5.1. Findings

Our data show no evidence that the NNI code-switched to Japanese because the NNI’s English is limited compared to the NIs’. NIs and the NNI appear capable of producing equal amounts of tokens per unit time (Table 1). $S_4$ and $S_7$ show that NIs and the NNI seem to share a common core vocabulary of both general English and CE (Fig. 4). $S_4$ and $S_7$ also show high correlation with the 2,000 most frequent word types in the benchmark (Fig. 7). The similarity between NIs and NNI lexical distributions (Figs. 5 and 8) suggests that the NIs ($S_1 + S_3$) and the NNI ($S_2 + S_4$) employ comparable vocabulary sizes in the classroom.

The lexical types and tokens do differ between NIs and the NNI (Fig. 6), but the semantic coverage of lexical types found with potentially high keyness (Figs. 9 and 10, and Tables 4 and 5) is consistent with classroom observations that the NIs and NNI perform different roles in the classroom (partly based on the NIs’ lack of teacher training and their infrequent presence in class), and that these different roles impose different interactions with students, resulting in different lexical choices.

At the present time, it appears that both NIs and NNIs share a sizeable core of vocabulary ($S_4$ and $S_7$), NIs teach vocabulary depth ($S_1$), and NNIs teach vocabulary breadth ($S_2$). If NIs and NNIs are equally effective—or, at least they are effective for separate teacher roles—then NIs and NNIs should be able to learn teaching content and techniques from each other to be better understood by learners. The popular presumption that NNIs cannot speak English and ought to unilaterally learn from NIs seems overly simplistic, as far as our data show.

5.2. Implications to Language Pedagogy

Both NIs and NNIs may benefit from the findings in our research. Perhaps NIs use lexical types in Table 4 more often because they are native speakers or because they engage in general-purpose conversations with students. The fact that the precise cause is unknown is irrelevant for practical purposes. NNIs desiring to become more NI-like (by either becoming more native-like or engaging in general-purpose conversations) may do so, as a crude approximation, by using lexical types in Table 4. Conversely, NIs might become NNI-like (in teaching textbook content or managing the class) by using lexical types such as those in Table 5. Indeed, all that may be needed to cross-train NIs and NNIs may be to change their word-selection habits. Instead of lecturing NNIs to engage in open-ended oral queries using such-and-such phrases, for instance, NNIs could be encouraged to use personal pronouns more often. Conversely, NIs could be given a list of content words that, due to the classroom actions associated with those words, would help NIs assume NNI-like roles. This seemingly superficial claim presumes that NIs and NNIs are capable of modifying their behavior using implicit instruction by virtue of their language and teacher training. Testing this hypothesis on experienced in-service teachers would be of interest.

Less experienced or pre-service teachers might benefit from focusing on the vocabulary shared between NIs and NNIs. We believe that this common ground is part of the
CE core vocabulary, with which teachers should become completely familiar. Developing sample phrases and examples of classroom interaction may prepare middle and high schools in Japan to teach English language using mostly or solely English.

5.3. Limitations
Our conclusions should be regarded as tentative because our data have limited size and variety. The tiny sample size obviously limits generalizing our conclusions. We are continuing efforts to expand our corpus. Unfortunately, developing teacher corpora from recorded data requires significant resources [25]. For this study, we analyzed 3 NIs, and showed that our sole NNI is comparable with 5 other NNIs who teach different classes using CLT.

Including NNIs with various levels of L2 proficiency would allow us to better understand NNI issues. NNIs with different L1s would allow more global comparison.

The NIs lacked training and experience as teachers, and were not present during every class period. These factors must have affected the quality and quantity of spoken language. The NIs were native speakers of 3 varieties of English, however, improving the sample variance somewhat.

We do not necessarily consider as a limitation the fact that none of our NIs fluently spoke the learners’ L1, although proficiency is certainly valuable. Even without knowing the learners’ L1, many NIs can be trained as to how learners learn L2, and how the learners’ L1 affects their L2.

5.4. Further Research
Our study focused on lexical tokens and types in isolation. We need analyses involving word collocations, such as n-grams, or concordance lines using high-key tokens as nodes. Context-dependent analyses may help train teachers by showing examples of classroom phrases and interactions.

Another approach might use a core vocabulary wordlist such as [29] as a benchmark for analysis or teacher training. The fact that NIs taught with a limited vocabulary raises hopes of using minimal vocabulary to teach content (i.e., conveying information via L2, instead of focusing on teaching L2 itself).

ACKNOWLEDGMENTS
The authors are grateful to our learners and colleagues who agreed to participate in our study. This research was funded in part by Grants-in-Aid for Scientific Research (a kaken grant) from Japan Society for the Promotion of Science, and by a research grant from the Canon Foundation of Europe.

REFERENCES
[1] MEXT, “Statistical Abstract 2007 edition,” http://www.mext.go.jp/b_menu/shingi/ouhou/shotou/082/shiryo/jcsFiles/afieldsfile/2011/02/18/1301726_03.pdf (accessed 19 Sep. 2012).
[2] JET Programme, “JET Programme Participant Numbers,” http://www.jetprogramme.org/e/introduction/statistics.html#detail (accessed 19 Sep. 2012).
[3] JET Programme, “ALT Eligibility,” http://www.jetprogramme.org/e/aspiring/eligibility.html (accessed 6 Sep. 2012).
[4] Berlitz Japan, Inc., “English Instructor Job Description,” http://jobs.gajinpot.com/index/view/job/id/54530/lang/en#.UEiZsWxrcr (accessed 6 Sep. 2012).
[5] MEXT, “English language education improvement survey 2008,” http://www.mext.go.jp/b_menu/houdou/2003/08/031920/006.htm (accessed 17 May 2012).
[6] Tokyo Shoseki Publishing, “2009 Course of study for senior high schools: Foreign languages (English),” http://ten.tokyoshoseki.co.jp/downloadf1/pdf/hey76322.pdf?PHPSESSID=f125e9b860d1500343e6f1804092e59 (accessed 17 May 2012).
[7] MEXT, “The course of study for foreign languages 2009 (in Japanese),” http://www.mext.go.jp/a_menu/shotou/now-cs/youyou/kou/kou.pdf (accessed 17 May 2012).
[8] MEXT, “Progress report on 5 suggestions and measures to improve skills of English language as an international common language.” (in Japanese), p. 5 http://www.mext.go.jp/component/a_menu/education/detail/jcsFiles/afieldsfile/2012/03/21/1318781_02.pdf (accessed 19 Sep. 2012).
[9] K. Sakui, “Wearing two pairs of shoes: Language teaching in Japan,” ELT Journal, 58, 155–163 (2004).
[10] D. Liu, G. Ahn, K. Baek and N. Han, “South Korean high school English teachers’ code-switching: Questions and challenges in the drive for maximal use of English teaching,” TESOL QUARTERLY, 38, 605–638 (2004).
[11] M. Wada, “Teacher education for curricular innovation in Japan,” in Interpreting Communicative Language Teaching, S. J. Savignon, Ed. (Yale University Press, New Haven/London, 2002), pp. 31–40.
[12] K. Sakui and S. J. Gaiies, “Investigating Japanese learners’ beliefs about language learning,” System, 27, 473–492 (1999).
[13] K. Sato, “Practical understandings of communicative language teaching and teacher development,” in Interpreting Communicative Language Teaching, S. J. Savignon, Ed. (Yale University Press, New Haven/London, 2002), pp. 41–81.
[14] J. Sinclair and M. Coulthard, Towards an Analysis of Discourse: The English Used by Teachers and Pupils (Oxford University Press, Oxford, 1975), pp. 19–60.
[15] M. Hajer, “Creating a language-promoting classroom: Content-area teachers at work,” in Second and Foreign Language Learning through Classroom Interaction, J. Hall and L. Verplaatse, Eds. (Lawrence Erlbaum, Mahwah, N.J., 2000), pp. 265–285.
[16] J. S. P. Nation, Teaching and Learning Vocabulary (Newbury House, New York, 1990), pp. 11–12.
[17] M. McCarthy, “What constitutes a basic vocabulary for spoken communication?” http://www.bearsite.info/Articles/Language/What constitutes a basic vocabulary.pdf (accessed 17 May 2012).
[18] P. A. Duff, “Repetition in foreign language classroom interaction,” in Second and Foreign Language Learning through Classroom Interaction, J. Hall and L. Verplaatse, Eds. (Lawrence Erlbaum, Mahwah, N.J., 2000), pp. 109–138.
[19] Japan Exchange and Teaching Programme, “The JET programme,” http://www.jetprogramme.org/ (accessed 17 May 2012).
[20] SIL International, “The ILR (FSI) proficiency scale,” http://www.sil.org/lingualinks/languagelearning/mangngyrlngglngprgrm/theirilrfspiproficiencyscale.htm (accessed 17 May 2012).

[21] Council of Europe, “Reference level descriptions (RLD) for national and regional languages,” http://www.coe.int/t/dg4/linguistic/dnr_EN.asp (accessed 17 May 2012).

[22] MEXT, “Foreign language activities and English lesson demonstration video clips based on the new course of study, senior high school version,” Vol. 3, 1–3 [DVD] (Tokyo, 2012).

[23] B. Gardner and F. Gardner, Classroom English (Oxford University Press, Oxford, 2000), pp. 6–17.

[24] T. Teranishi, Classroom English, Vol. 1–3 (Kenkyusha, Tokyo, 1950).

[25] A. O’Keeffe, M. McCarthy and R. Carter, From Corpus to Classroom: Language Use and Language Teaching (Cambridge Language Teaching Library, Cambridge, 2007).

[26] Japan Association of College English Teachers, 8000 Basic Words (JACET Publishing, Tokyo, 2003), pp. 58–100.

[27] R. Carter, Vocabulary (Routledge, London, 1998), pp. 34–49.

[28] J. K. Hall, “Classroom interaction and additional language learning: Implications for teaching and research,” in Second and Foreign Language Learning through Classroom Interaction, J. Hall and L. Verplaetse, Eds. (Lawrence Erlbaum, Mahwah, N.J., 2000), pp. 287–298.

[29] M. West, A General Service List of English Words (Longman, Essex, 1953).

Noriaki Katagiri received the B.A. degree in linguistics from International Christian University, and the M.A. degree in International Media and Communication from Hokkaido University. He is currently a Ph.D. candidate at Hokkaido University graduate school. His interests are developing classroom corpora, analyzing native and non-native instructor speech, and training pre-service and in-service language instructors.

Goh Kawai is an associate professor of educational engineering and English language at Hokkaido University. He designs environments for learning spoken language. He has conducted research in California and Oregon, USA and in Belgium. Goh has a BA in linguistics, an MA in education, and a Ph.D. in information and communication engineering. http://goh.kawai.com/