Investigating Japanese EFL Learners' Overuse/Underuse of English Grammar Categories and Their Relevance to CEFR Levels

Yasutake Ishii, Yukio Tono
Seijo University, Tokyo University of Foreign Studies
ishii@seijo.ac.jp, y.tono@tufs.ac.jp

Abstract

This study focuses on L2 learners' overuse/underuse of English grammar categories and examines distributions of grammar use by Japanese EFL learners across different CEFR levels. A revised version of the Japanese EFL Learner (JEFLL) Corpus, a collection of approximately 10,000 compositions written by Japanese lower and upper secondary school students, classified according to CEFR levels, was used for this study. In addition, each composition was proofread by a native speaker and a parallel set of original and corrected versions was prepared. An inventory of grammar items and their accompanying corpus query syntax was developed using combinations of lemmas and parts of speech, which yielded 263 grammar categories. Altogether 501 individual grammar items were extracted from both original and proofread versions of the JEFLL Corpus. Overuse/underuse of grammar categories was examined by the paired comparison between students' original essays and their corresponding proofread essays across each CEFR-level group.

Keywords: CEFR, overuse/underuse, grammar profile

1. Introduction

There is a growing awareness of the importance of teaching grammar not as a mere list of grammar knowledge, but as an integral part of communicative competence that comprises the combination of declarative and procedural knowledge. This view is also reflected in the approach of the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001). Performance objectives are described as "can do" descriptors and are stated in a language-neutral way, and vocabulary and grammar for each individual language are defined as a part of reference level descriptions (RLD), a group of inventories of language items featured in each CEFR level. With the growing influence of the CEFR, a series of studies has been conducted in order to investigate the method of developing a language framework such as the CEFR and how to make RLDs for the framework. The project in Japan is called the CEFR-J (Tono, 2013). Our primary interest is to construct a CEFR-based open framework for English language teaching especially for Japanese learners of English.

To this end, we have collected our own can-do descriptors and scaled descriptors using the same statistical approach used for the original CEFR (North, 2000). We also conducted a series of corpus analyses to identify lexical and grammatical items that should be introduced at each of the CEFR or CEFR-J levels (Tono, 2017). One of our unique approaches toward developing an RLD is to use Japanese EFL learner corpora to see how Japanese learners of English at each CEFR level can produce English in terms of vocabulary and grammar. This study is a part of such RLD projects, where we compare the characteristics of learner English across different CEFR levels and hope to identify so-called "criterial features" for a given CEFR level in terms of learners' overuse and underuse of particular grammatical structures in English.

2. Literature review

There is a long history in SLA research on the acquisition order of grammar items (cf. Ellis, 2008), but most of the studies focus on a particular grammar construction or a group of categories such as tense and aspect (Bardovi-Harlig, 2000) or negation (Meisel, 1997). Recently, more data-intensive approaches are becoming popular. The Educational Testing Service (ETS), for instance, has used NLP technologies to automatically identify and correct errors using large native speaker and non-native speakers' corpora (Leacock and Chodorow, 2003). Attempts have also been made to improve the accuracy of automated scoring and automatic error correction (Burstein, Tetreault and Madnani, 2013). Their primary aim is to contribute to automated scoring in language testing.

There are groups of people who are more interested in L2 learner profiling research. The English Profile (Hawkins and Filipović, 2012) examined the Cambridge Learner Corpus (CLC) to extract grammatical features that serve as criteria to distinguish a given CEFR level from the others. Murakami (2014) used large corpora (e.g. CLC) to re-examine the order of grammatical morphemes. In the same vein, the CEFR-J project team conducted corpus-based criterial feature extraction in terms of grammar (Ishii, 2016) and textual features (Mizushima, Arase and Uchida, 2016). Ishii and his colleagues in the CEFR-J project prepared a list of English grammar categories to be extracted from the CEFR Coursebook Corpus, specially compiled for the project (Ishii and Tono, 2016). The same analysis was made on the learner data (Notohara, 2016), including the JEFLL Corpus (Tono, 2007; The JEFLL Corpus Project, 2007) and the NICT JLE Corpus (Izumi, Uchimoto and Isahara, 2004; National Institute of Information and Communications Technology, 2012).

This paper is a follow-up study of the CEFR-J Grammar Profile project. In this study, we will focus on the patterns of overuse vs. underuse for given grammar constructions by comparing learners' original writings with proofread versions of the same writings.

3. Method

It is not straightforward to identify and count the frequencies of grammatical items used in a given text. We need to get over some obstacles such as what grammatical items are in the first place, what patterns those items can take, and how we can identify and count those items in large texts, each of which will be described below.

3.1 The JEFLL Corpus

The corpus used in this present study is the Japanese EFL Learner (JEFLL) Corpus, which is a collection of English essays written by junior and senior high school students in Japan. The total number of essays is 10,038 and the total size of the corpus is 669,304 running words. The original sampling frame for the JEFLL Corpus was the grade levels for each school type (junior and senior high): Junior High 1 (1,393 files; 51,160 words); Junior High 2 (2,635 files; 159,741 words); Junior High 3 (1,589 files; 117,764 words); Senior High 1 (742 files; 60,713 words); Senior High 2 (1,977 files; 170,557 words); Senior High 3 (1,189 files; 78,981 words). For data elicitation, the six essay tasks were controlled in terms of text types (argumentative vs. narrative) and possible time expressions (past, present and future). The participants were asked to write an essay without the use of dictionaries in 20 minutes in class. All the compositions were hand-written and digitized later.

Currently, the JEFLL Corpus has a parallel corpus version and the CEFR-based version. The former was compiled by asking native speakers to proofread all the essays so that the original and corrected versions were prepared as a parallel corpus. We also asked those who were familiar with the CEFR to re-evaluate essays according to CEFR levels, thus producing the CEFR-based version of original vs. corrected essays. We used this version for our analysis. The total size of the original and corrected versions is shown in Table 1:

	A1	A2	B1	B2
Original	131,525	309,561	212,158	8,658
Corrected	153,887	338,696	226,600	9,092

Table 1: The JEFLL-CEFR Corpus: subcorpus breakdown

As one can see, the size of the B2-level subcorpus is much smaller than the other three, which might affect the results of less frequent grammar items. The following results provide the statistics including the B2 level nonetheless, because there is some useful information for frequent grammar items therein. Further research will be needed to sort out the effects of the unbalanced corpus size.

3.2 Selection of Grammar Categories

While Japan is a country where we teach English as a foreign language (EFL) and we have developed a unique grammar-based curriculum for English language teaching, the CEFR is now becoming influential as an international standard for planning language policies, developing national curriculum, designing teaching materials and assessing the outcomes. Therefore, as to grammar instruction, we need to have a clearer idea of how Japanese EFL learners learn and use grammar items in terms of the CEFR-J levels. In this context, together with other members of the CEFR-J development team, we have selected 263 grammar categories that are widely recognized and accepted in Japan, drawing on some work in the previous literature including North et al. (2010), English Grammar Profile (2015), and a list of grammatical items developed by Professor Hiroshi Sano at Tokyo University of Foreign Studies and his colleagues (Sano Lab, 2005), which was compiled based on a survey of major authorized English textbooks and reference books widely used in Japan. This has yielded 501 different patterns in total, if we distinguish the same items in different sentence patterns such as the affirmative declarative and the negative question. The whole database we have created and made available on the web is called the CEFR-J Grammar Profile.

3.3 Building Query Patterns for Each Grammar Item

We have made query patterns for each item using regular expressions searching for combinations of word forms, lemmas and parts of speech. Table 2 shows some examples.

ID	Item	Pattern
26	INDEFINITE PRONOUN: none	\bnone_NN_none\b
49	COMPARATIVE and COMPARATIVE (the same adjective)	\b(\S+_(JJR|RBR)_\S+) and_CC_and \1
66	TENSE/ASPECT: PAST PROGRESSIVE (AFFIRMATIVE DECLARATIVE)	(was|were)_VBD_be(?! (going_VVG_go to_TO_to|gonna_VVG_gonna) \S+_V._\S+) \S+_V.G_\S+
145	AUX+PERFECT (AFFIRMATIVE DECLARATIVE)	(?!cannot\b)\S+_MD_\S+ have_VH_have \S+_V.N_\S+

Table 2: Part of the items adopted in the CEFR-J Grammar Profile

The texts to be analyzed were processed on TreeTagger (Schmid, 1994), by which each word was morphologically analysed and provided with its word form, lemma and part of speech. Patterns of all the 501 grammar items were automatically extracted and counted from the two sets of corpora: the original JEFLL and the corrected JEFLL, and exported into a CSV file.

3.4 Counting the Frequency of Each Grammar Category in Our Corpora

To analyze how grammar categories are used by Japanese EFL learners, we compiled a frequency table, part of which is shown in Figure 1. The resulting data reveal which grammar categories are frequently or infrequently used in the learners' writings and their proofread versions.

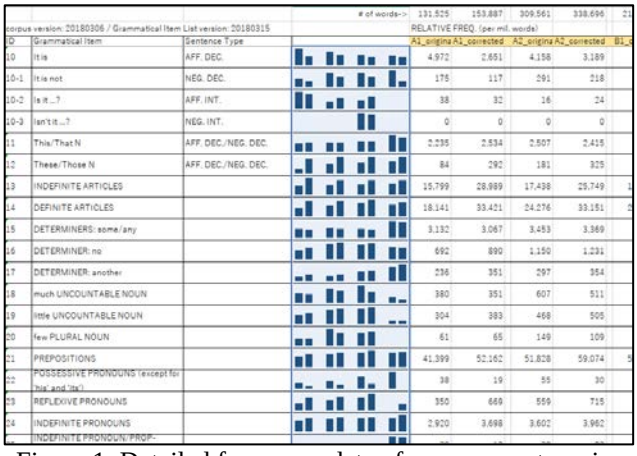

Figure 1: Detailed frequency data of grammar categories in the JEFLL Corpus.

4. Results and Discussion

The grammar items focused on in this study were chosen based on their frequencies in the learners' original writings, whereby out of 501 grammar items in our grammar profile,

193 items with the raw frequency of 20 or over were selected. The analysis was based upon the overall increase and decrease of the given grammar items that occurred in the essays across different CEFR levels. The frequency of each grammar item was defined as the relative frequency of that particular grammar item (per million words) in a subcorpus of the JEFLL-CEFR. The overuse/underuse was determined by calculating the ratio of the number of students' original uses over that of native speakers' corrections. For example, take the case of "I am ..." in Table 3:

I am ...	A1	A2	B1	B2
original	4,858	3,049	2,512	2,888
corrected	3,574	2,979	2,573	3,410
Ratio	1.36:1	1.02:1	0.98:1	0.85:1

Table 3: The distribution of "*I am ...*" across CEFR levels between the original and corrected versions of JEFLL

At A1 level, for example, 4,858 occurrences of "*I am ...*" were observed in the students' original writings, whereas in the corrected version, only 3,574 cases were found. (Note that the frequency figures given in this paper are relative frequencies per million words, except for those indicated otherwise.) It means that after native speakers' corrections, about a quarter of the use of "*I am ...*" was changed to some other constructions. Here the ratio of the original essays over the corrected ones was 1.36:1, which shows that A1-users tended to overuse "*I am …*" 136%, compared to the native speakers' corrected versions. This overuse tendency gradually decreases as the CEFR level increases. At the B1 level, for instance, the original essays only contained 2,512 cases of "*I am …*" compared to 2,573 cases in the proofread essays. Thus, the original to corrected ratio was 0.98 to 1, which means compared to the corrected version, 98% of the items occurred at the B1 level.

4.1 The Overall Tendency of Underuse

Table 4 shows the number of grammar items that are underused compared to the corrected version of the essays:

Underuse	A1	A2	B1	B2
0 to 49%	56	17	12	61
50 to 74%	43	38	34	27
75% & above	53	80	87	38
Total	152	135	133	126

Table 4: The number of grammar items underused[1]

The results show that at A1 level, there were 56 grammar items that were used less than half of their frequencies in the corrected version, but this rate rapidly decreased to 17 at A2 and 12 at B1 respectively. The figures for B2 level increased again, but this may not be very accurate due to the lack of B2-level data, which caused there to be many missing grammar items. Overall, the underuse phenomena will gradually disappear as the level goes up.

The twenty most underused grammar items in the JEFLL Corpus are shown in Table 5. It is worth noting that many items involve the combination of more than one grammar item and demand a quite heavy processing load. For instance, the most underused item is the construction "*being* + PAST PARTICIPLE," which is a complex combination of participle construction and passive voice. The original essays used this construction with the following frequencies: 0 (A1), 36 (A2), 71 (B1), 0 (B2), whereas in the corrected essays it occurred 201 (A1), 139 (A2), 212 (B1), 330 (B2) times. The correlation between the two essays was $r = -.434$, which means the usage pattern shows a negative correlation. Native speakers use this pattern in corrected essays, whereas learners constantly avoided it. The same kind of tendency is observed in the case of the second most underused item, "PREP + RELATIVE PRONOUN." This construction involves the raising of prepositions together with relative pronouns with oblique cases, such as "*of which*" or "*in which.*" The knowledge of appropriate choice of prepositions and relative pronouns poses a problem again, which leads to the constant underuse by the learners, while native speakers use this construction much more frequently, as shown in corrected essays.

Lexical forms	A1	A2	B1	B2	Average
being + PAST PARTICIPLE	0.000	0.256	0.334	0.000	0.147
PREP + RELATIVE PRONOUN	0.090	0.283	0.226	0.000	0.150
PASSIVE: AUX: NEGATIVE	0.130	0.259	0.467	0.000	0.214
TENSE/ASPECT: PRESENT PERFECT: NEGATIVE	0.268	0.398	0.415	0.000	0.270
RELATIVE PRONOUN: NONRESTRICTIVE	0.047	0.322	0.303	0.525	0.299
MODAL/AUX: *should*: NEGATIVE	0.000	0.547	0.691	0.000	0.310
RECIPROCAL PRONOUN: *each other*	0.731	0.421	0.115	0.000	0.317
TENSE/ASPECT: PRESENT PERFECT PROGRESSIVE	0.146	0.505	0.668	0.000	0.330
whether	0.351	0.469	0.534	0.000	0.338
WH- QUESTION: *When ...?*	0.585	0.377	0.420	0.000	0.345
PASSIVE: PAST PERFECT	0.123	0.729	0.551	0.000	0.351
PASSIVE: FUTURE	0.293	0.584	0.554	0.000	0.357
TENSE/ASPECT: PAST PERFECT	0.189	0.587	0.473	0.242	0.373
MODAL/AUX: *had better*	0.000	0.761	0.777	0.000	0.384
MODAL/AUX: *be able to*	0.178	0.384	0.417	0.630	0.402
AUX + PERFECT	0.251	0.452	0.534	0.420	0.414
be to DO	0.293	0.578	0.809	0.000	0.420
such (a/an) ADJ NOUN	0.439	0.752	0.577	0.000	0.442
know/wonder + WH-(CLAUSE) (except for *whether*)	0.334	0.729	0.765	0.000	0.457
MODAL/AUX: *will*: NEGATIVE	0.439	0.540	0.534	0.350	0.466

Table 5: The twenty most underused grammar items in the JEFLL Corpus[2]

[1] To avoid division by zero when calculating the ratios, we substituted 0's in the frequencies in the corrected versions with 1's. The same is applied to the data underlying Table 5.

[2] The scores indicate the ratio of the frequency of grammar items in the original to that of the corrected version.

Such combination of grammar items can be seen in other items such as PRESENT PERFECT PROGRESSIVE (present prefect + progressive), PASSIVE: PAST PERFECT (passive + past perfect), PASSIVE: FUTURE (passive + future), and AUX + PERFECT (auxiliary verb + perfect). One of the reasons for underuse of these items may be that the structure itself is very complex, involving more than one grammar item, which gives learners heavy burdens to process and leads to underuse phenomena.

Another possible factor causing underuse is related to how the item is taught. For example, there is another type of relative pronoun, ranked fifth among the underused items, which is a non-restrictive relative clause. This construction is usually introduced in a syllabus after teaching a set of restrictive relative clauses, but the treatment of non-restrictive relative clauses in class seems to be marginalized and not very systematic. Despite the fact that this construction is quite frequently used by native speakers, Japanese EFL learners find it difficult to use in writings.

Some grammar items might involve difficulty in acquisition due to functional-semantic problems. For instance, there are many TENSE/ASPECT categories such as present perfect, present perfect progressive, and past perfect, ranked in the top 20 underused items. These constructions are usually introduced around A2 level, so it is natural that A1-level essays do not include them. There is a tendency that learners started using the constructions from A2 to B1 levels, but still the rate of using the constructions was constantly lower than the native speakers' corrected version of the essays. These tense/aspect markers are known to be problematic for L2 learners, due to the gap regarding how to express tense/aspect between L1 and L2 (cf. Bardovi-Harlig, 2000).

4.2 The Overall Tendency of Overuse

There are not as many grammar items which are constantly overused across the CEFR levels. Table 6 shows some of those items.

Lexical forms	A1	A2	B1	B2	Average
These/Those are ...	5.265	2.188	2.594	1.050	2.774
he/she is ...	3.385	1.647	1.617	1.470	2.030
there + be ...: NEGATIVE	1.404	1.810	1.978	2.100	1.823
It is not ...	1.495	1.331	1.308	3.150	1.821
ELLIPTICAL ACCUSATIVE RELATIVE PRONOUN	1.209	1.245	1.434	2.071	1.489
much + UNCOUNTABLE NOUN	1.083	1.189	1.576	2.100	1.487
MODAL/AUX: *would*: NEGATIVE	0.439	0.574	0.518	4.201	1.433
COMPARATIVE OF SUPERIORITY: *more* + ADJ/ADV	0.794	0.802	0.867	3.150	1.403
It is ...	1.875	1.304	1.244	1.187	1.403
TENSE/ASPECT: PRESENT (BE)	1.791	1.320	1.248	1.214	1.393
TENSE/ASPECT: PRESENT (BE): NEGATIVE	1.390	1.346	1.258	1.470	1.366

Table 6: The most overused grammar items in the JEFLL Corpus[3]

Many overused grammar items involve the use of *be*-verb, as in "*These* [*Those*] *are ...*", "*he* [*she*] *is ...*", "*there is* [*are*] not ...", "*It is not ...*", "*It is ...*", and "*be*" in the present tense. Especially the distinction of copula *be* and lexical verbs is confusing to Japanese learners of English. The Japanese language has a topic-comment structure, which is often confused with a subject-predicate construction in English (e.g. *Boku wa* [TOPIC: 'as for me'] *Ramen da* [COMMENT: 'it's ramen']. vs. *Kare wa* [SUBJECT: 'He is'] *daigakusei da* [PREDICATE: 'a college student'].). The writings in the JEFLL Corpus at A levels contain many wrong choices of *be*-verb directly mapped from Japanese TOPIC-COMMENT structures, which actually should have been expressed with different lexical verbs in English.

4.3 Interaction of overuse/underuse and CEFR levels

Some overuse/underuse phenomena are more specifically related to a particular CEFR level. We will discuss some of the most noticeable ones in this section. One of the most striking differences was found in postpositive past participle construction (Table 7).

	A1	A2	B1	B2
Original	105	331	228	10
	(798.3)	(1069.3)	(1074.7)	(1155.0)
Corrected	337	337	216	7
	(2189.9)	(1113.1)	(953.2)	(769.9)

Table 7: The use of postpositive past participle[4]

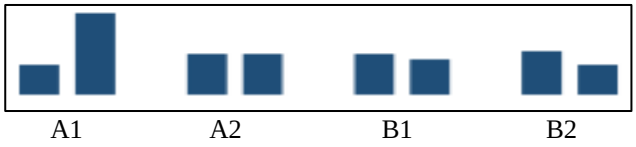

Figure 2: The use of postpositive past participle between the original (left) vs. corrected (right) essays.[5]

A1-level users used this construction 105 times in the original essays in total, whereas the corrected essays contain 337 occurrences, which shows that A1-level users either could not produce or avoided the construction. At A2 level the difference is much smaller, and at B1 and B2 levels the proofread versions have slightly fewer instances, which suggests that B-level learners slightly overused the construction, but native speakers corrected some of their uses without using the construction. However, the number of observations in raw data at B2 level was rather small, so the result of B2 levels is not very conclusive.

[3] Those items with at least one zero in the frequency are omitted from this data.

[4] The scores indicate raw frequencies and normalized frequencies in parentheses. The same is applied in Table 8 and thereafter.

[5] The scales of the y-axes are set differently among the different levels. The same is applied in Figure 3.

The same thing can be said about past perfect construction in the affirmative declarative (Table 8). In contrast to postpositive past participle construction, the past perfect construction shows the persistent tendency of underuse throughout A1 to B2 levels.

	A1	A2	B1	B2
Original	27	255	219	3
	(205.3)	(823.7)	(1032.3)	(346.5)
Corrected	167	475	494	13
	(1085.2)	(1402.4)	(2180.1)	(1429.8)

Table 8: The use of past perfect

Figure 3: The use of past perfect between the original (left) vs. corrected (right) essays.

It is not surprising that A-level users cannot use these constructions very well, but the result is suggestive in that there are many opportunities to use the postpositive past participle or past perfect constructions in even A1-level writings. This gap has to be filled by using alternative constructions, for example, splitting the participle construction into two independent constructions (e.g. simple noun and independent clause describing the noun). Identifying such gaps will shed light on the nature of more effective grammar instructions focusing on what learners can do with language at different CEFR levels.

So far, we have examined statistically most salient overused or underused items, many of which happen to be items to be introduced toward A2 or later stages. What about other more basic grammar items such as articles or prepositions? These two have been very popular grammar errors under study in NLP areas for automated scoring (Burstein, Tetreault and Madnani, 2013).

Tables 9 and 10 show the frequencies of definite and indefinite articles respectively.

	A1	A2	B1	B2
Original	2,386	7,515	6,049	228
	(18,141)	(24,276)	(28,512)	(26,334)
Corrected	5,143	11,228	8,216	279
	(33,421)	(33,151)	(36,258)	(30,686)

Table 9: The use of definite article *the*

	A1	A2	B1	B2
Original	2,078	5,398	3,963	164
	(15,799)	(17,438)	(18,679)	(18,942)
Corrected	4,461	8,721	5,855	228
	(28,989)	(25,749)	(25,838)	(25,077)

Table 10: The use of indefinite article *a/an*

In both cases, there is a marked tendency showing that Japanese EFL learners underused definite/indefinite articles, compared to the proofread versions. The average ratio of article use by the learners to that of native speakers is 0.70, which is about 30% less than the target corrections made by native proofreaders. The underuse ratio, however, was found to be very high when we looked at A1 level only, which is approximately 0.54 for both definite and indefinite articles. This clearly indicates that the learners at A1 level tend to omit articles in writings. The omission errors will decrease in number as the level goes up, but still there is a constant underuse phenomenon observed throughout all levels.

Table 11 shows the use of prepositions in the original and corrected versions of JEFLL.

	A1	A2	B1	B2
Original	5,445	16,044	12,304	495
	(41,399)	(51,828)	(57,995)	(57,173)
Corrected	8,027	20,008	15,009	554
	(52,162)	(59,074)	(66,236)	(60,933)

Table 11: The use of prepositions

Prepositions are quite common grammar items and the usage ratio between learners and native speakers is 0.87:1. As far as overuse/underuse phenomena are concerned, prepositions are much more readily supplied than the definite or indefinite articles. Preposition errors are more serious when it comes to the selection of proper prepositions in context. Thus, it is important to know what types of errors are most relevant, depending on types of grammar items.

5. Conclusions

In this study, the overuse and underuse of grammar categories were examined by comparing Japanese EFL students' original essays and their proofread versions. The analysis of the most underused items suggests important pedagogical implications. First, the underused constructions often involve a complex combination of grammar items, which is usually introduced one at a time throughout the course, but as the CEFR levels go up, learners are supposed to produce complex sentences by combining two or more constructions at the same time. Unless teachers are aware of the underuse of those complex items, learners may not have sufficient knowledge or opportunities to use those items. Hawkins and Filipović (2012) also pointed out a similar case of combination of different grammar knowledge, such as prepositions followed by verb-*ing* forms (gerund).

Another interesting finding is that many overused grammar items are related to the use of copula *be*, and for Japanese learners of English, the use of copula is quite problematic in a cross-linguistic sense. The function-form mapping involving subject-predicate vs. topic-comment structures is extremely complicated and Japanese A-level users produced errors related to mapping those two functions into proper constructions in English. This has much to do with English and Japanese verb semantics and their alternation patterns like the ones proposed by Levin (1993). It is inspiring that a bird's-eye-view comparison between the original and the corrected essays in terms of grammar item usage clearly shows those tendencies. Pedagogically, teachers should keep in mind that some of those overused items need special attention as learners attempt to use them in output activities.

Finally, the overall analysis of overuse/underuse phenomena showed that some of the previously well-known underused items such as definite/indefinite articles or prepositions turned out to be moderately underused, compared to those highly underused grammar items shown in Table 5. For these, however, a more complex picture emerged, suggesting that a closer attention needs to be paid

on the complex relationship between CEFR levels and the emergence of different error patterns (overuse, underuse, misuse). We hope that the results of the present study will provide the basic descriptive data which will shed light on more pedagogically relevant and theoretically sound approaches toward the treatment of errors.

6. Acknowledgements

This work is supported by JSPS KAKENHI Grant Numbers JP16H01935 and 16K16882.

7. Bibliographical References

Bardovi-Harlig, K. (2000). *Tense and Aspect in Second Language Acquisition: Form, Meaning, and Use*. Malden, MA: Wiley-Blackwell.

Burstein, J., Tetreault, J. and Madnani, N. (2013). The e-rater automated essay scoring system. In M. D. Shermis and J. Burstein (Eds.), *Handbook of Automated Essay Scoring: Current Applications and Future Directions*. New York: Routledge, pp. 55–67.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.

Ellis, R. (2008). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

English Grammar Profile. (2015). Available from English Grammar Profile Web site, http://www.englishprofile.org/english-grammar-profile

Hawkins, J. and Filipović, L. (2012). *Criterial Features in L2 English: Specifying the Reference Levels of the Common European Framework* (English Profile Studies). Cambridge: Cambridge University Press.

Ishii, Y. (2016). Grammatical items for creating the CEFR-J Grammar Profile. Paper presented at the Symposium on the CEFR-J RLD Project: Developing Grammar, Text and Error Profiles Using Textbook & Learner Corpora. March 5, 2016. Tokyo University of Foreign Studies.

Ishii, Y. and Tono, Y. (2016). CEFR-J Grammar Profile no tame no Bunpo koumoku hindo-chosa [A frequency analysis of grammar categories for the CEFR-J Grammar Profile]. Paper presented at the Symposium on the CEFR-J RLD Project: Developing Grammar, Text and Error Profiles Using Textbook & Learner Corpora. March 5, 2016. Tokyo University of Foreign Studies.

Izumi, E., Uchimoto, K. and Isahara, H. (2004). The NICT JLE corpus: Exploiting the language learners' speech database for research and education. *International Journal of the Computer, the Internet and Management*, 12(2): 119–125.

Leacock, C. and Chodorow, M. (2003). Automated grammatical error detection. In M. D. Shermis and J. Burstein (Eds.) *Automated Essay Scoring: A Cross-Disciplinary Perspective*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers, pp.195–207

Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.

Meisel, J. (1997). The acquisition of the syntax of negation in French and German: Contrasting first and second language development. *Second Language Research* 13: 227–263.

Mizushima, K., Arase, Y. and Uchida, S. (2016). CEFR junkyo-kyouzai ni okeru goi bunpou no tokuchou bunseki to level jidou bunrui [Lexico-grammatical feature analysis and automatic level classification using CEFR-based teaching materials]. *Proceedings of the Twenty-second Annual Meeting of the Association for Natural Language Processing*, pp. 789–792.

Murakami, A. (2014). *Individual Variation and the Role of L1 in the L2 Development of English Grammatical Morphemes: Insights from Learner Corpora*. PhD thesis. University of Cambridge, Cambridge, United Kingdom.

North, B. (2000). *The Development of a Common Framework Scale of Language Proficiency*. New York: Peter Lang.

North, B., Ortega, A. and Sheehan, S. (2010). *A core Inventory for General English*. British Council / EAQUALS, available at https://englishagenda.britishcouncil.org/sites/default/files/attachments/books-british-council-eaquals-core-inventory.pdf

Notohara, Y. (2016). English sentence patterns and criterial features for the CEFR levels. Paper presented at the Symposium on the CEFR-J RLD Project: Developing Grammar, Text and Error Profiles Using Textbook & Learner Corpora. March 5, 2016. Tokyo University of Foreign Studies.

Sano Lab. (2005). *Bunpou koumoku betsu BNC youreisyuu oyobi bunpou koumokusyuu* [Classification of examples collected from the BNC by grammatical items]. (Version 1.0). http://bnc.jkn21.com/nc_docs/.

Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. *Proceedings of International Conference on New Methods in Language Processing*, pp. 45–49.

Tono, Y. (Ed.). (2007). *Chukousei 1-man nin no eigo corpus: The JEFLL corpus* [A corpus of 10,000 Japanese secondary school students' writings: the JEFLL corpus]. Tokyo: Shogakukan.

Tono, Y. (Ed.). (2013). *The CEFR-J Handbook*. Tokyo: Taishukan Shoten.

Tono, Y. (2017). The CEFR-J and its impact on English language teaching in Japan. *JACET international convention selected papers*, 4: 31–52.

8. Language Resource References

National Institute of Information and Communications Technology. (2012). *NICT JLE Corpus*, Version 4.1, distributed at https://alaginrc.nict.go.jp/nict_jle/index_E.html

The JEFLL Corpus Project. (2007). The JEFLL (Japanese EFL Learner) Corpus, accessed via Shogakukan Corpus Network (https://scnweb.japanknowledge.com/).