

Investigating the frequency and dispersion of English grammatical items in textbooks and learner corpora: For more informed ELT practice
Yasutake ISHII (Seijo University, ishii@seijo.ac.jp)

It is clear that when we teach or learn a foreign/second language, we need to use materials which are appropriate for the target learners. One of the defining factors of the appropriateness is difficulty, which is typically dependent on vocabulary and grammar. While it is relatively easy to define the levels of vocabulary used in texts, 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.

Japan is an EFL country and has developed a unique grammar-based curriculum for English language teaching. However, the CEFR (Common European Framework of Reference for Languages: Learning, Teaching, Assessment) is now being introduced in the national curriculum, and we have been developing the CEFR-J to adapt and refine the CEFR to the Japanese EFL environment where most learners are at A1 or A2 levels. As to grammar, we need to have a clearer idea of how Japanese EFL learners learn and use grammatical items in terms of the CEFR-J levels. In this context, together with other members of the CEFR-J development team, we have created the CEFR-J Grammar Profile, which consists of a comprehensive list of grammatical items and their frequencies in different corpora. In this project, I have selected 263 grammatical items that are widely recognized and accepted in Japan, drawing on some work in the previous literature including North et al. (2010) and *English Grammar Profile* (2015). 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. I have also made regular expression patterns for each item making use of word forms, lemmas and parts of speech, some examples of which are shown in Table 1. The texts to be analyzed are processed on TreeTagger (Schmid, 1994), and each word is provided with its lemma and part of speech. As to the accuracy of the definitions of our grammatical items, they can be deemed accurate enough in general, as our manual checking of 207 items in a typical ELT coursebook in the previous version of our data yielded the average precision, recall and F-measure of 0.947, 0.891 and 0.892 respectively (Ishii, 2016).

To analyze how grammatical items are currently taught in Japan, I created a corpus of all government-authorized English textbooks used in junior and senior high schools (grades 7-12), and compiled a frequency table, part of which is shown in Figure 1. I also analyzed different corpora including EFL coursebooks based on CEFR levels and Japanese EFL learner corpora. The resulting data reveal which grammatical items are frequently and infrequently used in the input to learners as well as in their output, and shed light on such basic grammatical notions as definiteness, tense and aspect which are conceived of differently by Japanese EFL learners and native speakers. Our findings will help improve the curriculum and teaching methods in Japan.

This study composes a part of CEFR-J Grammar Profile which has been developed to be used by EFL learners, teachers and material makers in Japan. We are focusing on Japanese learners of English, but the methods and resources adopted in our study are open to the public and could be used in other EFL/ESL environments.

Table 1

Examples of grammatical items and their patterns

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+

corpus version: 4.0 (20180205) / Grammatical Item List version: 20180315

of words-> 31,838 45,032 47,022 229,397 314,168 309,041

RELATIVE FREQ. (per mil. words)

RANGE

ID	Grammatical Item	Sentence Type		Grade 7	Grade 8	Grade 9	Grade 10	Grade 11	Grade 13	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11	Grade 13
73	PASSIVE: PRESENT	AFF. DEC.		63	1,532	2,637	2,084	2,483	2,799	2	6	6	25	25	21
73-1	PASSIVE: PRESENT	NEG. DEC.		0	67	21	65	86	94	0	3	1	9	15	16
73-2	PASSIVE: PRESENT	AFF. INT.		31	44	64	17	3	3	1	2	2	3	1	1
73-3	PASSIVE: PRESENT	NEG. INT.		0	0	0	0	0	0	0	0	0	0	0	0
74	PASSIVE: PRESENT PROGRESSIVE	AFF. DEC.		0	0	0	57	105	97	0	0	0	5	12	14
74-1	PASSIVE: PRESENT PROGRESSIVE	NEG. DEC.		0	0	0	4	0	6	0	0	0	1	0	1
74-2	PASSIVE: PRESENT PROGRESSIVE	AFF. INT.		0	0	0	0	3	3	0	0	0	0	1	1
74-3	PASSIVE: PRESENT PROGRESSIVE	NEG. INT.		0	0	0	0	0	0	0	0	0	0	0	0
75	PASSIVE: PRESENT PERFECT	AFF. DEC.		0	0	43	214	341	408	0	0	1	12	22	21
75-1	PASSIVE: PRESENT PERFECT	NEG. DEC.		0	0	0	17	22	13	0	0	0	4	6	4
75-2	PASSIVE: PRESENT PERFECT	AFF. INT.		0	0	0	4	0	0	0	0	0	1	0	0
75-3	PASSIVE: PRESENT PERFECT	NEG. INT.		0	0	0	0	0	0	0	0	0	0	0	0
76	PASSIVE: PAST	AFF. DEC.		31	1,199	3,530	3,405	3,520	2,796	1	6	6	25	25	21
76-1	PASSIVE: PAST	NEG. DEC.		0	67	64	74	89	81	0	3	3	12	15	14
76-2	PASSIVE: PAST	AFF. INT.		0	44	106	35	13	0	0	2	3	5	4	0
76-3	PASSIVE: PAST	NEG. INT.		0	0	0	0	0	0	0	0	0	0	0	0

Figure 1. Detailed frequency data of some grammatical items in authorized English textbooks used in Japan.

References

- English Grammar Profile*. (2015). Available from English Grammar Profile Web site, <http://www.englishprofile.org/english-grammar-profile>
- Ishii, Y. (2016, September). The development of the CEFR-J Grammar Profile. Paper presented at the symposium "The development of the grammar/text/error profiles for the CEFR-J" conducted at the JACET (Japan Association of College English Teachers) 55th International Convention, Hokkai Gakuen University, Hokkaido, Japan.
- North, B., Ortega, A., & Sheehan, S. (2010). *A core inventory for general English*. British Council / EAQUALS.
- Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. *Proceedings of international conference on new methods in language processing*, 45–49.