THE REFLEXIVE CONSTRUCTION AND ELT: A CORPUS PRESPECTIVE

BARRY GROSSMAN, Ph.D.
Professor, Hachinohe Gakuin University

1. INTRODUCTION

The reflexive construction is typically described in Japanese high school English guidebooks by way of reflexive pronouns, called 'saikidaimeshi' in Japanese, of the following types:

a. as an object pronoun (*a~ of mine, this ~ of mine*)
b. as an emphatic (*I myself did it*)
c. reflexive idioms such as
    a. *in itself*
    b. *by oneself*
    c. *help yourself to~*
    d. *enjoy yourself*
d. as an emphatic possessive
    a. *one's own~* (*Use your own pen*)
    b. *A of one's own* (*This is a word of his own invention*)

(Koike, 2013, p. 118-119)

It is explained as an exception to typical SVO syntax; "When the subject and object are the same person or thing and the action is directed toward that person or thing, then the reflexive pronoun (-self/ -selves) is used" (translated) (Ou, Suzuki, & Kawasaki, 2017, p. 48). The example provided herein is, *I stood up and introduced myself* (ibid.) The second explanation from this source is, "When the target of action is not a different person or thing, but the same person or thing as the subject", with the single example, *The princess looked at herself in the mirror* (ibid., p. 532). The third explanation here is the emphatic, as in *You should do it yourself*, explained, "In the case of reflexive pronouns, the subject or pronoun can be emphasized" (translated. ibid, p. 532). A few 'idiomatic' expressions are also given:

a. *Help yourself to the salad.*

b. *I couldn't make myself understood in English.*
c. *I had to shout to make myself heard in the noisy class.*
d. *Please make yourself at home.*
e. *Take care of yourself.*
f. *Did you tie your shoelaces by yourself?*
g. *You should decide your course for yourself.*
h. *behave oneself*
i. *enjoy oneself*
j. *introduce oneself*　(Ou et al., 2017)

Shinoda & Yoneyama (2018) provide similar explanations and the following examples:

a. *enjoy oneself*
b. *hide oneself*
c. *dress oneself*
d. *seat oneself*
e. *help yourself to ~*
f. *Make yourself at home.*
g. *(all) by oneself*
h. *be beside oneself (with ~)*
i. *in itself*
j. preposition + oneself: (*(all) by oneself, for oneself, (all) to oneself, between ourselves, be beside oneself, in itself/ in themselves, in spite of oneself*) (p. 325)

Nakao (2014), in accord, gives the following examples:

a. preposition + oneself: *by onself, for oneself, beside oneself, in itself (themselves), between ourselves*
b. *help oneself to~*
c. *make oneself at home*
d. *by oneself (=alone)*
e. *I love myself*
f. *enjoy oneself*
g. *I myself said so = I said so myself.*　(p. 110-111)

The explanations and examples above provide no semantic consistency nor systematicity. Within the 'umbrella of reflexivity', examples of semantic reflexive, benefactive (as well as other thematic roles), emphatic and middle voice all occur, many with only the term 'idiomatic' given as an explanation. It is therefore completely understandable that students are often confused by the use of reflexivity in English (anecdotal). According to the results of a corpus investigation, many of the 'reflexive' examples in the high school English guidebooks instantiate semantically nonreflexive and infrequently-used cases. Perhaps frequently-used constructions are not presented in high school English guidebooks because a large number of these instances are metaphorical with very few systematic descriptions readily available. The present research aims to begin to rectify this deficiency by presenting the results of a corpus investigation that suggests that 'native' reflexive data is systematic and categorizable when analyzed semantically and when collocational data is taken into consideration, and therefore, applicable as a tool and/or resource for ELT pedagogy.

2. METHOD

It has been noticed that corpus retrieval of metaphors is "almost impossible for the simple reason that conceptual mappings are not linked to particular linguistic forms" (Stefanowitsch & Gries, 2007, p. 2). This refers to corpus-based metaphor studies that attempt to retrieve and analyze data using metaphoric SOURCE conceptions of conceptual metaphors for their corpus queries. The Metaphor Identification Procedure for Reflexives presented below overcomes these disadvantages, first, by being limited to the reflexive construction, an easily-searchable syntactic parameter. Second, the method allows for queries of metaphoric use by objectively identifying TARGET domain samples. Only then are SOURCE domain mappings proposed, analyzed and categorized according to semantic content and collocational patterning. In this way, this procedure provides an objective way to uncover previously undetermined and underspecified metaphoric SOURCE data.

For these purposes, both the COCA (Davies, 2008) and the British National Corpus ("The British National Corpus (BNC XML Edition), 2007) were utilized. The web interface used for both was the corpus query interface at Brigham Young University. Even though the COCA is a very large corpus in itself, the BNC was included because it contains spontaneous speech and is therefore more representative of natural language. Further, its inclusion allows for a data set inclusive of regional variation. Although this could not eliminate all idiosyncrasies of collected works inherent in any corpus, the strength of each corpus adds

statistical value to the research by including two regions whose native language is English as well as different registers contained within the corpora themselves.

This research proposes a unique method of corpus research for metaphor retrieval and analysis that is useful in both quantitative and qualitative terms. The results uncover a wealth of information and allow for coarse- and fine-grained analyses. This research begins with the query, "What verbs are instantiated within the reflexive construction?" In this respect, it is data-driven methodologically similar to *collostructional analysis* in that the search "always starts with a particular construction and investigates which lexemes are strongly attracted or repelled by a particular slot in the construction…" (Stefanowitsch & Gries, 2003, p. 214). This is followed by the questions, "Which verbs that appear in the reflexive construction are used metaphorically? Is this evidenced in the data?" When metaphorically-construed verbs are objectively identified, they are grouped into semantic categories. It is only at this point that *possible* metaphorically expressed verbs are identified. These verbs are then analyzed, token by token, with an effort to keep theoretical and semantic assumptions to a minimum.

The specific procedure adopted (and adapted) for this research for determining and analyzing possible metaphors mined from corpora is based on the Pragglejaz Group's *Metaphor Identification Procedure (MIP*). "…the purpose of MIP is to provide a procedure that starts from the actual discourse, and inductively builds the case for why a particular word was used metaphorically in context" (2007, p. 34). However, this procedure is not adopted in its entirety. Only the relevant steps are implemented here, providing an objective decision-making tool for distinguishing between metaphoric and literal instantiations within the reflexive construction. In other words, I have omitted some steps of the MIP (especially its reliance on propositions (Crisp, 2002; Heywood et al., 2002; Steen, 2002)) and have only incorporated the essential procedures that make data retrieval and analysis of reflexive construction-based metaphors more systematic and reliable. For this research, the most recent version of the MIP, called the *Metaphor Identification Procedure Vrije Universiteit* (aka MIPVU) was used, it being more encompassing and intuitive (Steen et al., 2010).

The following six steps were adhered to in order to delineate metaphorical use within reflexive events (once potential metaphorically-construed verbs have been identified). The method is coined here the *Metaphor Identification Procedure for Reflexives (*hereafter MIPR):

1. Reflexive construction parameters are input into the corpus query field, being 'abstract' (i.e., [v*]) or 'concrete' (i.e., [find]) as necessary and sufficient.

2. Check retrieved data for 1) antecedent-pronoun accuracy ($noun_1$ + verb + refl. pro_1), 2) genuine reflexive meaning (i.e., compared to emphatic, benefactive or logographic, etc.), and 3) missing antecedents or pronouns (ellipses).
3. Check data for metaphor-related words (MRWs) by examining the text on a word by word basis by referencing 'base' meanings in a corpus-based dictionary (as per the MIPVU, (Steen, 2010)).
4. If a word's use is considered metaphorical, analyze TARGET → SOURCE mappings.
5. Find collocational and/or contextual evidence corroborating the mapping in #4.
6. Analyze data statistically and confirm results.

Within these six steps, only number three is the same as the MIPVU. Original MIPVU procedures one and two were combined and cross-domain plausibility checks were limited to the reflexive verb, anaphoric NPs, and the reflexive's immediate adjuncts. Procedure number two was added to distinguish reflexive from non-reflexive anaphors as well as to confirm antecedent-pronoun agreement. The decision to perform this step at this time was more practical than theoretical. It was simply more efficient to weed out non-reflexive, non-anaphoric tokens before proceeding with the more time-consuming metaphoric identification and analysis. For procedure number four, various metaphorical interpretations were possible for many tokens in the data, and it was critical to remain open to all possible interpretations. This was sometimes difficult, and the expanded context of tokens were consulted frequently. Procedure number five was added in order to confirm collocational and broader contextual evidence when encountering ambiguous metaphorical interpretations. Although this was time consuming, it allowed for more objective judgements to be made for metaphorical interpretation. Finally, in step six, the data is calculated and the results are recorded.

3.1. RESULTS

For the first step of the investigation, the COCA and BNC were mined with the search parameters [v*] [ppx*], (i.e., any verb lemma followed by any reflexive pronoun). The 500 most frequent verb lemmas with a minimum hit value of 10 (i.e., ≥10) were yielded. The 500^{th}-ranked verb had a frequency of 16 in the COCA and 18 in the BNC. Next, reciprocal data were deleted, leaving 462 total hits in the COCA and 468 in the BNC. A cross-corpora comparison was then conducted and a list of common verbs was recorded. Each of the verb

lemmas from this data set were then queried in both corpora in order to retrieve tokens of that particular verb within the reflexive construction, e.g., [find][ppx*], as in:

1. ***Walter found himself*** *grasping for breath.* (COCA:2017.FIC.Bk.Mercy)

If more than 100 hits were retrieved for any verb-pronoun pair, a random sample (n=100) was chosen for the analysis. The results were then checked for metaphoric instantiation in a corpus-based dictionary, as per the MIPR method discussed in Section 2.

Out of a cross-corpora total of 121 verb lemmas, 66 metaphorically-construed verb lemmas concurred with the criteria described above. 44 verbs were judged non-metaphorical. There were 11 verbs whose 'base' meanings were difficult to assign due to the high number and variety of metaphorical and non-metaphorical senses, as well as a few of them functioning mainly as auxiliary verbs. These verbs were: *do, get, give, have, hold, keep, make, put, set, take*, and *turn*. These were eliminated from the analysis. The data were classified into four metaphorical and six non-metaphorical semantic categories, shown below. These categories were not predetermined but were organic to the data, although they occasionally overlap with established verb categorizations (see: Levin, 1993). Shown below are the four metaphorically-construed categories (verbs listed alphabetically, category marker in parenthesis).

1. Self-Perception (P): *be, catch, check, feel, find, identify, immerse, lose, see, watch*
2. Self-Causation (F): *assert, bring, catch, check, drag, draw, drive, fling, force, hang, haul, help, kill, launch, lock, pull, push, resign, set, shake, steel, throw, work (***** (F) stands for 'Force-dynamic' conception.)
3. Societal Interaction (SI): *align, attach, behave, call, commit, distance, distinguish, establish, excuse, expose, express, identify, involve, lend, lower, present, prove, raise, sell, show, suit*
4. Self-Maintenance (M): *brace, compose, feed, help, resolve, save, settle, shoot, support, treat, watch, wrap*

The verbs *catch* and *check* construed more than one category. They were included in each category due to the unique conceptions being construed.

The six literally-construed (i.e., non-metaphorical) categories are shown below (verbs listed alphabetically, category marker in parenthesis).

1. Sense Perception & Physicality (P'): *hear, know, manifest, perceive, regard, seat, sit*
2. Self-Causation (F'): *allow, busy, calm, let, steady, stop, will*
3. Social Interaction (SI'): *avail, concern, extricate, identify, introduce, reveal*
4. Self-Communication (C): *ask, blame, remind, repeat, teach, tell*
5. Self-Judgement (J): *believe, fancy, feel, hate, imagine, pride, think, trust*
6. Self-Maintenance (M'): *calm, ease, enjoy, prepare, rid, organize, protect, steady, surround, transform*

One noticeable trend in the data are the four equivalent semantic categories between the metaphoric and literal, e.g., Perception, Causation, Interaction and Maintenance. This might at first seem to be a possible point of ambiguity or confusion if it not for the mutual exclusivity of the verbs that comprise them, *identify* being the only redundancy. In other words, the number of semantic categories being limited and the verbs in the metaphorical vs. literal categories are different; therefore, systematic categorization of reflexive verbs based on corpus data is not only possible, but suggests a fairly straightforward systematicity that would be easily adaptable to ELT curricula (see Part 4).

3.2. REFLEXIVE + PREPOSITION

Another collocation (besides the above [verb][refl. pro]) that appears in the high school English guidebooks is the following: [verb][noun][prep][refl. pro] (i.e., [v*]_nn*_i*[ppx*] in corpus query syntax). For example,

2. *When do you **find time for yourself** except when other people are sleeping?*
(COCA:1998.NEWS.CSMonitor)

The most frequent (n=100) lemmas of these types of collocations are shown in Table 1 (COCA: n=3130, ($n \geq 5$, reciprocals deleted)).

Table 1. The most frequent 100 collocations for ([v*][nn*] _i [ppx*]) (COCA)

take care of ppx[1] (*n*= 2,224, ratio = 67.7%)	*lost control of ppx* (n=28, ratio =0.09%)
draw attention to ppx (n= 157, ratio = 5%)	*lost track of ppx* (n=19, ratio = 0.06%)
caught sight of ppx (n=31, ratio = 0.1%)	*find time for ppx* (n=15, ratio =0.05%)

The data in Table 1 suggests a similar kind of systematicity as the semantic categorization data for metaphorical vs. literal verbs in 3.1., i.e., that a limited number of variations manifest once frequency and schematicity are taken into consideration. More than 67% of the most frequent 100 sample are from one collocation, namely, [take][care][of][refl.pro.]. Again, this systematicity can only be gleaned once levels of schematicity are delineated and large amounts of data are analyzed, lending support for these types of corpus-based descriptions.

In more general terms, for much corpus data, lexemes can be grouped together to form larger, more schematic categories. Meaning is not only retrievable at the lexeme level, but is evident at the abstract schema level as well as the specific phraseme level. A very abstract collocation will allow numerous lexemes into its slots, the structure will be more flexible and its meaning will also be abstract. At the other end of the spectrum, a very concrete collocation (e.g., idiom) will have a very limited number of lexemes possible, have a more rigid structure and its meaning will be specific. Collocation schemata are thus a scalable phenomenon, from very abstract (e.g., [v*][nn*]_i*[ppx*]) to more specific (e.g., [p*]found[ppx*]in[n*]), to idiomatic (e.g., [n*]caught sight of[n*]). This schematicity refers to the level of abstractness of an analysis, i.e., the level of analytical granularity of one's research viewpoint (Sinclair, 1991), and depending on this viewpoint, the type of analysis will change because what the researcher is looking for changes. Looking through a simple magnifying glass will reveal different objects than that of the Hubble telescope, even if they are both looking in the same direction.

4. CONSIDERATIONS FOR ELT

A summary of the reflexive examples taken from the high school English guidebooks investigated here is provided in Table 2. Comparing the left column of this table with the data set in Table 3, which shows the most frequent verb lemmas (*n*=100) that are immediately followed by a reflexive pronoun (i.e., [v*][ppx*]), the dissonance between what is actually used by native speakers and what is presented in high school English textbooks becomes

[1] This is the only item that appears in the high school English guidebooks and corroborates corpus evidence.

apparent. Similarly, the upper right portion of Table 2 can be compared with Table 1 above. Whether by chance or design, the collocation [take][care][of][refl. pro.] from Table 1 is also found in Table 2, although the pronoun is limited to the second person singular there. Due to space constraints, however, all collocation patterns that appear in the English guidebooks could not be investigated here and further study is necessary to flesh out all the collocation patterns (as well as inconsistencies) inherent in this type of corpus analysis.

Table 2. Reflexive examples from four high school English curriculum guidebooks.

Verb + Reflexive Pronoun	Preposition + Reflexive Pronoun
introduce myself	*looked at yourself*
help yourself (to ~)	*take care of yourself*
make myself understood/heard	*by yourself*
make yourself at home	*for yourself*
behave oneself	*(all) by oneself*
enjoy oneself	*in itself*
hide oneself	
dress oneself	Verb + Object + Reflexive Pronoun
seat oneself	*do it yourself*
be beside oneself (with ~)	
love myself	

Table 3. Most frequent 100 lemmas for [verb + reflexive pronoun] ([v*][ppx*]) (COCA).

RANK	LEMMA	FREQ	PER MILLION
1	[FIND] [HIMSELF]	7323	12.96
2	[FIND] [THEMSELVES]	6053	10.71
3	[FIND] [MYSELF]	5686	10.06
4	[FIND] [HERSELF]	3982	7.05
5	[KNOW] [EACH]	3190	5.65
6	[TELL] [MYSELF]	2482	4.39
7	[FIND] [YOURSELF]	2346	4.15
8	[SEE] [EACH]	2261	4
9	[BE] [ITSELF]	1990	3.52
10	[CALL] [THEMSELVES]	1910	3.38
11	[ASK] [YOURSELF]	1903	3.37
12	[MAKE] [HIMSELF]	1872	3.31
13	[SEE] [THEMSELVES]	1867	3.3
14	[TELL] [HIMSELF]	1846	3.27
15	[KILL] [HIMSELF]	1797	3.18

RANK	LEMMA	FREQ	PER MILLION
16	[FIND] [OURSELVES]	1759	3.11
17	[FIND] [ITSELF]	1758	3.11
18	[CALL] [HIMSELF]	1751	3.1
19	[TELL] [HERSELF]	1739	3.08
20	[SEE] [MYSELF]	1521	2.69
21	[BE] [THEMSELVES]	1477	2.61
22	[SEE] [HIMSELF]	1436	2.54
23	[GET] [YOURSELF]	1402	2.48
24	[PROTECT] [THEMSELVES]	1385	2.45
25	[IDENTIFY] [THEMSELVES]	1383	2.45
26	[ASK] [MYSELF]	1380	2.44
27	[GET] [HIMSELF]	1374	2.43
28	[CONSIDER] [THEMSELVES]	1237	2.19
29	[LOVE] [EACH]	1218	2.16
30	[MANIFEST] [ITSELF]	1215	2.15
31	[GIVE] [YOURSELF]	1208	2.14
32	[MAKE] [MYSELF]	1145	2.03
33	[PULL] [HIMSELF]	1145	2.03
34	[BE] [HIMSELF]	1105	1.96
35	[MAKE] [YOURSELF]	1076	1.9
36	[DEFEND] [THEMSELVES]	1074	1.9
37	[MAKE] [HERSELF]	1065	1.88
38	[ALLOW] [HIMSELF]	1050	1.86
39	[MAKE] [THEMSELVES]	1043	1.85
40	[DEFEND] [HIMSELF]	1006	1.78
41	[INTRODUCE] [HIMSELF]	1001	1.77
42	[THROW] [HIMSELF]	998	1.77
43	[PUT] [YOURSELF]	992	1.76
44	[DESCRIBE] [HIMSELF]	966	1.71
45	[LEND] [ITSELF]	956	1.69
46	[PRESENT] [ITSELF]	944	1.67
47	[CONSIDER] [HIMSELF]	942	1.67
48	[PUT] [HIMSELF]	941	1.67
49	[CONSIDER] [MYSELF]	935	1.65
50	[FACE] [EACH]	919	1.63
51	[PROTECT] [YOURSELF]	914	1.62
52	[FORCE] [HIMSELF]	885	1.57
53	[GET] [MYSELF]	877	1.55
54	[GIVE] [HIMSELF]	854	1.51
55	[HELP] [EACH]	853	1.51
56	[FEEL] [HIMSELF]	831	1.47
57	[SEE] [HERSELF]	820	1.45

RANK	LEMMA	FREQ	PER MILLION
58	[LET] [MYSELF]	806	1.43
59	[EXPRESS] [THEMSELVES]	797	1.41
60	[LET] [HIMSELF]	785	1.39
61	[REMIND] [MYSELF]	780	1.38
62	[SEE] [YOURSELF]	771	1.36
63	[HELP] [THEMSELVES]	754	1.33
64	[PUT] [MYSELF]	754	1.33
65	[PRESENT] [THEMSELVES]	750	1.33
66	[ALLOW] [HERSELF]	747	1.32
67	[ASK] [THEMSELVES]	743	1.32
68	[PUT] [THEMSELVES]	736	1.3
69	[FORCE] [HERSELF]	727	1.29
70	[KILL] [EACH]	708	1.25
71	[HELP] [HIMSELF]	707	1.25
72	[GIVE] [MYSELF]	705	1.25
73	[ASK] [OURSELVES]	702	1.24
74	[PUSH] [HIMSELF]	699	1.24
75	[KILL] [THEMSELVES]	698	1.24
76	[CALL] [ITSELF]	697	1.23
77	[SHOOT] [HIMSELF]	692	1.22
78	[KILL] [HERSELF]	690	1.22
79	[PROVE] [HIMSELF]	685	1.21
80	[LET] [HERSELF]	678	1.2
81	[THROW] [HERSELF]	678	1.2
82	[FEEL] [MYSELF]	673	1.19
83	[IDENTIFY] [HIMSELF]	672	1.19
84	[REPEAT] [ITSELF]	658	1.16
85	[PULL] [HERSELF]	657	1.16
86	[REMIND] [HERSELF]	654	1.16
87	[TURN] [HIMSELF]	650	1.15
88	[PRESENT] [HIMSELF]	640	1.13
89	[FEEL] [HERSELF]	636	1.13
90	[DEFEND] [ITSELF]	631	1.12
91	[REMIND] [HIMSELF]	630	1.12
92	[ASK] [HIMSELF]	628	1.11
93	[BRING] [HIMSELF]	625	1.11
94	[CALL] [HERSELF]	611	1.08
95	[ALLOW] [MYSELF]	599	1.06
96	[KILL] [MYSELF]	590	1.04
97	[FORCE] [MYSELF]	582	1.03
98	[REVEAL] [ITSELF]	572	1.01
99	[GET] [THEMSELVES]	569	1.01

RANK	LEMMA	FREQ	PER MILLION
100	[GET] [HERSELF]	566	1

Only three constructions co-occur in Tables 2 and 3: ***make*** (*himself* (rank 12), *myself* (rank 32), *yourself* (rank 35), *herself* (rank 37), *themselves* (rank 39)) (see Table 2 above), ***help*** (*themselves* (rank 63) and *himself* (rank 71), no token for *yourself*), and ***love*** (reciprocal use only, rank 29). However, more important than pure frequency count or rank is occurrence 'per million'. This is a better indicator because the frequency of use can be put into overall context within the whole corpus and language in general. So, for example, the number one ranked lemma [FIND][HIMSELF] has a per million frequency of 12.96, meaning that for every one million words, *find* + *himself* occur together almost 13 times. Whether or not this is frequent depends on the parameters, depth and breadth of the investigation, but it is a useful metric for comparing different words, expressions or phrases, and can be useful when comparing lexemes or collocations across different corpora with different sample sizes.

To get a better sense of the frequency differences for the high school English curriculum guideline textbooks versus actual usage, we can look at each frequency per million of the high school English guidebook examples, shown in Table 4. The example with the highest ratio in the high school English guidebook is *enjoy yourself*, at 0.53 per million. This is about half that of the least frequent corpus item (rank 100) instantiating at one per million, a noticeable disparity.

Table 4. Frequency/per million ratio of reflexively-used verbs that occur in four high school English guidebooks.

Verb	Pronoun	Freq.	Per Mil	Verb	Pronoun	Freq.	Per Mil
INTRODUCE	myself	265	0.47	BEHAVE	yourself	111	0.20
	yourself	187	0.33		themselves	47	0.08
	himself	165	0.29		himself	23	0.04
	themselves	155	0.27		myself	20	0.04
	ourselves	98	0.17		herself	12	0.02
	herself	93	0.16		itself	7	0.01
	itself	7	0.01		ourselves	4	0.01
ENJOY	yourself	301	0.53	HIDE	himself	35	0.06
	themselves	164	0.29		yourself	35	0.06
	myself	119	0.21		themselves	30	0.05
	himself	99	0.18		herself	20	0.04
	ourselves	64	0.11		myself	20	0.04
	herself	51	0.09		itself	16	0.03

Verb	Pronoun	Freq.	Per Mil	Verb	Pronoun	Freq.	Per Mil
	itself	1	0.00		ourselves	3	0.01
DRESS	himself	47	0.08	SEAT	himself	22	0.04
	herself	40	0.07		itself	19	0.03
	myself	34	0.06		yourself	15	0.03
	yourself	31	0.05		herself	14	0.02
	themselves	31	0.05		themselves	12	0.02
	itself	14	0.02		myself	4	0.01
	ourselves	7	0.01		ourselves	2	0.00
BE BESIDE	myself	5	0.01	LOVE	yourself	143	0.25
	themselves	4	0.01		myself	100	0.18
	herself	3	0.01		itself	55	0.10
	yourself	2	0.00		ourselves	42	0.07
	himself	2	0.00		themselves	38	0.07
					herself	19	0.03
					himself	15	0.03
					thyself	1	0.00

The results of this reveal a number of important issues for ELT. The first is a matter of language ‘naturalness’ and frequency of occurrence (Leech, 2001). This investigation has uncovered inconsistencies between “what is taught” and “what is used” with regards to high school English language guides in Japan compared with native sources in America and Britain. This is not to suggest that all native sources are ideal for the EFL classroom. However, the implementation of frequency-based words lists into reading and four-skills EFL textbooks in recent years shows awareness of the issue addressed here. Specifically, this investigation has uncovered issues of ‘linguistic relevance’ for EFL pedagogy concerning the reflexive construction. By focusing on the reflexive object pronoun, such as the high school English guidebooks do, reflexive as well as non-reflexive examples manifest. However, a focus on collocations within the ‘true reflexive’ event (e.g., [nn*][v*]_i[ppx*]) would help systematize descriptions found in the English curriculum guidebooks.

Collocations can be thought of as ‘collections of expressions’ that share a schematically higher-level structure and have meaning. That type of ‘collection’ is coined here ‘meaning structure’. An example of the meaning structure for [nn*][v*][ppx*] is shown in Table 5. Under the [nn*] (i.e., any noun lemma) column, the subject/agent nouns that appear are organically categorized; in other words, the investigator finds semantic commonalities in the first noun corpus token list and chooses a category title that is representative of the data. In Table 5, the

category ‘People’ represents the nouns *people, bomber, American, man, mother*, *participant,* etc. For the category Groups, nouns such as *government, students, administration*, etc. are instantiated. This categorization is carried out for all the items in the data set as well as each category within the meaning structure (e.g., nouns, verbs, prepositions, etc.) until the chosen frequency criteria has been met, for example, the most frequent 100 from a random sample. (Ideally, this is a statistically significant portion of the total number of occurrences.) Individual tokens, complete sentences and expanded texts can all be used as possible data sources.

Table 5. Meaning structure for [nn*][v*][ppx*] (*n*=100; COCA&BNC).

[nn*]	**[v*]**	**[ppx*]**
People	present	reflexive pronouns
Groups	call	
Companies	repeat	
Proper Names	know	
Pronouns	identify	

Some example sentences for the meaning structure in Table 5 are:

3. *I liked the way **people presented themselves**. They were very honest.*
(COCA:2005.NEWS.Chicago)

4. *This teacher was able to explain how the area got its name and why the **students call themselves** “The Bottom Kids”.* (COCA:2013.ACAD.ResearchinMiddle)

5. ***Apple presents itself** as unique.* (COCA:2003.MAG.PsychToday)

Table 6. Meaning structure for [nn*][v*]_i [ppx*] (COCA&BNC).

[nn*]	**[v*]**	**_i**	**[ppx*]**
Data (records, facts, numbers, evidence)	speak	for	reflexive pronouns
People	talk	at	reciprocal pronouns
	look	to	
	divided	against	

6. *Concerning investigation into the conditions endured by animals in laboratories, the **facts speak for themselves**.* (BNC:B03.W_newsp_other_report)

7. *(**You**) Do not spend the whole time **talking about yourself**, your job, and your life.*
(BNC:EVF.W_commerce)

As another example, the meaning structure [nn*][v*] _i [ppx*], shown in Table 6, may be taught alone or contrastively with Table 5. Notice that the two have very different meaning structures even though the only difference is the addition of a preposition. Every item in the meaning structure affects the totality of the collocational. In other words, each item is lexically related to the others and these relations affect the meaning of the whole expression. When organized and described in this way using frequency-related data, the information becomes more manageable and is able to incorporate expressions that students are likely to experience when exposed to native English sources.

This is all part of a larger discussion about whether corpus data can and should be utilized for curriculum creation and design in Japanese ELT environments. For this kind of evidence-based data to be accepted and established, it must be presented in a way that is easily accessible to both teacher and student, what Braun calls “pedagogically-enriched” (2005, p. 55). For example, high school English textbooks and guidebooks already contain a multitude of lexically-based charts, and so students and teachers are familiar and comfortable working with them. Therefore, with refinement and the addition of various types of class activities (and teachers’ manuals) (ibid.), the meaning structures in Tables 5 and 6 could be adapted and used to supplement or even replace the current syntax-based models.

A more direct pedagogical approach is one in which students work with corpus data, called Data Driven Learning (DDL) (Friginal, 2018; Johns & King, 1991; Timmis, 2015). Under this method, students analyze concordance lines (or data presented in other forms) of the collocations under investigation in order to find patterns and meanings for themselves. Concordance lines can consist of raw corpus data, data that is simplified by the teacher (i.e., teaching-oriented corpora (Braun, 2005; Leech, 1997)) or data from a learner corpus (Granger, Hung, & Petch-Tyson, 2002) in an effort to teach students the skills of linguistic pattern detection and analysis. As for the classroom activities implementing such data analysis, “…DDL does not require radically different classroom activities: common pedagogic exercise types such as gap-fills, matching activities and multiple choice tasks can be generated from concordance output for DDL purposes” (Timmis, 2015, p. 135).

5. Conclusion

A Google Scholar search for “Data Driven Learning DDL Japan” yielded 1,870 hits, suggesting that DDL is starting to be ‘put to the test’. Whether the corpus investigation of language data by students and teachers gains wide acceptance in Japanese English education depends on a number of factors, including but not limited to: teacher training of corpus tools, increased importance of critical thinking skills in English education, creation of accurate but learner- and teacher-friendly corpus data sets and the organization and presentation of corpus linguistic data in a way that fits the specific pedagogical and cultural paradigms of Japanese English Education. It is indeed encouraging that many corpora now incorporate user-friendly interfaces and are available online for free. The only thing missing, perhaps, is the awareness that these tools exist and that no formal software programming is necessary to use them.

This investigation has revealed a number of important findings; 1) descriptions of the reflexive construction in high school English guidebooks do not correlate with native English corpus data, 2) metaphorical and literal instances of the reflexive can be categorized semantically, 3) corpus analyses of reflexive collocations uncover reliable patterns and 4) it is possible to implement corpus-based collocational data into ELT curricula. It is my sincere hope that this investigation has contributed to the raising of awareness of the emerging field of corpus-based curriculum design and DDL for ELT in Japan.

REFERENCES

Barlow, M. (1996). Corpora for theory and practice. *International Journal of Corpus Linguistics, 1*(1), 1-37.

Barlow, M., & Kemmer, S. (1994). A Schema-Based Approach to Grammatical Description. *The Reality of Linguistic Rules, 26*, 19.

Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents. *ReCALL*, *17*(01).

The British National Corpus (BNC XML Edition). (2007). Distributed by Bodleian Libraries, University of Oxford on behalf of the BNC Consortium
URL: http://www.natcorp.ox.ac.uk/

Collins Dictionary Online. Glasgow, Scotland: HarperCollins Publishers Limited.

Crisp, P. (2002). Metaphorical propositions: a rationale. *Language and Literature, 11*(1), 7-16.

Davies, M. (2008). The Corpus of Contemporary American English. From Brigham Young University. http://corpus.byu.edu/coca/

Deignan, A. (2007). The grammar of linguistic metaphors. In A. Stefanowitsch & S. Gries (Eds.), *Corpus-based Approaches to Metaphor and Metonymy* (pp. 106-122). Berlin: Walter de Gruyter.

Ebeling, J., & Ebeling, S. O. (2013). *Patterns in Contrast* (Vol. 58): John Benjamins Publishing.

Faltz, L. M. (1985). *Reflexivization : A Study in Universal Syntax*. New York: Garland Publishing.

Friginal, E. (2018). *Corpus Linguistics for English Teachers: Tools, Online Resources, and Classroom Activities*. Routledge.

Gibbs, R. W. (2002). Psycholinguistic comments on metaphor identification. *Language and Literature, 11*(1), 78-84.

Gibbs, R. W. (2007). Why cognitive linguists should care more about empirical methods. In M. Gonzalez-Marquez (Ed.), *Methods in Cognitive Linguistics* (pp. 2-18). Amsterdam: John Benjamins Publishing.

Gibbs, R. W. (2016). Seven Empirical Challenges for Cognitive Linguistics. *Journal of Cognitive Linguistics, 2*(March 2017), 13.

Gilquin, G. (2010). *Corpus, Cognition and Causative Constructions*. Amsterdam ; Philadelphia: John Benjamins Publishing.

Goatly, A. (2002). Text-linguistic comments on metaphor identification. *Language and Literature, 11*(1), 70-74.

Granger, S., Hung, J., & Petch-Tyson, S. (2002). *Computer Learner Corpora, Second Language Acquisition, and Foreign Language Teaching*. John Benjamins Publishing.

Grossman, B. (2017). Metaphorical Expressions of Self Awareness in English, *ELR Journal*, 2. 82-105. *University of Birmingham: Centre for English Language Studies*.

Group, P. (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol, 22*(1), 1-39.

Heywood, J., Semino, E., & Short, M. (2002). Linguistic Metaphor Identification in Two Extracts from Novels. *Language and Literature, 11*(1), 35-54.

Hunston, S., & Francis, G. (2000). *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*: John Benjamins Publishing.

Johns, T., & King, P. (1991). Classroom Concordancing: English Language Research Journal, 4. *University of Birmingham: Centre for English Language Studies*.

Kemmer, S. (1993). The Middle Voice. Amsterdam; Philadelphia: John Benjamins Publishing.

Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and Thought* (pp. 202-251). Cambridge: Cambridge University Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic books.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar: Theoretical Prerequisites* (Vol. 1). Stanford, California: Stanford university press.

LDOCE. (2014). Longman Dictionary of Contemporary English Online. Available from Pearson Education Limited, http://www.ldoceonline.com/

Leech, G. (1997). Teaching and language corpora: A convergence. In *Teaching and language corpora* (Vol. 123).

Leech, G. (2001). The role of frequency in ELT: New corpus evidence brings a re-appraisal. *Foreign Language Teaching and Research*, 2001.

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

MacArthur, F. (2015). On using a dictionary to identify the basic senses of words. *Metaphor and the Social World*, *5*(1), 124–136.

Nakao, T. (2014). *Scramble Basic* (2nd Edition). Tokyo: Obunsha Co. Ltd.

Ou, T., Suzuki, N., & Kawasaki, Y. (2017). *Evergreen Sougo Eigo*. IIzuna Shoten RT.

Radden, G., & Kövecses, Z. (1999). Towards a theory of metonymy. In A. Ortony (Ed.), *Metonymy in Language and Thought* (pp. 17-60). Cambridge: Cambridge University Press.

Shinoda, S., & Yoneyama, T. (2018). *Vintage* (3rd edition). Tokyo: Iizuna Shoten.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*: Oxford University Press.

Steen, G. (2002). Towards a procedure for metaphor identification. *Language and Literature, 11*(1), 17-33.

Steen, G. (2010). A Method for Linguistic Metaphor Identification: From MIP to MIPVU. John Benjamins Publishing.

Steen, G., Dorst, A. G., Hermann, J., Kaal, A. A., Krennmayr, T., & Pasma, T. (2010). *A Method for Linguistic Metaphor Identification* (Vol. 14). Amsterdam: John Benjamins Publishing.

Stefanowitsch, A., & Gries, S. T. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics, 8*(2), 209-243.

Stefanowitsch, A., & Gries, S. T. (Eds.). (2007). *Corpus-based Approaches to Metaphor and Metonymy*. Berlin: Mouton de Gruyter.

Timmis, I. (2015). *Corpus Linguistics for ELT: Research and Practice* (1 edition). London ; New York: Routledge.