A corpus-driven approach to identifying discipline-specific metaphorical
vocabulary for Japanese university English-medium instruction programs

Bennett, Phillip

900,000

5　　　　EMI　　　　200

4

4

2

EMI　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　4

Most vocabulary instruction is focused on the teaching of basic or literal word definitions, and
little attention is paid to how highly-frequent words often take on extended meanings. My hope is
that my research findings will help to encourage a broader understanding of lexical development.

During the time I have been receiving kaken funding, I have built a
2-million-word corpus of materials from English as a Medium of Instruction courses in 5 Japanese
universities. The corpus contains both spoken (16%) and written (84%) data, and is 4 times larger
than the corpus used in my prior study in this area.
I have taken samples of data from this corpus for four academic disciplines (Applied Linguistics,
Literature, Philosophy, and Political Science - 50,000-word samples each) and manually coded them
for the presence of metaphorical vocabulary. I enlisted a second rater who is also familiar with the
metaphor coding procedure to establish the reliability of the codings.
I then used these codings to calculate estimates of the rate at which learners in EMI programs are
likely to encounter metaphorical vocabulary. I also identified prominent metaphorical themes in each
of the four disciplines and estimated the proportion of all metaphors that these comprise.

Vocabulary acquisition

Corpora　EMI　Metaphor

１．研究開始当初の背景

Within the field of Applied Linguistics, it is recognized that vocabulary knowledge plays a significant role in learners' ability to comprehend text. Thus far, most research into the lexical demands of academic text has focused on identifying the words that are most likely to be encountered in academic discourse and using these findings to produce word lists that can be used to guide instruction.

While there is clear value in this approach, there are a number of shortcomings that may limit its effectiveness. Firstly, word lists often treat academic discourse as a single genre, and do not differentiate between the lexical demands of individual fields. As a result, such lists often contain words that are of little use to certain groups of learners (e.g., hard science vocabulary being presented to students in the social sciences or arts). A second issue is that of monosemic bias. This is the tendency to treat words as having only a single dominant sense. However, corpus-based studies have shown that metaphorical senses of lexical items are particularly frequent in academic discourse, meaning that learners would benefit from having an awareness of the typical ways in which lexical meaning is extended figuratively. A related concern is the finding that words can take on different meanings in separate fields, which suggests that investigations on the discourse patterns of lexis at the individual discipline level would be beneficial. A final concern is the fact that word lists usually deal only with individual words and neglect the role of phraseology in determining meaning. Word meaning is often determined by collocation, so there is clear value in helping learners to see not only single words that frequently appear in a field but also the collocations or phrases that are used to express common concepts.

２．研究の目的

There were three main objectives to this study. The first was to expand a corpus of materials drawn from English as a Medium of Instruction (EMI) courses being taught at universities in Japan. The second goal was to analyze samples of this EMI corpus for the presence of metaphorical vocabulary in several disciplines. The final goal was to identify common metaphorical themes in each discipline which might be used to guide vocabulary instruction for learners in these fields.

３．研究の方法

(1) The EMI corpus was based on that of a prior study. This original corpus contained 500,000 text tokens from 3 academic disciplines taught at 2 Japanese universities. The corpus consisted entirely of written texts. To expand the corpus, a call for participants who were willing to share EMI course materials was placed on the Facebook page of the JALT Vocabulary Special Interest Group. Several responses were received to this posting, and I was able to obtain materials from instructors at 5 universities and increase the number of disciplines represented in the corpus to 15. As the study coincided with the breakout of the Covid-19 pandemic, many instructors had video or audio recordings of their lectures available, which meant I was able to collect a larger amount of spoken data than had been anticipated. At present, the corpus has been expanded to 2,000,000 tokens (84% written, 16% spoken). Due to the need to complete other aspects of this project and the time-consuming nature of transcribing audio recordings, I still have several sets of materials that have not been added to the corpus. In the coming years, these will be transcribed and added to the EMI corpus.

(2) In order to analyze EMI course materials for the presence of metaphorical vocabulary, I first selected four disciplines which were relatively well-represented in the corpus. The disciplines each had over 150,000 tokens of text including both written and spoken materials and were drawn from more than one institution. The four disciplines selected were applied linguistics, philosophy, political science, and literature. I also

assembled a reference corpus of materials used to teach the same four disciplines in UK and US universities. This reference corpus was compiled from the British Academic Written English Corpus (BAWE), the British Academic Spoken English Corpus (BASE), the Michigan Corpus of Upper-Level Student Papers (MICUSP), and the Michigan Corpus of Academic Spoken English (MICASE). This reference corpus contained 2,270,000 tokens of text and was used to obtain a wider sample of specific metaphors in individual disciplines. From the EMI corpus, 50,000-word samples were taken from each of the 4 disciplines. These were analyzed for the presence of metaphor using the Metaphor Identification Procedure Vrije Universiteit (MIPVU). In brief, this procedure involves breaking a text into individual words or phrases, and then using the Macmillan Dictionary to identify the contextualized sense of each word or phrase and determine whether a more basic sense exists. If the contextualized sense was found to be one that is more abstract yet able to be understood through comparison with the basic sense, as well as being one that involves a crossing of domains (i.e., from the domain of physical movement to that of learning, as in the case of the literal and metaphorical meanings of the word 'approach'), then the word was coded as a metaphor. In total, 200,000 words of words were coded for the presence of metaphor. An experienced researcher who was familiar with the MIPVU process was then engaged to act as a second rater in order to provide reliability estimates. The rater coded 20% of the EMI corpus sample and inter-rater reliability was found to be substantial. This analysis allowed for estimates to be obtained for the proportion of all tokens in a discipline that were metaphors.

(3) Metaphorical themes were identified using a keyword analysis. For each of the 4 disciplines, 20 keywords (10 nouns, 10 verbs) were identified. Keywords are words that are significantly more likely to appear in a particular discipline than in other disciplines. Concordance lines were then inspected for each of the keywords in both the EMI and reference corpora. Metaphorical collocates of the keywords were identified using the MIPVU procedure. These metaphors were then grouped into metaphorical themes if the collocates belonged to similar semantic domains, as determined by dictionary definitions and semantic classification systems such as the USAS tags developed at the University of Lancaster in the United Kingdom.

４．研究成果

Using the MIPVU procedure to identify metaphors in discipline-specific texts, I was able to obtain estimates for the proportion of lexical items in each field that were metaphorical (table 1). These values are comparable to the proportion of metaphorical tokens in academic English written by and for an academic audience (18.50%), which suggests that metaphor still plays an important role in undergraduate-level academic discourse and is therefore valuable as a focus of instruction for learners.

Table 1: Estimated proportion of metaphorical tokens in four EMI disciplines

Discipline	Estimated % of tokens that are metaphors
Applied Linguistics	16.73%
Literature	13.25%
Philosophy	13.33%
Political Science	16.67%

The use of keyword analyses allowed for identified metaphors to be classified into common metaphorical themes. The frequencies of the identified themes could then be used to estimate the proportion of all metaphors in a given disciple that belonged to one of the themes (by comparison with the values in table 1). In total, between 10 and 12 metaphorical themes were identified in the 4 disciplines that accounted for a notable proportion of all metaphor use.

Table 2 provides the estimated percentage of all metaphorical words in each discipline that belonged to one of the identified metaphorical themes. It can be seen that in philosophy and political science, these metaphor themes account for the majority of all metaphors in the discipline. In applied linguistics and literature, meanwhile, the identified metaphor themes account for around 35% of all metaphor use. These differing

values are probably explained by the broader range of topics covered by the latter two disciplines. Clearly, all disciplines have a certain of breadth of focus, but philosophy and political science tend to concentrate more on a narrow range of concepts, while applied linguistics and literature are less limited in scope. Figures 1-4 provide the specific metaphorical themes identified for each discipline and the proportion of all metaphor use each represents.

Table 2: Estimated frequency of metaphorical themes in four EMI disciplines

Discipline	Estimated % of all metaphors that belong to identified metaphorical themes	One metaphorical theme encountered every … words
Applied Linguistics	35.13%	28
Literature	35.63%	35
Philosophy	70.34%	18
Political Science	65.21%	16

Figure 1: Prevalence of metaphorical themes in EMI applied linguistics materials
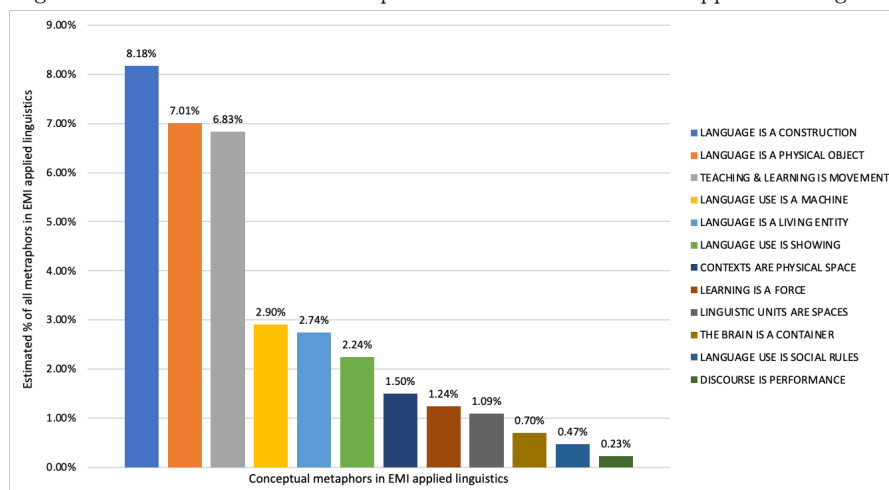
Figure 2: Prevalence of metaphorical themes in EMI literature materials
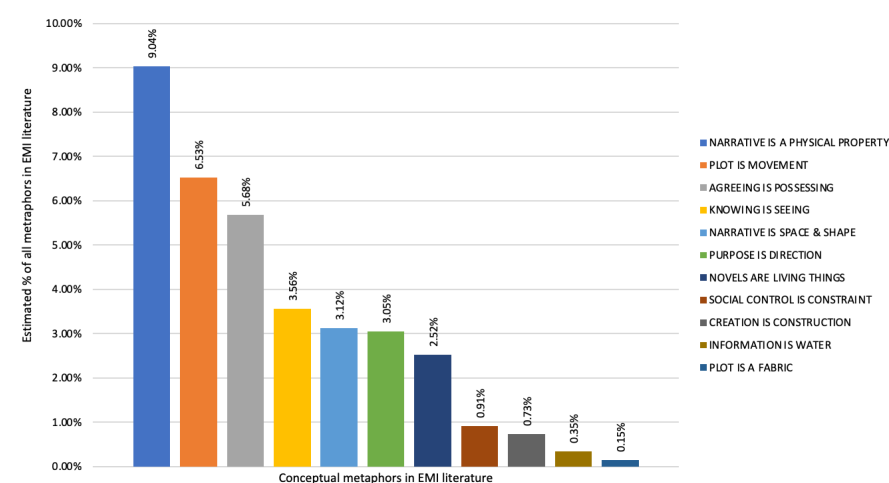

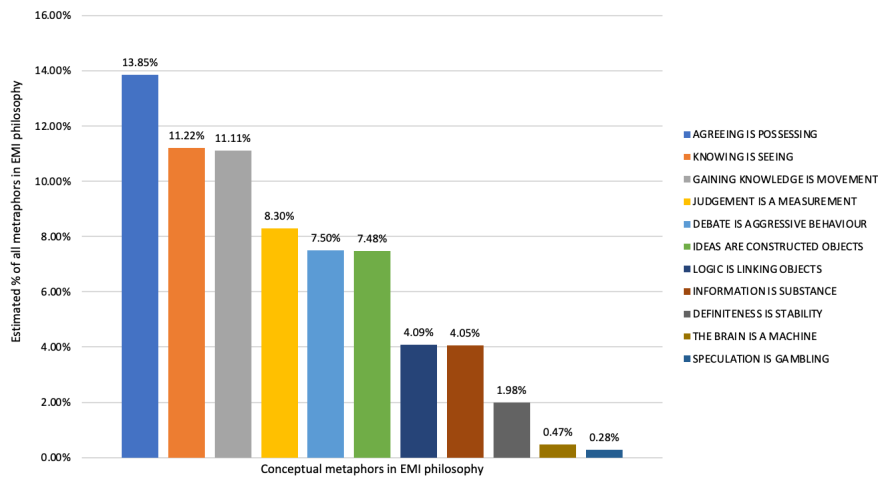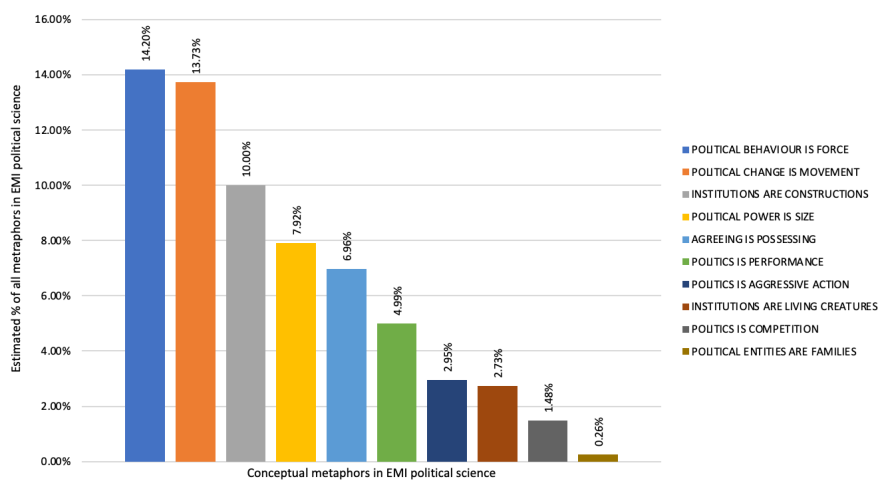Figure 3: Prevalence of metaphorical themes in EMI philosophy materials

Figure 4: Prevalence of metaphorical themes in EMI political science materials

In sum, I feel that the study has been quite successful. I was able to expand the corpus of materials significantly, and this corpus will continue to grow in the coming years. The corpus techniques I employed were able to identify individual metaphors and broader metaphorical themes. The results obtained offer support for the thesis that metaphorical vocabulary is frequent in EMI materials, differs somewhat in its patterning across disciplines, and yet is classifiable into a number of themes that might well be of value as instructional guides for learners.

The main challenge faced in the study has been the time-consuming nature of data collection and analysis. In order to build on the current findings, I plan to further expand the corpus and produce more detailed analyses of the behaviour of lexical items in particular fields, and this is something that I will work towards over the coming years.

One aspect of the analysis that has been quite challenging has been the use of semantic analyses to classify metaphors into themes. The approach used thus far has been to base decisions on dictionary definitions and previously published semantic classifications of vocabulary. However, I feel that there is scope to revisit these analyses and produce a more detailed account of the process of coding decisions. Hopefully, this will enable a greater degree of replicability for future studies in this vein.

As a final future goal in this line of research, I hope to develop teaching materials that can be used in EMI courses to help raise learners' awareness of metaphor patterns in their particular field. If these prove to be useful, then I intend to make contact with publishers to see if there is interest in developing discipline-specific textbooks.

0

0