Corpus Evidence on English Collocational Patterns in Scientific Writing
― Implications for Effective Writing Development ―

Hiroshi OHTAKE[1]・Brian MORREN[2]

1 Introduction

How may learners be helped to write acceptable academic articles? What type of assistance do they require in becoming more competent and effective writers? These are questions that language teachers have long been struggling to address. In this connection, the present study focuses on how to make Japanese learners more aware of word collocations in English, since such awareness has tended to be largely neglected in English teaching in Japan. When native speakers learn their own language in a natural setting, they can quickly pick up common word patterns through being immersed in the language and by relying on their intuitions. This type of implicit knowledge is acquired unconsciously through frequent exposure to the strings of words and set phrases encountered in everyday communicative interaction (Kirsner, 1994). In the case of non-native learners, however, exposure to the target language is not so readily available and opportunities for implicit learning are greatly reduced. They therefore do not have access to the significant store of core words and collocations that form the basis of native speaker fluency. Consequently, when second language learners are required to write in the target language, they usually try to use their grammar with imperfect understanding of vocabulary, which often results in skewed or awkward English such as *a high possibility.* They may also be hampered by the constraints of cognitive capacity in that they do not have ready-made language patterns at their disposal that can be automatically retrieved from long-term memory. In particular, they lack the ability for native-like selection (Pawley & Syder, 1983) in that they are less able to choose the preferred linguistic sequence from a number of grammatically acceptable alternatives. They therefore do not have the same level of fluency and pragmatic competence that they possess in their first language.

In this regard, it has been found that learners consider vocabulary as most important in relation to writing needs followed by grammatical accuracy (Gosden, 1996; Leki & Carson, 1994; Muncie, 2002). In particular, learners are frustrated by their inability to quickly retrieve or access relevant lexical/grammatical forms (Snellings et al., 2002).

[1] Department of Foreign Languages, Kyoto Prefectural University of Medicine
[2] Center for Languages, Arts, and Sciences, Fukui Prefectural University

Vocabulary deficits also account for imprecision in the expression of their thoughts and impede the flow of their writing in that they sometimes have to settle for an inappropriate word substitute which does not accurately express their intended meaning. Such a concern for vocabulary is also shared by those involved in the assessment of learner writing who tend to view an inadequate knowledge of vocabulary unfavorably. In this respect, it has been shown that lexical content has a great influence on their judgements in that mistakes in lexical selection are considered less acceptable than other types of error (Engber, 1995; Harley & King, 1989; Santos, 1988). Learners are aware of their deficiencies at the basic level and that this is holding them back from becoming more accomplished writers. In order to accelerate their language processing and develop greater efficiency, therefore, they need to have greater control over basic vocabulary and grammatical conventions.

2 Learning Vocabulary

Knowledge of vocabulary is an essential component in written production in that quality of writing is closely related to vocabulary proficiency (Astika, 1993). However, rather than the number of words a reader knows, it is the depth of word knowledge that is critical in the production of text (Liu & Shaw, 2001). If readers have only a partial knowledge of a particular word, this may not be sufficient to grasp the full range of meanings of the word as it is used in various contexts (Hunston & Francis, 1998). It is also important to know which particular patterns are common in a given register and which are not. Such information would certainly be useful to learners (Biber & Conrad, 2001). In the absence of such knowledge, learners of all levels of proficiency are likely to remain insensitive to the range of potential meanings expressed by a given word and experience difficulties in producing the appropriate sense of a word in their academic writing.

Instruction involving the completion of vocabulary exercises or the consulting of a dictionary may increase vocabulary knowledge but may have relatively little beneficial effect on student writing. In order to improve written production, it is necessary to have more than just a superficial understanding of words. In addition, there is a need to have ready access to that vocabulary if learners are to enhance their skills in language use. Without such lexical access, the production processes will not be able to proceed smoothly and the resulting text will be impaired (McCutchen et al., 1994). It is imperative, therefore, to ensure that learners have sufficient opportunities for developing their knowledge and skills in using vocabulary.

3 Lexical Relations

In recent years, the trend in vocabulary learning has been to view words not as independent, separate entities but as constituents of lexical phrases. As such, their range of meaning is determined by their relations with the accompanying words in the text and by their syntactic and collocational patterns (Beheydt, 1987; Leffa, 1998). In this respect, advances in corpus linguistics have allowed us to gather useful language data and provided insights into recurring lexical patterns that have hitherto gone undetected (Sinclair, 1991; Willis, 1990). We are now more aware of the importance of lexical relations and how these are governed by rules whereby words interact with each other not in random clusters but in a clearly principled way. Furthermore, it is now known that lexis has a lot in common with syntax and phonology in that there is a great deal that is automatic in the routine behavior of words. Namely, in normal language use, words form predictable and stable combinations which consistently occur in fixed linguistic patterns in conventional everyday discourse.

The view of language as a phraseological system differs radically from the traditional concept of a clear distinction between lexis and grammar (Lewis, 1993; Nattinger & DeCarrico, 1992). In the case of lexis, it is necessary to know about the syntactic patterns in which it appears, since in the absence of such information, we will not know how a particular lexical item is actually used in discourse, and accuracy and fluency will suffer. The same applies to grammar in that a certain syntactic pattern is dependent on particular lexical items that have the same pattern. In other words, lexis may not be described without reference to syntax, nor syntax without reference to lexis since words and patterns are inextricably linked. They are therefore interdependent and not separate features of language organization (Hunston & Francis, 1998).

In particular, styles of language are governed by certain norms and conventions. Knowledge and observance of such conventions are essential in creating conditions for clear articulation of meaning, and smooth and rapid processing. Moreover, in the case of academic writing, observing the conventional collocational forms is critical in achieving precision of expression (Howarth, 1998). Where word combinations deviate from the norm, this may cause difficulties in processing and comprehension. If readers have problems in understanding a writer's intent due to stylistic dissonance or imprecision, this could result in loss of clarity (Martin, 1984). Even though the underlying meaning may still be grasped, unconventional collocations are likely to distract from what the writer is trying to say. Such communicative deficiencies on the part of writers may ultimately lead to their marginalization within the academic community (Lennon, 1998).

4 Corpus Analysis

There are regularities underlying the way in which words co-occur. Words tend to be subject to certain constraints or preferences which affect language choice in discourse in a systematic way. Furthermore, Stubbs (2001) notes that "native speakers' unconscious knowledge of collocations is an essential component of their idiomatic and fluent language use and an important part of their communicative competence" (p. 73). A knowledge of collocations is exactly what distinguishes native speakers from foreign learners and evidence derived from corpus analysis may contribute to narrowing the gap between the two. For example, with regard to English synonyms, it has been found that one way to distinguish their subtle differences of meaning is to focus on their collocational behavior. In particular, any positive or negative nuances they exhibit may not be considered as an essential property of a particular lexical item acting alone but as the result of its association with particular collocates. In other words, an examination of a word's immediate collocational environment can provide great insight into the particular nuances expressed (Partington, 1998).

In view of this, there is a need to familiarize learners with frequent patterns by focusing their attention on the lexical items that immediately precede or follow a given word (in their immediate linguistic context). In certain cases, words at a further remove may also be examined in that the key word may appear in a fixed expression involving a cluster of words each of which must be included to convey a particular meaning. For example, prepositions and compounds are notoriously difficult for learners to master so that they might benefit from having an array of one-line samples of use. This might then help to highlight which words will typically co-occur with a given lexical item and make writing a more fluent, less laborious, and ultimately more successful task for learners.

5 Corpus Evidence

In this section, we present a number of case studies focusing on lexical or collocational errors/deviations, which are found in Japanese scientific research writing that has been published and is also available via the Internet. This type of error is known to be one of the most difficult to identify by non-native writers but is also one that is likely to be most detrimental to their efforts to express themselves adequately in English. We intend to show how certain words are used in a given context by taking advantage of corpus analysis, concordance, collocation, and statistical data. In particular, we illustrate what kinds of patterns are the norm among native speakers and then provide examples of

faulty expressions found in Japanese scientific papers so as to draw attention to certain stylistic infelicities that appear repeatedly in such writing and that need to be avoided.

For the purpose of gathering statistical data and concordance samples, we consulted the Life Science Dictionary Project corpus, which has been compiled by storing English texts in medical journals mainly through the public MEDLINE database that contains numerous scientific articles published all over the world. The corpus currently consists of over 29 million running words.

In this study, we focus only on the highlighted part of the selected sentences and do not concern ourselves with other types of error (e.g., incorrect use of determiners, singular/plural number agreement) that may appear in the sample sentences.

5.1 potential/potentiality/possibility/probability

Sample Sentences:

(1a) Free radical scavenger … will **have the possibility to** become the medicine for the dementia.

(1b) These findings clearly **suggest the** therapeutic **potentiality** of … as a novel approach for the treatment of … .

On reading the above sample sentences, native speakers of English are likely to say they prefer to use *potential* in place of *possibility* and *potentiality*. However, for most Japanese learners of English, knowing how to properly differentiate the meanings of *potential, potentiality, possibility*, and *probability* is an extremely difficult task since they can find one common Japanese equivalent *kanousei* (可能性) for these four English words in English-Japanese dictionaries. As a result, few Japanese students pay attention to the distinctive differences among them, especially their usages. Although they may read explanations given in dictionaries — according to which the likelihood of occurrence is considered higher when *probability* is used instead of *possibility*, and *potential* is defined as referring to future possibility — these explanations do not seem to be much help to non-native speakers of English. This is because they lack the information about how each of these words is actually used, namely with what other English words they typically occur. When learners are required to use these four words, they may consider only the degree of likelihood and not know how to properly discriminate among them in their writing. The above sentences produced by Japanese researchers are a good example of the inadequacies of grammar books and dictionaries. While learning *possibility*, they

may have neglected to learn the common and natural collocational patterns that accompany it (Table 1). The sample sentences introduced here can be translated into natural-sounding Japanese sentences, which makes this type of error not so salient from the perspective of Japanese learners of English.

(Table 1: Excerpt from concordance for *the possibility*)

120 ...	These results open	the possibility	of using the supF tRNA am...
121 ...es as both inputs and output, they open		the possibility	of communication between ...
122 ...	Microarray technology has opened	the possibility	of evaluating the express...
123 ...	Our finding opens	the possibility	of searching for the addi...
124 ...e identification of such peptides opens		the possibility	of using these sequences ...
125 ...	We can rule out	the possibility	of a negatively charged b...
126 ...	These data rule out	the possibility	of an egg effect in which...
127 ...	Our measurements rule out	the possibility	that chromosome stretchin...
128 ...	To rule out	the possibility	that HS3 is required for ...
129 ...	To rule out	the possibility	that membrane attachment ...
130 ...	We rule out	the possibility	that monoallelic expressi...
131 ...this smaller sample, we cannot rule out		the possibility	that small numbers of div...
132 ...	These findings effectively rule out	the possibility	that the presence of nonc...
133 ...ential entry of these fac tors rules out		the possibility	that they are transported...
134 ...extrin system was used, which precludes		the possibility	of cross-contamination be...
135 ...requiring consecutive responses prevent		the possibility	of using the most desirab...
136 ...	They also raise	the possibility	that disrupting multiple ...
137 ...	These results also raise	the possibility	that IL-7 therapy could b...
138 ...	Our observations also raise	the possibility	that neurons use a burst ...
139 ...NARE-mediated membrane fusion and raise		the possibility	that additional component...
140 ...-activated MAP kinase pathway and raise		the possibility	that environmental condit...
141 ...s effects of oxidative stress and raise		the possibility	that humans with heterozy...
142 ...new pathway of p53 regulation and raise		the possibility	that increased MDM4 level...
143 ...week of mammalian development and raise		the possibility	that multiple isozymes of...
144 ...nge in neurological diseases, and raise		the possibility	that neuromodulators may ...
145 ...n retinal capillary pericytes and raise		the possibility	that PGs may play a role ...
146 ...sion in activated macrophages and raise		the possibility	that PPARgamma ligands...
147 ...dependent protein phosphatase and raise		the possibility	that regulation of other ...

When asked what was wrong with the sample sentence (1a), some Japanese learners replied that *possibility* should be replaced by *probability*. In so doing, they thought they could convey the higher expectations of some possible future occurrence and were completely unaware of the fact that *have the possibility to ...* and *have the probability to ...* are not widely used collocational patterns in English and should be considered as deviating from the norm. In view of this, seemingly grammatically correct English sentences represent a serious problem for non-native speakers of English, even for those at a rather advanced level. It is not until they have been immersed in a sea of natural English expressions/sentences that they learn that *have the possibility to ...* is not

commonly used among native speakers.

Searching through the corpus data, it can be seen that in the concordance for *possibility* there are no instances where *have the possibility to ...* is used. Statistical results (Table 2) show that the most frequently occurring verb before *possibility* is *raise*, that *suggest the possibility* is also a quite common expression, and that *open the possibility* and *rule out the possibility* are also used although far less frequently. Thus, learners should focus on the conventional forms of expression when using *possibility* and become aware of its common collocations.

Table 2: Positional frequency for *the possibility* (2,362)

2nd left		1st left		1st right		2nd right	
we	221	raise	348	that	1641	the	292
and	166	raises	227	of	655	a	144
to	156	raising	216	for	20	this	76
results	116	investigated	77	is	8	these	62
findings	69	with	67	to	6	an	47
have	65	raised	66	remains	5	using	39
this	62	suggest	64	exists	5	it	35
data	54	examined	59	was	4	other	23
also	53	suggesting	51	existed	3	some	21

As for the expression *raise the possibility*, there still remains a problem concerning Japanese learners who are not accustomed to this expression since they may falsely assume that it means 'to increase the likelihood' of something, whereas it is actually similar in meaning to *suggest the possibility*. According to the *Oxford Advanced Learner's Dictionary*, the meaning of *raise* is (1) 'to lift or move something to a higher level,' with which most Japanese learners are familiar, and (2) 'to mention something for people to discuss,' which they are less likely to have come across.

Here in the sample sentence (1a), the expression *have the potential to ...* should have been used instead of the unnatural collocational pattern *have the possibility to ...* . The concordance results show that for the word *potential*, the most frequent verb forms appearing just before it are *have* and *has*, and that *have/has the potential to ...* is therefore a commonly used expression when referring to future possibility (Tables 3 & 4).

(Table 3: Excerpt from concordance for *the potential*)

23	...ly than unmodified oligonucleotides and have	the potential	to be used as diagnost...
24	...studied, and we believe that these loci have	the potential	to be used as polymorp...
25	...taining single amino acid substitutions have	the potential	to be used for modulat...
26	... These cells ha ve	the potential	to be used for restora...
27	... These primers have	the potential	to be used in a PCR as...
28	...venting the function of these receptors have	the potential	to be useful anti-canc...
29	...f substrates (rubbed films of proteins) have	the potential	to be useful in a vari...
30	...ls arising in fetal or neonatal tissues have	the potential	to become B-1a cells c...
31	...tems based on HIV or other lentiviruses have	the potential	to become important to...
32	...ation, intestinal transplant procedures have	the potential	to become the standard...
33	... patients with inflammatory myopathies have	the potential	to behave as antigen p...
34	... stigated whether quinol-GSH conjugates have	the potential	to behave as genotoxic...
35	... Such transplants have	the potential	to benefit retinal dis...
36	... workers are not well characterized and have	the potential	to bias the measures f...
37	... heparan sulfate pro teoglycans (HSPGs) have	the potential	to bind and directly r...

Table 4: Positional frequency for *the potential* (3,210)

2nd left		1st left		1st right		2nd right	
to	395	has	342	to	644	of	959
and	155	have	221	for	476	the	79
we	149	of	150	of	425	be	53
results	76	and	142	role	291	for	39
that	41	with	107	use	65	this	39
study	40	demonstrate	85	utility	55	form	34
have	37	investigate	81	importance	48	between	33

As for the sample sentence (1b), *suggest the potentiality* is not found in the corpus and does not seem to appear anywhere among commonly used English sentences. In this respect, since it may be difficult to recognize the definitional difference regarding *potential* and *potentiality* in English-Japanese dictionaries, Japanese learners should be strongly encouraged to make the most of concordance evidence. Such evidence shows that the number of instances of *potentiality* is only four, compared to 17,274 in the case of *potential*, which clearly indicates that *potentiality* occurs only rarely while *potential* is far more common.

Some Japanese learners may assume that *have the probability to ...* can be used when referring to something whose likelihood of occurrence is very high, but the expression is not accepted as a natural English expression as seen in the concordance results for *probability* (Table 5). We learn from the statistical analysis that the most frequent verb coming before *probability* is *increase* (Table 6). As the statistical results demonstrate, it may be useful to suggest to learners that they first get accustomed to using the most commonly used expression for each word: *have the potential to ...* , *raise*

the possibility of ... , and *increase the probability of ...* . Learners should be well aware that neither *raise the probability* nor *suggest the probability* is generally used, even though these awkward word patterns may sound natural when translated into Japanese. The corpus evidence also shows that *increase the possibility* is not very widely used.

(Table 5: Excerpt from concordance for *the probability*)

179 ...	These alterations increase	the probability	for abnormal thalamocorti...
180 ...ssay sensitivity, and they can increase	the probability	of a diagnosis by verifyi...
181 ...by no more than 2 kcal/mol can increase	the probability	of nucleation of disorder...
182 ...	Predators may increase	the probability	of prey extinction result...
183 ...	Such techniques often increase	the probability	of detecting linkage, but...
184 ...rs can reduce the power and/or increase	the probability	of obtaining false positi...
185 ...constraints, can substantially increase	the probability	of cell immortalization. ...
186 ...multisite enhancer elements to increase	the probability	of an interaction between...
187 ...e potential which may serve to increase	the probability	of neoplastic progression...
188 ...in-protein interactions, which increase	the probability	of establishing an active...
189 ...y of actin polymerization and increased	the probability	of further Cdc42 accumula...
190 ...n time of single channels and increased	the probability	of their failure to open....
191 ...creased HIV-1 replication and increased	the probability	to target HIV-1 IN in inf...
192 ...tion of the proviral enhancer increased	the probability	of maintenance-coupled de...
193 ...and radio/television) further increased	the probability	of supine placement (OR, ...
194 ...moval of external Ca(2+) ions increased	the probability	of channel opening (Po) s...
195 ...ephrins or mechanical probing increased	the probability	of lead growth cone retra...
196 ...red blood cells significantly increased	the probability	of pore formation by GPI-...
197 ... prolonged visual stimulation increased	the probability	of the up state. ...
198 ..., showing that PKC activation increases	the probability	of channel opening rather...
199 ...	We show that the cHS4 increases	the probability	that integrated proviruse...
200 ...he wild-type POR considerably increases	the probability	of photoactive state form...
201 ...	Mutation of CIN genes increases	the probability	that whole chromosomes or...
202 ...ain abnormal findings greatly increases	the probability	of breast cancer. ...
203 ...cially pericentric inversion, increases	the probability	of genetic isolation amon...
204 ...phosphorylation of VR1 by PKC increases	the probability	of channel gating by agon...
205 ...linked allele rather than by increasing	the probability	of establishing transcrip...
206 ...d DNA damage and, second, by increasing	the probability	of recombination with dec...
207 ...ory synaptic transmission by increasing	the probability	of transmitter release vi...
208 ... via heparin-like molecules, increasing	the probability	that the virus will bind ...

Table 6: Positional frequency for *the probability* (538)

2nd left		1st left		1st right		2nd right	
to	32	in	23	of	403	a	45
and	15	and	22	that	68	the	34
by	14	that	18	distribution	15	of	18
that	12	of	18	density	8	an	14
increase	10	increase	17	for	6	release	9
may	10	increases	17	and	5	survival	9
or	8	increasing	15	distributions	4	transmitter	7
estimate	6	increased	15	profile	4	detecting	6

5.2 may/should/possibly/probably

Sample Sentence:

> During the process of … , a loss of the anticoagulant property … **may probably** occur.

This sentence requires Japanese learners to be highly sensitive to how the English modal auxiliaries *may* and *should* interact with the adverbials *possibly* and *probably*. However, it is not surprising to encounter this type of error in Japanese writing, given the lack of sensitivity to these words on the part of Japanese learners (Ohtake & Morren, 2001). As is the case with the differentiation of *possibility* and *probability*, Japanese learners are generally weak at discriminating *may* from *should* when these words are used in referring to the possibility or probability of something happening. Especially when Japanese learners use *should*, they appear to exclusively use it to 'indicate obligation, duty, or correctness' (*Concise Oxford English Dictionary*: COD), as often found in the concordance results for *should* derived from an English corpus of Japanese writing. Few Japanese learners appear to effectively or successfully use *should* to 'indicate what is probable' (COD) and thus neglect to learn that *should probably* sounds natural while *may probably* does not (Table 7).

(Table 7: Excerpt from concordance for *should*)

40 ...	Our approach	should	be applicable to other membrane pr...
41 ...ed genetic and pharmacological approach		should	allow analysis of the specific rol...
42 ...	Our present approach	should	facilitate the generation of multi...
43 ...	This approach	should	be applicable for the creation of ...
44 ...	This approach	should	have significant future implicatio...
45 ...	The assay	should	be particularly useful for kinetic...
46 ...	This assay	should	serve as a useful criterion for as...
47 ...	DNA chip-based assays	should	play a valuable role in high throu...
48 ...	The coupled assays	should	be widely applicable since the pro...
49 ...ators and inhibitors in in vitro assays		should	clarify the role of individual mot ...
50 ...lanations as to why observed behaviours		should	differ from the optimal behaviour ...
51 ...	Characterization of URE3-BP	should	provide insight into transcription...
52 ...de range of attenuation for mouse brain		should	prove useful in fine-tuning recomb...
53 ...he view that hydrophobic surface burial		should	be commensurate with hydrogen-bond...
54 ... postentry steps of HIV-1 infection but		should	also help to enhance the efficacy ...
55 ...ocking drug efflux with fumitremorgin C		should	allow for functional analysis of t...

The most defective aspect may arise from their heavy dependence on bilingual dictionaries (Ohtake & Morren, 2002; Schmitt, 1997) and persistent belief that each and every English word has its exact counterpart in the Japanese lexicon. This results in

learners acquiring only a partial definition of English words and remaining unaware of the full range of meanings they can express. In the case of *should*, most Japanese learners seem to associate it with the Japanese phrase *suru beki da* (するべきだ), which is a relatively strong expression and not normally used in polite situations such as when giving advice to one's elders. Consequently, they assume that *should* is to be avoided in such situations and tend to use *had better* instead, which is often translated into Japanese as *shita hou ga yoi* (したほうがよい) and has a less commanding and much softer tone in Japanese. Owing to the false connotation derived from the Japanese translated expressions, it is not unusual to encounter Japanese learners who firmly believe that *should* is equivalent to *suru beki da* in Japanese and *had better* is the same as *shita hou ga yoi*. In view of this, it should be pointed out to learners that, on a scale of intensity, *should* appears more neutral than *had better* in that *should* suggests 'advisability' while *had better* implies a 'strong recommendation' and has a more forceful connotation. This might also account for learners' excessive use of *had better* rather than *should* in the mistaken belief that it behaves the same way as in Japanese and is the preferred choice in more formal contexts. Moreover, in terms of register, semi-modals such as *had better* have been found to appear far more frequently in conversation than modal verbs but are much less common in formal writing (Biber et al., 1999). In the case of *had better*, its use is therefore mainly confined to spoken English while it is almost non-existent in the conservative medium of academic writing. Conversely, the modal *should* is well-established in academic writing and is considered more stylistically appropriate.

(Table 8: Complete concordance for *may possibly*: 14)

1	...es in estrogen metabolism and, thereby,	may possibly	explain interindividual diff...
2	...rogesterone may stimulate breathing and	may possibly	improve symptoms of hypovent...
3	...to changes in iron content and form and	may possibly	be used as indicators of suc...
4	...or studying rapid channel movements and	may possibly	act as a fluorescent activit...
5	...t that mechanoperception in p lant cells	may possibly	be transduced through intrac...
6	...ectin further suggests how some domains	may possibly	be important for protein int...
7	... While various unknown factors	may possibly	give rise to selective activ...
8	... Microfilaments	may possibly	act by uncoupling Lyn from t...
9	...an unitary displacements, this mutation	may possibly	perturb the mechanical coord...
10	... A meaningful fraction of patients	may possibly	be cured when treated as agg...
11	... This local response	may possibly	assist in limiting the clini...
12	...d implicate them in CNS disorders, that	may possibly	be induced or exacerbated by...
13	... to trigger transcription and therefo re	may possibly	serve as a transcription act...
14	...ogs have not been fully explored, which	may possibly	have limited the scope of th...

(Table 9: Excerpt from concordance for *may occur*: 417)

104	...ctivation, PI 3-kinase-sensitive events	may occur	both upstream of Ras and betwee...
105	... Such events	may occur	in inflammatory bowel disease d...
106	...am in serial radiographic examinations)	may occur	more rapidly, and with less bet...
107	...ntalizing suggestion that base excision	may occur	by cleavage of the glycosidic b...
108	...nic disease, relevant dietary exposures	may occur	over decades. ...
109	...erve distinct functions; axon extension	may occur	predominantly in the outermost ...
110	... Bone marrow (BM) fibrosis	may occur	in myeloproliferative diseases,...
111	...y, dynamic protein backbone fluctuation	may occur	, enabling Cys532 to move within...
112	... Fractionation	may occur	in aqueous solution during equi...
113	...ate that endopin 2 inhibitory functions	may occur	in the regulated secretory path...
114	...iated repair of physiological functions	may occur	independently of integrin reloc...
115	... Intravascular gas	may occur	as a transient incidental findi...
116	...at the regulation of the rat MnSOD gene	may occur	not only at the transcriptional...
117	...s in the hepatitis B virus (HBV) genome	may occur	during therapy. ...

(Table 10: Excerpt from concordance for *probably occur*: 44)

25	... G. chilensis and the Galapagos lineage	probably occurred	6 to 12 million years a...
26	...adiation of major lepidopteran lineages	probably occurred	during the Late Jurassi...
27	...n regulation of perforin-mediated lysis	probably occurs	without direct interacti...
28	...is decrease in IAA within the mesocotyl	probably occurs	primarily by a change in ...
29	...pression pattern in the cortex, it most	probably occurs	independently from it, an...
30	...osomal breakage between Nubp2 and Nubp1	probably occurred	during the evolution of...
31	...ded from pre-human strata; three others	probably occurred	on 'Eua in pre-human ti...
32	...ns, indicating that the phosphorylation	probably occurs	on a conserved histidine ...
33	...e CA1 region suggests that this process	probably occurs	before CA3, possibly in t...
34	... Much of this new productivity	probably occurred	in microbial mats, whic...
35	...onal activation of this docking protein	probably occurs	through the IGF-1R. ...
36	... The latter two reactions	probably occur	within RPE phagolysoso...
37	...erefore, both CB(1) and CB(2) receptors	probably occur	throughout the vertebrat...
38	... Reinitiation	probably occurs	through a different pathw...
39	... Repositioning	probably occurred	by disassembly of the i...
40	... Most selection	probably occurs	at birth and will be grea...

(Table 11: Complete concordance for *should occur*: 13)

1	...han the free electron mass--exciton BEC	should occur	at temperatures of about 1 K...
2	...In acute infection, a beneficial effect	should occur	. ...
3	...pectations that morphological evolution	should occur	largely within Pleistocene r...
4	...n of phosphorylated [8-14C]-ganciclovir	should occur	almost exclusively in tissue...
5	...eration of treatment for this infection	should occur	in any patient with cystic f...
6	...beta-hydroxyacylthioester intermediates	should occur	during the reactions catalys...
7	...r accumulation of deleterious mutations	should occur	on polysomic chromosomes wit...
8	...ques than others--ie, irregular plaques	should occur	in multiple vascular beds in...
9	... This process	should occur	, for example, in periodicall...
10	...e picket fence model, maximal quenching	should occur	at two different levels in t...
11	... Consequently, amino acid replacements	should occur	at a higher rate in compleme...
12	...n stratification due to climate warming	should occur	. ...
13	...ankton-to-higher trophic level food web	should occur	when this ratio falls below ...

To correct the flaw in the sample sentence, if the writer wants to retain either one of the two originally used words, *may probably* should be replaced with either *may possibly* or *should probably* to conform to a natural standard English expression. However, the concordance results reveal that the frequency of *may possibly* in the corpus is rather low (Table 8), and no instances of *should probably* were retrieved, even though *Harrison's Principles of Internal Medicine* (1998) contains a considerable number of instances for this particular pattern. As for the corpus sentences that include *occur*, the most frequently appearing expression is *may occur*, followed by *probably occur*, and then *should occur* (Tables 9, 10, & 11). This may be partly because *may possibly* and *should probably* sound rather redundant in English, whereas rather straightforward expressions tend to be preferred in scientific journals.

5.3 study on/about, research on/about

Sample Sentences:

(3a) The localization of … warrants further **study about** the role of … in the induction of hepatocellular damage.

(3b) The epidemiological **research about** cataract frequency … has to take into account that … .

Prepositions are known to be a major obstacle in learning English for most Japanese students, especially when writing in English. It is not an exaggeration to say that few, if any, students are confident enough to choose an appropriate preposition from among others when there is room for another to be filled in. The following short passage consisting of 53 words includes as many as 11 prepositions. This original passage, with all but one preposition (*within*) deleted, was given to 252 Japanese university students

Fill in the following blanks with an appropriate preposition:

But here it gets complicated. Does higher testosterone produce more aggressive behavior? Or does the more aggressive male – whose aggression was learned, say, #1# home or #2# school or #3# the neighborhood or #4# the team or #5# the culture #6# large – call #7# a release #8# testosterone #9# within himself #10# assistance?

(Extract from 'Are You Man Enough?' in *Time Magazine* April 24, 2000, Vol. 155, No. 16. See original passage in Appendix.)

and a number of native speakers of English as a task in which they were required to fill in each blank with an appropriate preposition.

As for the native speakers of English, all their responses were perfectly correct and uniform except for one or two variants that were also possible substitutes in the context. However, in the case of the Japanese university students, their answers were far from satisfactory and they achieved an average score of only around 46%. These results clearly show that with regard to prepositions, native speakers will usually infer the correct ones because they are familiar not only with idioms but also natural collocations, while non-native speakers have difficulty here because of the tendency for prepositions to have multiple meanings and for the choice to be context-dependent.

In this connection, it is well known that English prepositional usage is exceedingly complicated and difficult to learn and may pose problems for even advanced learners. While there are generally few problems when dealing with prepositions that refer to simple spatial relations and movements (e.g., *on* the table; *to* the station), it is in other areas of usage that preposition selection may appear rather arbitrary (Rastall, 1994). In certain instances, the choice of preposition is determined by the context and the preposition itself may contain little or no information value. It is therefore often a matter of convention and the preposition makes little contribution to meaning. As such, they may be classified as either 'free' or 'bound' prepositions, where the former carry an independent meaning irrespective of other words in the context, while the latter have little independent meaning and are directly influenced by some word in the context (Biber, et al., 1999). Nevertheless, research has shown that prepositions do not merely serve a grammatical function but also have semantic content that may influence the words with which they are linked. That is to say, prepositions can have a salient spatial sense that shapes the accompanying words and imbues them with a certain shade of meaning. A preposition may also have an extended meaning which is derived from its underlying spatial characteristics (Lakoff & Johnson, 1980). They therefore have the potential to affect how a certain phrase is interpreted. In view of this, it is important for learners to be aware of the basic spatial meaning of prepositions and of how this may be extended into more figurative meanings. In other words, prepositions should not be considered simply as lexically empty grammatical categories since they have communicative value and add to the information in the text.

With regard to preposition entries in dictionaries, they appear to suffer from various defects including disorganized scattering of semantic information, confused presentation of sense information, lack of information on semantic contrasts, and omitted usages (Lindstromberg, 1996). Such deficiencies may be due to the tendency to list entries

according to their order of frequency. While this provides learners with important information on the most common uses of prepositions, it also tends to conceal information about semantic similarities and contrasts so that for learners who wish to examine a particular entry, they often have to scan a long undifferentiated list of examples if they wish to locate the object of their search. Consequently, learners may become discouraged and despair of ever finding the information they are looking for. They may thereby refrain from using dictionaries in their search for prepositional meanings. Through the use of concordances, however, learners may be able to pinpoint certain usages that either do not appear in dictionaries or are poorly presented. They may then be able to learn the relevant information inductively through careful scrutiny of the concordance samples. Although they may still have no clear idea about why certain prepositional collocations are allowed while others are not, this may gradually become apparent with sufficient practice and experience in using prepositions accurately. They may in time sharpen their intuitions and learn to distinguish the various senses of prepositions from each other and use them appropriately not only through memorization but also through applying their knowledge of a preposition's semantics and extended meanings. While information about frequency is useful in helping learners to collocate words correctly, there would also seem to be a place for helping learners become more aware of the semantic scope of prepositions and why they collocate in the way they do.

With regard to the sample sentences, Japanese learners usually have difficulty in discriminating *about* from *on* when used to mean *concerning*. The difference between these two prepositions does not appear to be very important, and indeed these two prepositions can be used interchangeably. However, when we examine the conventional patterns containing these prepositions, we find that the use of *about* just after *study* or *research* is rarely, if ever, found in the English corpus (Tables 12 & 13), even though a few instances of *a study/research about ...* were found when we examined a corpus of Japanese writing.

As for *information*, however, the frequency of the accompanying prepositions, *about* and *on*, is almost the same. This implies that even if the choice between the two seems to be rather arbitrary, researchers constantly favor the use of *on* with *study* and *research*, while they are not so consistent in choosing one of the two prepositions with *information* (Table 14). For the distinction between *study on* and *study about*, explanations and definitions given in dictionaries do not seem to be much help and instead may only serve to confuse learners (Lindstromberg, 1996, 2001). Particularly for Japanese learners, the distinction does pose a problem because there is one Japanese equivalent for these two prepositions, ... *ni tsuite* (～について). When we look at the

Table 12: Positional frequency for *study* (77,448)	
1st right	
of	6594
in	2905
with	1280
to	1023
on	895
by	808
for	320
at	231
from	174
under	51
:	:
about	1

Table 13: Positional frequency for *research* (2,713)	
1st right	
on	135
in	102
into	48
to	40
with	23
for	16
by	11
of	7
at	6
over	4
:	:
about	0

Table 14: Positional frequency for *information* (6,847)	
1st right	
about	795
on	732
from	338
to	276
for	255
on	237
of	48
with	46

evidence obtained in the form of concordance and statistics, we can see that *on* is the much preferred preposition occurring immediately after s*tudy* or *research*.

This kind of information may be useful in deciding which preposition to choose. Nevertheless, this does not explain the reason for the different collocational distribution of *on* and *about* when used to mean 'on the subject of' or 'concerning' a particular topic. To account for such differences, we should know that these two prepositions may also be distinguished in terms of formality in that *on* is used in more formal contexts while *about* is not (Quirk et al., 1985). Since academic writing may be considered a relatively formal affair, this would then account for the choice of *on* as the preferred collocate for *study* and *research* respectively. In addition, from the semantic perspective, *on* would seem to imply a more concentrated focus in contrast to *about* where the focus is more dispersed (Leech & Svartvik, 1975). This may be understood as derived from the literal sense of *on* as being 'positioned upon some particular base.' In this case, the extension of the spatial meaning of the preposition *on* serves to represent the image of the object of investigation (the topic) as the basis upon which *study* is positioned (with direct contact) for the purpose of elucidation (Lindstromberg, 1996). On the other hand, the preposition *about* has a wider, more scattered focus (without direct contact). It therefore has a less specific orientation than *on* so that the area of research is less clearly demarcated.

With respect to prepositions, therefore, it would seem sensible for learners to view the concordance lines and note the types of collocations that appear. The more common

patterns and collocates could then be committed to memory and treated simply as contextually determined forms. However, the apparently arbitrary nature of prepositional fixing would preclude the need for any semantic interpretation that might help to clarify the role of the preposition in assigning meaning to its particular collocate. This is unfortunate, since a clear understanding of the spatial senses may help us to explain why one preposition rather than another is chosen to express a given meaning. In this way, it may be useful for learners to view prepositions in terms of a continuum of meaning senses extending from prototypical to metaphorical (Maclennan, 1994). The use of this type of imagery may also promote deeper cognitive processing and thereby assist in enhancing learner recall (Boers, 2000; Boers & Demecheleer, 1998).

6 Pedagogical Implications

In the literature on the learning of vocabulary, it has been shown that learners need to have opportunities to encounter lexical items in multiple contexts in order to engage in deeper processing and thereby derive a fuller understanding of their various meaning senses (Laufer, 1990; Schmitt, 1997). This type of intensive exposure to representative patterns of language use may therefore enable learners to better understand the ways in which certain words behave. In this respect, the insights gained from concordances may be used in the design of appropriate classroom activities to promote a clearer understanding of the typical uses of particular linguistic expressions (Thurston & Candlin, 1998). Certainly, learners should be made aware of how such words are bound together with other words so that they may learn to express their ideas with greater clarity and fluency. In this regard, care must be taken to ensure that awkward, non-native expressions are eliminated. Care should also be taken to focus on the particular collocations that are relevant to learners' needs and that may be of some benefit to them. Learners should be encouraged to explore their various functions and uses and adopt those patterns that are relevant to their own particular communicative purpose in writing. In this regard, it is useful to select key words in their particular field (most common, frequently used words) that may pose problems for learners.

Concordances should therefore be considered as a dynamic resource in that learners may use them as another type of text whereby they may see a certain word exemplified in a number of usage examples, which may serve to provide further exposure to the array of meanings it may express in different situations. Furthermore, the ability to conduct corpus-based research independently would obviate the need to have access to native speakers (who are not always available) and pave the way toward further learner

autonomy. Language corpora may also be used by learners as a reference for the purpose of correcting their own lexical errors. By studying concordance samples of language use, they may derive the particular linguistic rules and patterns governing a given word's behavior (Johns, 1991; Makino, 1993; Todd, 2000). In this respect, it would appear that the use of concordances may be a valuable tool in encouraging learners to discover for themselves the syntactic and collocational properties of words.

At the same time, however, pedagogical decisions should not rely on frequency information alone since this might lead to the neglect of certain structures and expressions that, while relatively rare, might be of particular use to learners. Coupled with quantitative evidence, therefore, there is also a need for qualitative analysis to determine the particular communicative functions of recurring lexico-syntactic patterns (Conrad, 1999). It is also important to remember that concordance techniques are not completely objective in that subjective decisions are involved at various points regarding which particular items to analyze and certainly in the interpretation of results. Furthermore, it should be borne in mind that the patterns of language found through concordances do not necessarily offer suitable models that learners may reproduce in all circumstances. This is because the samples found cannot be expected to reflect typical usage in every case. Moreover, while the linguistic data in corpora are authentic, they are open to reinterpretation once they have been removed from their original context (Gavioli & Aston, 2001). Learners therefore have to understand that such data have a particular communicative function in a particular context and should not be considered as all-purpose items that can be used in completely different contexts (Milton, 1999). In view of this, rather than indiscriminately importing certain features found in concordances, learners should be encouraged to explore their various functions and uses and learn how to manipulate the patterns they find if they wish to achieve greater flexibility and balance in their writing.

References

Astika, G. 1993. Analytical assessments of foreign students' writing. *RELC Journal, 24(1)*, pp. 61-72.

Beheydt, L. 1987 The semantization of vocabulary in foreign language learning. *System, 15(1)*, pp. 55-67.

Biber, D., & Conrad, S. 2001. Quantitative corpus-based research: Much more than bean counting. *TESOL Quarterly, 35(2)*, pp. 331-336.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. Longman.

Boers, F. 2000. Metaphor awareness and vocabulary retention. *Applied Linguistics, 21(4)*, pp. 553-571.
Boers, F., & Demecheleer, M. 1998. A cognitive semantic approach to teaching prepositions. *ELT Journal, 52(3)*, pp. 197-204.
Conrad, S. 1999. The importance of corpus-based research for language teachers. *System, 27*, pp. 1-18.
Engber, C. 1995. The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing, 4(2)*, pp. 139-155.
Francis, G., & Sinclair, J. 1994). I bet he drinks Carling Black Label: A riposte to Owen on corpus grammar. *Applied Linguistics, 14*, pp. 190-200.
Gavioli, L., & Aston, G. 2001. Enriching reality: Language corpora in language pedagogy. *ELT Journal, 55 (3)*, pp. 238-246.
Gosden, H. 1996. Verbal reports of Japanese novices' research writing practices in English. *Journal of Second Language Writing, 5*, pp. 109-128.
Harley, B., & King, M. 1989. Verb lexis in the written compositions of young L2 learners. *Studies in Second Language Acquisition, 11*, pp. 415-440.
Howarth, P. 1998. Phraseology and second language proficiency. *Applied Linguistics, 19(1)*, pp. 24-44.
Hunston, S., & Francis, G. 1998. Verbs observed: A corpus-driven pedagogic grammar. *Applied Linguistics, 19 (1)*, pp. 45-72.
Johns, T. 1991. Should you be persuaded – Two examples of data-driven learning materials. *English Language Research Journal, 4*, pp. 1-16.
Kirsner, K. 1994. Second language vocabulary learning: The role of implicit processes. In N. Ellis (Ed.), *Implicit and Explicit Learning of Languages* (pp. 283-311). London: Academic Press.
Lakoff, G., & Johnson, M. 1980. *Metaphors We Live By*. University of Chicago Press. Chicago/London.
Laufer, B. 1990. Ease and difficulty in vocabulary learning: Some teaching implications. *Foreign Language Annals, 23 (2)*, pp. 147-155.
Leech, G., & Svartvik, J. 1975. *A Communicative Grammar of English*. Longman.
Leffa, V. J. 1998. Textual constraints in L2 lexical disambiguation. *System, 26,* pp. 183-194.
Leki, I., & Carson, J. 1994. Students' perceptions of EAP writing instruction and writing needs across the disciplines. *TESOL Quarterly, 28(1)*, pp. 81-101.
Lennon, P. 1998. Approaches to the teaching of idiomatic language. *IRAL, 36(1)*, pp. 11-30.
Lewis, M. 1993. *The Lexical Approach*. Hove: Language Teaching Publications.
Lindstromberg, S. 1996. Prepositions: Meaning and method. *ELT Journal, 50(3)*, pp. 225-236.
Lindstromberg, S. 2001. Preposition entries in UK monolingual learners' dictionaries: Problems and possible solutions. *Applied Linguistics, 22(1)*, pp. 79-103.
Liu, E., & Shaw, P. 2001. Investigating learner vocabulary: A possible approach to looking at EFL/ESL learners' qualitative knowledge of the word. *IRAL, 39*, pp. 171-194.
Maclennan, C. 1994. *Metaphors and prototypes in the teaching and learning of grammar and vocabulary. IRAL, 32(2)*, pp. 97-110.

Makino, T. 1993. Learner self-correction in ESL written compositions. ELT Journal, 47, pp. 337-341.
Martin, M. 1984. Advanced vocabulary teaching: The problem of synonyms. *The Modern Language Journal, 68(2)*, pp. 130-137.
McCutchen, D., Covill, A., Hoyne, S., & Mildes, K. 1994. Individual differences in writing: Implications of translating fluency. *Journal of Educational Psychology, 86(2)*, pp. 256-266.
Milton, J. 1999. Lexical thickets and electronic gateways: Making text accessibly by novice writers. In C. N. Candlin & K. Hyland, (Eds.), *Writing: Texts, Processes and Practices*, (pp. 221-243). London: Longman.
Muncie, J. 2002. Process writing and vocabulary development: Comparing lexical frequency profiles across drafts. *System, 30*, pp. 225-235.
Nattinger, J. & DeCarrico, J. 1992. *Lexical Phrases and Language Teaching*. Oxford University Press: Oxford.
Ohtake, H., & Morren, B. 2001. A corpus study of lexical semantics in medical English. *Studia Humana et Naturalia, 35*, pp. 15-45.
Ohtake, H., & Morren, B. 2002. Bridging the gap between dictionaries and learners: From bilingual to monolingual dictionaries. *Studia Humana et Naturalia, 35,* pp. 1-30.
Partington, A. 1998. *Patterns and Meanings: Using Corpora for English Language Research and Teaching*. John Benjamins: Amsterdam.
Pawley, A., & Syder, F. H. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication*, (pp. 191-227). Longman: London.
Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. 1985. *A Comprrehensive Grammar of the English Language*. Longman: London/New York.
Rastall, P. 1994. The prepositional flux. *IRAL, 32(3)*, pp. 229-231.
Santos, T. 1988. Professors' reactions to the academic writing of nonnative-speaking students. *TESOL Quarterly, 22*, pp. 69-88.
Schmitt, N. 1997. Vocabulary learning strategies. In N. Schmitt & M. McCarthy (Eds.), Vocabulary: *Description, Acquisition, and Pedagogy*, (pp. 199-227). Cambridge: Cambridge University Press.
Sinclair, J. M. 1991. *Corpus, Concordance, Collocation*. Oxford University Press: Oxford.
Snellings, P., van Gelderen, A., & de Glopper, K. 2002. Lexical retrieval: An aspect of fluent second language production that can be enhanced. *Language Learning, 52(4)*, pp. 723-754.
Stubbs, M. 2001. *Words and Phrases: Corpus Studies of Lexical Semantics*. Blackwell.
Thurston, J., & Candlin, C. N.1998. Concordance and the teaching of academic *English. English for Specific Purposes, 17 (3)*, pp. 267-280.
Todd, R. 2001. Induction from self-selected concordances and self-correction. *System, 29*, pp. 91-102.
Willis, D. 1990. *The Lexical Syllabus*. Harper: New York; Collins: London.

Dictionaries Cited

COD = *Concise Oxford English Dictionary.* J. Persall (Ed.). London: Oxford University Press, 2002.
OALD = *Oxford Advanced Learner's Dictionary of Current English.* S. Hornby (Ed.). London: Oxford University Press, 2000.

Text Cited

Harrison's Principles of Internal Medicine, 14th CD Edition. A. Fauci, E. Braunwald, K. Isselbacher, J. Isselbacher, & J. Wilson (Eds.). New York: McGraw-Hill, 1998.

Appendix

Original passage from 'Are You Man Enough?' in *Time Magazine* April 24, 2000, Vol. 155, No. 16:

> But here it gets complicated. Does higher testosterone produce more aggressive behavior? Or does the more aggressive male – whose aggression was learned, say, at home or in school or in the neighborhood or on the team or in the culture at large – call for a release of testosterone from within himself for assistance?

Notes

1 Information about the Life Science Dictionary Project can be obtained at http://lsd.pharm.kyoto-u.ac.jp.

2 Information about MEDLINE can be obtained at http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed.