From Collocational/Textual Perspectives -- How could a Learners' Dictionary Help Learners in their Real Encoding Process?

Hiroaki Otani
PhD student in the University of Tokyo
4-7-19, Takaidohigashi, Suginamiku, Tokyo, Japan
E-mail: hiroaki811@mail.goo.ne.jp

Abstract

It is so-called 'transparent,' 'compositional' collocations that often become problematic for intermediate and advanced learners. As such collocations are not marked in any way grammatically or semantically, learners usually have little problem 'decoding' such strings, whereas learners sometimes face difficulties in 'encoding' such patterns in their own utterances. In this paper, I assigned a text-completion task to native speakers and learners to see what kind of collocational strings they access in an actual context-specific text-generating process and what kind of information in learners' dictionaries could really help learners involved in such a process. The results of this experiment revealed that whereas native speakers accessed several mostly 'transparent' but highly predictable collocational strings, learners had little access to such strings. It also turned out that some of such transparent but recurrent strings of words observed in this particular study were more or less paralleled in more general data (a large corpus). In the light of this experiment, learners' dictionaries need to do more to help learners to establish access to such collocations.

0 Background

A recent focus on phraseological elements of language has resulted in an increasing effort to describe multi-word units such as idioms and collocations in learners' dictionaries. However, the problem would be, as Howarth [1998: 186] points out, that 'teachers of EFL themselves, both native-speaker and non-native, have little understanding of the phraseological mechanisms of the language.' Here, we might be able to substitute 'teachers' for lexicographers. It is hardly surprising that it will be impossible for learners' dictionaries to accommodate real needs of learners without understanding this mechanism at work in actual language use. In this paper, I will look at what difficulties learners may face in an actual contextualized text-generating ('encoding') process, focusing on learners' accessibility to what is known as transparent, compositional but more or less restricted collocations. I will also analyse the result against a large corpus to see to what extent such collocations are predictable and how learners' dictionaries could extend a helping hand to learners in this regard.

1 Methodology

I gave both native speakers and EFL learners at Japanese senior high school the following task to complete the final part of a very short article[1]:

Read the following passage and complete the rest of the last paragraph:

The bath was invented before the bath plug. The bath plug could not have been invented before the bath, except as a small object with which to play ice hockey. The order in which inventions are made is very important, much more important than has ever been realized because we tend automatically to think that later inventions are better than earlier ones. A moment's thought will show this is not so. If, for example, a solution to today's urban traffic problems was proposed in the shape of a small man-powered device, a space rocket trying to tackle suburban problems, we would greet it as a great technological breakthrough. 'Bicycle makes car obsolete!' we would cry. Unfortunately, the bike came first, so we shall always unconsciously see it as a cruder version of the car

Other things which may have been invented too early are the airship, the radio, the railway train, the piano-roll player and the cuff-link

Consider also the zipper. [Please complete – around 50 words.]

This method has some advantages in that, by narrowing down the potential content, it will become easier to watch the ways in which native speakers and learners access their vocabulary when verbalizing a particular content. The point the original author is making is very simple: 'Later inventions do not necessarily supersede their earlier counterparts,' which is summarized in the first paragraph. The rest of the text is obviously devoted to various examples supporting this assertion. Therefore, when looking at the opening sentence ('Consider also the zipper') of the final paragraph, we could quite easily guess that this paragraph, too, may well be allocated for yet another exemplification, with the *zipper* on the one hand and something else on the other. A basic framework will be to claim supremacy of an older invention over its newer counterpart by highlighting the advantages of the former and the disadvantages of the latter. As it turned out, the pattern citing buttons as an example of an earlier, but superior, invention was the most frequent in both groups; 39% of native speakers and 62% of learners followed this path. The observation of the texts of this category will therefore enable us to consider the following three points: (1) What kind of words (co-occurring with the word *zip(per))* do our informants access to verbalize the disadvantages of the zipper? Do any intertextually recurrent, preferred collocations appear within each group, or across two groups? (2) Is such a tendency (if there is any) also paralleled in a large, general corpus? Or to what extent could this pattern appearing in this particular experiment be predictable from a corpus? (3) How could such information be included in learners' dictionaries?

2 Results

Table A-1, A-2 show how the recurrent (i.e. accessed by more than one native speakers) keywords are shared among different informants, complete with other 'unique' expressions each informant employed.

informant no.	break/broken	stick/stuck	jam/jammed	repair	replace	come undone	complex	others
n3		1						
n47	1				1			
n5	1			1	1			
n6	1	1			1			
n7	1				1			
n23	1							add to the cost of clothing
n24						1		awkward moments, catching fibres and threads in its 'teeth', it cannot be done
n8	1							attach, infliction of injury
n10		1	1			1		expensive, surprising
n48	1		1					
n13	1	1	1		1			inefficient
n14	1							force to work
n18			1					frustratingly closed, embarassingly open
n49	1							mend
n51					1		1	dangerous potentially disfiguring connotations
n53	1		1		1			expense, nuisance
n56	1		1				1	go wrong
n54								look too modern
n9								(only the advantages of buttons)
n19								(only the advantages of buttons)
n26								(only the advantages of buttons)
n55								(only the advantages of buttons)
total(/22/58)	12	4	6	1	7	2	2	

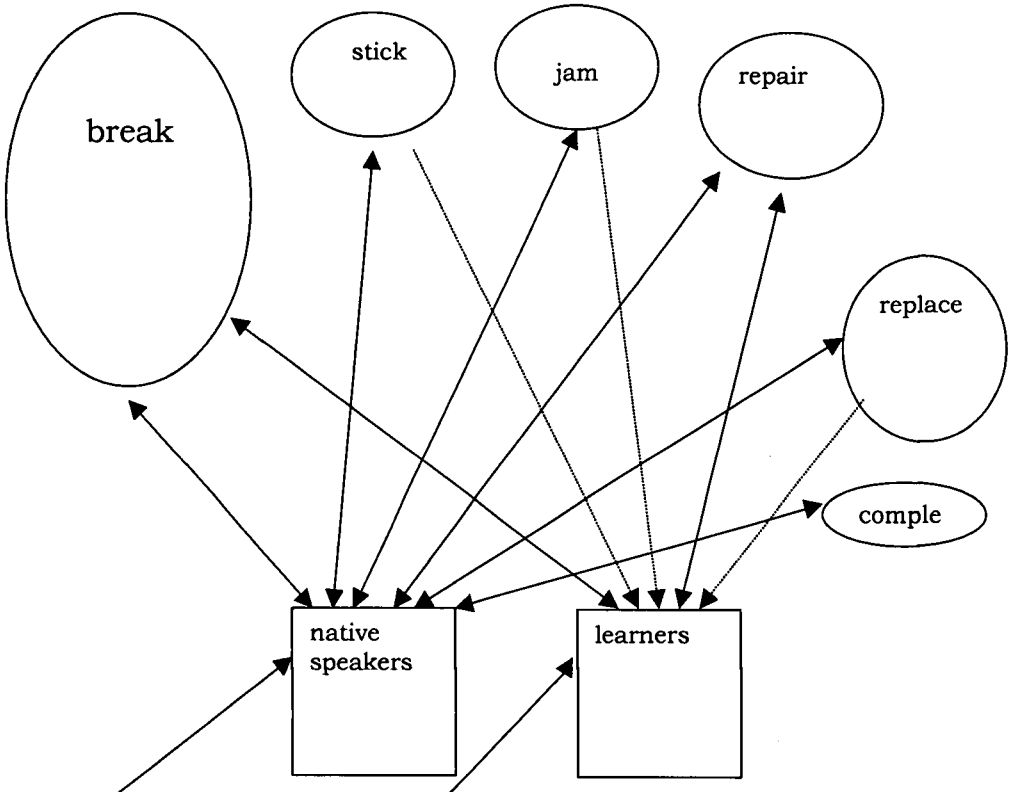
Table A-1: Native Speakers

informant no.	break/broken	stick/stuck	jam/jammed	repair	replace	come undone	complex	others
s5	1							change
s14	1			1				
s6	1			1				dangerous, be injured, made of metal
s21	1			1				bok worse, cost a bt
s17	1							recover
s19	1							
s28	1							
s42	1							cant withstand heavy powers
s30	1			1				must buy new one
s13								doesnt bok well
s7								costs more moeny, does not always move smoothly
s8								is almost inpossible to make at home, injure
s23	1							is difficult to wear or change the clothes
s18								coulbn t fit cottons and silk etc
s15								doesnt bok nice,
s3								isnt always good
s41								is not always better than
s25								is unconvenient(sic)
s26								(only the advantages of buttons)
s26								(only the advantages of buttons)
s16								(only the advantages of buttons)
s24								(only the advantages of buttons)
s10								(only the advantages of buttons)
s9								(only the advantages of buttons)
s20								(only the advantages of buttons)
s37								(only the advantages of buttons)
total (/26/42)	10	0	0	4	0	0	0	

Table A-2: Learners

	break/broken	stick/stuck	jam/jammed	repair	replace	come undone	complex
word (t-score MI-score) span:l-r=4	broken (1.41 0.00), broke (1.4 1 0.00)	stuck (1.73 10.81), sticks (0.9 9 0.00)	jammed (1.41 0.00)	repairing (1.00 0.00), repaired (1.00 0.00), repairs (0.99 0.00)	replace (0.99 0.00), replaced (0.99 0.00), replacing (1.73 15.94), replacement (0.99 0.00)	undone (1.00 0.00) (undo (2.81 16.80))	(0.99 0.00)

Table A-3: Corpus (Cobuild full-corpus)



add to the cost of clothing
awkward moments, catching fibres and threads in its 'teeth', it cannot be done up.
attach, infliction of injury
expensive, surprising
inefficient
frustratingly closed, embarrassingly open
etc.

dangerous, be injured, made of metal
look worse, cost a lot
must buy a new one
doesn't look nice,
dangerous, be injured, made of metal
etc.

Figure A (based on Table A-1, 2): Access to potential collocates of *zip(per)*

Looking at Table A-1, we notice that around half of the native speakers adopting this discourse pattern have accessed the lemma *break*. Besides, lemmas such as *stick, jam* and *replace* (in the context of 'the zipper is hard to *replace* once broken') have also been accessed by four to seven native speakers. Apparently, in discussing disadvantages of the zipper, native speakers access a couple of typical, somewhat shared set of keywords and their choice seems to show 'family resemblances' in Wittgenstein's [1953: 32] well-known term. Learners (see Table A-2 above), on the other hand, have equal access to the relatively frequent and versatile words (lemma), *break* and *repair,* whereas they have no access to other words typical in native speakers' writing. Apparently, while such words as *jam, stick, replace* might have been decodable at least to some learners (indicated by an one-way broken arrow in Figure A above), an interesting fact is that none of my EFL learners could 'encode' the patterns formed by these words in an actual text-producing process. In other words, although such combinations as *zipper* and *get stuck*, or *replace* and *zipper* can be seen as grammatically as well as semantically unmarked and 'transparent,' so many learners seem to bang their heads against this transparent language-learning 'glass-ceiling,' as it were. Ironically, it could be argued that the more grammatically and/or semantically marked ('opaque') a word combination is, the easier it will be for learners to learn it (see also Howarth 1998); first of all, it manifests itself as a stumbling block in a normal decoding process. Moreover, dictionaries, leaning/testing materials and teachers are always ready to help learners regarding such marked patterns. Our experiment seems to suggest, it is, if anything, more transparent word-combinations that really impede learners' encoding, sometimes causing them to get 'stuck' in their utterances. To circumvent these obstacles, some of the learners renounced detailed accounts altogether by saying *(be/look) good, better, worse,* while others resorted to somewhat circumlocutory expressions such as *(does)not always move smoothly*, or *doesn't work well.* We will come back to this point later.

Finally, let's look at to what extent native speakers' choice of keywords (collocates of *zip(per)*) in our particular experiment is paralleled and thus predictable in a large, general corpus. The corpus used here is the full corpus of Cobuild Direct (Cobuild, hereafter). At the bottom of Table A-3 above are the data concerning the strength of collocational link (t-score and MI-score) between each collocate and *zip(per)(s)*. From this, we can observe that such word-forms as *broken, stuck, jammed, replacing* have relatively high t-scores and *stuck* and *replacing* have high MI-scores as well. Such words as *repair, (be) complex* and the *(come) undone,* on the other hand, have relatively lower t-scores, indicating that they are less strongly (i.e. more freely) connected to *zip(per)* . (*Undo* (transitive active verb) has a strong link, but this word-form never appeared in our experiment.) Although it might be open to discussion that these t-scores are statistically 'significant,' it could be also argued that these combinations appear to be too 'significant' to ignore in our experiment; the accessibility of word-strings with this level of general frequency does matter in a real text-producing process. Finally in passing, let's look at how the original text goes:

Zips represent a technological advance on buttons, being faster and more complete. They are also more liable to come adrift , **break, jam**, malfunction, **stick** and catch. Buttons can only go wrong if the thread is faulty. Even then, buttons can be mended by the user. Zips rarely can.
(My emphases)

The author has tightly packed all pros and cons of both the zipper and buttons and strikingly,

he accesses three of our common keywords, namely *break, jam,* and *stick* .

3 Conclusion -- How can a Learners' Dictionary Help Learners?

Based on the findings above, then how could learners' dictionaries really help learners establish access to those items that were formerly denied to them? Despite that such communication strategies as avoidance or circumlocution [Tarone 1977] employed by our learners are highly effective in real-time communication, it would sometimes also be necessary to 'compensate' such compensative strategies in some way or other. Unless our learners can get an instant feedback from their interlocutors or teachers, they may well consult their dictionaries for this kind of encoding information. (The function of a learners' dictionary to help learners in such an encoding process is pointed out by Cowie [1981, 1989, 1998] and Carter [1987]). In this light, it would be intriguing to take a look at the entry of *zip(per)* in various monolingual/bilingual learners' dictionaries. What concerns us here is whether learners' dictionaries can provide any information that might help our learners to overcome their encoding difficulties. (See Table B below.)

Category	dictionaries(publisher,title,year)	entry	break/broken	stick/stuck	jam/jammed	repair	replace	come/be undone	complex
bilingual(English-Japanese) learners' dictionary	Taishukan: Unabridged Genius English-Japanese Dictionary; 2001	zip							
		zipper							
	Taishukan: Genius EJD. 1994	zip							
		zipper							
	Sanseido: The New Global EJD 2001	zip							
		zipper							
	Sanseido: The New Crown EJD	zip							
		zipper							
	[illegible] Lighthouse EJD 1995	zip							
		zipper							

Category	dictionaries(publisher,title,year)	entry	break/broken	stick/stuck	jam/jammed	repair	replace	come/be undone	complex
monolingual English learners' dictionaries	Oxford UP: Oxford Advanced Learner's Dictionary 6th ed. (OALD) 2000	zip		My zip is stuck.					
		zipper							
	Longman: Dictionary of Contemporary English. - 3rd ed. with new words supplement (LDCE) 2001	zip	The zip on my skirt had broken.					Your zip is undone at the back.	
		zipper							
	HarperCollins: Cobuild English Dictionary for Advanced Learners 3rd ed. 2001	zip							
		zipper							
	Longman: Language Activator 1993	zip							
		zipper							

Category	dictionaries(publisher,title,year)	entry	break/broken	stick/stuck	jam/jammed	repair	replace	come/be undone	complex
monolingual specialized collocation dictionary	John Benjamins: The BBI Combinatory Dictionary of English 1986	zip		a zip gets stuck					
		zipper							
bilingual(English-Japanese) specialized collocation dictionary	Kenkyusha: Dictionary of English Collocation on CD-ROM (KDEC) 1996	zip		The zip stuck./The zip got stuck./The zip is stuck open [shut].				The zip came undone.	
		zipper	The zipper on his sleeping bag was broken.					Your zip is open [undone].	

Table B: Dictionaries

Among major mono-/bi-lingual learners' dictionaries as well as some specialized collocation dictionaries I have consulted, OALD, LDCE, BBI and KDCE seem to have useful information to our learners. No wonder, however, none of them covers the whole range of native speakers' schema (zipper-dysfunctional schema, as it were) instantiated in our experiment (*a zipper sticks/jams/breaks, replace a (broken) zipper).* As regards the

presentation of such schema-forming collocates, OALD appears to be the most systematic: the collocational examples are apparently grouped together according to a coherent schema (*to do up/undo/open/close a zip*). As Cowie [1986] proposes that collocates of related meanings be put together within each grammatical class under superordinate terms, this method of presentation will not merely help learners imagine overall lexical/semantic 'profile' of a word [Stubbs 2001] but also serve as a learners' encoding resource, enabling them to establish quick access to (even transparent) strings underlying each word. Since it is only at this moment that an individual, isolated word begins to appear as an integral part of the entire linguistic network and thus becomes really available to learners in real communication, this kind of knowledge will be no less important than that of grammatical parts-of-speech. (We may even call this information semantic/lexical 'parts-of-speech'.) Of course, there is a great deal of discussion as to the role and efficacy of examples in a dictionary (see [Nesi 2000], [Cowie 1989], for example), and it will require more research to see whether or how learners can (or can be trained to) successfully take advantage of such information in a dictionary in their encoding process. In any event, it could be said that whereas many learners will be able to encode, without difficulty, a past-tense suffix of a 'transparent' regular verb, they cannot necessarily encode 'transparent' collocations without necessary information. It will be worth considering, therefore, what kind of information in a dictionary can really help learners and how it should be presented, especially now that various electronic formats (CD-ROM, Web) are available which can contain vast amount of data in a three-dimensional way.

Endnote

1. The informants and conditions under which they completed the text are as follows:
58 Native speakers (British, Australian, American): average age: 35.6 They did the task by answering my questionnaire by mail or e-mail.
42 EFL learners: 18 year-old high school students in Tokyo who have studied English for five years (about five hours a week) in an EFL setting. They did their task, without any use of referential materials, as part of their regular 50-minute test.

References

[Carter, R. 1987] *Vocabulary.* London: Allen&Unwin

[Cowie, A.P. 1981] 'The treatment of Collocations and Idioms in Learners' Dictionaries' *Applied Linguistics.* 2/3

[Cowie, A.P. 1986] 'Collocational Dictionaries – a Comparative View' in Murphy, M. D. (ed.) *Proceedings of the Fourth Joint Anglo-Soviet Seminar on English Studies.* London: The British Council

[Cowie, A.P. 1989] 'The language of examples in learners' dictionaries.' in James, G. (ed.) Lexicographers and Their Works. Exeter Linguistic Studies vol.14 Exeter: Exeter University Press

[Cowie, A.P. 1998] 'Phraseological Dictionaries: Some East-West Comparisons' in Cowie, A.P. (ed.) *Phraseology Theory, Analysis, and applications.* Oxford: Clarendon Press

[Howarth, P. 1998] 'The Phraseology of Learners' Academic Writing' in Cowie, A.P. (ed.) *Phraseology Theory, Analysis, and Applications.* Oxford: Clarendon Press

[Nesi, H. 2000] *The Use and Abuse of EFL Dictionaries.* Tubingen: Max Niemeyer

[Stubbs, M. 2001] *Words and Phrases.* Oxford: Blackwell

[Tarone, E. 1977] 'Conscious communication strategies in interlanguage.' in H. D. Brown *et al.* (ads.) *On TESOL. 1977* Washington, D.C.: TESOL

[Wittgenstein, L. 1953] *Philosophical Investigations/* (translated by G.E.M. Anscombe) Oxford: Blackwell