

Word-Focused Extensive Reading with Guidance

Chin-Chuan Cheng (鄭錦全)

Academia Sinica

chengcc@gate.sinica.edu.tw

A few years ago a reading method called “word-focused extensive reading” was proposed for adults learning the fine details of word usage (Cheng 1998a, 1998b, 1998c). In this mode of reading the computer retrieved from a large corpus and displayed the word in question in a sentence. In reading similar sentences the reader would hopefully gain the knowledge of the word usage. That is the idea of word-focused extensive reading. However, not every reader will easily see the significance of word collocation. An automated guidance can be added to this reading program. The guidance could take the form of pre-processed statistics of co-occurring words in various categories such as subject, object, preposition, modifier, etc. (Kilgarriff and Tugwell 2001). But in this way the reader does not participate in the construction of the lexical knowledge. With the guidance added to the program when a sentence is shown the computer tabulates the collocating words, displays them in relevant locations, and waits for the user to press a key for the next sentence. The significant collocating words of the next sentence are then added to the list. In this way the reader sees the step by step build up of collocating words and thus participates in the building of the knowledge of word usage.

Word-Focused Extensive Reading for Fluency

An adult native speaker of a language had at least a dozen years of social interaction and reading to gain language fluency. An adult learner of a non-native language, on the other hand, does not have a score of years to study to reach a native-like competence. However, extensive reading is required to learn a language well. Therefore a method for accelerating learning is needed for adult learners. A few years ago I proposed a method called “word-focused extensive reading” (針對一詞廣泛閱讀) to help learners in word learning (Cheng 1998a, 1998b, 1998c). In this mode of learning using the word-usage software the user typed in a word in question to see how it was used in texts. The sentences with the word were shown on the screen one at a time. The texts included 201 great English books of the past. The user could add other texts for search. The package was called EWUSAGE.

As many sentences containing the word in question were displayed, the user was supposed to read and understand how the word was used, especially how it collocated with others. Instead of spending years to read all the 201 books to gain the knowledge of word usage, now the user focused on a specific word to read its usage in the books. Thus the extensive reading had the focus of a particular word. That was why the method was called word-focused extensive reading. The program still works in the current Windows environment. For example, using the British National Corpus (BNC) (<http://www.natcorp.ox.ac.uk>), it showed the following sentences in response to the input word ‘dawn’ one at a time for the user to read:

I decided to abandon my earlier plan of just walking around until dawn and instead to try and find somewhere where I could get some sleep.

Waves of sleep then started to wash over me and I dozed fitfully until dawn.
An idea suddenly dawned on me.

I became conscious again at about dawn.

Lucy filled her glass with the palest gold, firelight making a star in the bowl, the label on the wine was olive-green and gold and Lucy's eyes were spring leaves dancing in the light of dawn.

The rose garden at dawn.

Glossy surfers' hair, eyes blue as a robin's egg and a dreamy far-off ocean at dawn.

Did the program really help the user? It depended on the ability of the user to make a synthesis of the data. As we look at the examples above, we are aware that the preposition before 'dawn' is 'at' and 'until'. These words appeared twice there. As we continue to read the sentences so displayed we see a very high frequent combination of 'at dawn'. That will make us understand that 'in dawn' is not a grammatical phrase. However, some users might just read on without such an awareness of the collocation. In order to heighten awareness of the forces of numbers in fluency and grammaticality, recently I rewrote the usage program to incorporate usage guidance. It is called CCWUsage.

Cultivating Awareness of Quantitative Information

The CCWUsage program can be used for Chinese and English texts. It uses a database for English collocation and another for Chinese word collocation. The English collocation database was established using a portion of the BNC corpus. The Chinese database was built on the basis of Balanced Corpus of Academia Sinica and other texts. In this writing we will discuss only English word usage. The program now provides guidance to help the user gain the lexical knowledge of the word in question. At this moment the guidance displays the word in question and the high-frequency words one, two, and three words before and after it. For example, when the user types in the word 'dawn', it displays the list of words with frequency of occurrence given in parentheses as in Table 1 initially for the user to have an overall picture of the collocation. The information given in Table 1 shows the words with higher frequency of occurrence. There are many other words that co-occur with 'dawn', but we feel that listing of many words in the summary will defeat the purpose of assisting the user to see the salient features of the collocation. In building the collocation database we went through two processes. First a computer program collected the words above certain frequency relative to the frequency of the key word. Then we examined the collocation data for each word to determine the linguistic significance of the collocation. For example, numbers occurring around a key word might be eliminated from the automated compilation. Another example of treatment by hand was the addition of lower frequency but significant words. For example, in Table 1 the verbs 'came' and 'broke' were added so that the user could understand what verbs to use with 'dawn', even though their frequency was not higher than 20.

Table 1. Collocating Words for 'dawn'.

3	2	1	Word	1	2	3
the (85)	a (84)	the (255)	dawn 1234	of (119)	the (93)	the (92)
of (27)	the (60)	at (218)		and (72)	dusk (42)	a (24)
and (26)	just (44)	before (116)		Before (70)	a (38)	
in (25)	crack (30)	of (91)		on (48)	and (30)	
was (23)	of (29)	after (40)		to (47)		
to (22)	in (25)	a (39)		was (25)		
a (20)	light (21)	from (27)		the (21)		

	up (20)	by (25)		to (21)		
	and (20)	as (22)				
	at (20)	until (20)		<VBD>was (25)		
				<VVD>came (15)		
				<VVD>broke (14)		

While the guidance program shows as a summary the three words before and after the key word, during the reading process it tabulates only the word before and after the key word for clarity. The guidance does not tabulate the words not in the summary list. In this way the more salient features are captured for synthesis. For example, the sentences given above in actual reading process were given with tabulation of the cumulating frequency.

I decided to abandon my earlier plan of just walking around until dawn and instead to try and find somewhere where I could get some sleep.

until 1 dawn 1 and 1

Waves of sleep then started to wash over me and I dozed fitfully until dawn.

until 2 dawn 2 and 1

An idea suddenly dawned on me.

until 2 dawn 3 and 1
on 1

I became conscious again at about dawn.

until 2 dawn 4 and 1
on 1

Lucy filled her glass with the palest gold, firelight making a star in the bowl, the label on the wine was olive-green and gold and Lucy's eyes were spring leaves dancing in the light of dawn.

until 2 dawn 5 and 1
of 1 on 1

The rose garden at dawn.

until 2 dawn 6 and 1
of 1 on 1
at 1

Glossy surfers' hair, eyes blue as a robin's egg and a dreamy far-off ocean at dawn.

until 2 dawn 7 and 1
of 1 on 1
at 2

As the reading continues, the word 'at' will appear more frequently used. Eventually the user is led to gain the knowledge that 'at dawn' is the right way to express time. A curious user will want to know the difference between 'dawn' and 'morning'. The summary of 'morning' is given in Table 2.

Table 2 shows clearly that 'at' does not appear immediately before 'morning'. That is, 'at morning' is not normally used. Thus the contrast 'at dawn' and 'in the morning' is clear. The guidance can be very useful for differentiating near synonyms.

Kilgarriff and Tugwell (2001) propose to build lexical knowledge as "word sketch". They use BNC tagged corpus to determine grammatical relations in word collocation. For example, they give the word sketch for 'bank' as follows. The numbers are the absolute frequency of

occurrence of the words and their salience. Silence is a relative frequency calculated from the occurrence relations of words. The relations are “subject of”, “object of”, “modifier”, etc. Here the frequency and salience numbers are presented for “subject of” only. In the other categories they are omitted here for ease of reading.

Subject of -- lend (95 21.2), issue (60 11.8), charge (29 9.5), operate (45 8.9), step (15 7.7), deposit (10 7.6), borrow (12 7.6), eavesdrop (4 7.5), finance (13 7.2), underwrite (6 7.2), ...

Object of -- burst, rob, overflow, line, privatize, defraud, climb, break, oblige, ...

Modifier -- Central, Swiss, commercial, grassy, royal, far, steep, issuing, confirming, ...

The purpose of word sketch is to serve as the basis for compilation of lexical entries for a dictionary. It is not explicitly stated as guidance for learning.

Table 2. Collocating Words for ‘morning’

3	2	1	Word	1	2	3
a (665)	in (3476)	the (5105)	morning (18218)	and (1132)	the (913)	the (622)
the (558)	the (1720)	this (3823)	as a noun	of (422)	and (536)	and (293)
on (500)	on (938)	next (1134)	I (351)	I (334)	a (316)	<VBD> was (279)
o'clock (405)	of (564)	that (656)		<VVD>came (27)	<VBD>was (255)	a (262)
in (264)		early (476)				
and (257)		Sunday (453)				
up (252)		Saturday (407)				
		every (401)				
		good (400)				
		tomorrow (369)				
		following (365)				
		Monday (352)				
		yesterday (326)				

Our program of word usage with guidance does not give grammatical relations. The grammatical relations Kilgarriff and Tugwell (2001) established automatically from the tagged corpus are not without errors. In our guidance errors should not be introduced to the learner. Moreover, not all texts are tagged. In our program the user can add any texts to use. It does not require texts that are tagged. We hope that our guidance will heighten awareness of quantitative information in word collocation and lead the user to language fluency.

Naturally, it is not always easy to different word senses and usage from collocation alone. The guidance cannot take the place of dictionaries. Dictionaries are still needed to find the meanings of a word. Following is a challenge to differentiate ‘stir’ and ‘stimulate’.

Challenges

Nowadays MsWord is ubiquitous. Its synonym function helps many users in diction. In some cases it is justified to oversimplify the matter. For example, it gives the following synonyms for the word 'stimulated':

enthused, inspired, stirred, moved, encouraged, motivated

A good writer in search of hints of words can use the synonym function for word variation. However for a somewhat unsophisticated reader, as 'stirred' is a synonym of 'stimulated', it might be taken as correct to say:

Mrs. Jones stimulated the soup.

In fact the sentence was what Miller and Gildea (1991) give to illustrate the lack of fine distinction among synonyms by school children. The word 'stimulated' in this context should be replaced by 'stirred'. How does a collocation database help in this case? Before we actually built the collocation summary, we actually examined collocation words occurring more than once in the BNC corpus. Since the matter is soup after the verb. We examined the words after the 'stimulate' and 'stir'.

The words occurring immediately after 'stimulate' with numbers in parentheses indicating frequency were the following:

the (279), by (217), a (69), and (52), interest (29), to (29), an (18), discussion (17), in (15), their (15), growth (13), with (13), more (12), gastric (11), new (11), your (11), demand (10), economic (10), acid (9), debate (9), further (9), her (9), his (9), competition (8), investment (8), them (8), ideas (7), local (7), employment (6), gastrin (6), it (6), or (6), sales (6), some (6), consumption (4), industrial (4), my (4), other (4), such (4), thought (4), transcription (4), aggregate (3), circulation (3), him (3), industry (3), innovation (3), into (3), its (3), not (3), PLC (3), rather than (3), speculation (3), action (2), active (2), aggressive (2), aminopyrine (2), but (2), calcium (2), cells (2), children's (2), clarity (2), DNA (2), either (2), employer (2), energy (2), exports (2), females (2), first (2), home (2), immunity (2), improvements (2), in part (2), in vivo (2), insulin (2), jobs (2), limb (2), management (2), migration (2), muscle (2), our (2), pancreatic (2), partly (2), perhaps (2), PLC- (2), private (2), production (2), proliferation (2), public (2), research (2), technical (2), this (2), those (2), through (2), you (2)

The second words occurring after 'stimulate' were the following:

the (81), and (51), in (38), to (31), of (29), interest (26), growth (30), a (19), development (17), economy (17), by (15), production (13), secretion (13), activity (11), acid (10), for (10), release (9), circulation (8), demand (8), an (6), on (6), this (6), body (5), curiosity (5), proliferation (5), vortex (5), about (4), appetite (4), awareness (4), great (4), his (4), within (4), arachidonic (3), body<POS 's (3), but (3), calcium (3), cells (3), discussion (3), economic (3), flow (3), free (3), from (3), imagination (3), investment (3), more (3), or (3), produce (3), search (3), sense (3), study (3), support (3), synthesis (3), their (3), these (3), thrombin (3), use (3), work (3), British (2), CAT (2), G (2), absorption (2), accumulation (2), already (2), among (2), artistic (2), at (2), between (2), both (2), build (2), cell (2), changes (2), children (2), contact (2), creation (2), develop (2), emotional (2), encourage (2), erotic (2), expansion (2), fiscal (2), forces (2), fresh (2), greater (2), healing (2), hearing (2), housing (2), ideas (2),

industry (2), into (2), it (2), learning (2), lower (2), market (2), mind (2), minds (2), motivate (2), national (2), olfactory (2), only (2), opportunities (2), other (2), own (2), phospholipase (2), reaction (2), research (2), reverse (2), sexy (2), sufficiently (2), thought (2), through (2), traffic (2), wider (2), with (2)

The third words occurring after 'stimulate' were the following:

of (134), and (81), the (67), in (54), to (40), by (28), a (21), for (16), activity (10), secretion (10), interest (9), cells (7), economy (7), or (7), with (7), mixing (5), about (4), from (4), growth (4), ideas (4), it (4), literature (4), presence (4), these (4), this (4), Economic (3), acid (3), addition (3), an (3), are (3), as well as (3), but (3), deal (3), debate (3), development (3), discussion (3), encourage (3), ionophore (3), own (3), publication (3), response (3), students (3), that (3), transcription (3), activities (2), against (2), as (2), at (2), business (2), create (2), demands (2), dirty (2), engineering (2), enquiry (2), even (2), extent (2), into (2), itself (2), large (2), market (2), more (2), movement (2), need (2), new (2), output (2), photographs (2), political (2), produce (2), protein (2), provision (2), range (2), rapid (2), research (2), studies (2), thought (2), through (2), way (2), which (2), will (2), write (2)

It is a challenge to find the contrast or difference in the words occurring after 'stir'. The first words occurring after 'stir' were the following:

up (202), the (97), in (67), and (52), by (36), it (33), her (24), a (23), into (17), restlessly (15), to (13), his (12), them (12), himself (11), from (9), at (8), him (8), things (8), with (8), myself (6), for (5), herself (5), my (5), something (5), against (4), inside (4), again (3), as (3), but (3), out of (3), sugar (3), themselves (3), uneasily (3), when (3), your (3), about (2), all (2), gently (2), like (2), memories (2), not (2), once more (2), out (2), slightly (2), their (2), this (2), until (2), us (2), was (2), yourself (2)

The second words occurring after 'stir' were the following:

the (78), a (42), and (32), his (24), her (23), by (21), to (18), racial (17), trouble (17), up (16), in (11), , he (8), thereby (7), coffee (6), their (6), emotions (5), fire (5), into (5), of (5), tea (5), action (4), blood (4), him (4), life (4), memory (4), with (4), an (3), as (3), at (3), contents (3), hair (3), it (3), leaves (3), my (3), opened (3), senses (3), some (3), soup (3), your (3), Gloucester (2), I (2), Steve (2), against (2), air (2), all (2), another (2), any (2), ashes (2), at last (2), controversy (2), curtains (2), darker (2), embers (2), first (2), flinging (2), for (2), gently (2), groaning (2), hatred (2), imagination (2), its (2), made (2), masses (2), me (2), new (2), on (2), our (2), out (2), renewed (2), shaken (2), so (2), stretched (2), this (2), turned (2), until (2), vigorously (2), waters (2), when (2)

The third words occurring after 'stir' were the following:

and (39), the (32), of (27), a (18), hatred (16), in (16), his (13), her (11), against (7), by (6), he (6), into (6), for (5), hornet's (7), sleep (5), up (5), with (5), animosity (4), from (4), it (4), on (4), to (4), an (3), arms (3), as (3), before (3), him (3), memories (3), a bit (2), action (2), again (2), aggregometer (2), between (2), body (2), but (2), chair (2), flour (2), gentle (2), grey (2), house (2), life (2), made (2), may (2), memory (2), mixture (2), or (2), others (2), out (2), said (2), she (2), spot (2), that (2), they (2), thought (2), was (2), way (2)

With the large amount of words occurring after ‘stimulate’ and ‘stir’, it is not easy to help the reader understand the difference between the two words. Since the matter is soup, we will look at substantive words and eliminate functional words to see if a clearer picture will emerge. The substantive words occurring immediately after ‘stimulate’ were the following:

interest (29), discussion (17), growth (13), gastric (11), demand (10), economic (10), acid (9), debate (9), competition (8), investment (8), ideas (7), employment (6), gastrin (6), sales (6), consumption (4), industrial (4), thought (4), transcription (4), aggregate (3), circulation (3), industry (3), innovation (3), PLC (3), rather than (3), speculation (3), action (2), aminopyrine (2), calcium (2), cells (2), clarity (2), DNA (2), employer (2), energy (2), exports (2), females (2), home (2), immunity (2), improvements (2), insulin (2), jobs (2), limb (2), management (2), migration (2), muscle (2), PLC- (2), production (2), proliferation (2), research (2)

The first substantive word occurring after ‘stir’ were the following:

something (5), sugar (3), memories (2)

The second substantive words occurring after ‘stimulate’ were the following:

interest (26), growth (30), development (17), economy (17), production (13), secretion (13), activity (11), acid (10), release (9), circulation (8), demand (8), body (5), curiosity (5), proliferation (5), vortex (5), appetite (4), awareness (4), arachidonic (3), calcium (3), cells (3), discussion (3), flow (3), imagination (3), investment (3), produce (3), search (3), sense (3), study (3), support (3), synthesis (3), thrombin (3), use (3), work (3), absorption (2), accumulation (2), build (2), cell (2), changes (2), children (2), contact (2), creation (2), develop (2), expansion (2), forces (2), healing (2), hearing (2), housing (2), ideas (2), industry (2), learning (2), market (2), mind (2), minds (2), olfactory (2), opportunities (2), phospholipase (2), reaction (2), research (2), thought (2)

The second substantive words occurring after ‘stir’ were the following:

trouble (17), coffee (6), emotions (5), fire (5), tea (5), action (4), blood (4), life (4), memory (4), contents (3), hair (3), leaves (3), senses (3), soup (3), Gloucester (2), Steve (2), air (2), ashes (2), controversy (2), curtains (2), embers (2), groaning (2), hatred (2), imagination (2), masses (2), waters (2)

The third substantive words occurring after ‘stimulate’ were the following:

activity (10), secretion (10), interest (9), cells (7), economy (7), growth (4), ideas (4), literature (4), presence (4), acid (3), addition (3), deal (3), debate (3), development (3), discussion (3), encourage (3), ionophore (3), publication (3), response (3), students (3), transcription (3), activities (2), business (2), demands (2), engineering (2), enquiry (2), extent (2), market (2), movement (2), need (2), output (2), photographs (2), produce (2), protein (2), provision (2), range (2), research (2), studies (2), thought (2), way (2)

The third substantive words occurring after ‘stir’ were the following:

hatred (16), sleep (5), animosity (4), arms (3), memories (3), action (2), aggregometer (2), body (2), chair (2), flour (2), house (2), life (2), memory (2), mixture (2), spot (2), thought (2), way (2)

The word 'stimulate' could go with 'ideas', and 'stir' could combine with 'memory'. One could be lead to conclude that both words can be used to agitate mental matters. And this conclusion might lead to the "stimulation" of soup.

The guidance provides a final resort. If in doubt, the user can type in the entire phrase 'stir soup' and 'stimulate soup' one at a time to see their occurrences. It allows the user to search the phrase to see if it has been used before. For example, the search of 'stir soup' found 13 sentences in the BNC corpus. Some of them are given below:

Stephan hummed a little tune and stirred his soup.
But mother uses the spoon to stir the soup.
Hence the scullions having to use small stools to stand on when they wanted to do anything like wash up dishes, stir soups or work controls.

With the search string 'stir soup' the guidance program appropriately found 'stirred his soup', 'stir the soup', 'stir soups', and others with any inflectional form of 'stir' preceding 'soup' or 'soups'. On the other hand, the search for the phrase 'stimulate soup' did not find any instances occurring in the corpus. The extensive reading and search should make the user aware that one can stir but not stimulate soup.

It is not our intention to make the usage guidance program the only resource for building word knowledge. Dictionaries can serve well in giving the senses of words. For example, Liang (1996) *Far East English-Chinese Dictionary* gives the Chinese equivalents of 'stir' as 移動, 攪和, and 激動 and 'stimulate' as 刺激, 激勵, and 增進. These Chinese words make it clear that the two words are not identical synonyms.

Hornby (2000) *Oxford Advanced Learner's Dictionary* gives "mix", "move", "feel", and "cause trouble" as the sense categories for the word 'stir'. The senses for 'stimulate' are "to make something develop or become more active", "to make somebody interested and excited about something", and "to make a part of the body function". Sinclair (2001) *Collins Cobuild English Dictionary for Advanced Learners* gives 'stir' the senses related to the action of "mix", "move", "move to react", "think or feel", and "cause excitement". On the other hand to 'stimulate' is "encourage to begin or develop", "to feel full of ideas and enthusiasm", and "to cause to move". These dictionaries provide good resources for word learning.

Conclusion

Instead of asking our students to spend many years to read many books to understand word usage, we propose to use the computer to search and display word collocation of specific words. The word usage program provides guidance to focus on word occurrence frequency so that the user can make proper syntheses of word collocation. The reader sees the step by step build up of collocating words and thus participates in the building of the knowledge of word usage. The compilation of the database for guidance is a fairly significant and challenging task for establishing word knowledge.

REFERENCES

- Cheng, Chin-Chuan. 1998a. "一詞泛讀：英文詞語用法檢索軟體(Focused reading: English word usage retrieval program)". In 戴維揚 ed. 超倍速英語學習年代 S1-S11. Taipei: Crane Publishing House.
- Cheng, Chin-Chuan. 1998b. "針對一詞廣泛閱讀：電腦輔助的詞語學習 (Extensive reading for individual words: computer-assisted word learning)". 華文世界 (The World of Chinese Language) 87:30-44.

Chin-Chuan Cheng (鄭錦全)

- Cheng, Chin-Chuan. 1998c. *英語用法寶典 (English Word Usage)*. Taipei: Crane Publishing House.
- Hornby, A. S. 2000. *Oxford Advanced Learner's Dictionary*. Oxford: Oxford University Press.
- Kilgarriff, Adam and David Tugwell. 2001. *Proceedings of Collocations Workshop* 32-38. ACL 2001, Toulouse, France.
- Liang, Shiqiu 梁實秋. 1996. *Far East English-Chinese Dictionary*. Taipei: Far East Book Co.
- Miller, George A. and Patricia M. Gilda. 1991. "How Children Learn Words". In William S-Y. Wang (ed.) *The Emergence of Language Development and Evolution* 150-158. New York: W. H. Freeman.
- Sinclair, John. 2001. *Collins Cobuild English Dictionary for Advanced Learners*. Glasgow: HarperCollins Publishers.