



## **A Study of Collocation Behaviors on Lexical Pragmatics**

**Ching-Ying Lee**

*National Taiwan Normal University & Kang Ning Junior College, Taiwan*

### **Bio Data**

Ching-Ying Lee currently is a doctoral student in the Department of English, National Taiwan Normal University. She is also an English lecturer at a vocational college of medical care and management. Most of her research interests are central to computer-assisted language learning and corpus-based approaches to EFL learning, in particular concerning lexical choices and writing. She has published many papers on Call-based EFL learning as well as corpus-assisted English writing.

### **Abstract**

Lexical pragmatics explores the meaning conveyed by a word underdetermined by its semantics, and the process of bridging the encoded and communicated meaning of words. Due to complication of word interpretations and pragmatic felicities, lexical pragmatics is considerably sophisticated. In other words, it examines “the processes by which linguistically specified word meanings are modified in use” (Blutner, 1998, p.115; Wilson & Carston, 2007, p.1560). The interaction between lexicon and pragmatics is often overlooked by most L2 learners because they rely on dictionaries and thesauri to provide denotational meaning of a lexical item which lack ample collocational information. The purpose of the study is to describe how collocational behaviors of near synonyms can be recognized from contextual usage data in corpora to improve L2 lexical learning. It also emphasizes that pragmatic misuse of near synonyms in the L2 context calls for more attention and treatment. The present study used computational techniques to identify target words which typically co-occur with a lexical item under investigation. Further, it generally expanded with software that elicited usage data within a corpus and matched it with statistical information for the purpose of analysis. The findings suggested that it could be beneficial for L2 learners to observe the collocates of near synonyms so that recognition of pragmatic characteristics help improve their lexical usage.

**Key Words:** collocational behaviors, corpus analysis, lexical pragmatics, near-synonym

## Introduction

To most native speakers, variability is inherent in their language production of native speakers and they can perform different language uses such as in lexical choice and speech acts. In general, they maintain the equilibrium of syntax, semantics and pragmatics. However, it is relatively difficult for L2 learners to make appropriate adaptations and to juggle with these sophisticated fields of language use. In particular when L2 learners do not develop pragmatic awareness, their pragmatic infelicities often cause misunderstanding. To attain appropriate communication, L2 learners need to develop the awareness of linguistic variability, flexibility and the interaction between lexicon and its pragmatic functions. The problems of appropriate lexical choice and near-synonym distinction are especially daunting for second language learners (Mackay, 1980, p.17). The majority of lexical misuses made by advanced language learners reflect learners' confusion among similar lexical items in the second language. Even for native speakers, it is extremely difficult to discriminate the near-synonyms that dictionaries of synonyms are compiled to reveal and describe the nuances. For instance, *Merriam Webster's New Dictionary of Synonyms* provides approximately 2500 entries of English synonymous words. *Choose the Right Word* (Hayakawa, 1994) contains approximately 900 clusters of more than 5000 near-synonyms of English. The language of explanations in dictionaries is somewhat arcane for second language learners such that it becomes limited in accessibility and practical usefulness in the L2 context.

Some recent studies on lexical pragmatics (Stubb, 2001, pp.153-154; Dam-Jensen & Zethsen, 2007, p.1608) have shown that acquiring the evaluative pragmatic meanings of building blocks and lexemes helped appropriate lexical usage. The interaction between lexicon and pragmatics here refers to the context-dependent lexical meaning which is constructed in an interpretive process by language users. In brief, lexical pragmatics is to examine "the processes by which linguistically specified word meanings are modified in use" (Blutner, 1998, p.115; Wilson & Carston, 2007, p.1560). The purpose of the study is to describe how pragmatic meanings in near synonyms can be derived from contextual usage data in corpora. It also emphasizes that pragmatic misuse of near synonyms calls for more attention and treatment. Typically, near synonyms operate in different contexts and are not collocationally interchangeable (Tognini-Bonelli, 2001, p.34). For example, while *announce* and *express* are considered as synonyms, the collocates of the two words cannot be mutually changed in the

following contexts – “Mr. Smith *announced* his candidacy for governor.” and “Mr. Johnson *expressed* some concerns about the plan.” It is helpful to compare collocates of synonyms in context so that pragmatic characteristics can be more concise and comprehensible for L2 learning.

### **Literature Review**

According to Levinson (1983, p. 226), it is of great significance to notice who said what and in what situation in dynamic language behaviors. This is echoed by Barron’s (2003, p.10) notion that “knowledge of the appropriate contextual use of the particular languages’ resources” is a crucial part of pragmatic competence. L2 learners often face difficulties in understanding subtle and elusive nuances of appropriateness (Dewaele, 2008, p.248). They thus need both linguistic and pragmatic competence to perform grammatical utterances and sociocultural norms for appropriate expressions. In this respect, appropriateness becomes one of the primary concerns. In addition, recent research (Simon-Vandenberg, 2008, p.1533) indicates that core lexical competence is characterized by appropriateness of word choices. As such, research of lexical pragmatics is of significance to provide a systematic and descriptive interpretation of pragmatic phenomena concerning the semantic underspecification of words. Blutner (2004) elaborates, suggesting that ‘lexical pragmatics is a particular account of the division of labor between lexical semantics and pragmatics. It connects the notion of semantic underspecification in lexicon with a theory of pragmatic strengthening (i.e. conversational implicature)’ (p.3). The mechanisms have been undergone in a bidirectional type of optimality theory that combines expressive and interpretive optimization (Blutner, 2004, p.21).

Lexical pragmatics can be drawn through a general mechanism of conceptual interpretation. It is grounded on two simple principles: (1) “An adequate representation of lexical items has to be given in a semantically underspecified format, and (2) their actual interpretation has to be derived by a pragmatic strengthening mechanism” (Blutner, 2004, p.17). Such a basic pragmatic mechanism allows a clear explication of notions such as conversational implicature. The concept of Gricean implicature can be used to account for the divergences between meaning and natural language interpretation of lexical units. This also provides the rationale to analyze the words with conceptual and contextual settings and to go beyond the aspects of meaning investigated by a collocational analysis of lexemes in

pragmatics.

Recent studies on collocation (Yeh, et al., 2007, pp.131-135; Chang, et al., 2008, pp. 283-289) has exploited concordance and collocation for advising L2learners in lexical choice. The collocational information is interpreted through the proximity of a consistent series of collocates. Its main function is to convey speaker or writer's attitude or evaluation (Louw, 2000, p.49). Many lexical items carrying pragmatic meanings are often negative, with a relatively small portion of them bearing affectively positive meanings (Channell, 2000, p.38). In this regard, Blutner (2004, p.25) shows how the bidirectional optimality theory accounts for the effects of negative strengthening. They construe the possible states of happiness. For instance, there are still gradual nuances between the two sentences, "*I'm not happy.*" and "*I'm unhappy.*" The latter indicates a stronger negative strengthening. As for pedagogical research, Martin (1984, p.131) discussed instructional approaches to synonym teaching and suggested the importance of providing common collocates to students.

## **Methodology**

With concordancing and computational techniques, over decades empirical studies have attested that pragmatic phenomena can be revealed by analyzing collocates in corpora. Corpora have generally expanded with computational techniques and software to elicit usage data and to match them with statistical information for the purpose of analysis. Recent development in concordancing tools has advanced to web-based systems that provide online access to query and results. Research issues concerning lexical behavior, collocational pattern, syntax, semantics, and pragmatics can all be facilitated by the language data access capability and the statistical summarization functions of these state-of-the-art concordancing tools.

For pragmatic grounding, the present study used *British National Corpus (BNC)* (Burnard, 1995) and *Corpus of Contemporary American English (COCA)* (Davies, 2008) to explore the potential of lexical collocation information. The procedures included collocational patterning, listing, sorting, counting, etc., conducted to search and categorize information as well as displaying the relations between the target word and adjacent words in contexts. In this study collocates were grouped by POS categories and graded by association strength with target words. There were a number of statistical tests used to measure collocate association strength, for example, the MI (mutual information), z, t, log-likelihood scores. The statistical

measure chosen to gauge association strength was the mutual information (MI) measure (Church & Hanks, 1990). The profile was compiled to describe three transitive verbs (*cause*, *promote* and *commit*) and was then analyzed to observe their collocational behaviors and pragmatic implications.

### **Analysis and Discussion**

The focus of the data analysis is on corpus evidence for the existence of the inherent evaluative meaning and pragmatic evaluative meaning. Moreover, it attempted to incorporate the analysis done by some researchers (Sinclair, 1987; Louw, 1993; Partington, 1998; Tognini-Bonelli, 2001; Stubbs, 1995, 2001) to describe the phenomenon of lexical pragmatics. The researcher examined whether significant collocates in different preference helped reveal evaluative orientation inherent in the target words and verified the results with the evaluative meanings suggested in previous literature.

Stubbs (1996, pp.138-141) provides evidence, based on manual analysis on the *Cobuild Corpus*, of the strongly unfavorable prosody of the lemma, *cause*. The collocates are distinctly unpleasant, including *accident*, *concern*, *damage*, *death*, *trouble*. By contrast, the word, *provide*, shows a favorable prosody with collocating words of *care*, *food*, *help*, *money*, and *shelter*. For this study, three target words, *cause*, *promote* and *commit* all transitive verbs, were chosen, based on the degree of difficulty for L2 learners and their fitness in serving as a representative lexical semantics learning task. Instead of repeating the analysis on, *cause*, a similar activity verb, *promote*, was chosen to serve as a complementary comparison. A third word, *commit*, was analyzed to strengthen the evidence on evaluative examination. The definitions given by both Merriam-Webster Dictionary and Webster's New Dictionary of Synonyms were adopted in contrast with collocational analysis in the following.

*Cause*: The definition of *cause* (as a transitive verb) given by the *Merriam-Webster Dictionary* is: (1) to serve as a cause or occasion of <cause an accident>; (2) to compel by command, authority, or force <caused him to resign>. The denotational meaning of *cause* is lightly sketched and its evaluative orientation is obscure. To examine the collocation profile of *cause*, POS categories and graded association strength of collocates were prepared. However, the lists of subject and adjective collocates exhibited only weak association and provided little

indication. Only object collocates were distinct. In each corpus column, two lists of lemma in the decreasing sequence of occurrence frequency and MI score were displayed. The analysis was done as follows:

- (1) The two object collocates lists of *cause* were largely consistent between the two corpora. 70% (14 out of 20) of collocates were shared, and 60% (6 out of 10) of the top 10 frequency lists and the top 10 MI lists were common.
- (2) The five collocates manually identified by Stubbs (2001, pp.163-165), *accident*, *concern*, *damage*, *death*, *trouble*, were covered in the BNC list, and only *accident* was not included in the COCA list.
- (3) The collocating word *damage* ranked highest at the 2nd in the frequency list and at the 1st in the MI list of COCA corpus. The lowest ranked of the five previously identified collocates in COCA corpus was *concern* at the 15th both in the frequency and MI lists.

Table 1 Object Collocates of *cause* (as transitive verb)

BNC				COCA			
lexis	freq	lexis	MI	lexis	freq	lexis	MI
problem	1001	distress	6.51	problem	2497	damage	5.89
damage	722	harm	6.40	damage	1592	harm	5.78
death	412	damage	5.70	death	1018	pain	4.80
trouble	304	confusion	5.46	pain	903	injury	4.78
harm	288	injury	5.16	people	848	cancer	4.71
injury	270	trouble	4.77	trouble	848	illness	4.70
concern	259	delay	4.70	cancer	807	trouble	4.61
difficulty	236	pain	4.38	disease	806	disease	4.33
change	204	cancer	4.37	harm	578	stir	4.29
loss	193	death	4.13	loss	546	problem	3.88
people	187	difficulty	4.11	change	527	reaction	3.83
pain	177	problem	4.09	injury	509	loss	3.83
distress	156	concern	4.04	lot	445	death	3.81
confusion	136	accident	3.71	effect	434	blood	3.07
delay	124	disease	3.65	concern	396	concern	2.99
increase	118	loss	3.58	blood	367	effect	2.50
disease	117	lot	2.08	stir	311	heart	2.22
accident	114	increase	1.86	heart	304	change	1.53
lot	107	change	1.58	reaction	298	lot	1.50
cancer	102	people	0.52	illness	295	people	0.66

(dark area marks dominant collocates, shaded area indicates strong collocates)

Collocates ranked in the MI list were a better indication of pragmatic meaning than those ranked in the frequency list of both corpora. Neutral words, such as *people*, *change*, *lot*, were ranked in the middle of the frequency list, but were downgraded to the bottom of the MI list. Dominant and strong collocates were highlighted and drew a vivid picture of strongly negative prosody. The list of dominant and strong collocates commonly appeared in both

corpora included *harm, damage, injury, trouble, pain, cancer, death, disease, problem, loss*. The dominant collocates, *distress, harm, damage*, seemed to be especially indicative of the semantic features and the pragmatic implications of *cause*. Among them, only *damage* was previously identified by Stubbs (1996, pp.138-141).

The two lists of dominant and strong collocates provided obvious and convincing evidence of the negative evaluation of *cause*. The systematic process of identifying significant (dominant and strong) collocates was shown to be effective in revealing semantic features and pragmatic implications of the target word. In addition, the result seemed to be more complete than manual analysis of concordance lines by Stubbs (1996, pp.138-141).

*Promote*: The definition of *promote* given by the Merriam-Webster Dictionary is: (1) to help or encourage to exist or flourish; further <to promote world peace>; (2) to advance in rank, dignity, position, etc. (opposed to *demote*). The denotational meaning of *promote* is concisely indicated and its evaluative orientation seems to have a positive inclination. To examine the collocation profile of *promote*, four POS categories and graded association strength of collocates were prepared. Similarly, the lists of subject and adjective collocates exhibited only weak association and provided little indication. Only object collocates and adjective collocates were distinct and shown in tables.

Based on Table 2, the following observations are made as follows:

- (1) 50% (10 out of 20) of top 20 frequent object collocates were common in the two corpora, while 70% (7 out of 10) of the top 10 frequent lists and 60% (6 out of 10) of the top 10 MI lists were shared.
- (2) With respect to the two top 20 lists of adjective collocates, 70% (14 out of 20) were common. When the lists were reduced to top 10, 70% of frequency lists and 50% of MI lists overlapped.
- (3) There were seven and six strong object collocates, respectively, in the BNC and COCA. Among them, *growth, development, and understanding*, were mutual. There was no dominant object collocate. The numbers of strong adjective collocates were reduced to two in BNC and three in COCA. The joint list included *healthy* and *economic*. There was no dominant adjective collocate, either.

The list of strong collocates that included *awareness, efficiency, democracy, welfare, growth,*

*development, understanding, learning, competition, peace* as objects, suggested abundant semantic features of *promote*, and clearly indicated its positive evaluative orientation.

Object collocates provided a better indicative function of pragmatic meaning than adjective collocates. The result was consistent with that of the *cause* lemma. The overall strength of positive prosody of *promote* was not as significant as that of the negative prosody of *cause*.

*Commit*: To gather more evidence, this study examined another verb *commit* known to show strongly negative prosody. The number of frequent collocates was still 20, but to save space in the manuscript, the finding showed only the top 10 of the total number of significant collocates. Subject collocates and adverb collocates are not indicative. Table 3 only showed object collocates of *commit*. Most observations were similar to the previous results. 60% of the top 10 frequent object collocates were shared, and 70% of the top 10 MI scored object collocates were common. As for adjective collocates, the percentages of overlapping in the top 10 frequency lists and MI lists are 50% and 30%, respectively.

Table 2 Object Collocates of *promote*

BNC				COCA			
lexis	freq	lexis	MI	lexis	freq	lexis	MI
development	161	awareness	5.29	development	727	democracy	5.12
interest	109	efficiency	5.24	growth	476	growth	4.69
growth	96	welfare	4.97	health	385	development	4.46
use	94	growth	4.46	use	356	understanding	4.44
health	85	understanding	4.23	student	349	learning	4.31
education	69	competition	3.90	democracy	329	peace	3.89
idea	68	development	3.68	interest	315	health	3.23
people	56	health	3.36	education	299	product	3.22
product	56	interest	3.08	idea	280	interest	3.15
work	55	education	2.97	understanding	244	education	2.88
service	54	product	2.93	product	237	idea	2.69
change	52	activity	2.66	book	234	value	2.61
competition	51	idea	2.65	peace	226	book	2.06
welfare	51	policy	2.09	school	219	student	1.94
policy	50	service	1.52	woman	217	program	1.49
activity	49	change	1.24	right	215	school	1.00
child	49	child	1.06	learning	205	use	0.94
understanding	49	use	0.84	program	202	woman	0.83
efficiency	48	people	0.42	child	198	child	0.73
awareness	47	work	0.35	value	198	right	0.28

(shaded area indicates strong collocates)

There were 12 significant object collocates in BNC and 13 in COCA. Among them, *suicide, adultery, crime, offence, sin, murder, act*, were common. The numbers of adjective collocates were only 2 in the BNC and 3 in the COCA, none was shared. The list of dominant collocates included *suicide, adultery, perjury, atrocity, crime, offence, sin, murder, fraud*, as objects, and *heinous*, as adjective.

As revealed by the list of significant collocates, the evaluative orientation of *commit* was as negative as *cause*, if not more. These significant collocates also provided highly marked links to the evaluative meaning of *commit*.

Table 3 Object Collocates of *commit*

BNC				COCA			
lexis	freq	lexis	MI	lexis	freq	lexis	MI
offence	322	suicide	9.05	crime	2753	adultery	8.70
suicide	315	adultery	8.23	suicide	1778	suicide	8.50
crime	279	offense	7.28	murder	837	perjury	8.40
murder	99	crime	6.55	act	815	atrocities	7.80
act	86	sin	5.46	sin	239	crime	7.46
trial	74	murder	5.44	offense	229	sin	6.24
people	58	breach	5.18	troop	227	murder	6.13
policy	52	act	5.18	people	212	offense	5.89
breach	51	government	4.89	perjury	211	fraud	5.86
government	48	trial	4.66	year	205	error	5.26
--		prison	4.44	--		act	4.65
		life	3.64			troop	4.26
		--				violence	3.60

(dark area marks dominant collocates, shaded area indicates strong collocates)

The examination of three lemma, *cause*, *promote*, *commit*, as transitive verbs, produces convincing evidence to support the use of collocation profile as an effective instrument in recognizing and learning semantic meaning and pragmatic implications of lexical items. With object collocates as primary and adjective collocates as secondary references, L2 learners were provided with abundant and obvious semantic cues to appreciate and underpin the evaluative orientation of the target item. Overall, we derived the following distinction based on collocational profile evidence. *Commit* implied mostly negative impact or disturbance caused by a strong stimulus. *Promote* assumed some entity that had a power to exert subtle control over the object and it overall indicates a positive evaluative meaning.

## Conclusion

L2 learners are often briefed using abstract definitions and left with vague lexical comprehension when self-learning with dictionary consultation. In self-learning with dictionary consultation, second language learners are frequently briefed using abstract definitions and close entries of near synonyms. Rather, they are also left with vague lexical

comprehension, which causes misuse of near-synonyms especially when they lack of pragmatic awareness. Collocational profiles, together with denotational meaning in dictionaries, give a solid conceptual grounding of target word for L2 learners in getting fuller understanding of the lexical pragmatics. A categorized collocational profile with graded association strength can filter and organize salient pragmatic features. It serves as a guided process to help develop concrete conceptual links such that evaluative pragmatic meanings of lexical items become more easily accessible to L2 learners.

There is, however, a gap between L2 learners' proficiency and the powerful investigative functions provided by these concordancing tools. It is suggested that the problem of how the linguistic resources and the computational functions, as provided by current concordancing tools, can be further built upon to support L2 near synonym learning. In addition, the process of constructing collocational profiles that we manually simulated on top of VIEW can be automated by a computer program and can be potentially developed as an online lexical query instrument for L2 learners in pedagogical and self-learning contexts. The development of such a software system, however, is not within the scope of the paper.

Evaluative pragmatic meaning of near synonyms has been an intriguing issue particularly in the field of corpus phraseology. Most L2 learners are unaware of the subtle pragmatic distinctions among near-synonyms. The approach proposed by this study can potentially fill the gap, improving L2 learners' lexical recognition and reducing pragmatic infelicities. In addition, pedagogical instruction, as an enhancement to the use of a collocational profile, is beneficiary for L2 learning. Last, the boundary of lexical semantic and pragmatics is not clear cut as Levinson (1983) questions the soundness of the dichotomy. Stubbs (2001) depicts the main aspects of the problem, "Amongst the many dualisms which plague linguistics is the question of how much meaning is expressed in the text as opposed to how much is in the mind of the hearer or reader. How much do we get out of a text and how much do we read into it" (p.438)? He further elaborates, "Traditionally, semantics has often been seen as a theory of those aspects of meaning which are expressed by sentences independent of context, whereas pragmatics is a theory of those aspects of meaning which are intended by speakers in context"(p.438). Evidently, it is potentially beneficial for L2 learners themselves to make pragmatic meanings from collocational information for near synonym learning.

## References

- Barron, A. (2003). *Acquisition in interlanguage pragmatics*. Amsterdam: Benjamins.
- Blutner, R. (1998). Lexical Pragmatics. *Journal of Semantics*, 15(2),115-162.
- Blutner, R. (2004). Pragmatics and the lexicon. In L. Horn & G. Ward (eds.), *The Handbook of Pragmatics*. Oxford: Blackwell. Partially retrieved from [www.blutner.de/pragmatics.pdf](http://www.blutner.de/pragmatics.pdf)
- Burnard, L. (1995). British National Corpus: User's reference guide for the British National Corpus. Oxford: Oxford University Computing Service.
- Chang, Y. C., Chang, J. S., Chen, H. J., Liou, H. C. (2008). An automatic collocation writing assistant for Taiwanese EFL learners: A case of corpus-based NLP technology. *Computer Assisted Language Learning*, 21(3), 283-299.
- Channell, J. (2000). Corpus-based analysis of evaluative lexis. In S. Hunston & G. Thompson (Eds.), *Evaluation in Text* (pp. 38-55). Oxford University Press.
- Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1):22-29.
- Dam-Jensen H. & Zethsen, K. (2007). Pragmatic patterns and the lexical system—A reassessment of evaluation in language. *Journal of Pragmatics*, 39(9) 1608–1623.
- Davies, M. (2008). The Corpus of Contemporary American English: 385 million words, 1990-present. Retrieved from <http://www.americancorpus.org>
- Dewaele, J. (2008). Appropriateness in foreign language acquisition and use: Some theoretical, methodological and ethical considerations. *International Review of Applied Linguistics in Language Teaching*, 46(3), 245-265.
- Kilgarriff, A., Rychly, P., Smrz, P. and Tugwell, D., (2004). The Sketch Engine. In *Proceedings of the Eleventh EURALEX International Congress*, 105-116. Lorient.
- Levinson, S. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Louw, B. (2000). Contextual prosodic theory: Bring semantic prosodies to life. In C. Heffer, H. Sauntson & G. Fox (Eds.), *Words in Context: A Tribute to John Sinclair on his Retirement*. Birmingham: University of Birmingham.
- Mackay, S. (1980). Teaching the syntactic, semantic, and pragmatic dimensions of Verbs.

*TESOL Quarterly*, 14(1), 17-26.

Marin, M. (1984). Advanced vocabulary teaching: The problem of synonyms. *The Modern Language Journal*, 68(2), 130-137.

Partington, A. (2004). Utterly content in each other's company: Semantic prosody and semantic preference. *International Journal of Corpus Linguistics*, 9(1), 131-156.

Simon-Vandenberg, A. (2008). Almost certainly and most definitely: Degree modifiers and epistemic stance. *Journal of Pragmatics*, 40(9), 1521-1542.

Sinclair, J. (2003). Corpora for lexicography. In Sterkenberg, P. Van (Ed.), *A Practical Guide to Lexicography*. Amsterdam: John Benjamins.

Stubbs, M. (1996). *Text and corpus analysis: Computer-assisted studies of language and culture*. Oxford: Blackwell.

Stubbs, M. (2001). Text, corpora, and problems of interpretation: a response to Widdowson. *Applied Linguistics*, 22(2), 149-172.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.

Wilson, D., & Wharton, T. (2006). Relevance and prosody. *Journal of Pragmatics*, 38(10), 1559-1579.

Yeh, Y., Liou, H. C., & Li, Y. H. (2007). Online synonym materials and concordancing for EFL college writing. *Computer Assisted Language Learning*, 20(2), 131-152.