



A Corpus-Based Collocation Study and Move Patterns in Online Smartphone Press Releases

Jiang Wei
Thammasat University
jwthl@hotmail.com

Chanika Gampper
Thammasat University
chanivad@yahoo.com

Abstract

This study is an investigation of fixed expressions or prefabricated adjective and verb collocations that occur in online smartphone press releases (OSPR). OSPR is a sub-genre of the press release. Previous studies reveal that different genres favor different linguistic structures and elements. Collocations of high frequency words in specific genres have indicated a prototypical phraseology. First, this study looked into the corpus-based collocation of adjective+ noun and verb+ noun in 230 OSPRs. Second, the sequences of moves of selected five OSPR were determined. Finally, the purpose(s) of each move was/were identified through adjectives used in each sentence. The findings reveal that while ostensibly informative, press releases also carry an implicitly self-promotional purpose. It is suggested that, for foreign language learners, focusing on collocation patterns within genre may provide a more productive way to study vocabulary. The move patterns found in this study enable learners to be aware of the move sequences in press release writing. Also, recommendations are made on how to use the research findings to improve English for Specific Purposes (ESP) teaching.

Keywords: collocation, corpus, press release, smartphone, move analysis

Introduction

Press releases are articles written by companies to promote new products, services or policies; their primary objective is to provide information for journalists. However, thanks to the development of internet marketing and PR centers accessible on company websites, press release readership has recently shifted towards the general public (Catenaccio, 2008), “The unpaid publicity that press releases are said to be geared at seems to keep the middle ground between advertising and news reporting” (Jacobs, 1999, p. 307). Linguistic research has shown that language patterns vary by genres, since each has its own unique rhetorical tradition. Patterns of communication in each genre are influenced by their purpose and by the language communities in which they arise. There are many previous studies about collocation in specific genres, such as advertisements, research papers, and news reports; however, online press releases have not been studied. Therefore, this study aims to 1) identify the collocation patterns found in OSPR focusing on adjective + noun and verb + noun collocations; 2) determine the sequences of moves in OSPR; and 3) determine through adjectives the purpose(s) of companies when they write each move in OSPR.

Literature Review

This section provides some background to the concepts of adjective classification, collocation, corpus, genre, move analysis and related studies.

Classification of adjectives

An adjective is a word that describes, identifies, modifies, or quantifies something (a noun or a pronoun), giving more information about the object signified (Dixon, 1982).

In this study, adjectives will be coded according to their functional roles into two types: 1) fact adjectives and 2) opinion adjectives. Fact adjectives are used to describe an object, providing information such as age, number, color, time, and source of power, shape, origin, purpose, material, brightness, speed etc., and in the case of the OSPR, giving informative details of the product. Opinion adjectives are used to describe the opinion and emotion of the writer. In this promotional-commercial OSPR corpus, opinion adjectives are used to promote the products and to motivate readers to purchase. For example, *white* is a fact adjective, whereas *stylish* is an opinion adjective, when they describe a smartphone.

Collocation

Collocations are particular words that consistently co-occur with other words in a certain situation with significant frequencies. Halliday and Hasan (1976) stated that since lexicon items in collocation provide cohesion to the text as a whole, they regularly co-occur in collocation chains in similar environments and suggest that collocation is often found in a specific register or genre. With the rise of computers in the 20th century, corpus linguistics has received a lot of attention from researchers. Many recent studies of collocation are corpus-based and focused on the contribution of collocation in the creation and maintenance of representation within genre (Williams, 2002), as well as the textual and discourse function of collocation for that genre (Callison, 2003).

The concept of “collocation” is still fuzzy and subject to debate. However, according to Halliday and Hasan (1976, p. 248), collocation is “lexical items that regularly co-occur”, or “a tendency for words to occur together” according to Sinclair (1991, p. 71).

According to the BBI combinatory dictionary of English (Benson, Benson, and Ilson, 2010), there are two main types of collocation:

1. Lexical collocations comprise two or more content or open-class words, which are words carrying content or the meaning of a sentence, such as a noun, verb, adjective, and adverb, as in the examples below:

Adjective+ Noun	:	historic sites
Verb+ Noun	:	make a cake
Adverb+ adjective	:	truly remarkable

2. Grammatical collocation consists of open-class words (content words) and closed-class words, which are function words or grammatical words, such as articles, prepositions, conjunctions, auxiliary verbs, and pronouns, as in the examples below:

Noun+ Preposition	:	an increase in
Verb+ preposition	:	participate in
Adjective + Preposition	:	keen on
Preposition+ noun	:	on foot



The study of collocation can provide sources for EFL pedagogy and methodology, especially within specific genres. As Halliday and Hasan (1976) mentioned, collocation is often found in specific registers or genres. Frequent collocations extracted from specific genres can give insights for specific genre writing course pedagogy. Language teachers can have a high-frequency collocation list as a mean of focusing students' vocabulary learning. High-frequency collocations are part of the lexicon that learners need to acquire, and some attempts have now been made to generate pedagogically-oriented listings of high-frequency collocations.

Corpus

This section outlines corpus linguistics and specialized corpus in corpus design and construction that will later be used in this study. A corpus is a collection of texts, written or spoken, which is stored on a computer (O'Keeffe, McCarthy, and Carter, 2007). A corpus may contain texts in a single language (monolingual corpus) or text data in multiple languages (multilingual corpus); it is principally a collection of texts available for qualitative and quantitative analysis. Large general corpora offer a huge amount of data and different kinds of genres, which allows linguists to investigate important facts about language.

Specialized corpora are narrower in scope, containing sample texts that share a particular genre or register. Specialized corpus studies focus on the influence of context on language choice and linguistic characteristics, which could be applied to English for Specific Purposes (ESP) and English as Foreign Language (EFL) pedagogy. Nowadays, researchers have paid attention to more specific sub-genres, such as medical research papers (Marco, 2000), travel advertisements (Ding, 2009), giving more specific insights to the characteristics of sub-genres.

Genre

According to Bhatia (1993), genres are recognizable communicative events, characterized by a set of communicative purposes with identified, highly structured and conventionalized constructs. Established members of discourse communities regulate genres; however, their communicative intentions are also socially recognized.

Bhatia (1993) has pointed that scholars often identify genre as pure form. However in reality, genres are always seen as hybrid and mixed forms. The press release is one such example. Press releases aim to inspire cooperation between organizations, journalists and the general public. They are short texts resembling news stories and product promotion information; thus the press release is a typical example of "hybrid genre" which contains informative and promotional purposes (Catenaccio, 2008).

The online smartphone press release (OSPR) is a hybrid genre with informative and promotional discourse; its purpose is to announce and promote new products. This study will focus on analyzing collocation patterns in OSPR, and the relationship between hybrid genre and collocation.

Move Analysis

The notion of move (Swales, 1990) is used to identify the textual regularities in certain genres, and to describe the function of particular portions in texts. Moves are semantic and functional units of text that can vary in length and size from several paragraphs to one

sentence. Move analysis is a helpful tool to examine the communicative purpose and linguistic boundaries of each portion of texts, and help to locate places for change of topics and themes.

Catenaccio (2008, p. 24) has summarized the move structure of the press release genre in light of its hybrid nature, as shown below:

1. Press / news release caption (frame)
2. Headline
3. Summary of main points (optional-can be conflated with move 2)
4. "For immediate release" formula (optional)
5. Lead: Announcing newsworthy information
 - Launching a new product/service
 - Announcing results and other company-internal information
 - Describing other types of company activity
6. Justifying the product or service or simply the newsworthiness of the information
 - Indicating the need for the product/service etc.
 - Referring to the advantages for potential beneficiaries
 - Qualifying the result as positive
7. Detailing product/service/company/other event which is the object of the release
8. Explicit promotional component: Attributed to company officials, or standard users:
 - Emphasizing positive results
 - Indicating reliability of company/product
 - Independently endorsing company results
 - Independent expert opinion
9. Boilerplate description(s) / establishing credentials
10. Contact details
11. Company logo

Methodology

The following sections will outline the methodology used in this study.

OSPR Corpus Construction

The corpus in this study consists of 230 online press releases of all new smartphone launch-related press releases from January 2008 to September 2012 from 11 companies, as follows:

- 1) Apple (www.apple.com)
- 2) Samsung (www.samsung.com)
- 3) LG (www.lg.com)
- 4) Vodafone (www.vodafone.com)
- 5) Motorola (www.motorola.com)
- 6) Nokia (www.nokia.com)
- 7) HTC (www.htc.com)
- 8) Blackberry (www.blackberry.com)
- 9) Asus (www.asus.com)



- 10) Acer (www.acer.com)
 11) Microsoft (www.microsoft.com)

These texts were saved as a *.txt readable in UTF-8 text format. The number of press releases forming the corpus can be found in table 1 below.

Table 1: Number of text and range of word numbers in the OSPR corpus

Company Name	No. of Press Releases (by Rank)	Word Number Range
Motorola	61	310~1,235
Blackberry	33	260~1,597
HTC	29	411~4,225
LG	27	153~902
Samsung	25	123~1,613
Apple	21	309~1,478
Nokia	16	267~1,168
Asus	8	485~1,063
Microsoft	4	948~1,461
Vodafone	3	405~489
Acer	3	458~1,181
Total	230	154,912

Remarks: Since Motorola issues different press releases for different countries and regions, it has the highest the number of press releases.

Analysis of OSPR Collocation

A corpus program Antconc (<http://www.antlab.sci.waseda.ac.jp/>) was used to analyze the corpus data, which includes keyword extraction and collocation identification. First, the top 20 nouns were identified with the keyword function of AntConc, using the criterion of frequency order. All the words that refer to smartphone products such as iPhone were replaced with XXYY. It was found that XXYY is the most frequent keyword.

Second, the adjective and verb that precede XXYY was determined using Antconc collocation identification function. For example,

adjective + noun: the *thinnest* mobile tablet

verb + noun: *presents* iPhone 4

Span is the key consideration in identification of collocation, with suggestions generally ranging from one word either side of the node to four (Sinclair, 1991). This study looked into the collocation of attributive adjectives and collocation of verbs that precede the product, using the left span 3:0 with a minimum occurrence of 3. The reason to choose Minimum frequency= 3 was to try to avoid the low frequency collocates.

Move analysis

Out of 230 press releases, one press release from each of the five companies with the highest numbers of press releases in the corpus (i.e., Apple, HTC, LG, Motorola, Samsung) was randomly chosen for the analysis. In other words, five press releases were analyzed using the code adapted from Catenaccio's framework (2008).

Move	Description	Remarks
Headline	the summary of the main point of the article and eye-catcher	Mandatory
Subheadline	placed between the headline and the text of the article, traditionally it gives the name, and/or position of the writer of the article; sometimes it also includes the article summary	Optional
Lead	the first part of body copy, aims to answer "5 Ws"(who, what, when, where, why) and how	Mandatory
Detailing of products	details of the product	Mandatory
Quotation from CEO	direct quotation from the CEO about the new product	Mandatory
Boilerplate	description that ends the release with information about the issuing organization	Optional
Legal claim	legal claim about trademark and patent	Optional

Text Purpose

To determine whether OSPR is a hybrid genre with the purposes to both give information and to promote the product, first, each adjective was coded as either fact (underlined) or opinion (italicized) adjective. Next, each sentence was coded for its purpose as follows:

- 1) Informative: Informative sentences are the sentences using only fact adjectives, aiming to give the fact of products. For example
iPad2 runs almost all of the over 350,000 apps available on the app store.
- 2) Promotional: promotional sentences are the sentences using opinion adjectives, aiming to impress the potential buyers and encourage the purchase action. For example
"We're *excited* for customers to get their hands on this *magical* and *revolutionary* product and connect with their apps and content in a *more intimate, intuitive* and *fun* way than ever before."
- 3) Mixed: mixed sentences are the sentences using both fact and opinion adjectives in the same sentences.
The Galaxy Tab 2 Series, available in both 10.1" and 7.0" screen sizes, are *powerful dual-core* tablets running the latest version of Android.
The purposes are later reported for each move.



Findings and Discussion

The following sections present the results of the adjective and verb collocation in online smartphone press releases (OSPR).

Adjective Collocations

This section presents frequent adjectives that come before *XXYY* (product). Also, whether the adjective is fact or opinion is determined. This reveals several fact collocates and opinion collocates that are used to describe the products.

Table 1: Top 20 left-span adjective collocates of the word XXYY (product): (span = 3:0, min=3)

Rank	Collocate	Frequency	Type
1	new	385	fact
2	first	147	fact
3	powerful	47	opinion
4	powered	39	fact
5	next	37	fact
6	revolutionary	31	opinion
7	thinnest	28	fact
7	latest	28	fact
9	premium	24	opinion
10	affordable	21	opinion
11	stylish	20	opinion
12	white	20	fact
13	innovative	20	opinion
14	friendly	19	opinion
15	slim	19	opinion
16	original	18	opinion
17	best	16	opinion
18	unique	13	opinion
19	smarter	10	opinion
20	fashionable	9	opinion

From the above table, the top two most frequent adjectives *new* and *first* occurred more than 500 times in the corpus of 154,912 words. They occurred more often than the remaining 18 adjectives combined. 12 out of 20 are opinion adjectives. Apparently, even though fact adjectives occurred more often than opinion adjectives, more varieties of opinion adjectives were used in OSPR because they implicitly carry a self-promotional purpose.

Verb Collocations

Table 2 presents the verbs that frequently come before XYYY (product). A common sentence pattern found is: company name + verb + product name.

Table 2: Top 10 left-span verb collocates of the word XYYY (product) (span= 3:0, min=3)

Rank	Collocate	Frequency
1	introduce	44
2	launch	42
3	announce	41
4	display	33
4	make	33
6	unveil	30
7	use	27
8	have	27
9	release	22
10	purchase	15

Example:

Apple launches ipad

HTC Unveils HTC Explorer

Move Analysis of OSPR

The sequences of moves in five selected texts are shown in the table below. The numbers 1-9 represent the sequences of moves found in each press release. For example, Apple, Samsung and HTC begin their OSPR with a headline while LG starts with Date formula.

Table 3 Move Sequences of OSPR

Remarks	Move	Apple	LG	Samsung	HTC	Motorola
Mandatory	Headline	1	2	1	1	2
Mandatory	Lead	3	5	4	3	3
Mandatory	Detailing of product	4,6	6,8	5,7		4,6
Mandatory	Quote	5	7	6	4	5
Optional	Press release formula					1
Optional	Date formula		1			
Optional	Sub-headline	2		2	2	



Remarks	Move	Apple	LG	Samsung	HTC	Motorola
Optional	Download link		3			
Optional	Picture		4	3		
Optional	Pricing and availability	7		8	5	7
Optional	Boilerplate	8	9	9	6	
Optional	Legal claim	9				

It can be seen that headline, lead, quotation and product detailing are mandatory moves in OSPR while the rest are optional. For example, LG includes a picture while Apple doesn't.

Text Purpose at Sentence Level

This part presents the purpose(s) of companies when they write each particular move in OSPR. The moves that are included in this part occur in at least four press releases. The date formula and download link are excluded because they are not written in sentences, while the picture is not a text.

Table 4: Headline

	Informative	Promotional	Mixed
Apple	1		
LG			1
Samsung			1
HTC	1		
Motorola			1
Total	2	0	3

Table 5: Sub-headline

	Informative	Promotional	Mixed
Apple	1		
LG			
Samsung			1
HTC			1
Motorola			1
Total	1	0	3

Table 6: Lead

	Informative	Promotional	Mixed
Apple	3		
LG	1		2
Samsung	2		2
HTC			3
Motorola	2		2
Total	8	0	9

Table 7: Quote

	Informative	Promotional	Mixed
Apple		2	
LG		1	
Samsung		1	
HTC		2	
Motorola		1	
Total	0	7	0

Table 8: Detailing products

	Informative	Promotional	Mixed
Apple	1		11
LG	6	5	11
Samsung	8	1	16
HTC	3	1	7
Motorola	6		8
Total	24	7	53

Table 9: Availability and pricing

	Informative	Promotional	Mixed
Apple	8		3
LG	6	5	11
Samsung			9
HTC	2		
Motorola	3		
Total	19	5	23

Table 10: Boilerplate

	Informative	Promotional	Mixed
Apple			3
LG	4	4	2
Samsung	6	1	6
HTC	2	2	1
Motorola	3		
Total	15	7	12

From Tables 4-10 above, it can be seen that the headline, sub-headline and lead are either informative or mixed. On the other hand, quotations from CEOs are promotional only. The adjectives found in product detailing, availability and pricing, and boilerplate make those sentences both informative and promotional.

Discussion and Implications

Discussion

As a corpus-based study, this study reveals collocation of adjective + noun and verb + noun to discover the linguistic patterns favored by OSPR. From the top 20 attributive



adjectives of the XYYY (product) list, the top two most frequent adjectives *new* and *first*. They occurred more often than the remaining top 18 adjectives combined. Of these 18 adjectives, 13 are opinion adjectives. Therefore, it can be concluded that while giving information through frequent use of fact adjectives, OSPR writers also use more varieties of opinion adjectives to promote the products. It is apparent that although “new” and “first” are classified as fact adjectives, they connote something modern and exciting and may serve to promote the product as well.

Compared with Catenaccio's (2008) framework, this study found four mandatory moves in OSPR: headline, lead, product detailing and quotation. Inclusion of other moves such as the sub-headline, download link and picture are at the preference of the writer. One unique move of OSPR is the download link because it is not applicable in offline texts.

Previous studies suggested that different genres favor different linguistic structures and elements. Two of the most prominent features of the press release are that it is prepatterned and a hybrid genre. While ostensibly informative, press releases also carry an implicitly self-promotional purpose as can be seen from the type of sentences in OSPR. However, different moves favor different sentence types because they have different purposes. For example, headline, sub-headline and lead are not primarily intended to promote the product. Like news, they introduce the launch of the new product. In addition, they also describe the product in an attractive manner. It is evident that the quote from CEO is absolutely promotional. The CEOs try to impress the readers and encourage further action. Unsurprisingly, detailing of product, pricing and availability and boilerplate are mixed between informative and promotional messages.

Implications and Suggestions

First, OSPR readers should be aware that OSPR does not only give product-related facts but also aims to promote the product. Readers should read it bearing the dual purposes in mind.

Second, for English teachers and material developers, the adjective and verb collocation patterns found in this study give English teachers and textbook writers well-informed word lists for teaching collocation in writing press releases. The move sequences of OSPR determined in this study can also give some guidelines for press release writing as they are written by professionals.

For the pedagogy of ESP, language teachers should take a genre-based approach to teaching writing. The data of corpus-based studies in specific genres will enable students and novice writers to write more naturally and native-like.

References

- Benson, M., Benson, E., & Ilson, R. F. (2010). *The BBI combinatory dictionary of English: Your guide to collocations and grammar*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Bhatia, V. J. (1993). *Analysing genre: Language use in professional settings*. London: Longman.
- Callison, C. (2003). Media relations and the internet: How Fortune 500 company websites assist journalists in news gathering. *Public Relations Review*, 29, 29-41.
- Catenaccio, P. (2008). Press releases as a hybrid genre: Addressing the informative/promotional conundrum. *Pragmatics*, 18(1), 9-31.

- Ding J. (2009). Collocation: Experiential grammar and conventionalization of travel advertising. *Journal of Cambridge studies*, 4(1), 35-49.
- Dixon, R. M. (1982). *Where have all the adjectives gone? And other essays in semantics and syntax*. Berlin: De Gruyter Mouton.
- Halliday, M.A.K., & Hasan, R. (1976). *Cohesion in English*. London: Longman
- Jacobs, G. (1999). *Preformulating the news: An analysis of the metapragmatics of press releases*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Marco, M. J. L. (2000). Collocational frameworks in medical research papers: A genre-based study. *English for specific purposes*, 19, 63-86.
doi:10.1016/S0889-4906(98)00013-1
- O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.
- Sinclair, J (1991). *Corpus, concordance and collocation*. Oxford: Oxford University Press
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Williams, G. (2002). In search of representativity in specialised corpora: Categorisation through collocation. *International Journal of Corpus Linguistics*, 7(1), 43-64.
doi:10.1075/ijcl.7.1.03wil