

A Corpus-based Study of the Use of Adjectives in Spoken and Written Registers

Guifen Tang

Department of Fundamental Teaching, Shanghai Publishing and Printing College, Shanghai, 200093, China

Abstract: *The present study is designed to investigate the use of adjectives in spoken and written registers. The word “strong” is taken as an example and its synonyms are retrieved from TV/Movies Corpus and Academic Journals Corpus for analysis in terms of word frequency and semantic preferences. The results indicate that there are significant differences in the word frequency and semantic features of the adjectives between oral and written registers. 1) Adjectives in written register generally occur with higher frequency than those in spoken register. 2) The polysemy of adjectives is not determined by the word frequency, but consistent with linguistic registers. The adjectives used in written English carry a wider range of extended and metaphorical meanings than those in spoken English. 3) Nouns in adjective-noun collocations tend to be abstract and indirect in written register, but concrete and direct in spoken register. 4) Nouns influence the negative, neutral or positive semantic prosody of the collocations. As a corpus-based empirical research, the study also offers a few constructive suggestions for improving the efficiency of vocabulary acquisition and teaching by using authentic data from balanced corpora.*

Keywords: *spoken register, written register, word frequency, lexical collocation, semantic feature*

1. Introduction

“Adjectives belong to one of the major form classes in any of numerous languages and typically serve as a modifier of a noun to denote a quality of the thing named, to indicate its quantity or extent, or to specify a thing as distinct from something else”[1]. It is the third largest part of speech, second to noun and verb, and plays a significant role in the study of the syntax and semantics[2]. Biber (2006) studied the distributions of part of speech in spoken and written registers based on TOEFL 2000 Spoken and Written Academic Language (T2K-SWAL) corpus. His research reveals that adjectives are used most frequently in written academic English and least in spoken English[3].

Quite a few Chinese researchers have carried out research to explore learners acquisition of English adjectives. Inspired by previous research, Chen and Liu (2010) also made a comparative study on the distribution features of parts of speech in Chinese learners’ spoken and written language and that of the native speakers. Their findings indicate that learners show far less significant difference in spoken and written language, and they tend to mix the register features of lexis[4]. Through a contrastive study of the corpora LONESS and SECCL, Gao and Zhao’s research (2020) reveals that English majors in China produce lower diversity and complexity of adjectives, lower correlation of high frequency adjectives in spoken and written English, and less word types[5].

To find out causes of learner’s lag in adjectives learning, Wu and Chen (2016) reflected on the breadth and depth of adjectives in the English textbooks universally used by high school students in China. Their paper points out that in comparison with *Collins COBUILD English Course*, some adjectives in Chinese textbooks are either too high or too low in word frequency, and the adjectives are lacking in meaning multiplicity and expressive vitality[6]. Li (2019) presents the evidence that Chinese is a language with no morphological signs of parts of speech, so the negative transfer of mother tongue occurs during lexis acquisition, resulting in morphological errors and syntactic errors in adjectives use[7].

It is an indisputable fact that there is a certain gap between L2 learners and native speakers in language output. Objectively speaking, any textbook has to some extent reflected the stagnation and limitation of knowledge since its publication, which is beyond reproach. Fortunately, with profound linguistic theories, empirical data and appropriate teaching pedagogy, researchers and teachers have been guiding second language learners to understand and master the use of words accurately and

efficiently. Based on COCA corpus, this paper extracts synonyms of *strong* in two registers and illustrates their semantic features so as to offer constructive suggestions for adjective learning and teaching.

This paper is set to answer the following two research questions:

1. Is there variation in the word frequency of adjectives in oral and written registers? If yes, what is the difference?
2. Does the same adjective embody identical semantic meanings in the two registers? If not, what are the distinctive features?

2. Methodology

2.1 Research data

COCA corpus contains about 1 billion words in nearly 500,000 texts from 1990 to 2019, which are nearly evenly divided between spoken, fiction, magazines, newspapers, academic journals, blogs, other web pages, and TV/Movie subtitles (120-130 million words in each genre)[8]. In this study, two sub corpora, TV/Movies and Academic Journals (abbreviated as Academic) from COCA, are chosen to extract the data for research. Two reasons may account for the choice of the data sources. For one thing, English in TV dramas and movies belongs to the authentic oral texts and its media form is usually favored by students of second English learning to foster study interest and understand culture of English speaking countries. For another, Academic covers the entire range of American Library of Congress classification system, including philosophy, psychology, world history, education, technology, business, and religion. It provides authentic written material that English learners are supposed to acquire before they can produce decent interlanguage. Therefore, a comparative study of the adjectives features in TV/Movies and Academic may prove enlightening and illuminating to second language learners, which may shed some light on learners' weakness in language production.

Table 1 Corpus information

Data	Word tokens	Sources	Topics
TV/Movies Corpus	128,074,534	American television and movie subtitles	action, adventure, animation, comedy, crime, documentary, drama, game, horror, reality, science fiction, musical
Academic Journals Corpus	119,790,456	approximately 100 peer-reviewed academic journals	education, history, geography, social sciences, law, humanities, philosophy, science and technology, medicine, and business

2.2 Research instrument and data analysis

With online corpora search engine, 53 synonyms of *strong* are retrieved by entering “=strong” and selecting TV/Movies and Academic in Sections column. Word frequency (tokens) is one of the important indicators to measure the importance of a word. The study lists the top 20 synonyms in two corpora respectively in the order of word frequency.

In table 2, Word is a rank-ordered list of adjectives, which refers to the synonyms for *strong* (including *strong*) in the study. Tokens1 and Tokens2 represent the total number or cross word frequency of adjectives in each corpus. PM1 and PM2 stand for the tokens of adjectives per million words in each corpus. Ratio1 is the result of PM1 divided by corresponding PM2 of the same adjective, which stands for the relative percentage of adjectives in the TV/Movies. And vice versa, ratio2 stands for the relative percentage of adjectives in Academic. In the case of *great*, there are 10,7836 tokens in total or 842 tokens per million words in TV/Movies, while there are 34,522 tokens in total or 288.2 tokens per million words in Academic. *Great* occurs 2.9 times as much in TV/Movies as in Academic.

Table 2 Top 20 synonyms of strong in the corpus

	TV/Movies: 128,074,534 words				Academic: 119,790,456 words			
	Word	Tokens1	PM1	Ratio1	Word	Tokens2	PM2	Ratio2
1	great	107836	842	2.9	great	34522	288.2	0.3
2	hot	23639	184.6	5.7	effective	28807	240.5	19.3
3	clear	22046	172.1	0.8	clear	25584	213.6	1.2
4	strong	12771	99.7	0.5	strong	23105	192.9	1.9
5	tough	11190	87.4	6.6	powerful	11539	96.3	2.1
6	deep	10421	81.4	1.1	deep	8694	72.6	0.9
7	powerful	5818	45.4	0.5	firm	8682	72.5	2.6
8	bright	4561	35.6	1.7	solid	5554	46.4	2.1
9	brilliant	4418	34.5	3.6	intense	4104	34.3	2.4
10	firm	3611	28.2	0.4	hot	3877	32.4	0.2
11	solid	2872	22.4	0.5	dedicated	3496	29.2	3.2
12	sharp	2725	21.3	0.8	sharp	3027	25.3	1.2
13	intense	1857	14.5	0.4	compelling	2845	23.7	6.7
14	effective	1596	12.5	0.1	robust	2829	23.6	28
15	dedicated	1184	9.2	0.3	concentrated	2750	23	9.9
16	stark	1173	9.2	0.9	bright	2542	21.2	0.6
17	convincing	982	7.7	0.6	eager	1708	14.3	1.9
18	passionate	969	7.6	1	tough	1579	13.2	0.2
19	eager	967	7.6	0.5	convincing	1501	12.5	1.6
20	keen	692	5.4	0.7	persuasive	1360	11.4	4.7
	total	221,328	1728.3	29.6		178,105	1487.1	91

3. Result and Discussion

3.1 Shared adjectives in the two registers

As is shown in table 2, the top 20 adjectives in TV/Movies include great, hot, clear, strong, tough, deep, powerful, bright, brilliant, firm, solid, sharp, intense, effective, dedicated, stark, convincing, passionate, eager and keen; the top 20 adjectives in the Academic include great, effective, clear, strong, powerful, deep, firm, solid, intense, hot, dedicated, sharp, compelling, robust, concentrated, bright, eager, tough, convincing and persuasive. On closer examination, 16 of the top 20 adjectives are shared by both registers, which are great, hot, clear, strong, tough, deep, powerful, bright, firm, solid, sharp, intense, effective, dedicated, convincing, and eager.

These shared adjectives are frequently used in both colloquial and written registers, and can be used as substitutes for strong in certain contexts. Among them, the frequency ratio 1 of clear, deep and sharp is 0.8, 1.1 and 0.8 respectively, all close to 1, which means they occur more or less equally as adjectives in spoken and written texts.

3.2 Variation in the frequency of adjectives in the two registers

T-test of PM is conducted to further investigate the holistic word frequency distribution of shared adjectives in spoken and written registers. The result reveals that there is very significant difference ($P < 0.01$) in the word frequency of adjectives between the two groups.

A bar chart is created to clearly illustrate the discrepancy of word frequency. *Great* is removed from the chart because it obviously takes an absolute lead in the word frequency of the two registers. To be specific, the word frequency of *great* is 107,836, nearly equal to one half of the total word frequency of

top 20 adjectives (221,328) in the spoken register. It is also more than three times that of *great* (34,522) in the written register. In view of its excessive use, to include *great* in the chart may weaken the positions of other adjectives.

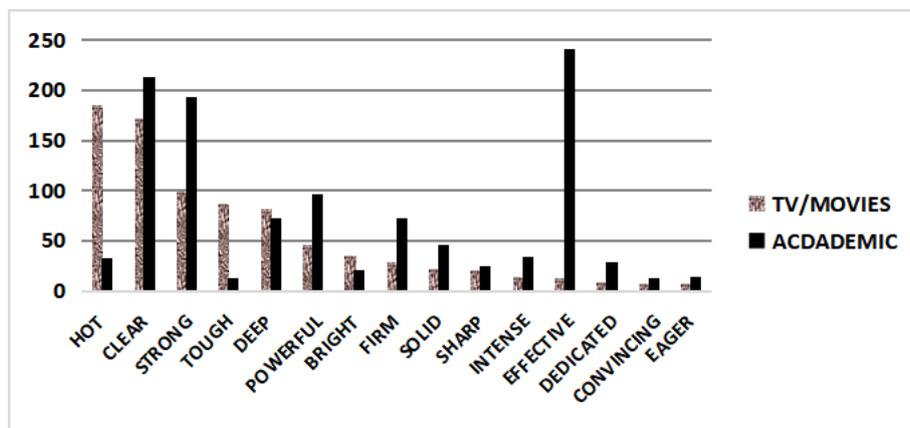


Figure 1. PMs of shared adjectives in two corpora

As is illustrated in figure 1, the three words “hot” “tough” and “bright” demonstrate distinctive styles of informal speech, with the PMs in TV/ Movies far exceeding those in Academic, while the nine words “effective”, “strong”, “powerful”, “firm”, “solid”, “intense”, “dedicated” “convincing” and “eager” see obvious superiority in PM 2 and they are more frequently used in Academic. Among the top 20 non-shared adjectives (table 2), “brilliant” is typically used in informal register. Conversely, other adjectives including “robust”, “concentrated”, “compelling”, and “persuasive” highlight their written stylistic characteristics.

After further analysis, there are 113,492 tokens for the top 20 adjectives in TV/ Movies and 144,183 tokens for the top 20 adjectives in Academic, excluding the excessively overused word *great*. In other words, the total word frequency of top 20 adjectives in spoken register is lower than its corresponding number in written register. Considering the total word tokens in TV/ Movies being higher than Academic, we may draw a safe conclusion that the word frequency of adjectives is generally higher in Academic than that in TV/ Movies. Take a look at the relative percentage of adjectives, also the ratios. The total of ratio 2 (the value is 91) is more than three times that of ratio 1 (the value is 29.6). We can arrive at the same conclusion as above.

As modifiers of nouns, adjectives occur in proportion to nouns. Therefore, the number of adjectives increase in registers where nouns occur in higher frequency and with higher degree of complexity, as is in the case of written academic journals.

3.3 Semantic features of adjective-noun collocations in the two registers

One of major grammatical functions that adjectives perform is to serve as a noun phrase modifier. This section compares and examines the semantic features of the adjectives in spoken and written registers by analyzing adjective-noun collocations.

The term collocation represents a multi-word construct, a co-occurrence or combination of words on the lexical, syntactic and grammar level. There are two types of lexical collocations, which are verb-noun and adjective-noun. This paper will deal with word combinations consisting of an adjective and a noun.

The collocations of the higher frequency spoken adjectives *great* and *hot*, higher frequency written adjectives *solid* and *firm*, as well as equal frequency adjectives *deep* and *sharp* in each corpus are extracted respectively and displayed separately from table 4 to table 6. The data is retrieved by following three rules. 1) The nouns coexist in the first and second place on the right side of the adjective. 2) To show typical semantic prosody, the collocation frequency is set to be higher than 10. 3) Only collocations with minimum ratio at 2 are chosen for analysis in the study. A maximum of 20 nouns will be listed in the table due to the limited space.

Table 3 illustrates the information of nouns retrieved from the corpora by following the rules mentioned above. In TV/Movies, the first six nouns meet the requirements while there is more than a dozen in the Academic.

According to co-selection theory, under few circumstances do words exist in isolation. In most cases, words realize their exact meaning by functioning in a lexical pattern. In other words, when a word occurs with another one, it displays one aspect of its meaning and abandons others[9]. The following sections are intended to discuss about the semantic preferences of adjectives and nouns when they co-select each other in specific register and context.

Table 3 Nouns information for the adjective-noun collocations of strong

TV/Movies:128,074,534 words					Academic: 119,790,456 words			
	Word	Tokens1	Tokens2	Ratio1	Word	Tokens2	Tokens1	Ratio2
1	man	193	25	7.2	support	518	18	30.8
2	word	104	10	9.7	evidence	452	35	13.8
3	woman	97	26	3.5	sense	449	63	7.6
4	point	45	19	2.2	relationship	346	15	26.4
5	arm	29	13	2.1	ties	217	25	9.3
6	people	23	11	2	leadership	168	12	15
7	will	29	15	1.8	desire	142	15	10.1
8	words	27	17	1.5	position	94	11	9.1
9	line	14	10	1.3	role	89	10	9.5
10	team	22	17	1.2	presence	108	23	5
11	body	20	15	1.2	force	103	24	4.6
12			

In the case of higher frequency spoken adjectives, a closer analysis of *great* and *hot* may well explain their semantic features in the two registers. As we can see from table 4, in TV/ Movies, meanings of *great* include: a nice feeling to see, to hear or to enjoy something, like *great day*, *great gifts*, *great idea*, *great food*; an admiring feeling for something or somebody, like *great job*, *great warriors*, *great country*. Usually, the feelings for things or for people are positive and encouraging. In Academic, the concept of *great* is extended to express a large quantity, like *great number/ majority*; the degree of something abstract, like *great concern*, *great challenge*, *great importance*. Various notions reflect its positive, negative and neutral semantic features depending on the lexical collocations in the genre.

Orally, *hot* is often used to describe the temperature of food that human body can directly touch or taste. By contrast, *hot* is not touchable in phrases like *hot issues*, *hot topics* or *hot spots*. The collocation *hot issues* obviously does not mean topics or places that an individual can touch with his hands and feel the heat. It's a metaphor, implying that a topic attracts much attention or arouses heated controversy. Here is the original context extracted from the Academic Journal *Knowledge Quest*. "In the 'What's Hot, What's Not' Literacy Survey for 2013, 75 percent of respondents said adolescent literacy and comprehension were 'very hot.' One hundred percent were in agreement that both topics should be hot issues in teaching and learning (Cassidy and Grote-Garcia 2012, 10)." *Hot issues* in this sentence implies subjects quite popular with the majority of readers and learners.

Table 4 Semantic features of higher frequency spoken adjectives

Adj.	Semantic feature in TV/Movies	Modified nouns in TV/Movies	Semantic feature in Academic	Modified nouns in Academic
great	people, things, country, idea, feeling	day, job, time, idea, things, friends, story, state, warrior, father, gift, place, friend, honor, man, start, country, feeling, things, food	number, situation, importance, difficulty and change	majority, lakes, books, works, plains, powers, concern, importance, numbers, significance, society, promise, changes, cost, debate, challenge, Depression
hot	food	dog, chocolate, sauce, potato, coffee	climate, topic, place	weather, spots, days, issue, topic

Solid and *firm* are representative of higher frequency adjectives in Academic. In table 5, their modified nouns are displayed to reveal the semantic preferences. In TV/ Movies, typical function of *solid* is to modify concrete objects, such as gold, rock, steel, which all denote the hardness of the object. Sometimes, it also describes the quality of a plan or people, implying that the plan is feasible and people are dependable. In Academic, except for a few phrases like *solid wood* and *solid line*, *solid* is mostly used in a figurative way to evaluate the quality of abstract nouns such as base/foundation, research, understanding, support, footing, system.

Table 5 Semantic features of higher frequency written adjectives

Adj.	Semantic feature in TV/Movies	Modified nouns in TV/Movies	Semantic feature in Academic	Modified nouns in Academic
solid	objects, people	gold, rock, case, steel, plan, citizen	things, feeling, action, system	waste, line, base, foundation, state, research, wood, fuel, core, understanding, support, footing, system
firm	action, human body	handshake, hand, grip	base, evidence, belief	foundation, career, evidence, hold, belief

Likewise, *firm* in TV/ Movies is used in such collocations as *firm hand*, *firm handshake*, and *firm grip*, expressing a strong feeling of real touch, while it is commonly seen followed by nouns with abstract concepts in the Academic, such as foundation, evidence, belief. It is not easy to explain how strong the foundation is or how convincing the evidence is without the semantic context.

Finally, noun collocations of the shared adjectives *deep* and *sharp* also demonstrate distinctive semantic functions, as shown in table 6. In TV/Movies, *deep* specifies long and measurable distance from the top surface to the bottom in the phrase *deep sea*; a *deep breath* uses or fills the whole of lungs; by *deep heart* we mean that we feel it very strongly with our whole heart. In the Academic, *deep* frequently occurs with nouns like division, understanding, concern, knowledge, learning, recession, involvement, suspicion, distrust, crisis and faith in the context. For example, here is the original text of deep division retrieved from *Law and Psychology Review in Academic*. "Although less publicly visible than the insanity defense, the issues of civil commitment and involuntary medication have far greater practical importance in the lives of people who are seriously mentally ill. Additionally, deep divisions exist among both patients and treatment providers regarding when, if ever, these measures are appropriate." *Deep divisions* denote disagreement and dispute between patients and medical workers. These abstract nouns cannot be measured or directly perceived as strong or weak. Rather, they indirectly express the seriousness, power, importance, or degree of something in a metaphorical sense.

Table 6 Semantic features of adjectives with equal frequency

Adj.	Semantic feature in TV/Movies	Modified nouns in TV/Movies	Semantic feature in Academic	Modified nouns in Academic
deep	human body, action, things	breathe, throat, end, heart, sea	difference, relationship, feeling	division, understanding, concern, knowledge, learning, recession, involvement, suspicion, distrust, crisis, faith
sharp	objects, people, ideas, force	stick, force, man, instrument, dresser, shock, knife, suit, mind	relationship, change, things, feeling, action	contrast, drop, rise, increase, focus, criticism, relief, changes, lines, transition, images, items, tip, curve

Not coincidentally, *sharp* shows similar differences in semantic meanings. In TV/ Movies, the sharpness of stick, knife and instrument can all be seen or touched; in Academic, the extended meaning is complicated, indirect and expressive. Strong and determined tone of one's voice: *sharp criticism*; clearly discernible images or sounds: *sharper images/ lines*; the strong impact of a sudden change on one's visual senses: *sharp drop* or *sharp increase*.

4. Conclusion and Implications for teaching

4.1 Evaluation of the research questions

The previous section examines the use of adjectives regarding the word frequency, the register features and semantic preferences in spoken (TV/Movies) and written (Academic) registers. Based on the discussions above, we may pin down the research result into four points.

Firstly, there are significant difference in the word frequency of shared adjectives between oral and written registers. Except for the overused word *great*, adjectives in written register generally occur with higher word frequency in written register than in colloquial register.

Secondly, meaning variety and complexity of adjectives are not in proportion to the tokens or word frequency, but in consistence with the linguistic register. Adjectives in written language carry a greater variety of extended and metaphorical meanings than those in spoken language. Although some adjectives occur in higher frequency in colloquial register, they do not show more multiplicity in meanings.

Thirdly, meanings of the modified nouns in spoken register are simple and direct, specifying the features of concrete things, people or their behaviors and immediate feelings. As for those in written register, the nouns tend to be abstract and indirect, evaluating the state, result, degree of something happening, interpersonal relationship or human emotions.

Last but not least, semantic prosody of adjectives share one thing in common in the two registers. The tone of adjectives is not decided by adjectives alone, but influenced by the positive, neutral or negative sense of nouns in the collocation, like the different tones expressed between *great significance* and *great cost*, *deep faith* and *deep suspicion*, *hot weather* and *hot issue*, *sharp relief* and *sharp criticism*, *solid waste* and *solid support*.

4.2 Implications for teaching

The present empirical study makes an attempt to elaborate lexical features in different registers. A few suggestions will be offered for the current teaching and learning practice to improve our future work.

Teachers need to be clearly aware of the distinctive language features in different registers, and share the knowledge with students. In English classroom, teachers are often heard encouraging students to improve English skills by cultivating interest first. As an alternative approach to serious classroom teaching and learning, some teachers may recommend classical television dramas and movies for students to appreciate in and out of class. However, the present study reveals that it is likely to overuse colloquial or neglect written vocabulary if we rely on oral English learning resources in the long run. Secondly, the words in spoken register are informal with less complexity. Excessive focus on spoken English materials may give rise to an illusion that language learning is simple and fun. Consequently, close and extensive reading might be overlooked as essential way to improve language levels.

Therefore, it is highly advisable to provide students with access to various topics and registers in English learning. Word frequency and meanings vary greatly in different registers. Only with a good understanding of the semantic features can we achieve accuracy and complexity in language acquisition. What's more, lexical meaning depends on co-selected words, grammatical structure and the context. It is hardly possible for a single vocabulary book or textbook to list all the concepts of words. With abundant and authentic language data, corpus with balanced registers may serve as a valuable platform and a powerful reference tool for language learning and teaching.

References

- [1] The definition of "adjective", Webster's Online Dictionary. [EB/OL] <https://www.merriam-webster.com/dictionary/adjective>.
- [2] D. Biber, S. Johansson, G. Leech, S. Conrad, and E. Finegan (1999). *Longman grammar of spoken and written English*. Foreign Language Teaching and Research Press: Beijing.
- [3] D. Biber (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.

- [4] J.S. Chen and X.C. Liu (2010). *A corpus-based study of distribution of parts of speech in Chinese learners' spoken and written English*. *Journal of Hebei University of Engineering*, vol. 27, n.4, p.90-92.
- [5] Y.J. Gao and X.D. Zhao (2020). *A corpus-based quantitative analysis of the use of English adjectives by Chinese English majors*. *Foreign Languages Research*, vol.1, n.2, p.24-31.
- [6] Y.W. Wu and B.H. Chen (2016). *A study on the breadth and depth of adjectives in high school English textbooks*. *Journal of Ningbo University*, vol.38, n.5, p.119-125.
- [7] X.J. Li (2019). *A study on word class acquisition of adjectives by Chinese EFL learners*. *Journal of Northeast Normal University*, vol.1, n.5, p.111-117.
- [8] *English Corpus online retrieval website [EB/OL]*. <https://www.english corpora.org/>
- [9] N.X. Wei (2011). *Theory of co-selection and the corpus-driven phraseology*. *Journal of PLA University of Foreign Languages*, vol.35, n.1, p.1-6, 74.