Original Paper

Exploring Semantic Nuances in English Synonyms through Multimodal Analysis

Lei Yu[1*]

[1] Faculty of Foreign Studies, Beijing Language and Culture University, Beijing, China

[*] Lei Yu, E-mail: yulizzy@126.com

Abstract

In foreign language education, distinguishing English synonyms has long posed a formidable challenge due to context-dependent usage, often leading ESL learners to perplexity. Learners commonly attempt differentiation based on subtle emotional nuances, evaluative connotations, and semantic intensity. This study employs the Corpus of Contemporary American English (COCA), along with image and audio data, to scrutinize the semantic nuances of the synonyms: "reveal", "disclose", and "expose". The investigation unveils the subsequent pivotal discoveries: "Reveal" exhibits greater prevalence relative to "disclose" and "expose", frequently associating with abstract nouns, whereas the latter two demonstrate a proclivity for specific terms. These terms display collocational variations, signifying distinctive contextual preferences. Through the integration of image and audio data, the study further underscores the subtle distinctions among these synonyms on both visual and auditory levels. This research contributes a novel methodology for a nuanced comprehension of these synonyms, allowing for a multidimensional understanding facilitated by the amalgamation of multimodal data. It provides educators with a more enriched pedagogical approach to assist learners in employing these synonyms with greater precision.

Keywords

Synonyms, COCA, Multimodal Analysis, Semantic Nuances

1. Introduction

In the foreign language pedagogy, the precise distinction of English synonyms has persistently presented a formidable challenge. This challenge stems from the intricate and context-dependent nature of these synonymous terms, frequently resulting in learners of English as a second language grappling with nuances that defy straightforward categorization. Learners often undertake the task of

distinguishing these terms by discerning subtle emotional inflections, evaluating connotations, and recognizing varying degrees of semantic intensity. However, the integration of a multimodal approach, which encompasses a diverse range of data modalities, holds the potential to provide a fresh and comprehensive perspective on this linguistic complexity. This study undertakes an exploration that combines the robust analytical tool of the Corpus of Contemporary American English (COCA) with the incorporation of image and audio data. Through this innovative methodology, the study investigates the semantic variants of three synonyms: "reveal", "disclose", and "expose". The objective is to elucidate the intricate distinctions and contextual variations that characterize these terms. Additionally, the study will shed light on the collocational preferences of these terms, further delineating their distinct contextual proclivities. Furthermore, through the integration of image and audio data, the study aims to present a nuanced understanding of these synonyms, elevating our comprehension to encompass both visual and auditory dimensions.

This research not only aims to enhance our comprehension of these synonyms but also equips educators with a pedagogical approach strengthened by the integration of multimodal data. In doing so, it strives to introduce an innovative research perspective into the extensive field of linguistics. The novelty of this study lies in its utilization of a multimodal approach, which integrates text, images, and audio data, to offer a new viewpoint and methodology for addressing the challenge of distinguishing English synonyms. Concurrently, this research introduces a state-of-the-art research perspective into the expansive realm of linguistics by incorporating multimodal data, thereby augmenting our understanding of these synonyms, and refining pedagogical approaches in foreign language education.

2. Literature Review

Synonyms have long stood out as one of the most formidable challenges in the second language acquisition (Shahzadi, Asghar, & Javed, 2019). Chinese learners frequently grapple with distinguishing synonyms of English, a difficulty attributed to a range of factors including inadequate pedagogical methods, limited learning resources, interference from their native language, and cultural disparities, among others (Ren, 2008). This phenomenon of synonymy not only manifests within the lexical system but also extends to the syntactic and discourse levels, intricately linked to grammatical structures, and intimately connected with word meanings. In English, synonyms serve the purpose of conveying identical or similar concepts, and they are also employed to modify specific verbs or adjectives for communicative precision. It is estimated that synonyms account for more than 60% of the English vocabulary (He, 2003). This diverse assortment of synonyms empowers English speakers to articulate their ideas with greater exactitude and efficacy (Liu & Espino, 2012).

Corpus linguistics has emerged as an indispensable research tool with the rapid development of computer technology. It encompasses various avenues, including lexical statistics, dictionary compilation, and language testing, all hinging on the collection of authentic language data from extensive textual sources. While the process of corpus compilation often prioritizes lexical meanings,

other facets tend to be overlooked. Notably, frequency and collocation stand out as two prominent features of corpora. These aspects play a pivotal role in comprehending the associations between words and their variations across diverse contexts. This concept aligns with Firth's proposition that a word's meaning can be inferred from its typical collocates (Firth, 2020). Within the linguistic framework, a specific mechanism of association, known as semantic prosody, comes into play. This implies that evidence for specific usages can be gleaned through frequency analysis, phrase distribution, and collocational environments. Language learners often struggle with instances of "semantic ambiguity", where their interpretation of a word significantly deviates from its intended meaning. The conclusions drawn from probabilistic statistics generally carry a high degree of credibility. In recent years, theories pertaining to synonyms have found widespread application in vocabulary instruction. Corpus-based methods have demonstrated efficacy in discerning synonyms, as corpora provide a wealth of real-language data that learners can employ for the analysis and differentiation of synonym usage (Wang, 2009).

Semantic prosody is a linguistic means utilized primarily to convey the speaker's or author's attitude (Louw, 2000). Native speakers often select words with specific semantic prosodies to serve communicative purposes (Stubbs, 2009). These prosodies, formed by lexical collocations, can be broadly categorized into three types: negative, neutral, and positive (Stubbs, 1996). Essentially, semantic prosody gauges the level of association between a word and other entities or individuals. When a word carries a negative semantic tone in conjunction with other words, it possesses a negative semantic prosody, indicating a state of dissonance. Conversely, a positive semantic prosody emerges when there is a favorable semantic atmosphere between a vocabulary item and its collocates. The tone, whether negative or positive, imparts a corresponding semantic prosody (Wu & Jiang, 2016). Semantic prosody shapes the overall semantic ambiance and results from the combinations of node words and their typical collocates. The semantic choices inherent in these collocations tend to establish specific semantic prosodies (Partington, 2004; Lu, 2010). It is precisely due to the existence of semantic prosodies that only vocabulary items with distinct semantic features can naturally gravitate towards one another and co-occur in the same context with a specific node word. The presence of any disparate words in context may lead to conflicts in semantic prosody (Wei, 2006).

The application of multimodal analysis in linguistic research has garnered significant attention in recent years. Multimodality refers to the integration of multiple modalities, such as text, images, and audio, gestures, eye gaze (Poria et al., 2017). Within this framework, particular emphasis has been placed on its potential in investigating semantic distinctions among synonymous terms. Traditionally, linguistic analysis primarily relied on textual data, which sometimes proved insufficient in capturing subtle semantic nuances. Multimodal analysis, however, presents a paradigm shift by incorporating additional sensory dimensions. For instance, when examining synonyms like "reveal", "disclose", and "expose", integrating image and audio data allows for a more nuanced exploration of their contextual usage. Visual cues, in conjunction with auditory elements, afford researchers a holistic view of how these

terms operate in various contexts. Moreover, the integration of multiple modalities affords researchers a broader methodological toolkit, enabling them to address questions that may have posed challenges to traditional approaches. By leveraging the strengths of each modality, researchers can yield more comprehensive and accurate insights, thereby opening new avenues for linguistic inquiry. In conclusion, the application of multimodal analysis in linguistic research, particularly in the investigation of semantic distinctions among synonymous terms, holds great promise. Its ability to provide a multi-dimensional perspective enhances our understanding of language phenomena, offering a robust framework for future linguistic investigations. This methodology stands as a testament to the evolving nature of linguistic research, paving the way for innovative approaches in the field.

3. Method

The study employed the Corpus of Contemporary American English (COCA) as the primary source of textual data, while integrating image and audio data to explore the semantic nuances of the synonyms "reveal", "disclose", and "expose" through multimodal analysis. During the data collection phase, the study obtained textual data containing these synonyms from COCA, and selected images samples relevant to these concepts to provide visual and auditory dimension support. The study used some images and audio samples that are relevant to these words to cover as many scenarios as possible. Following this, the study applied a multimodal data analysis approach to explore the semantic usage patterns of these synonyms from the perspectives of text and image. By integrating these diverse data modalities, we gained a more comprehensive and profound understanding of the meanings of these synonyms, offering a broader and deeper perspective for our research. The following are the descriptions of some images and audio:

"Reveal":

Scenario 1: A magician reveals a delightful surprise by elegantly unveiling a gift box during a live performance.

Scenario 2: An archaeologist carefully removes the layers of dirt from the surface of an artifact, exposing a remarkably well-preserved ancient relic.

Scenario3: A gesture of lifting the stage curtain to reveal a new performance.

Audio sample: A sudden "click" sound, as if unveiling a hidden secret.

"Disclose":

Scenario 1: The government publicly discloses and confers an award upon the donor.

Scenario 2: An entrepreneur openly shares the company's financial statements and performance data with the public.

Scenario 3: A person formally submits documents for review, it implies the disclosure and sharing of information.

Audio sample: A recording including people announcing important information.

"Expose":

74

Scenario 1: A series of photos illustrate the environmental damage to a wildlife habitat, exposing the severity of an ecological issue.

Scenario 2: A journalist uncovers a scandal during an investigation, exposing improper behavior by an institution.

Scenario3: A biologist uses magnifying glass to see series of animal footprints, suggesting the concept of exposure.

Audio sample: The sound of the wind, accompanied by an object laid bare to the natural elements.

4. Results and Discussion

4.1 Corpus Analysis

From Table 1 and Figure1, it is evident that the total frequency of "reveal" reaches an astonishing 18,590 occurrences, while "disclose" has a total frequency of 3,950 occurrences, and "expose" stands at 5,139 occurrences. Impressively, the frequency of "reveal" surpasses the combined total of the other two words by more than double. This clearly indicates that, in terms of frequency, "reveal" holds an absolute advantage among these three words. The reason for the widespread familiarity with this vocabulary is due to its high societal impact, outstanding expressiveness, and extensive application in various domains. Its high frequency reflects its broad usage in different contexts. Based on the analysis of these high-frequency words, it can be concluded that English learners need to proficiently master this commonly used word. "Reveal" frequently appears in academic papers, followed by frequent occurrences in magazine articles. This suggests that this word is more commonly used in formal register genres. In the United States, "reveal" is also commonly used in news reporting, literary works, and other genres. However, "disclose" tends to appear more often in American writing in newspapers and academic articles, with relatively fewer instances in novels. This may be because people tend to perceive novels as an artistic form used for appreciating plots and characters, rather than for expressing the author's viewpoints or themes. Furthermore, the word "expose" is also relatively common in American academic writing and magazines. In summary, the usage frequency, and preferences of these words in different genres reflect their contextual characteristics, and gaining a deeper understanding of these characteristics can assist English learners in better comprehending and using these words.

Table1. Genre Distribution

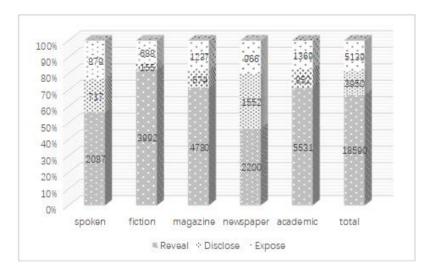| genre	spoken	fiction	magazine	newspaper	academic	total
Reveal	2087	3992	4780	2200	5531	18590
Disclose	717	155	674	1552	852	3950
Expose	879	688	1237	966	1369	5139

Figure 1. Genre Distribution

Several statistical tests are available to measure the strength of collocations. This study employs Mutual Information (MI) as a metric for collocational strength. Mutual Information quantifies the mutual attraction between a node word and its collocates, with its value reflecting the degree to which a vocabulary item possesses strong collocational tendencies. A higher Mutual Information value indicates a stronger collocational relationship between two words (Liang et al., 2010). Through extensive statistical analysis of a large vocabulary, it was found that Mutual Information values among nouns tend to be relatively low. Generally, when the Mutual Information value exceeds 3, it signifies a close collocational relationship (Hunston, 2002). In Tables 2, 3, and 4, I list the top ten nouns that are closely related to each word.

From the Mutual Information values in Table 2, it can be observed that the collocational strength between "reveal" and "secrets" is the highest. This collocation carries a forward-looking semantic aspect, as "reveal" emphasizes the disclosure of concealed content. When this word is used, it often implies that people are about to learn some unknown information. This semantic prosody can be either positive or negative, and it may be used for explaining or speculating about an event, as well as for incorrect predictions or inferences. Typically, "reveal" is collocated with abstract nouns such as "secrets", "truth", and "facts", which denote the occurrence of something under specific conditions or circumstances. Since these nouns are semantically neutral, the semantic prosody of "reveal" remains neutral as well. In English, people usually determine its specific meaning based on the context when using these vocabulary items.

Table 2. Noun Collocates of "Reveal"

No.	Context	Freq.	Total	MI
1	secrets	207	10810	7.2
2	truth	33	3296	6.26
3	workings	17	1995	6.03
4	facts	12	1571	5.87
5	whereabouts	13	1868	5.74
6	identities	36	6418	5.43
7	details	201	39032	5.3
8	similarities	22	4881	5.11
9	clues	24	5990	4.94
10	shortcomings	10	2503	4.94

Table 3 reveals the closest collocational relationship between "disclose" and "donors". Additionally, we can observe that in this collocational context, there is a certain relationship and connection between the entities described by these two words, giving them an asymmetrical characteristic. Semantically, we can see that "disclose" is closely related to words such as "donors", "finances", "holdings", "fees", "names", "compensation", and "status". These words each represent different aspects of entity characteristics or describe the relationships between them. These terms involve issues related to personal property and can be used to express someone's possession of certain assets or items at a specific moment. It's worth noting that the term "donors" carries a positive connotation semantically; Therefore, "disclose" in this context also carries a positive semantic prosody as shown in Figure 2. This illustrates that in English, there are verbs that can convey a positive sense, especially when it comes to matters related to property or related issues.

Table 3. Noun Collocates of "Disclose"

No.	Context	Freq.	Total	MI
1	donors	54	6316	8.27
2	finances	14	4953	6.67
3	information	490	183701	6.59
4	details	95	39032	6.46
5	holdings	10	5096	6.15
6	identities	12	6418	6.08
7	names	56	39956	5.66
8	compensation	13	9891	5.57
9	fees	19	16182	5.41

| 10 | status | 55 | 48352 | 5.36 |

Figure 2. The Collocation of "Disclose"

As shown in Table 4, the combination of "expose" with "weaknesses" exhibits the strongest collocation strength. However, when these two words are combined, their lexical meanings undergo varying degrees of change. From a semantic prosody perspective, it is evident that 80% of the ten most common collocations in these examples carry a negative semantic prosody. In these collocations, we can observe that "expose" is frequently paired with words such as "weaknesses", "hypocrisy", "wrongdoing", "flaws", "corruption", "abuses", "fraud", and "risks". These words are typically used to describe or reveal negative, unethical, or flawed situations. For instance, "expose weaknesses" implies showing shortcomings or vulnerabilities, while "expose corruption" means uncovering corrupt behavior. These collocations are often associated with negative events, unethical behavior, or potential dangers. In terms of grammatical meaning, these collocations signify uncertain or unpredictable situations or events. When people use "expose", it usually implies the disclosure or publicizing of certain adverse information or circumstances, which may lead to negative consequences. Therefore, in this context, the term "expose" carries a negative semantic prosody, emphasizing that the revelation of information may have adverse effects on society, individuals, or organizations. This in-depth semantic analysis aids English learners in better understanding and using the word "expose", particularly when describing negative situations or events.

Table 4. Noun Collocates of "Expose"

No.	Context	Freq.	Total	MI
1	weaknesses	33	4281	7.74
2	hypocrisy	13	2258	7.32

3	wrongdoing	14	2626	7.21
4	flaws	14	3773	6.69
5	corruption	39	11044	6.61
6	abuses	14	4461	6.44
7	secrets	16	10810	5.36
8	audiences	12	9635	5.11
9	fraud	14	11702	5.05
10	risks	19	19970	4.72

Within the corpus analysis, notable distinctions among "reveal", "disclose", and "expose" have been observed. Specifically, it was noted that "reveal" exhibits a markedly higher frequency of occurrence in the text corpus compared to "disclose" and "expose". Furthermore, "reveal" tends to collocate with abstract nouns, indicating a propensity towards unveiling abstract concepts. Conversely, "disclose" and "expose" more frequently align with specific terms, implying a tendency towards unveiling concrete information or facts.

4.2 Image Data Analysis

In the analysis of image data, it is evident that each synonym is associated with distinct visual characteristics. For instance, images linked with "reveal" typically depict a process of gradual revelation or display, whereas those associated with "disclose" and "expose" may portray a more direct scenario of revelation or presentation. The study employed carefully selected images to exemplify the nuanced semantics of the synonymous terms "reveal", "disclose", and "expose". For example, "reveal", the study chose an image portraying the dramatic gesture of lifting a stage curtain, symbolizing the intentional act of unveiling something new, such as a performance. In contrast, for "disclose", the selected image featured a person receiving an award, representing an official process of disclosing recognition and appreciation to the recipient. Lastly, this is an image depicting a journalist investigating a scandal, thereby exposing the institution's improper behavior. This exposed the institution's improper behavior. This natural setting underscores the act of bringing something to light. These visual representations not only capture the essence of each term but also provide valuable insights into their subtle distinctions, enriching our understanding of their semantic nuances.

4.3 Audio Data Analysis

In the audio data, distinctive acoustic features were observed in samples corresponding to each synonym. These features encompass variations in pitch, speech rate, emotional tonality, among others, providing insights into the diverse usages of these synonyms at the auditory level. Based on the analysis of the provided sound data, specific conclusions have been reached for each of the synonymous terms. The sound associated with "reveal" is characterized by a sudden "click" sound, evoking the notion of unveiling a hidden secret. In contrast, "disclose" is represented by recordings of individuals making significant announcements, conveying a formal and authoritative disclosure. On the

other hand, "expose" is distinguished by the sound of the wind, complemented by the imagery of an object laid bare to the natural elements. This auditory exploration has enabled a deeper comprehension of the nuanced distinctions between these synonyms, offering a comprehensive understanding of their contextual usage. By integrating both visual and auditory dimensions, the study provides a novel perspective that enriches research in the realms of language acquisition and education.

Through the integration of findings derived from the analysis of text, image, and audio data, a comprehensive understanding of the semantic nuances and usage disparities among "reveal", "disclose", and "expose" has been attained at the multimodal level. This not only furnishes an enriched grasp of these synonyms, but also introduces novel research perspectives for future investigations in teaching and language acquisition. These research findings underscore the significance of multimodal data analysis in synonym studies, concurrently offering learners of language more precise guidance in the application of these synonyms.

5. Conclusion

In summary, the investigation has yielded significant insights into the usage nuances of the synonymous terms "reveal", "disclose", and "expose". The analysis revealed distinct patterns in their prevalence and collocational preferences, shedding light on their contextual nuances. Additionally, the examination of image data unveiled visual characteristics associated with each term, providing a multi-modal perspective on their semantic distinctions. Moreover, the analysis of audio data highlighted unique acoustic features corresponding to each synonym, further enhancing our understanding of their auditory connotations. This comprehensive approach, integrating both visual and auditory dimensions, has enriched our comprehension of these synonyms. It has not only deepened our grasp of their subtle differences but also contributed to the broader discourse on language acquisition and education. These findings underscore the importance of considering multi-modal approaches in linguistic studies, offering a fresh perspective for future research in this domain.

References

Bill, L. O. U. W. (2000). Contextual prosodic theory: Bringing semantic prosodies to life. *Hef fer Chris, & Saunt son Helen (eds.)*, 48-94.

Firth, J. R. (2020). *Papers in linguistics, 1934-1951*. London: Oxford University Press.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524773

Partington, A. (2004). "Utterly content in each other's company": Semantic prosody and semantic preference. *International journal of corpus linguistics*, *9*(1), 131-156. https://doi.org/10.1075/ijcl.9.1.07par

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information fusion*, *37*, 98-125.

https://doi.org/10.1016/j.inffus.2017.02.003

Shahzadi, A., Asghar, A., & Javed, S. (2019). Effectiveness of corpus in teaching English synonyms. *CORPORUM: Journal of Corpus Linguistics*, *2*(1), 51-66.

Stubbs, M. (2009). The search for units of meaning: Sinclair on empirical semantics. *Applied linguistics*, *30*(1), 115-137. https://doi.org/10.1093/applin/amn052

He, X. D. (2003). *A Bilingual English-Chinese Dictionary of Synonyms*. Beijing: Business Book Printing Place International Co., Ltd.

Liang, M. C., Li, W. Z., & Xu, J. J. (2010). Using Corpora: A Practical Coursebook. *Foreign Language Teaching and Research Press*, 90-110.

Lu, J. (2010). A corpus-based study on collocational behavior and semantic prosody of near synonyms in Chinese learner English. *Modern Foreign Languages*, (03), 276-286+329-330.

Ren, P. H. (2008). Teaching and Learning of Synonyms: With Common and Ordinary as Example. *Journal of Pla University of Foreign Languages*, (04), 57-60.

Wang, X. F., Mark, D., & Liu, G. H. (2008). A Good Platform for English Teaching Researchers and Learners: the Corpus of Contemporary American English (COCA). *Media in Foreign Language Instruction*, (05), 27-33.

Wang, C. Y. (2009). A Corpus Based Study of Distinction of English Synonym for Chinese Learners.. *Foreign Languages and Their Teaching*, (06), 27-31.

Wei, N. X. (2006). *Definition and Research Framework of Collocations*. Shanghai: Shanghai Jiao Tong University Press.

Wu, L., & Jiang, H. (2016). Synonym Discrimination Based on COCA—Taking an example of choose and select. *Journal of Hunan Industry Polytechnic*, (06), 93-97.