



SEMANTIC PREFERENCE IN “DIARY OF A WIMPY KID”

Do Truong Nhat Thinh; Truong Bach Le*

University of Foreign Languages and International Studies, Hue University

57 Nguyen Khoa Chiem Str, Hue, Vietnam

* Correspondence to **Truong Bach Le** < mtletruong@gmail.com >

(Received: January 16, 2025; Accepted: December 29, 2025)

Abstract: This paper reports on part of a study investigating how the representation of the father character, Frank Heffley, is constructed through semantic preference in the first five volumes of the Diary of a Wimpy Kid series by Jeff Kinney. By using concordance analysis to analyze corpus data with the use of Sketch Engine, the study identifies semantic preference reflecting recurring themes associated with this parent. Results reveal six key themes which illustrate Frank Heffley as a complex figure who is caring and protective but also strict, and occasionally disconnected from his children due to generational differences. The paper ends with implications that are relevant to English language study and teaching, as well as literary analysis.

Keywords. representation, semantic preference, concordance analysis, Diary of a Wimpy Kid

1. Introduction

Semantic preference is a relationship between a word or phrase and a set of semantically related words (Sinclair 1996, 2004; Stubbs, 2001; Partington, 2004). In other words, a lexical item tends to co-occur with other items from a particular semantic field. It is a feature of the collocates, not between individual words. To illustrate, in his seminal work, Stubbs (2001) reported the high frequency of the word “large” collocating with words for “quantities and sizes”, including numbers, scales, parts, amounts, and quantities. From the perspective of linguistic analysis, semantic preference plays a significant role in understanding how words function within language and how they contribute to the construction of meaning, evaluation, and representation. That is to say, semantic preference empowers not solely revealing semantic relationships and meaning within text or discourse but also performing lexical priming and corpus analysis. In the context of a literary corpus, semantic preferences contribute to constructing representations in discourse by highlighting the collocates of people, places, ideas,

and events. According to Heritage and Taylor (2024), observing a word's collocates is useful for the analysis of representation because collocates can reveal how a word is used, uncover its meaning, and show how the phenomenon it denotes is evaluated in discourse.

Diary of a Wimpy Kid by Jeff Kinney, is a famous American series that is written in a diary format, combining handwritten text, dialogues, and comic illustrations to show the misadventures of the protagonist, Greg Heffley. He is a teenager who struggles every day with the challenges of a middle school student. Despite the humor and relatability to most readers, many comments on Goodreads.com and Commonsensemedia.org spark divisive debate due to the characters' questionable and unlikeable personalities and behaviors, particularly Greg's father, Frank Heffley. Moreover, it can be said that semantic preference remains underexplored in understanding character representation among fictions. Existing research, while illuminating, primarily focuses on other areas. For instance, Zuraida and Simatupang (2023) explored the semantic preferences and prosody of "Mom" and "Dad" in young adult novels, reporting distinct associations with family. However, their study concentrated on parental terms in a specific genre, rather than delving into individual character representation within a fictional narrative. Similarly, Selmistratis (2020) investigated semantic preference in synonymous adjectives and academic writing, highlighting its role in lexical choice and genre conventions. Begagić (2013) examined the collocation "make sense" across registers, demonstrating the influence of context on semantic preference. Even Won et al. (2025)'s work on AI and language assessment, while innovative, focuses on evaluating non-native writing rather than analyzing character portrayal. These studies collectively underscore the need for research focusing on semantic preference analysis in popular fiction, specifically in the representation of characters in "*Diary of a Wimpy Kid*".

These have prompted the current study to explore semantic preference to provide a detailed way to understand the main character's father by exploring semantic contexts, including key themes and characteristics, to which he is typically referred.

Hence, this study addresses the following research question: *What are the dominant themes related to the character Frank Heffley as revealed through collocates co-occurring with the node word "dad" in Diary of a Wimpy Kid?*

2. Semantic preference

Semantic preference has become increasingly important in corpus linguistics over the last two decades with the help of large corpora development. It assists in uncovering hidden meanings and usage patterns limited by the subjective interpretations of traditional dictionary definitions. Semantic preference refers to the tendency by which a word is used with a set of words that share a similar semantic field (Ghozlen, 2022; Partington, 2004; Begagić, 2013). In

other words, a certain number of words or node items show a preference to co-occur with words or other items that share equal semantic features (Partington, 2004; Stubbs, 2007). Much of the literature on semantic preference pays particular attention to the relationship between a word and a group of words that share similarities. For instance, the word *rising* is often found to co-occur with economic terms such as *incomes*, *prices*, *unemployment*, etc. (Baker, 2023). These words share a common semantic field owing to their frequent appearance in the context of economics and finance. That is to say, this connection elucidates a semantic preference for economic context. Similarly, the word *glass* typically appears with words such as *milk*, *water*, *sherry*, and *champagne*, indicating a semantic preference for the acts evolving around drinking (Baker, 2023). As a result, there is a consensus among linguists that semantic preference is defined as the relationship between a word and a set of semantically related words. The origin of semantic preference can be attributed to Sinclair's (1996) model of extended units of meaning.

Semantic preference can be said to be context-dependent, as it can change based on the specific context in which it is used. The reason is that the collocates of a word (what semantic preference is about) are different depending on the genre or the domain in which they are used. For example, the node words (words whose semantic preference is being analyzed) and their surrounding words can be different from the genre of news reports to the genre of scientific articles (Partington, 2004). From the relationship between the node word and other items surrounding it, the study of semantic preference is the identification of lexical sets. For example, in his study in 2023, Baker reported *refugees* have a semantic preference for quantification, as this word co-occurs with numerical data such as population. These lexical sets not only emphasize underlying meanings and potential categories that the speakers aim to represent but, more importantly, allow researchers to examine patterns in language use that reflect representations, attitudes, and perceptions in discourse analysis. As a result, semantic preference can be identified with both corpus analysis and researcher interpretation (Bednarek, 2008). This can be explained by the fact that the corpus provides information concerning word occurrences, whereas it is the researcher's job to analyze and categorize these collocates to determine the semantic preference of the node word. Another point to notice is about collocation, semantic preference is closely related but the two concepts are different in terms of their implications. Collocation and semantic preference are two separate but related language concepts. While collocation measures the statistical frequency of word co-occurrence (e.g., how often *America* pairs with specific terms like *policy* or *leadership*), semantic preference identifies the shared meaning categories of these co-occurring words. *America* exhibits a semantic preference for agentive roles, reflected in its tendency to pair with verbs like *shape* or *impact* (Baker, 2023). This contrast highlights how word combinations reflect both quantitative patterns (collocation) and thematic associations (semantic preference).

Apart from the relation to collocation, semantic preference is also a matter of comparison with discourse prosody. The definition is relatively the same as that of semantic preference where patterns can be identified between a word or lemma and a set of related words in discourse. This leads to the challenges in making a clear distinction and according to the study of Baker (2023) and Stubbs (2001), there are two main features to set these two concepts clear which are the openness of the categories and the presence of evaluative language. Stubbs (2001) took a list of the words for *drinks* as an example for semantic preference while he mentioned a more open-ended category such as unpleasant things as discourse prosody. Subsequently, discourse prosody is regarded as being more concerned with the speakers and listeners' perceptions and emotions and their choices of language. Finally, there are linguists who regard semantic preference as a manifestation of lexical priming, which describes the process of the repeated exposure to a word in a specific context affects how we understand and use that word. This means that this process shapes our expectations of how that word is used and our interpretation of it.

3. Methodology

The study, first, created a corpus of Diary of a Wimpy Kid, spanning from volume 1 to volume 5, including Diary of a Wimpy Kid (2007), Diary of a Wimpy Kid: Rodrick Rules (2008), Diary of a Wimpy Kid: The Last Straw (2009), Diary of a Wimpy Kid: Dog Days (2009), and Diary of a Wimpy Kid: The Ugly Truth (2010). The texts from 5 books are extracted by OCR (Optical Character Recognition) provided by Google Lens and stored in PDF form before being loaded onto Sketch Engine to compile a specialized corpus. The final resultant corpus can be named and loaded as the chosen corpus for further study. Second, in this stage of the study, the node word "dad" is decided as the chosen search term based on a pilot study with a small sample of the corpus (a small-scale preliminary analysis of the corpus to determine suitable search terms). This is because, as a parent, Frank Heffley, whom Greg mentions as "dad" throughout his diary, has a vital role to play in the relationship with not only Greg but also the whole family. Third, a concordance of the search term "dad" is obtained with 559 concordance hits. Fourth, a random sample of 100 lines was generated (see Figure 1) and the thinning process reduced this number to 77 (irrelevant hits of those related to other characters than dad, e.g., Rowley and Chirag, are passed over). Fifth, the study of these randomized samples continues with sorting the whole concordance of each node word to the left and right. The final step is manual annotation, which can be understood as annotating concordance lines with labels to categorize and group concordance lines for further analysis of thematic categorization (see Figure 2). From this, semantic preference is identified for providing insights into the representation of the examined character.

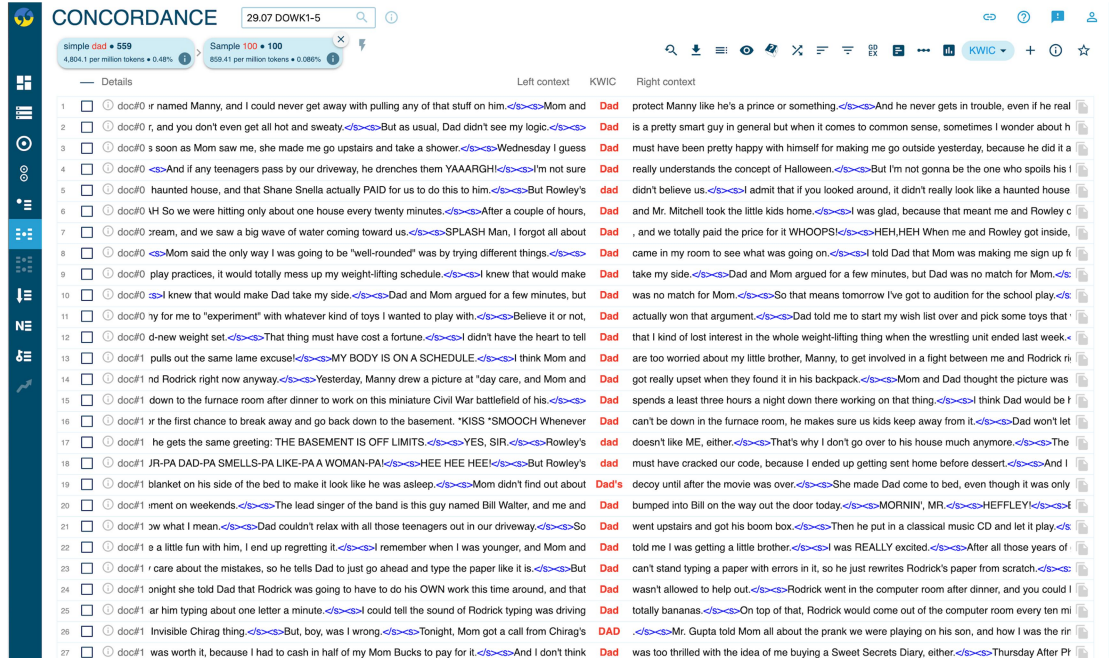


Figure 1: A random sample concordance lines of the node word “dad”

No.	Random concordance lines			Collocates	Word form	THEME
1	never get away with pulling any of that stuff on him. </s></s> Mom and	Dad	protect Manny like he's a prince or something. </s></s> And he never	protect	verb	Caring and supportive father
2	get all hot and sweaty. </s></s> But as usual, Dad didn't see my logic. </s></s>	Dad	is a pretty smart guy in general but when it comes to common	pretty smart	adj	Caring and supportive father
3	, she made me go upstairs and take a shower. </s></s> Wednesday I guess	Dad	must have been pretty happy with himself for making me go	pretty happy	adj	Caring and supportive father, delighted
4	pass by our driveway, he drenches them YAAARGH! </s></s> I'm not sure	Dad	really understands the concept of Halloween. </s></s> But I'm not	understand	verb	Strained father-son relationship
5	one house every twenty minutes. </s></s> After a couple of hours,	Dad	and Mr. Mitchell took the little kids home. </s></s> I was glad,	took	verb	Caring and supportive father
6	of water coming toward us. </s></s> SPLASH Man, I forgot all about	Dad	, and we totally paid the price for it WHOOPS! </s></s> HEH,HEH When me	forgot	verb	Strained father-son relationship
7	going to be "well-rounded" was by trying different things. </s></s>	Dad	came in my room to see what was going on. </s></s> I told Dad that Mom was	came	verb	Caring and supportive father
8	mess up my weight-lifting schedule. </s></s> I knew that would make	Dad	take my side. </s></s> Dad and Mom argued for a few minutes, but Dad was	take argued	verb verb	Caring and supportive father
9	Dad take my side. </s></s> Dad and Mom argued for a few minutes, but	Dad	was no match for Mom. </s></s> So that means tomorrow I've got to	no match	noun	
10	kind of toys I wanted to play with. </s></s> Believe it or not,	Dad	actually won that argument. </s></s> Dad told me to start my wish list	told	verb	Caring and supportive father

Figure 2: Manual annotation of the first 10 concordance lines of the node word “dad”

4. Findings and discussions

As can be seen from the random sample of 77 concordance lines, the examination of the node word “dad” and the immediate words around it contribute to the representation of this

character. As a result, a total of 6 main themes were categorized, which are listed with examples of concordance lines shown from Section 4.1 to Section 4.6.

4.1. THEME 1: Caring and supportive father

By analyzing the surrounding words and phrases, dominant themes that emerge offer insights into the writer's portrayal of the father figure. Chief among these is Dad being caring, delighted, and well-meaning while encouraging his children to be strong and well-rounded.

Example 1

Line 1: [...] never get away with pulling any of that stuff on him. Mom and Dad protect Manny like he's a prince or something. And he never [...]

Example 2

Line 47: [...] freeze to my legs. Every time Mr. Litch called a huddle, Dad made me get off the bench and go join the rest of the team. Have [...]

It is learned from examples 1 and 2 that Greg associates his dad with collocates such as *protect* and *made me get off the bench and go join the rest of the team*, and *stay in touch*. These collocates show the preference for words and phrases associated with guidance, encouragement (to be physically active for better health), and practical care (to show concern for his son's well-being). Dad also shows his caregiving toward his children when he is *too worried*. This suggests that these collocates paint a picture of a caring, delighted father who provides emotional support.

Example 3

Line 12: [...] the same lame excuse! MY BODY IS ON A SCHEDULE. I think Mom and Dad are too worried about my little brother, Manny, to get [...]

The concordance analysis with example 3 also provides more evidence that Frank's caring nature can sometimes lead to conflicts and his effort to maintain a positive relationship with other family members.

4.2. THEME 2: Parental authority and responsibility

Another emerging theme from the concordance analysis is that Dad's roles in this family involve the elements of decision-making, responsibility, and discipline. The following examples 4 and 5 providing collocates, such as *told me I was getting a little brother* and *switched me over to private school* suggest that together with mom, dad bears a certain responsibility in making decisions, giving information, and enforcing consequences in raising the children.

Example 4

Line 19: [...] up regretting it. I remember when I was younger, and Mom and Dad told me I was getting a little brother. I was REALLY excited.

Example 5

Line 75: [...] make me wonder if I'd be getting a better education if Mom and Dad switched me over to private school. Then Nurse Powell said [...]

In the picture of family dynamics, he is inferior compared to his mom's authority in some situations, particularly when Greg thinks his dad is *no match for mom*.

Example 6

Line 9: [...] Dad take my side. Dad and Mom argued for a few minutes, but Dad was no match for Mom. So that means tomorrow I've got to [...]

The patterns in collocates found in this theme demonstrate a language preference that is related to responsibility and authority in the family.

4.3 THEME 3: Personal interests

A recurrent theme of the node word "dad" in the concordance is the father's hobbies and interests, which he engages deeply in outside his life of parenting. His hobbies are described with collocates, such as *at least three hours a night* and *was working on* in examples 7 and 8.

Example 7

Line 14: [...] to work on this miniature Civil War battlefield of his. Dad spends at least three hours a night down there working on that [...]

Example 8

Line 29: [...] in what Dad was working on, So Dad told him all about it. Dad gave Uncle Joe this big speech about the 150th regiment and [...]

He sometimes shows his playfulness and delights in the collocate of *decoy*.

Example 9

Line 16: [...] to make it look like he was asleep. Mom didn't find out about Dad's decoy until after the movie was over. She made Dad come to bed [...]

These collocates tend to be linked to Frank's personal interests or eccentric behaviors, representing this character with his unique interests and relatively quirky behaviors to entertain him and those around him and make him a more relatable figure in the series.

4.4. THEME 4: Strict parenting

Dad maintains firm rules and high expectations in his house, where he raises his children. Particularly revealing is the study of concordance lines pictured the father as being strict with collocates, such as *never let me bring anything to church when I was his age* and *'d ground me if they found out we watched a horror movie* (see Examples 10 and 11).

Example 10

Line 44: [...] of stuff with us to keep him entertained. Believe me, Mom and Dad never let me bring anything to church when I was his age. Mom [...]

Example 11

Line 58: [...] more freaked out than I was. I knew we couldn't go to Mom and Dad this time because they'd ground me if they found out we [...]

He is also a serious, overactive, and occasionally psychotic father, as the node word is found being surrounded by collocates, such as *the people at fast- food restaurants are idiots* and *driving dad totally bananas* (see Examples 12 and 13). These negative traits can be explained by the fact that his strictness is motivated by his desire to provide guidance and ensure his children's well-being.

Example 12

Line 21: [...] minute. I could tell the sound of Rodrick typing was driving Dad totally bananas. On top of that, Rodrick would come out of [...]

Example 13

Line 61: [...] Dad wouldn't pull over, because he said the people at fast-food restaurants are idiots.

Collectively, these collocates are inclined to emphasize expressions associated with Dad's role as a strict and authoritative figure who enforces rules and limits children's freedom.

4.5. THEME 5: Strained father-son relationship

Greg's father is proven to have a poor relationship with all of his sons when the node word is surrounded with collocates such as *aren't cut out to be for that kind of father-son stuff* and *'s kind of an unpredictable guy*.

Example 14

Line 59: [...] or something. Or play catch in the... But the thing is, me and Dad just aren't cut out to be for that kind of father-son stuff.

Example 15

Line 68: [...], because it looked an awful lot like his handwriting. But Dad's kind of an unpredictable guy, so I guess wait until Sunday to [...]

The semantic preference drawing from collocates relates to generational differences and communication challenges, illustrating Dad as a parent who falls short of his children's expectations. Further evidence from the concordance analysis shows the concerns for his power that lead to Greg's frustration, and even there is the case in which Frank projects his own fears and insecurities onto his son.

4.6. THEME 6: Dislike for teen culture and heavy metal

Frank Heffley is also a father who generally disapproves of some juvenile behaviors and detests heavy metal music played by his first son, Rodrick Heffley. These are presented clearly as the node word "dad" is surrounded with collocates, such as *'s starting to wish he had worded Rodrick's punishments little differently* and *was standing in the back of the gym with cotton balls*.

Example 16

Line 33: [...] over every day and practices down in the basement. I think Dad's starting to wish he had worded Rodrick's punishments a little differently.

Example 17

Line 36: [...] me while Rodrick's band played, so I looked around for him. Dad was standing in the back of the gym with cotton balls [...]

These collocates often align with words and phrases that convey discomfort and avoidance, highlighting Dad's detest for heavy metal music, and concerns for adolescents' negative behaviors.

From the identification and analysis of the six dominant themes related to the father character, this paper adds to our understanding of how semantic preference operates in the representation of fictional characters, building upon previous work in corpus linguistics (Baker, 2023; Cheng, 2012; Fischer-Starcke, 2010; Heritage & Taylor, 2024). The application of semantic preference analysis is seen in this study's character representation, social perceptions, and broader discourse analysis, as Baker (2023) and Heritage and Taylor (2024) discuss how this method can be employed to study representation, argumentation, and other aspects of language use.

The present study found the node word "dad" collocates with semantic fields of a caring and supportive father, parental authority, and responsibility. Similarly, Mautner (2007) discovered that "elderly" frequently appears with words showing disability, care, and vulnerability. This also accords with the capacity of semantic preference in revealing implicit meanings and societal attitudes being embedded in language use. The findings can be said to

provide a rich source of perspectives on the themes of parents and children in literature. Parental authority and discipline are acknowledged as a significant aspect of family in both this study and Zuraida and Simatupang's (2023) and Knowles and Malmkjær's (1996). Similar to Greg's father, the mom character in Alec Greven's novel is represented by setting rules and expectations, while the father's role in a Victorian society includes upholding moral standards and requiring children's obedience. Moreover, it is noteworthy that the study recognizes the complexities in family relationships, such as the strained father-son dynamic. It reflects the challenges of fatherhood, including moments of frustration found in the work by Zuraida and Simatupang (2023). Regarding the specificity of character representation, one surprising variable that was found to be significantly associated with this finding was the identification of multiple, sometimes contradictory, themes concerning each character. This may be explained by the methodology and focal points of this study compared to the broader characterization found in other works (Baker, 2023; Fischer-Starcke, 2010; Heritage & Taylor, 2024; Knowles & Malmkjær, 1996).

5. Conclusions

Taken together, the semantic preference analysis of the node word "dad" uncovered a total of six themes, namely *strict parenting*, *strained father-son relationship*, *dislike for teen culture and heavy metal*, *supportive father*, *personal interests*, and *parental authority*. In other words, the word "dad" in the first five volumes of *Diary of a Wimpy Kid* shows a semantic preference for collocates related to caregiving, authority, strictness, control, and generational gaps. These topics and themes collectively depict Frank Heffley as a complex figure who is caring and protective but also strict and occasionally disconnected from his children due to generational differences.

As a key concept in contributing to the analysis of how characters are represented, semantic preference implies several key points of significance in language analysis for identifying representations, language teaching and literary analysis. One of the major implications is identifying dominant themes associated with other characters. By identifying how words coexist, semantic preference analysis contributes to a deeper understanding of how these characters are represented. This, in turn, confirms the theories and applications of semantic preference in corpus linguistics literature. The research also suggests a link between language learning and teaching, particularly for English-majored students. This is because corpus-based analysis can enhance comprehension of sophisticated English, improve vocabulary usage, and reduce inappropriate language use. This approach provides authentic and rich sources of text for tasks like discussions, dialogues, and essay writing, improving language skills. This study also highlights the importance of obtaining empirical linguistic data for literary analysis, eliminating subjective interpretations.

References

1. Baker, P. (2023). *Using corpora in discourse analysis* (2nd ed.). Bloomsbury Academic.
2. Bednarek, M. (2008). Semantic preference and semantic prosody re-examined. *Corpus Linguistics and Linguistic Theory*, 4(2), 119-139. <https://doi.org/10.1515/CLLT.2008.006>
3. Begagić, M. (2013). *Semantic preference and semantic prosody of the collocation make sense*. *Jezikoslovlje*, 14(2-3), 403-416.
4. Ben Ghazlen, B. B. (2022). 'A man is known by the company he keeps': A critical overview of semantic prosody. *Legon Journal of the Humanities*, 33(2), 59-79. <https://doi.org/10.4314/ljh.v33i2.3>
5. Cheng, W. (2012). *Exploring corpus linguistics: Language in action*. Routledge. <https://doi.org/10.4324/9780203802632>
6. Kinney, J. (2007). *Diary of a wimpy kid*. Amulet Books.
7. Kinney, J. (2008). *Diary of a wimpy kid: Rodrick rules*. Amulet Books.
8. Kinney, J. (2009a). *Diary of a wimpy kid: The last straw*. Amulet Books.
9. Kinney, J. (2009b). *Diary of a wimpy kid: Dog days*. Amulet Books.
10. Kinney, J. (2010). *Diary of a wimpy kid: The ugly truth*. Amulet Books.
11. Fischer-Starcke, B. (2010). *Corpus linguistics in literary analysis: Jane Austen and her contemporaries*. Continuum.
12. Heritage, F., & Taylor, C. (Eds.). (2024). *Analysing Representation: A Corpus and Discourse Textbook* (1st ed.). Routledge. <https://doi.org/10.4324/9781003350972>
13. Knowles, M., & Malmkjær, K. (1996). *Language and control in children's literature*. Routledge, Taylor & Francis Group.
14. Mautner, G. (2007). Mining large corpora for social information: The case of elderly. *Language in Society*, 36(01), 51-72. <https://doi.org/10.1017/S0047404507070030>
15. 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>
16. Selmistraitis, L. (2020). Semantic preference, prosody and distribution of synonymous adjectives in COCA. *GEMA Online Journal of Language Studies*, 20, 1-18.

17. Sinclair, J. (1996). The search for units of meaning. *Textus*, 9, 75–106.
18. Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. Routledge.
19. Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Blackwell.
20. Stubbs, M. (2007). On texts, corpora and models of language. In M. Hoey, M. Mahlberg, M. Stubbs, & E. Teubert (Eds.), *Text, discourse and corpora: Theory and analysis* (pp. 127–161). Continuum.
21. Won, Dong-Ok & Shin, Yu & Kim, Ho-Jung & Yoo, Isaiah. (2025). Advancing Language Assessment with GPT: Is It Nonnative-Language Friendly? *Language Assessment Quarterly*, 22,1-28.
22. Zuraida, I., & Simatupang, E. C. (2023). The collocation of the words “mom” and “dad” in Alec Greven’s novels: A corpus linguistic analysis of semantic preferences and prosody. *English Review: Journal of English Education*, 11(3), 667–674. <https://doi.org/10.25134/erjee.v11i3.8818>