

Text Analysis for Classroom Resources and Materials: Considerations for English Teachers

Saurabh Anand

Many English teachers see the benefit of analyzing texts. Through text analysis, the structure, vocabulary, and words used in fiction or non-fiction texts can be organized and leveled. Text analysis is especially important for English teachers, as it involves reviewing classroom resources to match students' age, interests, and language skills. This helps students who are just starting to learn English by adjusting the texts selected to their reading level. This approach also fosters a positive attitude towards reading and writing (Häikiö et al., 2011), which can directly improve students' understanding of content (Perfetti & Stafura, 2014) and lead to better assessment results (De Zeeuw, 2019). However, some teachers might not know how to effectively carry out text analysis (Accurso & Meg Gebhard, 2021). In hopes of creating a balance between content and language goals, this article guides English teachers in how to conduct a thorough text analysis for the materials they use to ultimately benefit students' learning outcomes.

Considering and Utilizing Text Analysis

Before children begin school, they often learn creative ways to understand text and communication at home by, for example, reading together and sharing stories. Understanding how early readers develop their foundational skills is crucial (Eagleton & Dobler, 2007). These skills include decoding, fluency, and vocabulary knowledge. Reading experts also stress the importance of primary school in shaping a person's lifelong reading abilities in their native language (Verhoeven & Perfetti,

2001). English teachers help children become part of a complex reading community (Gee, 2015).

Text analysis empowers English teachers to dig deeper into the material they bring to the classroom. It allows them to understand how practical and useful a text resource might be (Biemiller et al., 2014). This knowledge ensures that classroom resources align with the learning environment teachers want to create (Kim, 2017). Text analysis plays a role in helping English teachers maintain a balance between content and language goals (Williams et al., 2009). Without it, teachers might unintentionally focus more on content, leaving learners disconnected due to limited vocabulary and decoding skills.

To build comprehension skills from an early age and instill long-term language self-sufficiency (Alsaif & Milton, 2012), English teachers could use individualized text materials. By individualizing materials, English teachers can help build linguistic competence in a second language (Canga Alonso, 2015). Text analysis ensures that the resources English teachers plan to use, align with their teaching goals (Cobb, 2021). Text analysis also prompts English teachers to reflect on their teaching methods and other classroom instruction materials. It helps ensure that materials offer understandable content (Krashen, 1982) and are suitable for the students they teach. This is especially beneficial for teachers working with diverse student populations, including pre-readers and students with limited or interrupted formal education (SLIFE). By tailoring content to students' comfort levels, teachers can better prepare today's readers to become successful academics in the future (Krashen, 1982).

Text Analysis Process and Example

PebbleGO, a Hi-Lo (high interest-low level text) resource, was used for the text analysis process. Hi-Lo resources are designed to provide "highly engaging age-appropriate subject matter at a low reading level for struggling readers" (Reading Rocket, 2016, paras. 1). Throughout this detailed text analysis, the aim is to empower English teachers to ensure that the content they bring into their

classrooms is accessible to all students, including pre-readers. The next section provides an example of how a text analysis can look.

Step 1: Format

Access PebbleGO's dataset (Earth and Spaces Science sub-module and Life Sciences sub-modules), saving the information in Plain Text format (.txt). This format is readable by various software applications.

Step 2: Use text analysis tools

To gain a deeper understanding of the text, the following software tools can be used:

1. AntConc: This software is freely available and is a tool used to help generate a list of words along with their frequencies. This was completed for both sub-modules.
2. Text Lex Compare Software: This tool is useful in identifying the shared vocabulary between the two sub-modules.
3. Vocabprofiler: This software is used to sort words based on how often they show up in English text (Friginal, 2018). The software classifies words broadly into four categories:
 - *Category 1:* The most frequent 1000 words of the English language (K1 words).
 - *Category 2:* The most frequent 2000 words of the English language (K2 words).
 - *Category 3:* Academic Word List (AWL) from the English language.
 - *Category 4:* Words that do not fall in the above categories, such as content words. Gardner (2008) has referred to this specific kind as a specialized words.

Step 3: Calculate Norming Value

In this example, the two sub-modules did not have the same text length. Therefore, to make them suitable for further analysis, 1000 words were used from each text. The formula recommended by Friginal

(2018) was used: $N(f) = (\text{number of occurrences} \div \text{total number of words}) \times 1000$.

Step 4: Summarize the Data

The percentages of each vocabulary type (K1, K2, AWL, and content words) are quite similar. Notably, the AWL vocabulary is the least prevalent. Instead, we see a higher presence of high-frequency vocabulary (K1 and K2). This inclusion of common words helps make the content more accessible, especially for readers who may need additional support (Schmitt & Schmitt, 2014). It is evident that these texts are not ideal for promoting academic language growth among students. They contain much K1 and K2 vocabulary, which makes them easy to read and boosts fluency. However, there is also a large amount of science vocabulary. Readers who have access to these resources should learn about the science vocabulary before reading the text. This exposure can help them read other texts in the same content area and participate in discussions (Gardner, 2004, 2008).

Table 1 lists 133 content words that are common to both sub-modules. This repetition means that students who read across these modules encounter these content words multiple times. Many researchers suggest that this recycling of vocabulary can be highly beneficial for language learners, especially during the early stages of language acquisition. It also supports reading and fluency development without causing undue stress (Chang and Pang, 2020).

Table 1*Total Shared Off-list Words between Earth and Space and Life Sciences Sub-Modules*

acorn	burrow	dinosaur	hawk	monarch	prey	soak
Africa	butterfly	dioxide	hibernate	moss	raspberry	spice
ant	cactus	dwarf	huge	munch	reef	sponge
Antarctica	camel	earthworm	hummingbird	nectar	reptile	sprout
arctic	Canada	equator	hump	oak	robin	squid
Asia	canopy	evergreen	India	oats	sac	squirrel
asparagus	carbon	evergreens	kelp	opossums	salmon	swamp
bacterium	cattails	fern	lawn	owl	saltwater	tentacle
beaver	cauliflower	finger nail	Lianas	oxygen	seaweed	tiny
bee	climate	fog	lichen	pacific	seep	toad
beetle	continent	forever	lily	pea	shark	tropics
blend	coral	fox	lion	petal	shrimp	tundra
bloom	crab	freshwater	lodge	pine	shrub	turtle
blowhole	crocodile	frog	mammal	polar	sip	underground
bog	crunch	fungus	mangrove	pole	skeleton	understory
bud	cypress	giant	maple	pollen	slippery	underwater
bug	dam	gills	marsh	pond	snack	vegetable
bulb	den	grassland	Mexico	prairie	snail	vine
bump	dew	hatch	microscope	predator	sneak	wetland

Conclusion

Text analysis can play an important role in making sure resources meet both content and language objectives. For example, depending on the frequency of a word, teachers could find strategies to ensure dedicated instruction is provided about those words before students read. Alternatively, teachers can look

for alternative or background text to prepare students before the actual text, depending on their students' language and reading abilities.

References

- Accurso, K., & Gebhard, M. (2021). SFL praxis in US teacher education: A critical literature review. *Language and Education, 35*(5), 402-428.
- Alsaif, A., & Milton, J. (2012). Vocabulary input from school textbooks as a potential contributor to the small vocabulary uptake gained by English as a foreign language learners in Saudi Arabia. *The Language Learning Journal, 40*(1), 21-33.
- Anthony, L. (2019). AntConc (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University.
<https://www.laurenceanthony.net/software>
- Biemiller, A., Rosenstein, M., Sparks, R., Landauer, T. K., & Foltz, P. W. (2014). Models of vocabulary acquisition: Direct tests and text-derived simulations of vocabulary growth. *Scientific Studies of Reading, 18*(2), 130-154.
- Canga Alonso, A. (2015). The receptive vocabulary size of Spanish 5th grade primary school students in CLIL and Non-CLIL instruction. ES. *Revista de Filología Inglesa. 36, 63-85.*
<https://uvadoc.uva.es/bitstream/handle/10324/22679/ES-2015-36-ReceptiveVocabularySizeSpanish.pdf?sequence=1&isAllowed=y>
- Chang, A. C., & Pang, J. C. (2020). Doing Narrow Reading through Graded Readers: Possibilities and Difficulties. *Journal of Extensive Reading, 5*, 206-218.
- Chujo, K., Anthony, L., Oghigian, K., & Uchibori, A. (2012). Based, computer-based, and combined data-driven learning using a web-based concordancer. *Language education in Asia, 3*(2), 132-145.
- Cobb, T. (2021). Corpus for courses: Data-driven course design. *Bulletin suisse de linguistique appliquée, 2*, 11-30.
- Cobb, T (n.d.). Text Lex Compare v.4 [computer program]. https://www.lexutor.ca/cgi-bin/tl_compare/
- Cobb, T. *Vocabprofilers VP-Kids v.1.2* [computer program]. <https://www.lexutor.ca/vp/kids/>

- Children Literacy Initiative. (n.d.). Data driven instruction. <https://cli.org/2015/04/02/what-is-data-driven-instruction/>
- De Zeeuw, M., Grootjen, F., Kootstra, G. J., & Tellings, A. (2019). Lexical characteristics of written language input across primary grades: An analysis of a Dutch corpus based lexicon. *Linguistics and education, 49*, 11-21.
- Eagleton, M. B., & Dobler, E. (2007). *Reading the Web: Strategies for Internet Inquire Solving Problems in the Teaching of Literacy*. Guilford Publications.
- Gardner, D. (2004). Vocabulary input through extensive reading: A comparison of words found in children's narrative and expository reading materials. *Applied linguistics, 25*(1), 1-37.
- Gardner, D. (2008). Vocabulary recycling in children's authentic reading materials: A corpus based investigation of narrow reading. *Reading in a Foreign Language, 20*(1), 92-122.
- Häikiö, T., Bertram, R., & Hyönä, J. (2011). The development of whole-word representations in compound word processing: Evidence from eye fixation patterns of elementary school children. *Applied Psycholinguistics, 32*(3), 533-551.
- Kim, Y. S. G. (2017). Multicomponent view of vocabulary acquisition: An investigation with primary grade children. *Journal of Experimental Child Psychology, 162*, 120-133.
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Longman.
- McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- PebbleGo (2014). PebbleGo: Science [web application]. *Capstone*. www.pebblego.com.
- Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific studies of Reading, 18*(1), 22-37.
- Reading Rockets (2016). High/Low Books for Children By: Reading Rockets. *Reading Rockets*. <https://www.readingrockets.org/article/highlow-books-children>

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(4).

TXTcollector. (n.d.). TXTcollector. <https://txtcollector.en.softonic.com>

University of Duisburg-Essen. (n.d.). Word type. <https://www.uni-due.de/ELE/LinguisticGlossary.html>

Verhoeven, L., & Perfetti, C. A. (2011). Morphological processing in reading acquisition: A cross-linguistic perspective. *Applied Psycholinguistics*, 32(3), 457-466.

Williams, J. P., Stafford, K. B., Lauer, K. D., Hall, K. M., & Pollini, S. (2009). Embedding reading comprehension training in content-area instruction. *Journal of Educational Psychology*, 101(1), 1.

Disclaimer

Capstone Publishing (<https://www.capstonepub.com>) provided the content access of PebbleGo (<https://www.pebblego.com>) to the author of this study.

Acknowledgments

The author would like to dedicate this piece to his mentors, Mr. Robert Coughlan (Bob), Principal at Coughlan Publishing, and Dr. Nancy Drescher, a Professor of English at Minnesota State University, Mankato. She was also the author's MA thesis advisor.