

Analyzing 3rd- and 5th-Grade Morpheme Types and Tokens

Christopher Horak, Audrey Hendrix, Sana Tibi

Introduction

Research Aim

- Word parts called morphemes act as units of meaning in academic vocabulary (Nagy & Anderson, 1986).
- Studies have shown that morphological instruction improves vocabulary and decoding (e.g., Goodwin & Ahn, 2013; Colenbrander et al., 2024)
- Some studies have analyzed the amounts and trends of morphemes across text genres (e.g., Hiebert et al., 2024)
 - Few contrast morpheme types and tokens across grades.
- Grade contrast would clarify morphological demands of students' formative years.
- To investigate morphological changes, this research compares the characteristics of unique morphemes in 3rd and 5th grade texts.

Research Question

- Do morpheme type and token counts differ significantly between 5th and 3rd grade educational readings?

Corpus
300+ Books

3rd grade
27 Texts

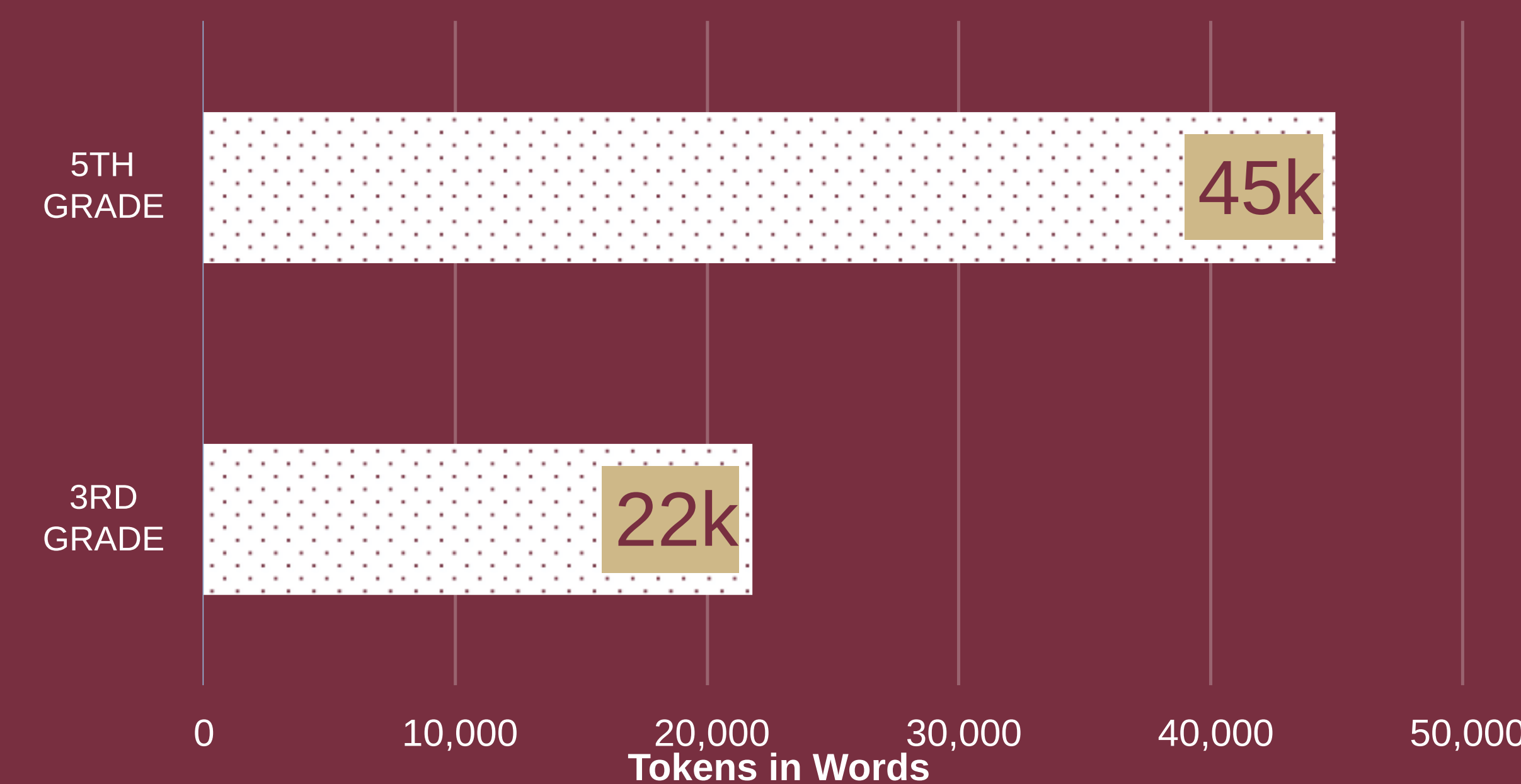
Word Count*
270k

*Estimated total word count for each grade

5th grade
26 Texts

Word Count*
610k

AVERAGE TEXT TOKENS BY GRADE (TOP 10)



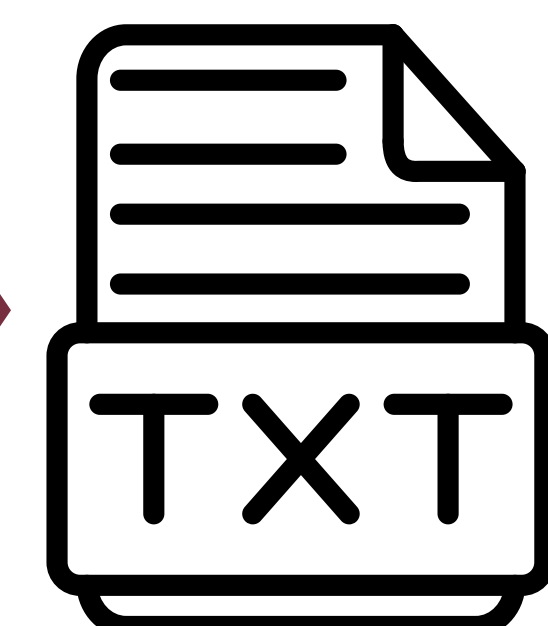
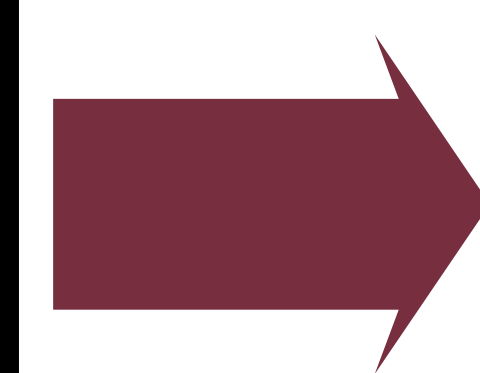
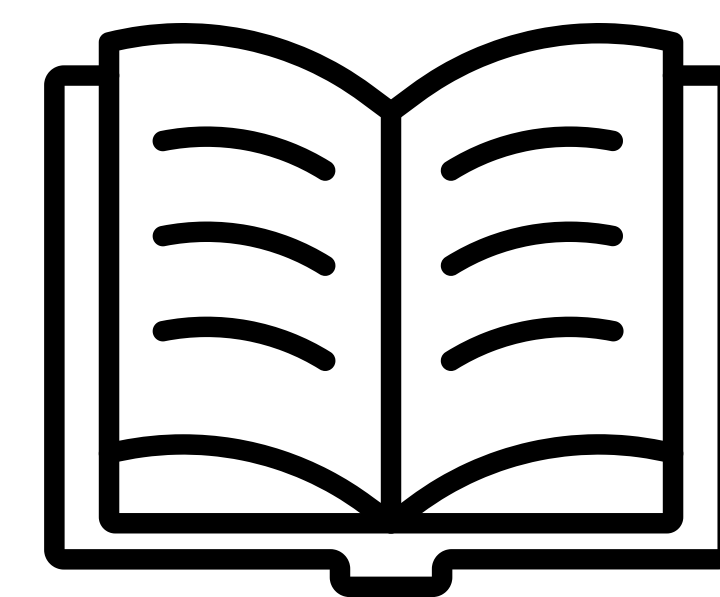
Results

Expectations

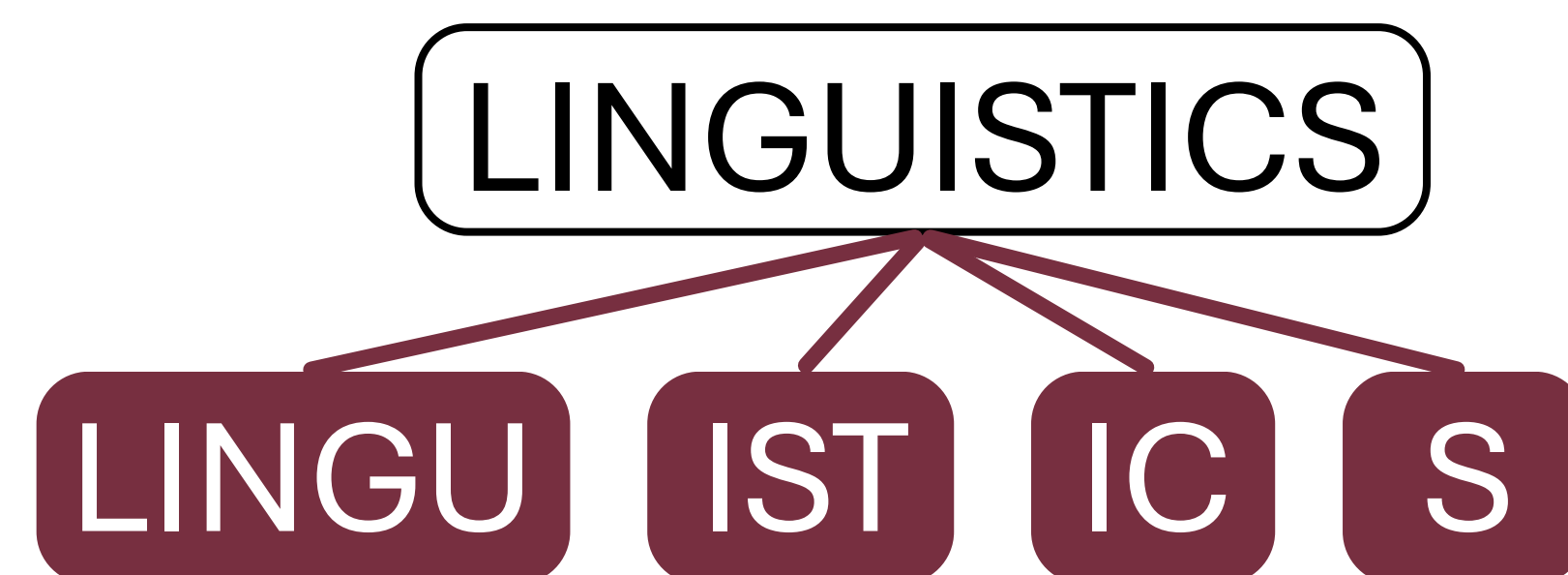
- There will be more 5th grade tokens than 3rd grade tokens
 - Despite the relative equality of the texts for each grade, 5th grade texts are substantially longer in length.
 - Higher word count implies higher morpheme tokens
 - Research indicates increases in morpheme types and tokens with increasing reader age (e.g., Dawson et al., 2023).
- 5th grade texts should be more complex, yielding a higher type count and percentage of morphological complex words.
 - 5th grade is likely to have a higher type to token ratio
 - May not be as significant as expected difference in raw type and token counts between the grades.

Methods

Books are converted to text files



Measurement of morphemes



Data Gathering

- Recommended readings from the Florida ELA Best Standards
 - a collection of books endorsed by Florida to be used in primary and secondary education
- Converted books into text (.txt) files, removing titles, section headers, and footnotes, etc.
 - Books and poems are transferred into a corpus with only relevant words.

Data Analysis

- The corpus of 3rd and 5th grade texts will be tokenized (converted from running text into word segments) at the morpheme level.
- Measuring type (unique morphemes) and token (total morphemes) frequency.
 - ex. The cat in the hat. → 5 tokens, 4 types (cat, hat, in, the)
- Tokenized words will be compared to an existing morphological database, MorphoLex, for measurement of morphological complexity (Sanchez-Gutierrez et al., 2017).
- This will allow the comparison between grade levels.
 - Morpheme type to token ratio will be calculated by grade level.



Logo for Florida B.E.S.T Initiative

References



Discussion

Significance

- If morpheme types and tokens are higher in 5th grade, it could indicate that the complexity of the language that students are reading increases between 3rd and 5th grade.
- Studies have investigated morphemes in relation to age, but this study is unique by focusing on the relationship with morphemes and academic grade level (Hiebert et al., 2024).
 - Other studies have used academic book lists like Common Core without exploring affix intricacies (Hiebert et al., 2017)
- If morphological complexity increases between 3rd and 5th grade, students may benefit from targeted morphological instruction for better reading and comprehension.

Future

- The readings come from Florida's Department of Education and may not be relevant to learners in other states, but a corpus is adjustable.
 - Future expansions of the corpus could include books from Common Core or TEKS.
- Only two grades are being compared in this study, but the process can be applied to all grade levels (K-12th).
- It would be possible to find morphological trends based on grade level readings and inform possible changes in morphological content between grade levels.
 - In doing so, the grades with significant increases in morphological complexity can be identified.
- Knowing these differences has implications for allowing instructors and interventionists to support a student's transition into higher level texts with greater ease and precision.