Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Evaluating Lexical Diversity and Vocabulary in Textbooks for Students with Moderate to Severe Intellectual Disabilities

Huzeyfe Bilge
Kafkas University
Hüseyin Sayin
Ministry of National Education
Cankurt Bozkurt
Ministry of National Education
Vural Yildiz
Ministry of National Education
Burçin Ugurlu
Ministry of National Education
Zeynep Olcay Gunaydinli
Ministry of National Education
Steve Graham
steve.graham@asu.edu

Arizona State University

Research Article

Keywords: special education, intellectual disability, textbook analysis, corpus linguistics, vocabulary repetition, Yule's K, lexical diversity

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11461046/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Evaluating Lexical Diversity and Vocabulary in Textbooks for Students with Moderate to Severe Intellectual Disabilities

Abstract

Adapted textbooks serve as a primary vehicle for vocabulary acquisition for students with intellectual disabilities. In this study, a corpus-based analysis was used to examine the lexical diversity and vocabulary repetition within reading passages and reading-focused tasks across all six official language arts textbooks designed for students with moderate to severe intellectual disabilities in Türkiye. Using a total corpus of 10,998 word tokens, a two-stage manual lemmatization process was executed, followed by computational evaluation via Simple Concordance Program (SCP) 5.0. Our evaluation revealed structural limitations across grade levels in the pedalogical adequacy of the materials. Although total token count generally expanded across textbook levels, the findings demonstrated an absence of systematic lexical progression. Notably, aggregated Yule's K values continuously decreased across advancing levels, uncovering a critical flaw in pedagogical adequacy. Instead of systematically providing vocabulary repetition, the textbooks exposed students to high lexical diversity consisting of distinct, low-frequency words without the necessary repetition for mastery. Over 98% of unique words across the entire corpus occurred 50 times or fewer, with more than 92% appearing only 1 to 10 times, highlighting insufficient vocabulary repetition. An analysis of high-frequency words indicated a heavy reliance on concrete nouns and essential daily verbs. Overall, the findings demonstrate that while the materials appear simplified on the surface, pedagogical adequacy was compromised by imposing a hidden cognitive load due to unmanaged lexical diversity and inadequate vocabulary repetition. The study highlights the urgent need for curriculum developers to utilize evidence-based frequency lists and ensure pedagogical adequacy through systematic vocabulary repetition strategies and controlled lexical diversity in special education textbook design.

Keywords: **s**pecial education; intellectual disability; textbook analysis; corpus linguistics; vocabulary repetition; Yule's K; lexical diversity.

Evaluating Lexical Diversity and Vocabulary in Textbooks for Students with Moderate to Severe Intellectual Disabilities

There is a strong correlation between individuals' levels of vocabulary knowledge, their proficiency in language skills, and their cognitive development (Suskind, 2020; Weisleder & Fernald, 2013). Indeed, theoretical frameworks in literacy position vocabulary knowledge at the very core of reading comprehension (Perfetti & Stafura, 2014). It is well established that a positive relationship exists between vocabulary knowledge and the level of reading comprehension achieved for typically developing students (Bilge & Kalenderoğlu, 2022). This relationship between vocabulary knowledge and comprehension skills holds equally true for students with intellectual disabilities (Wang et al., 2023). Consequently, enhancing the vocabulary of individuals with intellectual disabilities is essential for improving their reading comprehension and facilitating their overall access to the curriculum.

Research has established that developing language skills, including vocabulary, in individuals with intellectual disabilities contributes to better comprehension, enhanced psychological well-being, and improved socialization (Ervin, 2024; Munir et al., 2018; Robledo, 2017; Subasno et al., 2022). Conversely, insufficient vocabulary development for these individuals may lead to disadvantages in employment and social relationships (Burt et al., 2020). Therefore, it is essential that efforts to enhance these students' vocabulary development be a central part of their education and that such efforts are systematic and methodical. This necessity stems from the fact that the acquired vocabulary of individuals with intellectual disabilities is significantly smaller than that of their neurotypical counterparts (Bober et al., 2024), with this gap being particularly pronounced in individuals with severe intellectual disabilities (Ervin, 2024). Not surprisingly, major international reports (e.g., OECD, 2018) stress the critical importance of implementing targeted measures to eliminate educational inequalities among students in terms of vocabulary knowledge.

Unfortunately, previous investigations demonstrate that teachers typically allocate minimal time to explicit vocabulary instruction within their two-hour language arts blocks (Cunningham et al., 2009). While this needs to change if we are to maximize the vocabulary development of students with and without disabilities, there are other avenues or increasing the breadth and depth of students' vocabulary. One potentially effective avenue for advancing vocabulary knowledge involves subject matter textbooks.

Textbooks function as the primary instructional medium for mastering subject matter in most schools (Aldahmash & Omar, 2021), rendering them vital resources for vocabulary acquisition. Consequently, accounting for students' existing vocabulary during textbook development is of paramount importance, as an adequate vocabulary is a prerequisite for accessing curriculum content (Li & Deng, 2026). For students with intellectual disabilities in particular, the vocabulary embedded within adapted textbooks constitutes one of the most critical avenues for lexical enhancement. Repeated exposure to target vocabulary through structured reading in textbooks provides contextualized input, reinforcing word retention and deepening semantic understanding for learners with moderate to severe intellectual disabilities who require explicit, sustained practice (Browder et al., 2006). Therefore, a rigorous examination of various statistical metrics regarding textbook vocabulary is warrantedly required.

Research on the vocabulary development of neurotypical students has advanced significantly, yielding findings detailed enough to establish these students require at least ten exposures to learn a target word (Webb, 2007). Although such precise empirical parameters have not yet been established for individuals with intellectual disabilities, it is well documented that they require substantially more repetition and reinforcement to acquire new vocabulary compared to their neurotypical peers (Burt et al., 2020). For neurotypical populations, various corpus-based initiatives (Coxhead, 2000) and structured vocabulary

categorization frameworks (Beck et al., 2013) have been extensively developed. However, scholars frequently emphasize the scarcity of research concerning both the intrinsic vocabulary levels of individuals with intellectual disabilities and the vocabulary profiles of textbooks designed for them (Tūbele & Landrāte, 2021). Yet, the selection of vocabulary in textbooks should not occur haphazardly; rather, it ought to be informed by evidence-based resources (Hadley & Mendez, 2021). Nation (2022) stressed that word frequency and range of use must guide decisions regarding which words to teach and in what sequence. Furthermore, to optimize vocabulary instruction within textbooks, Schmitt (2019) recommended integrating core pedagogical principles into textbook design and expanding such research beyond widely studied languages like English (Liu et al., 2024) to other languages. To ensure a sound decision-making process in this regard, a rigorous analysis of the current landscape in existing textbooks is an essential first step.

The present study examines, from various perspectives, the vocabulary in primary reading passages and reading-focused task texts across official instructional materials used for students with moderate to severe intellectual disabilities in Türkiye. This corpus includes both core language arts (Turkish) textbooks (covering primary Level 1 and lower secondary Level 2) and specialized thematic textbooks developed for Level 2, specifically My Social Life (Sosyal Yaşamım) and My Daily Life (Günlük Yaşamım). These thematic materials were deliberately included because functional communication and daily living skills form the core of reading instruction for students with intellectual disabilities, making their vocabulary exposure in these targeted domains essential to evaluate alongside standard language arts texts. In alignment with OECD's (2018) recommendation regarding cross-cultural comparisons, this study seeks to provide statistical metrics that will serve as a baseline for comparative analysis with future international studies. To this end, the primary purpose of this study is to determine how vocabulary is presented and supported across these comprehensive

reading passages and task texts, analyzing the strengths and limitations of these materials for supporting vocabulary development.

Literature Review

Curriculum Accessibility, Inclusivity, and the Role of Vocabulary

One broad aim of special education services and systems in Türkiye and other countries is to foster the academic development of students with intellectual disabilities (Ministry of National Education [MoNE], 2018a, 2018b). Educational institutions are held accountable for the academic performance of all students, including those with intellectual disabilities (Evmenova & Behrmann, 2011), with stakeholders increasingly expected to demonstrate tangible outcomes in this regard (Karvonen et al., 2013). Consequently, all special education students must have direct access to the general education curriculum and undergo alternative assessments tailored to their grade levels (Browder et al., 2007). However, empirical research reveals that in core subjects such as language arts, students with intellectual disabilities are frequently not exposed to the full range of academic content they are expected to master (Karvonen et al., 2013). This structural limitation raises significant concerns regarding educational equity.

Within the frameworks of educational accessibility and inclusion, instructional materials designed for individuals with intellectual disabilities play a pivotal role (Cote & Banks, 2025). The design of textbooks, in particular, requires meticulous execution, given that students with intellectual disabilities exhibit highly heterogeneous profiles and distinct skill impairments, even when sharing the same clinical diagnosis (García Muñoz, 2012). Current practice indicates these students face substantial difficulties in fully accessing curriculum content (Karvonen et al., 2013), underscoring the presence of severe structural barriers to curriculum access compared to their neurotypical peers. Consequently, textbooks developed for individuals with intellectual disabilities necessitate adapting grade-level texts used by

neurotypical peers, differentiating content in terms of scope and depth, and identifying effective pedagogical pathways to teach academic concepts (Browder et al., 2007). In this regard, the vocabulary embedded within instructional texts offered to these students must be carefully curated and systematically analyzed. Research focusing on students' vocabulary knowledge can yield critical insights that inform both the assessment of curriculum accessibility and the evidence-based design of textbooks (Li & Deng, 2026).

Language Development and the Critical Role of Vocabulary in Individuals with Intellectual Disabilities

Vocabulary knowledge represents one of the foundational assets positioned at the core of academic achievement. Indeed, prominent theoretical models of reading comprehension explicitly place vocabulary at their center (Perfetti & Stafura, 2014), while major national frameworks similarly highlight it as a vital component of reading proficiency (National Reading Panel, 2000). This principle applies equally to individuals with intellectual disabilities; empirical evidence demonstrates that as their vocabulary repertoire expands, their performance on higher-order cognitive and academic tasks increases accordingly (Stevens et al., 2025).

However, language development in individuals with intellectual disabilities follows a trajectory that diverges significantly from that of neurotypical individuals. Reading and comprehension skills in this population are generally less developed (Munir et al., 2018), and their total vocabulary size also lags substantially behind that of their neurotypical peers (Bober et al., 2024). This limitation and developmental delay become markedly more pronounced, particularly among individuals with severe intellectual disabilities (Ervin, 2024).

This limitation in vocabulary knowledge directly undermines both the academic achievement and social adaptation of individuals with intellectual disabilities. Indeed, a restricted vocabulary repertoire is associated with communication difficulties, an inability to

follow instructions, comprehension deficits, reduced autonomy, and even challenging behaviors (Ervin, 2024; Pratiwi & Fithriani, 2026). Conversely, fostering vocabulary knowledge and general language skills enhances individuals' comprehension capacities and self-esteem, reduces self-injurious behaviors, facilitates social participation by enabling meaningful connections with others (Ervin, 2024; Robledo, 2017; Subasno et al., 2022; Wang et al., 2023), and strengthens their overall readiness for the curriculum (Li & Deng, 2026). In this context, Tūbele and Landrāte (2021) emphasize that expanding the vocabulary of individuals with moderate and severe intellectual disabilities supports their cognitive capacity, thereby enabling them to explore their environment and actively participate in social life.

Despite its critical importance, research specifically targeting the vocabulary knowledge of students with intellectual disabilities remains markedly scarce (O'Connor et al., 2019; Tūbele & Landrāte, 2021). The existing, albeit limited, literature includes various intervention studies aimed at expanding the vocabulary of students with intellectual disabilities, yielding promising outcomes (Brako-Hiapa et al., 2024; Hua et al., 2013; Rapti et al., 2023; Taşkın, 2023).This study adds to this existing body of research by examining if materials used to teach students with intellectual disabilities provide a viable resource for increasing these students' vocabulary.

Textbooks and Vocabulary for Individuals with Intellectual Disabilities

Textbooks serve as the foundational resource for attaining curricular goals (Aldahmash & Omar, 2021). Beyond functioning as a primary instructional tool for developing the vocabulary of students with intellectual disabilities, they constitute a critical instrument enabling teachers to assess students' vocabulary knowledge (Love et al., 2017). However, given the substantial individual heterogeneity in the developmental profiles of individuals with intellectual disabilities, materials developed for these students need to be structured so they are effective with this population. Indeed, Universal Design for Learning

(UDL) principles advocate that curricula and instructional materials must adapt to learners' individual needs (CAST, 2011). Conversely, the literature indicates that the readability and difficulty level of materials accessible to students with intellectual disabilities often fall below their grade level, limiting their opportunities to acquire vocabulary independently (O'Connor et al., 2019).

Effective vocabulary instruction is built upon pedagogical principles that require immersing students in extensive reading to ensure high-density word exposure, structuring instructional materials, and presenting target words across varied contextual frameworks (National Reading Panel, 2000). One of the most challenging phases of this instructional process lies in determining the selection criteria for target vocabulary (Coxhead, 2000). For neurotypical students, various frameworks have been established to guide the selection of words embedded within instructional texts. For example, Hadley and Mendez (2021) highlight the inherent complexity of vocabulary selection in early childhood education, proposing key principles such as pre-selecting target words, avoiding both overly simplistic and highly infrequent words, consulting empirical evidence-based resources for word selection, and delivering vocabulary in conceptually related clusters.

Another widely endorsed approach to vocabulary selection is the Three-Tier Model developed by Beck et al. (2013). According to this framework:

- Tier 1: Encompasses basic, high-frequency words routinely encountered in everyday oral communication and conversation.
- Tier 2: Comprises high-utility, rich vocabulary that appears less frequently in spoken language but occurs regularly across written texts, serving a critical role in verbal sophistication and literacy development.
- Tier 3: Consists of domain-specific, academic vocabulary intrinsic to distinct disciplines, which ought to be taught on an as-needed basis within relevant contexts.

While neurotypical students naturally acquire Tier 1 basic vocabulary through daily social interactions, individuals with intellectual disabilities may be deprived of these implicit learning opportunities due to their inherent limitations in social interaction and communication. Consequently, even for basic words relative to their peers' development—which are typically acquired naturally by neurotypical individuals—students with intellectual disabilities require textbooks developed through structured, cautious, and highly meticulous instructional designs.

Although the literature highlights various principles for designing textbooks and reading texts to foster vocabulary acquisition in neurotypical students (Sprenger, 2014), it is evident that not all of these principles are directly applicable to learners with moderate to severe intellectual disabilities. Therefore, to assess the pedagogical adequacy of instructional materials designed for this population, this study posits three core principles:

Pedagogical adequacy: Vocabulary repetition and lexical diversity. One of the primary principles for designing instructional texts for students with moderate to severe intellectual disabilities concerns the structural interplay between vocabulary repetition and lexical diversity. Prominent literacy synthesis reports emphasize that high-frequency repetition and multi-contextual exposure are critical components of effective vocabulary instruction (National Reading Panel, 2000). For learners with moderate to severe intellectual disabilities, systematic repetition is even more vital than for their neurotypical peers, as these students require significantly more frequent practice opportunities to consolidate new learning (Burt et al., 2020). By embedding frequent word repetitions within textbooks, instructional materials enable these learners to encounter target vocabulary repeatedly across varied textual contexts, thereby fostering long-term retention.

Directly linked to vocabulary repetition, the level of lexical diversity within instructional texts must also be carefully regulated. Research indicates that as the severity of

intellectual disability increases, learning efficiency typically declines (Schuchardt et al., 2010). Consequently, introducing novel words prior to providing sufficient exposure and reinforcement of previously encountered vocabulary can create a substantial barrier to vocabulary expansion. To prevent cognitive overload and support steady vocabulary development, instructional texts for this population should intentionally maintain low to moderate lexical diversity rather than exposing learners to an unmanaged abundance of unique words.

In this study, vocabulary repetition is operationalized through Yule's characteristic (K), where lower K values signify lower vocabulary repetition (higher lexical diversity), while total word count (N) reflects overall textual volume. Conversely, lexical diversity is operationalized through unique word count (V) and Type-Token Ratio (TTR), measuring the degree of lexical variation within the texts.

Lexical progression across grade-levels. National special education framework guidelines emphasize that instructional content must be structured along a continuum from simple to increasingly complex patterns across educational levels (MoNE, 2018a). Applied to a corpus-based perspective, this pedagogical progression logically implies that vocabulary repetition and lexical diversity should demonstrate a gradual expansion as grade levels advance. This incremental progression is particularly crucial because acquiring prerequisite target vocabulary at earlier stages is essential for comprehending the novel content introduced in subsequent grade levels (Yopp et al., 2009). However, this lexical progression must be calibrated with exceptional care for learners with moderate to severe intellectual disabilities. This necessity stems from the profound individual heterogeneity within this population, as these students exhibit widely disparate developmental profiles and learning capabilities (García Muñoz, 2012).

Grammatical distribution and cognitive load. Research on neurotypical development indicates that children acquire nouns more rapidly than verbs, a phenomenon partly attributed to the higher imageability of nominal concepts (McDonough et al., 2011). Especially for young learners, word concreteness significantly facilitates vocabulary acquisition. Grammatical categories—such as nouns, verbs, and adjectives—serve as critical determinants of a word's relative concreteness or abstractness. This distinction between concrete and abstract vocabulary is exceptionally sensitive for learners with moderate to severe intellectual disabilities, who experience profound difficulty in processing and acquiring abstract terms (Fabio et al., 2023). In neurotypical populations, empirical evidence demonstrates that as the cognitive resource demands of a word increase, its acquisition becomes substantially more challenging (Oshima-Takane et al., 2011). Consequently, given the heightened cognitive demands faced by learners with moderate to severe intellectual disabilities, selecting and structuring target vocabulary requires meticulous attention to grammatical distribution and associated cognitive load.

The aforementioned body of literature indicates that while vocabulary research concerning neurotypical students has reached advanced stages, empirical inquiry into the vocabulary needs of students with intellectual disabilities lags significantly behind (Tūbele & Landrāte, 2021). Although numerous vocabulary selection frameworks have been developed for neurotypical students, these criterion-based lists may not be directly generalizable to individuals with intellectual disabilities, given their distinct special educational needs. Developing specialized vocabulary criteria tailored for individuals with intellectual disabilities can mitigate the risk of arbitrary word selection in textbook design. This is critical because the words embedded within instructional textbooks must be functionally high-utility words for the learners (Coxhead, 2000). Furthermore, as previously emphasized, national special education framework guidelines dictate that instructional content should adhere to a

continuum from simple to increasingly complex structures across levels (MoNE, 2018a)—a pedagogical principle that, from a corpus-based perspective, implies a gradual expansion in word count and lexical diversity.

Although textbook materials and vocabulary repetition/lexical characteristics have been extensively examined in general education contexts, empirical research investigating textbook vocabulary repetition/lexical characteristics specifically for individuals with moderate to severe intellectual disabilities remains strikingly absent. To the best of our knowledge, this is the first study to examine textbook vocabulary material and lexical characteristics (including repetition, diversity, and grammatical distribution) within this specific target population. To inform the establishment of evidence-based criteria, a foundational step requires analyzing the lexical characteristics of textbooks designed for this target population across various linguistic metrics. To address this gap, the present study aims to examine the vocabulary profiles within reading passages and reading-focused task texts—specifically excluding instructional prompts and focusing strictly on text segments students are directly expected to read—across the complete corpus of six official instructional materials developed for students with moderate to severe intellectual disabilities in Türkiye.

Research Questions

Well-designed textbooks for students with moderate to severe intellectual disabilities are expected to optimize vocabulary acquisition by maintaining high vocabulary repetition, ensuring systematic lexical progression across grade levels, and offering cognitive accessibility. Based on these underlying assumptions, this study addresses the following research questions:

Pedagogical Adequacy: Vocabulary Repetition and Lexical Diversity

Given that students with moderate to severe intellectual disabilities rarely acquire new vocabulary without considerable repetition and exposure, the pedagogical adequacy of

instructional materials relies heavily on systematic word repetition and controlled lexical diversity. Rather than exposing learners to an overwhelming volume of unrepeated words, effective textbooks must ensure high levels of vocabulary repetition—indicated by higher Yule's characteristic (K) values—alongside balanced TTR. Evaluating these metrics together reveals whether reading passages and task texts offer the repeated exposure necessary for long-term retention without imposing excessive cognitive load.

RQ1. What are the patterns of vocabulary repetition and lexical diversity—as measured by total word count (N), unique word count (V), TTR, and Yule's K—across the reading passages and reading task texts of official textbooks designed for students with moderate to severe intellectual disabilities?

Lexical Progression Across Grade Levels

In a well-designed curriculum, instructional materials are expected to demonstrate a systematic, gradual expansion in total word count and lexical diversity as students advance from lower to higher grade levels (from Level 1 to Level 2). Operationalizing this principle, higher-level textbooks should progressively increase lexical diversity to facilitate continuous vocabulary development.

RQ2. To what extent does the vocabulary embedded in instructional materials demonstrate systematic, gradual textual expansion—as measured by total word count (N), unique word count (V), and frequency distribution tiers—across successive textbook levels (Level 1 vs. Level 2) and thematic materials?

Grammatical Distribution and Cognitive Accessibility

Given the socio-emotional and functional communication needs of individuals with intellectual disabilities, introductory instructional materials are expected to prioritize concrete, high-utility vocabulary (e.g., concrete nouns) while pacing abstract categories (e.g., abstract verbs or function words) that impose higher cognitive processing demands.

RQ3. How are the most frequent words distributed across grammatical categories (e.g., concrete nouns, verbs, adjectives, function words), and does this structural distribution align with the cognitive accessibility requirements of learners with moderate to severe intellectual disabilities?

Methods

Research Design

This study examined the vocabulary contained within the complete universe of six official language arts textbooks developed for students with moderate to severe intellectual disabilities. The study used a corpus-based descriptive research design to systematically analyze the vocabulary embedded in reading passages and reading tasks across these language instruction textbooks.

Materials

The study analyzed language arts (Turkish) textbooks prepared by the Ministry of National Education (MoNE) for students with moderate to severe intellectual disabilities, approved for classroom use from the 2020–2021 through the 2025–2026 academic years (MoNE Journal of Announcements [Tebliğler Dergisi], 2020) (see Table 1). Specifically, the materials comprise two core language arts textbooks for primary school (Level 1), two core language arts textbooks for lower secondary school (Level 2), and two specialized thematic volumes for Level 2 centered on functional daily domains—My Social Life (Sosyal Yaşamım) and My Daily Life (Günlük Yaşamım). Rather than being external supplementary materials, these two thematic volumes serve as integral components of the official Level 2 instructional series. Accordingly, all six volumes were incorporated into the corpus analysis to evaluate the full scope of official curricular materials designed for this student population.

It is important to contextualize that education in Türkiye is highly centralized, with MoNE exercising sole authority over national curricula, educational policies, and instructional

materials across special education. Specifically regarding the Turkish special education system, curriculum design and instructional material development for students with moderate to severe intellectual disabilities are fully managed under this centralized structure. Commercial or private publishing houses do not produce alternative textbooks for this target population. Consequently, these six MoNE-approved textbooks do not represent a randomized sample; rather, they constitute the entire complete universe (population) of official language arts instructional books distributed and utilized across all special education institutions in Türkiye for this student group.

The inclusion criteria for the corpus were defined strictly to encompass all primary reading passages and all activity texts explicitly designed for student reading across all six volumes of the official instructional series(see Table 1). Specifically, an activity segment was included if its prompt directly instructed the student to read (e.g., 'Read the text below'). Target sentences and text segments embedded within reading activities were systematically incorporated into the dataset, provided that these specific segments were designed for the students themselves to read independently (e.g., target practice sentences within tasks). Conversely, non-reading components—such as instructional prompts for listening, speaking, writing, visual-matching tasks, teacher scripts, and general unit instructions or comprehension questions—were systematically excluded. General activity instructions, teacher prompts, and comprehension or assessment questions were also systematically excluded to ensure the corpus strictly reflects direct student reading exposure. By applying these criteria universally across all six textbooks, the resulting corpus comprises 100% of the student-facing reading materials available in the official curriculum for this population.

Data Analysis

For data analysis, the textual content extracted from the data sources was first transcribed into Microsoft Word. Subsequently, all words across these texts were counted

based on the Current Turkish Dictionary (*Güncel Türkçe Sözlük*) of the Turkish Language Association. Turkish belongs to the family of agglutinative languages. Consequently, words derived from the same root may take various suffixes and appear as distinct headwords (lemma) in the dictionary. For instance, the root *yemek* (to eat / food) can function as a noun (*yemek* - meal), a verbal noun (*yemek* - eating), or another nominalized form (*yeme* - consumption), depending on its syntactic role within a sentence. Each of these word forms is identified based on its specific contextual usage.

To address the highly agglutinative nature of the Turkish language and ensure maximum analytical precision, a rigorous two-stage manual lemmatization process was executed prior to feeding the corpus into Simple Concordance Program (SCP) 5.0. In the first stage, a certified native Turkish language teacher manually disambiguated and lemmatized all word forms in context, stripping inflectional suffixes to align each token with its corresponding dictionary headword (lemmata) as defined in the Current Turkish Dictionary. In the second stage, a senior expert in Turkish language education independently reviewed and cross-verified all morphological analyses.

Once the target words were compiled as dictionary headwords (lemmata), they were analyzed using Simple Concordance Program (SCP) 5.0 in a multi-stage process. First, to establish baseline data, each of the six textbooks was analyzed individually and independently.Second, to evaluate outcomes across educational levels and content types, the textbooks were aggregated into three primary corpus groups: Level 1 Textbooks (combining the two primary-level textbooks), Level 2 Core Language Arts Textbooks (combining Textbook-1 and Textbook-2), and Level 2 Thematic Functional Literacy Textbooks (combining My Daily Life and My Social Life). Finally, all six textbooks were merged into a single, unified dataset to conduct an overarching corpus analysis. Across all analytical stages—individual, grouped, and overall—key quantitative metrics were computed, including

lexical diversity (Type-Token Ratio and unique word count), total word count, and vocabulary repetition (Yule's K index). In addition, absolute frequency counts, percentages, and word frequency distributions were systematically calculated for each individual textbook, grouped category, and the consolidated corpus.

To systematically evaluate the text characteristics, four interrelated metrics were operationalized into two major pedagogical constructs: Lexical diversity and vocabulary repetition. Lexical diversity represents the breadth and variation of the vocabulary presented to students and was operationalized using unique word count (V) and the Type-Token Ratio ($TTR = V/N$). Total word count (N) was calculated to reflect the overall textual exposure. However, because TTR is systematically confounded by text length (N)—decreasing as N expands—it fails to offer a reliable baseline for cross-text comparison. To address this limitation, Vocabulary repetition was operationalized using Yule's characteristic (K), a length-independent probabilistic metric. In Yule's K, higher values signify higher structural vocabulary repetition (a more restricted lexical diversity), whereas lower values denote lower repetition rates and higher lexical diversity.

To enhance the comparability of the lexical characteristics analyses in this study with existing and future empirical research in this domain, Yule's (1944) K was employed alongside the widely recognized TTR. This approach aligns with international methodological frameworks recommending the cross-cultural comparison (OECD, 2018). TTR is conventionally defined as the ratio obtained by dividing the unique word count (V) by the total word count (N) in a given text (Karadağ, 2018). However, TTR values exhibit a systematic inverse relationship with corpus length, decreasing as total word count increases. Richards (1987) noted that as sample sizes expand, high-frequency words recur more frequently, leading to a diminished proportion of lexical diversity. Consequently, TTR can yield misleading estimates of lexical diversity, often artificially inflating diversity metrics in

smaller samples. Text length thus functions as a confounding variable that directly influences TTR, making cross-text comparisons methodologically problematic. Conversely, Yule's *K* statistic is a mathematical measure specifically designed to evaluate lexical diversity independently of text length. As Miranda-García and Calle-Martín (2005) demonstrated, Yule's *K* measures the probability of vocabulary repetition regardless of text length, thereby enabling consistent, valid, and meaningful comparisons across diverse text types and empirical studies.

To evaluate grammatical distribution and cognitive accessibility, a high-frequency cutoff threshold was established based on raw word occurrences. Specifically, all unique words occurring with a frequency of 21 or more were selected for grammatical distribution analysis. This frequency criterion yielded exactly 99 words, representing the most robust, high-utility vocabulary core in the corpus while aligning closely with the conventional top-100 high-frequency lexical threshold (selecting a lower frequency threshold of 20 would have arbitrarily expanded the set to 104 words). To classify these 99 most frequent words according to their grammatical distribution, the following procedure was adopted: Similar to English, a single word in Turkish can function across different grammatical categories—such as a noun or an adjective—depending on its contextual usage. For instance, the word yeşil (green) can function as either a noun or an adjective within a sentence. In classifying words by grammatical distribution, experts were requested to indicate all potential functional categories for each word among verbs, nouns, adjectives, adverbs, prepositions, conjunctions, and interjections. These grammatical categories align with the official categorization framework established by the Ministry of National Education (MoNE, 2026).

To establish inter-rater reliability between two independent coders (one certified Turkish language teacher and one graduate researcher in Turkish language education), Cohen's Kappa (κ) coefficients were calculated for each grammatical category. Kappa values

were interpreted according to the benchmarks defined by Landis and Koch (1977). The analysis revealed perfect agreement for verbs, conjunctions, and prepositions ($\kappa = 1.000, p < .001$); almost perfect agreement for nouns ($\kappa = .898, p < .001$) and adverbs ($\kappa = .853, p < .001$); substantial agreement for interjections ($\kappa = .657, p < .001$); slight agreement for adjectives ($\kappa = .203, p < .05$); and poor agreement for pronouns ($\kappa = .097, p > .05$).

The relatively low Kappa coefficients observed for adjectives ($\kappa = .203$) and pronouns ($\kappa = .097$) warrant a specific methodological clarification. In corpus linguistics, Cohen's Kappa is highly sensitive to marginal category frequencies (Sim & Wright, 2005). As reflected in Table 4, pronouns ($n = 10$) and adjectives ($n = 9$) represented a very small subset of the top 99 high-frequency words analyzed. Under such low-frequency conditions, the 'Kappa paradox' occurs: even minor, context-dependent coding variations between raters artificially inflate the expected agreement (P_e), driving the resulting Kappa value down toward zero despite high absolute percent agreement (Cicchetti & Feinstein, 1990; Di Eugenio & Glass, 2004).

To preserve the methodological objectivity of the dataset and resolve discrepancies, a formal arbitration protocol was implemented. All words exhibiting divergent codings between the two initial raters in these particular categories were isolated into a separate inventory. These disputed items were independently reassessed by a third senior researcher (one the authors of this article). Based on this final evaluation, a majority rule consensus (agreement between two out of three raters) was applied to resolve all conflicting codings, thereby establishing the definitive grammatical category for each target word.

Results

To address the overarching aims of this study and evaluate the pedagogical adequacy of language arts instructional materials for learners with moderate to severe intellectual disabilities, the findings are presented in direct alignment with the three specific research

questions—evaluating vocabulary repetition, lexical diversity, lexical progression across grade levels, and grammatical distribution.

Pedagogical Adequacy: Vocabulary Repetition and Lexical Diversity (RQ1)

To determine the lexical characteristics (vocabulary repetition and lexical diversity) of the reading passages and reading-focused task texts across the textbooks, the descriptive statistics and lexical diversity indices obtained from the corpus are presented in Table 2.

The data summarized in Table 2 reveal the lexical metrics across the special education textbooks analyzed for the first research question (Principle 1: Vocabulary Repetition and Lexical Diversity). The empirical data confirm that the TTR is highly sensitive to total word count (N). As text length increases (e.g., $N = 10{,}998$ in the Grand Total), the proportion of vocabulary repetition naturally expands, causing the TTR to drop to 16.9%. Computed to overcome this methodological limitation, Yule's K statistic—which operates independently of total word count—reveals a critical pattern regarding vocabulary repetition. Higher K values indicate higher vocabulary repetition (a more restricted vocabulary core), whereas lower K values denote lower repetition rates and higher lexical diversity. The mean K value of Level 1 textbooks ($K = 46.2$) demonstrated a decline from Level 1 Textbook 1 ($K = 57.1$) to Level 1 Textbook 2 ($K = 50.6$). This exact trajectory is mirrored between the individual K values and aggregated totals of Level 2 textbooks ($K = 38.1$). Notably, this pattern becomes even more pronounced in the aggregated K value for all textbooks combined ($K = 35.4$).

Because a decrease in K reflects lower vocabulary repetition alongside a higher lexical diversity, word overlap across textbooks is minimal, and target words are not frequently recycled. Consequently, the progressively declining K values across aggregated levels point to a heavy load of infrequent words that appear sparingly within the texts without sufficient repetition or pedagogical reinforcement.

The thematic textbooks *My Daily Life* (*Günlük Yaşamım*) and *My Social Life* (*Sosyal Yaşamım*) exhibit higher Yule's K values ($K = 48.4$ and $K = 50.0$, respectively) compared to Level 2 academic core textbooks ($K = 38.1$). The elevated K values of these thematic textbooks, coupled with their relatively lower TTR percentages (28.0% and 24.9%, respectively), demonstrate that their textual structure was designed to systematically recycle target vocabulary. Consequently, while the thematic textbooks present a structure characterized by higher K values and lower TTRs that prioritize vocabulary repetition, the general language arts textbooks display a more variable lexical diversity with lower vocabulary repetition rates.

Lexical Progression Across Grade Levels (RQ2)

To examine the distribution ratio and lexical progression across individual textbooks and grade levels, data from Table 2 and Table 3 were analyzed (Principle 2: Lexical Progression Across Grade Levels).

Contrary to a structured linear expansion across educational levels, the empirical data in Table 2 indicate marked fluctuations in word count. In Level 1 Textbook 1, the total word count ($N = 1{,}632$) experiences a decline of 43.38% in Textbook 2 of the same level ($N = 924$). Conversely, an increase of 55.67% in total word count is evident between Level 2 Textbook 1 ($N = 1{,}252$) and Level 2 Textbook 2 ($N = 1{,}949$). Furthermore, the thematic textbooks ($N = 2{,}289$; $N = 2{,}952$) account for 47.65% of the total word count analyzed across the entire corpus. In summary, with the exception of the decline observed in Level 1 Textbook 2, the total word count embedded within reading passages and reading-focused tasks increased across textbooks; however, these increases lacked a systematic linear progression.

Table 3 provides the breakdown of unique word count frequencies across individual books and aggregated levels. According to the statistics presented in Table 3, the number of

unique word count (V) with a frequency count between 1 and 10 in Level 1 textbooks was 344 and 383, respectively. In Level 2, this figure exhibited a substantial increase, reaching 499 in Textbook 1 and 695 in Textbook 2. In the thematic textbooks *My Daily Life* (*Günlük Yaşamım*) and *My Social Life* (*Sosyal Yaşamım*), the unique word count occurring between 1 and 10 times was 599 and 680, respectively. Regarding words with a frequency count between 11 and 20, Level 1 textbooks contained 20 and 4 words, respectively, whereas Level 2 textbooks included 8 and 20 words. In the thematic readers *My Daily Life* and *My Social Life*, this figure increased slightly to 32 and 35 words, respectively. The number of target words repeated more than 20 times across the individual textbooks ranged between 2 and 20.

In percentage terms, words occurring with a frequency count between 1 and 10 accounted for at least 92% of all unique words count (V) in each textbook. This indicates that the vast majority of words embedded in the textbooks were not recycled sufficiently, demonstrating a critical lack of vocabulary repetition. When analyzing the entire corpus (Grand Total), words occurring 50 times or fewer totaled 1,835, constituting 98.6% of all unique word count (V) across all six textbooks. Conversely, the remaining 1.4% (comprising 26 words) exhibited a frequency count of 51 or higher. The three most frequently occurring words were *bir* ("a/one"; $n = 271$), *ve* ("and"; $n = 197$), and *çok* ("very/much"; $n = 187$). These three words alone accounted for 0.16% of all unique word count across the entire corpus and fell into the highest frequency band of 151+ occurrences.

Grammatical Distribution and Cognitive Accessibility (RQ3)

To evaluate grammatical distribution and cognitive accessibility (Principle 3: Structural Distribution and Cognitive Accessibility), the 99 most frequent words ($f \geq 21$) representing the high-utility core vocabulary across the corpus were categorized into grammatical classes, as presented in Table 4. This structural breakdown was analyzed to

determine whether the high-frequency vocabulary prioritizes concrete, high-utility vocabulary over abstract categories that impose higher cognitive processing demands.

The distribution of the top 99 most frequent words analyzed within the scope of this study across grammatical categories is presented in Table 4. According to the results of the frequency analysis, nouns constituted the largest proportion among words with an occurrence frequency of 21 or higher ($n = [52]$, 52.5%). Conversely, abstract categories and function words—which generally impose higher cognitive processing demands—exhibited varying levels of representation: verbs constituted the second largest category ($n = [24]$, 24.2%), followed sequentially by adverbs ($n = [11]$, 11.1%) and pronouns ($n = [10]$, 10.1%). Structural function words exhibited the lowest representation within the high-frequency word inventory, specifically conjunctions ($n = [5]$, 5.1%), postpositions ($n = [4]$, 4.0%), and interjections ($n = [4]$, 4.0%).

A closer qualitative inspection of the predominant noun category ($n = 52$) reveals a distinct concentration of concrete nouns directly anchored in the immediate physical, domestic, and social environments of the target population. Specifically, the most frequently occurring nouns across the corpus include *anne* ("mother"; $f = 147$), *çocuk* ("child"; $f = 91$), *levha* ("sign/plate"; $f = 74$), *baba* ("father"; $f = 67$), *yer* ("place/ground"; $f = 67$), *ev* ("home/house"; $f = 61$), *yemek* ("food/meal"; $f = 53$), *arkadaş* ("friend"; $f = 42$), *gün* ("day"; $f = 41$), and *park* ("park/playground"; $f = 41$). The heavy reliance on highly concrete lexical items directly reflects an intentional pedagogical effort to reduce cognitive processing load.

Discussion

This study presents a corpus-based analysis evaluating the pedagogical adequacy of introductory language arts textbooks designed for special education students with moderate to severe intellectual disabilities. Rather than evaluating textbook vocabulary through general

corpus linguistics metrics alone, our empirical findings are contextualized within three core pedagogical principles of instructional design tailored for this learner population: (1) Vocabulary repetition and lexical diversity (Principle 1; RQ1), evaluated through the Type-Token Ratio (TTR) to capture lexical diversity and Yule's K statistic to measure the rate of target vocabulary repetition independently of total word count; (2) Lexical progression across grade levels (Principle 2; RQ2), assessed through total word count (N) to determine overall textual exposure and unique word count (V) to measure lexical breadth and frequency distribution; and (3) grammatical distribution aligned with cognitive accessibility (Principle 3; RQ3), analyzed via the frequency and proportional distribution of grammatical categories ($f \geq 21$) to determine whether high-utility core vocabulary prioritizes concrete, high-utility words over abstract function categories.

Principle 1: Pedagocigal Adequecy: Vocabulary Repetition and Lexical Diversity

Computed for a deeper analysis of lexical diversity and repetition frequency, Yule's K statistic presents highly critical areas of discussion regarding textbook design and the learning characteristics of the target population. The findings indicate that as educational levels progress, lexical diversity expands while vocabulary repetition declines. It was also established that vocabulary repetition in thematic readers such as *My Daily Life* (*Günlük Yaşamım*) and *My Social Life* (*Sosyal Yaşamım*) was higher than in the other two core textbooks of Level 2. This pattern may stem from the fact that these thematic readers lack activities, consisting solely of reading passages with longer textual structures. An increase in text length naturally entails a higher frequency of specific recurring lexical items (conjunctions, prepositions, and primary verbs). Consequently, at first glance, the expansion of lexical diversity across progressing levels might appear positive, suggesting that students with intellectual disabilities are provided with texts designed to enrich their vocabulary.

However, the declining trend observed in Yule's K values when textbooks are aggregated by level (Table 2) reveals a fundamental problem. The analytical results demonstrate a consistent reduction in aggregated K values per level. For example, while the individual K values of Level 1 textbooks were 57.1 (Textbook 1) and 50.6 (Textbook 2), respectively, the combined K value for these two textbooks dropped to 46.2. A similar declining trajectory was observed in the Level 2 core textbooks and the thematic textbook groupings. This systematic decline in Yule's K indicates that, rather than repeating the lexical inventory of preceding volumes, the aggregated textbooks continuously introduce new and distinct words to students. Indeed, the finding that the Grand Total K value—calculated by evaluating all six textbooks combined—was lower than all individual and level-grouped values further corroborates this issue. This pattern suggests that reading passages in Turkish language arts textbooks prepared for students with moderate to severe intellectual disabilities exhibit excessive lexical diversity, featuring distinct words while failing to provide sufficient repetition necessary for mastery. This outcome implies that each textbook exposes students to novel vocabulary while dropping previously introduced words, posing a distinct pedagogical challenge for this student group, who rely heavily on frequent target word exposure (Browder et al., 2006).

Even among students with mild intellectual disabilities, vocabulary acquisition requires repeated exposure to target words across diverse contexts (Burt et al., 2020). Furthermore, as the severity of intellectual disability increases, learning efficiency diminishes markedly (Schuchardt et al., 2010). Research on the vocabulary development of neurotypical students shows they require at least ten exposures to acquire a target word (Webb, 2007). When evaluated against this baseline—and assuming instructional exposure relies primarily on these curricular texts—the finding that 92% or more of the words in each individual textbook occurred only 1 to 10 times indicates that the vast majority of the lexical inventory is

exceptionally difficult for students with moderate to severe intellectual disabilities to acquire. Because words occurring 50 times or fewer accounted for 98.6% of the aggregated six-volume corpus, these textbooks lack the systematic word recycling and frequency threshold required to foster vocabulary development. Therefore, rather than continuously introducing novel lexical items, textbooks designed for students with moderate to severe intellectual disabilities must provide texts that ensure adequate recycling of previously introduced vocabulary, as systematic word repetition constitutes a cornerstone principle of effective vocabulary instruction (Nation, 2022; Schmitt, 2019).

Principle 2: Lexical Progression Across Grade Levels

The empirical findings revealed that when the textbooks were analyzed individually, total word counts (N) did not demonstrate a linear increase. However, when aggregated by educational level, a gradual upward trajectory emerges. The lowest total word count was observed in Level 1 Textbook 2, whereas the highest count was concentrated in the Level 2 thematic readers. Crucially, the total word count in Level 1 Textbook 1 surpassed that of Level 1 Textbook 2, underscoring a fundamental lack of systematicity in the vocabulary load across individual textbooks. The most pronounced percentage jump in total word count occurred during the transition from Level 2 Textbook 1 to Level 2 Textbook 2 (a 55.7% increase).

When evaluating how this volume expansion impacts vocabulary structure, Type-Token Ratios (TTR) exhibit a fluctuating pattern when evaluated individually. The lowest TTR was identified in Level 1 Textbook 1, whereas the highest TTR was observed in Level 1 Textbook 2. However, because TTR is inherently sensitive to text length (N), Yule's K statistic provides a more reliable indicator of underlying lexical repetition independent of document size. In terms of Yule's K, the highest level of vocabulary repetition was concentrated in Level 1 Textbook 1. As textbooks progress toward Level 2 Textbook 2,

Yule's K values steadily decline—indicating a systematic decrease in vocabulary repetition and a corresponding expansion in lexical diversity. Interestingly, this trajectory shifts within the thematic readers (*My Daily Life* and *My Social Life*), where vocabulary repetition exhibits a distinct resurgence.

Across the entire corpus, the cumulative number of unique words (V) was established as 1,861. The increase in unique word counts between textbooks was lowest within Level 1 textbooks (a 4.0% increase) and highest within Level 2 textbooks (a 41.4% increase), featuring a substantial difference of 96.5% (a margin of 361 unique words) between the textbook with the lowest and highest unique word count.

This empirical finding aligns with prior literature examining language arts textbooks prepared for neurotypical students in Türkiye, which similarly demonstrates an disproportionate and unsystematic expansion in total word counts (N) and unique word counts (V) across grade levels (Doğan, 2016; Kargın, 2019; Author). Analogous patterns are frequently reported in international studies investigating textbooks across various subject areas designed for neurotypical populations (Hiebert et al., 2024; Kearns & Hiebert, 2022; Mo & Bi, 2025).

The total word count (N) in Level 1 (Grades 1–4) Turkish language arts textbooks designed for students with moderate to severe intellectual disabilities is approximately 300 words lower than that of the Turkish language arts textbooks utilized by their neurotypical 1st-grade peers in Türkiye (Arslan & Bay, 2023). Moreover, when compared to the textbooks used by their neurotypical 4th-grade peers, the proportions of both total word count (N) and unique word count (V) in these materials drop to merely one-third (Kan & Karadavut, 2021). Similarly, the total word count (N) in Level 2 (Grades 5–8) textbooks—including thematic readers—is less than half of the total word count found in 8th-grade Turkish language arts textbooks prepared for neurotypical peers, with the gap in lexical diversity reaching

approximately one-fourth (Şimşek, 2021). These sharp disparities demonstrate that the Turkish language arts textbooks tailored for students with moderate to severe intellectual disabilities are substantially more simplified compared to those designed for their neurotypically developing peers, a degree of simplification that may be pedagogically appropriate given the reduced acquisition pace of this population (Tūbele & Landrāte, 2021).

Principle 3: Grammatical Distribution and Cognitive Accessibility

The distribution across grammatical categories of the top 99 most frequent words (words with an occurrence frequency of 21 or higher) provides crucial insights regarding the content orientation of these materials. Analytical results demonstrate that nouns represent the most predominant grammatical category encountered by students (52.5%), followed sequentially by verbs (24.2%), adverbs (11.1%), and pronouns (10.1%). The most frequent nouns—such as *anne* ("mother"), *çocuk* ("child"), *levha* ("sign/plate"), *baba* ("father"), *yer* ("place/ground"), *ev* ("home/house"), and *yemek* ("food/meal")—comprise concrete nouns and concepts directly tied to nuclear family life, the domestic environment, and the immediate social surroundings. The predominance of concrete nouns among top-frequency words represents a positive pedagogical alignment, as students with intellectual disabilities encounter significant difficulties in processing abstract concepts (Fabio et al., 2023). Similarly, the most frequent verbs were identified as high-utility primary verbs essential for daily functioning, including *olmak* ("to be/become"), *yapmak* ("to do/make"), *gitmek* ("to go"), *almak* ("to take/get"), *gelmek* ("to come"), *demek* ("to say"), and *görmek* ("to see"). The high density of these core items supports the premise that the national special education curriculum in Türkiye is designed to integrate functional daily living skills into instructional programming (Ministry of National Education, 2018a).

When comparing the top 10 most frequent words in textbooks designed for students with moderate to severe intellectual disabilities with those in materials prepared for their

neurotypical peers, functional words such as *ve* ("and"), *bir* ("a/one"), and *olmak* ("to be/become") typically overlap across corpora (Göz, 2019; Author; Şimşek, 2021); however, one distinct word diverges significantly. In our study, the prominent emergence of *anne* ("mother") as the fourth most frequent word overall constitutes a highly noteworthy finding that diverges from previous literature (Göz, 2019; Author, 2024; Şimşek, 2021). The pervasive occurrence of *anne* serves as a critical indicator reflecting the central role of the maternal figure, upon whom students with moderate to severe intellectual disabilities remain substantially dependent in their daily lives (Kerr et al., 2023). The most frequent words in these books suggest that across all six analyzed textbooks, reading passages were deliberately tailored to support functional, daily living skills aimed at facilitating the social adaptation and integration of students with moderate to severe intellectual disabilities.

Conclusions

The findings derived from this study deepen and extend the current understanding in the literature regarding the lexical structure of textbooks designed for students with moderate to severe intellectual disabilities. In alignment with international comparability standards established by the OECD (2018), length-independent parameters employed in this investigation—such as Yule's K statistic—provide empirical baseline data that enable direct cross-linguistic and cross-cultural comparisons with special education instructional materials globally.

Evaluated holistically, the results demonstrate that Turkish language arts textbooks prepared for students with moderate to severe intellectual disabilities are constructed with a significantly simpler lexical framework in terms of both quantitative word count (N) and lexical diversity (V) when compared to materials designed for their neurotypically developing peers. At a superficial level, this structural simplification might be interpreted as an appropriate pedagogical choice for a target population characterized by reduced learning rates

and limited cognitive capacities. However, our deeper analysis powered by Yule's *K* points directly to the "submerged iceberg problem." Rather than demonstrating an upward trajectory at individual levels, the decline in aggregated *K* values across advancing levels reveals that, instead of reinforcing and carrying over previously introduced vocabulary across textbooks, students are continuously exposed to a heavy load of novel and infrequently recycled words. Indeed, the finding that 98.6% of all unique words across the six textbooks occurred only 1 to 50 times—and more than 92% occurred only 1 to 10 times—demonstrates that the core pedagogical principles of "repeated exposure and systematic reinforcement," which are paramount for vocabulary acquisition in this population, are currently being overlooked.

Overcoming this fundamental bottleneck necessitates key practical and pedagogical implications for policymakers and textbook developers alike. First and foremost, textbook developers must adhere to pedagogically validated vocabulary lists as structural frameworks to eliminate unsystematic word selection in material design (Coxhead, 2000; Hadley & Mendez, 2021; Mo & Bi, 2025; Nation, 2022). In parallel, educational policymakers responsible for designing special education curricula should prioritize the development of native-language-based special education vocabulary frequency lists to guide instructional content creation.

Beyond relying strictly on frequency metrics, textbook engineering must be guided by pedagogical utility and gradual lexical progression. It is well established that infrequently occurring words in texts may represent high-utility vocabulary used frequently in students' daily lives, while conversely, certain high-frequency words may impose high cognitive demands on learners (Hiebert et al., 2024). Accordingly, the lexical inventory embedded in instructional materials should be carefully selected based on functional utility and pedagogical value for the target population (Coxhead, 2000). In the present study, the emergence of *anne* ("mother") as the fourth most frequent word exemplifies a beneficial vocabulary choice that

directly reflects the social-emotional reality of the target audience. However, the pervasive presence of hundreds of rarely occurring words that fail to afford adequate reinforcement constitutes a major barrier to vocabulary acquisition. To address this challenge, material developers should initially focus on immediate environmental concepts with which children are most familiar during early literacy stages. In subsequent educational levels, they should ensure a measured, systematically reinforced, and smooth transition toward Tier 2 academic vocabulary, as delineated by Beck et al. (2013).

Limitations and Directions for Future Research

Despite its significant empirical contributions, the present study contains certain methodological limitations that should be taken into account when interpreting the findings, alongside corresponding directions for future inquiry:

Limitations

Despite its contributions, several limitations of the present study should be acknowledged.

First, regarding scope constraints, this investigation was confined exclusively to the reading passages and reading comprehension activities embedded within six Turkish language arts textbooks prepared for students with moderate to severe intellectual disabilities (MSID). Instructional rubrics, explanatory texts designed for teachers, and materials belonging to other core subject areas (e.g., Mathematics, General Science) were beyond the scope of this study.

Second, in terms of the level of linguistic analysis, the lexical evaluation focused primarily on foundational frequency metrics and grammatical categorization (e.g., nouns, verbs, pronouns). Deeper linguistic dimensions—such as lexical concreteness or abstractness, morphological complexity (simple, derived, and compound words), average sentence length, and formulaic language constructs (idioms and proverbs)—were not systematically evaluated.

Finally, a key constraint lies in the corpus type. The analysis relied strictly on a written textbook corpus; consequently, it does not encompass the receptive and expressive oral vocabulary inventories utilized by the target student population in daily interactions.

Directions for Future Research

Based on the empirical findings and limitations of this study, several critical avenues for future research are proposed. First, researchers are encouraged to expand this analytical framework to encompass special education instructional materials across other academic disciplines, such as Mathematics and Social Skills. Additionally, spoken corpus investigations—specifically comparative empirical studies—should be conducted to evaluate the degree of alignment between the oral vocabulary (spoken corpus) utilized by students with MSID in daily life and the written vocabulary (written textbook corpus) embedded in their instructional materials.

Second, future research should integrate deeper semantic (e.g., concreteness and abstractness) and morphological analyses into vocabulary evaluations. Furthermore, while a minimum threshold of 10 repeated exposures is established for vocabulary acquisition in neurotypical children (Webb, 2007), experimental and longitudinal studies are needed to determine optimal exposure thresholds and repetition benchmarks for students with MSID, who often experience severe working memory impairments (Burt et al., 2020). Although individual variation among learners with intellectual disabilities is substantial, establishing these baseline repetition benchmarks remains critical for informing instructional material design.

Lastly, cross-cultural and cross-linguistic investigations are strongly needed. Applying length-independent metrics, such as Yule's K and lexical diversity indices, to conduct international comparative studies across special education textbooks in diverse languages will provide valuable contributions to the field. Scholars are particularly encouraged to expand

vocabulary frequency studies in languages other than English (Schmitt, 2019) to address its historical dominance in the textbook corpus literature (e.g., Liu et al., 2024).

Author contributions The contribution of each author are as follows: Author (Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Project administration), Author (Writing – original draft, Writing – review & editing), Author (Data curation [digitization of textbook texts, morphological segmentation, and data quality control])), Author (Data curation [digitization of textbook texts, morphological segmentation, and data quality control], Formal analysis), Author (Data curation [digitization of textbook texts, morphological segmentation, and data quality control]), Author (Conceptualization [study idea and selection of textbook materials], Methodology, Formal analysis), Author (Conceptualization, Methodology, Supervision, Writing – review & editing). We take public responsibility for the study and have agreed to have our name listed as a contributor.

Data availability The Corpus Word Frequency Lists and Descriptive Statistics for this study are presented as Supplementary Material

Conflict of interest The authors declare no conflict of interest

References

Aldahmash, A. H., & Omar, S. H. (2021). Analysis of activities included in Saudi Arabian chemistry textbooks for the inclusion of argumentation-driven inquiry skills. *Studies in Educational Evaluation*, *68,* Article 100968. https://doi.org/10.1016/j.stueduc.2020.100968

Arslan, K., & Bay, Y. (2023). Examination of primary school Turkish textbooks in terms of vocabulary. *Turkish Journal of Primary Education*, *8*(1), 14–27. https://doi.org/10.52797/tujped.1282167

Beck, I. L., McKeown, M. G., & Kucan, L. (2013*). Bringing words to life: Robust vocabulary instruction*. Guilford Press.

Bilge, H., & Kalenderoğlu, İ. (2022). The relationship between reading fluency, writing fluency, speaking fluency, reading comprehension, and vocabulary. *Education and Science*, *47*(209), 156–171. https://doi.org/10.15390/EB.2022.9609

Bober, A., Kosenko, Y., & Korol, O. (2024). Development of an interactive website with didactic games to enhance vocabulary in young students with intellectual disabilities. *Information Technologies and Learning Tools*, *102*(4), 39-50. https://doi.org/10.33407/itlt.v102i4.5711

Brako-Hiapa, G., Amissah, E. K., & Mensah, D. N. A. (2024). Enhancing reading comprehension: Visual documentation approaches for children with reading difficulties and disabilities. *International Journal of Innovative Research and Development*, *13*(3), 18-39. https://doi.org/10.24940/ijird/2024/v13/i3/mar24003

Browder, D. M., Wakeman, S. Y., Flowers, C., Rickelman, R. J., Pugalee, D., & Karvonen, M. (2007). Creating access to the general curriculum with links to grade-level content for students with significant cognitive disabilities: An explication of the concept. *The*

Journal of Special Education, *41*(1), 2-16. https://doi.org/10.1177/00224669070410010101

Browder, D. M., Wakeman, S. Y., Spooner, F., Ahlgrim-Delzell, L., & Algozzine, B. (2006). Research on reading instruction for individuals with significant cognitive disabilities. *Exceptional Children*, *72*(4), 392-408. https://doi.org/10.1177/001440290607200401

Burt, C., Graham, L., & Hoang, T. (2020). Effectiveness of computer-assisted vocabulary instruction for secondary students with mild intellectual disability. *International Journal of Disability, Development and Education*, *69*(4), 1273-1294. https://doi.org/10.1080/1034912X.2020.1776849

CAST. (2011). *Universal design for learning guidelines version 2.0*. https://udlguidelines.cast.org/static/udlg_fulltext_v2-0.doc

Cicchetti, D. V., & Feinstein, A. R. (1990). High agreement but low kappa: II. Resolving the paradoxes. *Journal of Clinical Epidemiology*, 43(6), 551-558. https://doi.org/10.1016/0895-4356(90)90159-M

Cote, A., & Banks, L. M. (2025). *Global disability inclusion report: Accelerating disability inclusion in a changing and diverse world*. UNICEF.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213-238. https://doi.org/10.2307/3587951

Cunningham, A. E., Zibulsky, J., Stanovich, K. E., & Stanovich, P. J. (2009). How teachers would spend their time teaching language arts: The mismatch between self-reported and best practices. *Journal of Learning Disabilities*, *42*(5), 418–430. https://doi.org/10.1177/0022219409339063

Di Eugenio, B., & Glass, M. (2004). The kappa statistic: A second look. *Computational Linguistics*, *30*(1), 95–101. https://doi.org/10.1162/089120104773633402

Doğan, F. (2016). *Türkçe dersi ortaokul öğrenci çalışma kitaplarındaki etkinliklerin söz varlığı üzerine bir araştırma* [A research on the vocabulary of activities in middle school Turkish course student workbooks] (Publication No. 429751) [Master's thesis, Siirt University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr

Ervin, L. (2024). *Applied behavior analysis picture to text word comprehension intervention for children with moderate to severe impairments due to autism spectrum disorder*. [Doctoral dissertation, The State University of New York]. ProQuest Dissertations & Theses Global.

Evmenova, A. S., & Behrmann, M. M. (2011). Research-based strategies for teaching content to students with intellectual disabilities: Adapted videos. *Education and Training in Autism and Developmental Disabilities*, *46*(3), 315-325. https://doi.org/10.1177/215416471104600302

Fabio, R. A., Bhattacharya, U., Wei, X., & Canegallo, V. (2023). Reading and writing in severe intellectual disability: A systematic review. *Current Developmental Disorders Reports, 10*(1), 1-13. https://doi.org/10.1007/s40474-023-00269-5

García Muñoz, O. (2012). *Lectura fácil: Métodos de redacción y evaluación*. Real Patronato sobre Discapacidad.

Göz, İ. (2019). *Yazılı Türkçenin kelime sıklığı sözlüğü [Dictionary of word frequency in written Turkish]*. Türk Dil Kurumu.

Hadley, E. B., & Mendez, K. Z. (2021). Learning words that matter: Selecting vocabulary words for young children. *Reading Teacher*, *74*(5), 595-605. https://doi.org/10.1002/trtr.1978

Hiebert, E. H., Pugh, A., & Kearns, D. M. (2024). The presence and progression of rare vocabulary in texts across elementary grades and between genres. *Education Sciences*, *14*(12), 1314. https://doi.org/10.3390/educsci14121314

Hua, Y., Woods-Groves, S., Kaldenberg, E. R., & Scheidecker, B. J. (2013). Effects of vocabulary instruction using constant time delay on expository reading of young adults with intellectual disability. *Focus on Autism and Other Developmental Disabilities*, *28*(2), 89-100. https://doi.org/10.1177/1088357613477473

Kan, M. O., & Karadavut, Z. (2021). A corpus-based comparison: Examination of 4th grade Turkish textbooks in terms of vocabulary. *Inonu University Journal of the Faculty of Education*, *22*(2), 1597–1610.

Karadağ, Ö. (2018). A variable suggestion for language education research: Vocabulary coefficient. *Journal of Mother Tongue Education*, *6*(2), 532–537. https://doi.org/10.16916/aded.372078

Kargın, F. (2019). *6, 7 ve 8. sınıf Türkçe ders kitaplarındaki söz varlığı ögeleri üzerine bir inceleme* [A study on vocabulary items in 6, 7 and 8th grade Turkish textbooks] (Publication No. 615021) [Master's thesis, Van Yüzüncü Yıl University]. Council of Higher Education Thesis Center. https://tez.yok.gov.tr

Karvonen, M., Flowers, C., & Wakeman, S. Y. (2013). Factors associated with access to the general curriculum for students with intellectual disability. *Current Issues in Education*, *16*(3), 1-19.

Kearns, D. M., & Hiebert, E. H. (2022). The word complexity of primary-level texts: Differences between first and third grade in widely used curricula. *Reading Research Quarterly*, *57*(1), 255–285. https://doi.org/10.1002/rrq.429

Kerr, J., Sharry, J., & Wilson, C. (2023). Parents' experiences of raising adolescents with intellectual or developmental disabilities. *Journal of Intellectual and Developmental Disability*, *48*(2), 206-214. https://doi.org/10.3109/13668250.2022.2057843

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Li, Y., & Deng, J. (2026). Curriculum accessibility through vocabulary readiness: An assessment-informed analysis of senior-secondary EFL textbooks. *Language Teaching Research*. https://doi.org/10.1177/13621688261448626

Liu, M., Li, Y., Su, Y., & Li, H. (2024). Text complexity of Chinese elementary school textbooks: Analysis of text linguistic features using machine learning algorithms. *Scientific Studies of Reading*, *28*(3), 235-255. https://doi.org/10.1080/10888438.2023.2244620

Love, M. L., Spies, T. G., & Morgan, J. J. (2017). Using e-books to acquire foundational academic vocabulary. *Intervention in School and Clinic*, *53*(2), 88-93. https://doi.org/10.1177/1053451217693368

McDonough, C., Song, L., Hirsh-Pasek, K., Golinkoff, R. M., & Lannon, R. (2011). An image is worth a thousand words: why nouns tend to dominate verbs in early word learning. *Developmental Science*, *14*(2), 181–189. https://doi.org/10.1111/j.1467-7687.2010.00968.x

Ministry of National Education. (2018a). *Orta-ağır zihinsel engeli ve otizm spektrum bozukluğu olan öğrenciler için okuma - yazma dersi öğretim programı I. kademe (1, 2, 3 ve 4. sınıf)* [Literacy curriculum for students with moderate to severe intellectual disabilities and autism spectrum disorder, Stage I (Grades 1, 2, 3, and 4)].

Ministry of National Education. (2018b). *Orta-ağır zihinsel engeli ve otizm spektrum bozukluğu olan öğrenciler için okuma - yazma dersi öğretim programı II. kademe (5, 6, 7 ve 8. sınıf)* [Literacy curriculum for students with moderate to severe intellectual disabilities and autism spectrum disorder, Stage II (Grades 5, 6, 7, and 8)].

Ministry of National Education. (2020). *Journal of Announcements of the Ministry of National Education*, *83*(2751), 927–947. General Directorate of Support Services.

Ministry of National Education. (2026). *Türkçe konu anlatımı* [Turkish subject explanations]. General Directorate of Special Education and Guidance Services. https://orgm.meb.gov.tr/ekpssmebozel/turkcekonular.html

Miranda-Garcia, A. & Calle-Martin, J. (2005). Yule's characteristic K revisited. *Language Resources and Evaluation, 39*(3), 287-294. https://doi.org/10.1007/s10579-005-8622-8

Mo, J., & Bi, P. (2025). Evaluation of vocabulary use in EFL textbooks: Evidence from curriculum words. *English Teaching and Learning*, *49*(1), 1-16. https://doi.org/10.1007/s42321-024-00170-3

Munir, Setiawan, W., Nugroho, E. P., Kusnendar, J., & Wibawa, A. P. (2018). The effectiveness of multimedia in education for special education (MESE) to improve reading ability and memorizing for children with intellectual disability. *International Journal of Emerging Technologies in Learning*, *13*(8), 254-263. https://doi.org/10.3991/ijet.v13i08.8291

Nation, I. S. P. (2022). *Learning vocabulary in another language*. Cambridge University Press.

National Reading Panel (US), National Institute of Child Health, & Human Development (US). (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. National Institute of Child Health and Human Development, National Institutes of Health. https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf

O'Connor, R. E., Beach, K. D., Sanchez, V. M., Kim, J. J., Knight-Teague, K., Orozco, G., & Jones, B. T. (2019). Teaching academic vocabulary to sixth-grade students with

disabilities. *Learning Disability Quarterly*, *42*(4), 231-243. https://doi.org/10.1177/0731948718821091

OECD. (2018). *Equity in education: Breaking down barriers to social mobility*. OECD Publishing.

Oshima-Takane, Y., Ariyama, J., Kobayashi, T., Katerelos, M., & Poulin-Dubois, D. (2011). Early verb learning in 20-month-old Japanese-speaking children. *Journal of Child Language*, *38*(3), 455–484. https://doi.org/10.1017/S0305000910000127

Özmen, E. R. (2020). *İşlevsel okuma-yazma metin kitabı: Günlük yaşamım [2. kademe (5, 6, 7 ve 8. sınıf)]* [Functional literacy textbook: My daily life (Stage 2: Grades 5, 6, 7, and 8)]. Millî Eğitim Bakanlığı.

Özmen, E. R., Doğanay Bilgi, A., Er, Z., & Açıkgöz, G. (2020). *Türkçe okuduğunu anlama, yazılı ifade, sözlü iletişim, dinlediğini anlama [2. kademe (5, 6, 7 ve 8. sınıf) ders kitabı]-2* [Turkish reading comprehension, written expression, oral communication, listening comprehension (Stage 2: Grades 5, 6, 7, and 8 textbook)-2]. Millî Eğitim Bakanlığı.

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*(1), 22-37. https://doi.org/10.1080/10888438.2013.827687

Pratiwi, A. S., & Fithriani, R. (2026). Flashcards for vocabulary development in students with intellectual disabilities: EFL teachers' perceptions. *Acuity: Journal of English Language Pedagogy, Literature, and Culture*, *11*(2). https://doi.org/10.35974/qsrm9858

Rapti, D., Gerogiannis, D., & Soulis, S. G. (2023). The effectiveness of augmented reality for English vocabulary instruction of Greek students with intellectual disability. *European*

Journal of Special Needs Education, *38*(2), 185-202. https://doi.org/10.1080/08856257.2022.2045816

Richards, B. (1987). Type/token ratios: What do they really tell us? *Journal of Child Language, 14*(2), 201–209. https://doi.org/10.1017/S0305000900012885

Robledo, J. (2017). Facilitating local understanding and literacy development for students with autism spectrum disorder through teacher training. *International Journal of Whole Schooling*, *13*(2), 47-62.

Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching*, *52*(2), 261-274. https://doi.org/10.1017/S0261444819000053

Schuchardt, K., Gebhardt, M., & Mäehler, C. (2010). Working memory functions in children with different degrees of intellectual disability. *Journal of Intellectual Disability Research*, *54*(4), 346-353. https://doi.org/10.1111/j.1365-2788.2010.01265.x

Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, *85*(3), 257-268. https://doi.org/10.1093/ptj/85.3.257

Sprenger, M. (2014). *Vocab rehab: How do I teach vocabulary effectively with limited time?* ASCD.

Stevens, E. A., Tanner, E., & Mowbray, M. H. (2025). A systematic review of mathematics vocabulary ınterventions for students with or at-risk for mathematics difficulty. *Remedial and Special Education, 46*(6), 484-500. https://doi.org/10.1177/07419325241265972

Subasno, Y., Nini, K., & Densi, C. (2022). Improving vocabulary reading skills with word card and picture card for moderate ıntellectual disabilities. *Journal of ICSAR*, *6*(1) 101-109. https://doi.org/10.17977/um005v6i12022p101

Suskind, D. (2020). *Otuz milyon kelime [Thirty million words]*. Buzdağı.

Şimşek, R. (2021). An analysis of the vocabulary of the 8th grade Turkish textbook. *Kesit Akademi*, *7*(28), 346–355. https://doi.org/10.29228/kesit.52271

Taşkın, B. (2023). *Mind deficiency of direct teaching method vocabulary of students with impact on development* [Master's thesis, Biruni University]. YÖK Thesis Center.

Tūbele, S., & Landrāte, E. (2021). Development of vocabulary in children with intellectual disabilities. *Listy Klinické Logopedie*, *5*(1), 76-88. https://doi.org/10.36833/lkl.2021.001

Turkish Language Association. (2026). *Türk Dil Kurumu sözlüğü* [Turkish Language Association dictionary]. https://sozluk.gov.tr/

Wang, H. H., Swanson, E., & Vaughn, S. (2023). Understanding the role of academic vocabulary in content acquisition for middle school students with and without disabilities. *Journal of Special Education*, *57*(1) 3-12. https://doi.org/10.1177/00224669221097944

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*(1), 46-65. https://doi.org/10.1093/applin/aml048

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143-2152. https://doi.org/10.1177/0956797613488145

Yopp, H. K., Yopp, R. H., & Bishop, A. (2009). *Vocabulary instruction for academic success*. Shell Education.

Yule, G. U. (1944). *The statistical study of literary vocabulary*. Cambridge Press.

Table 1.
The Textbooks Included in the Study Corpus

Level	Name of the Books	Author(s)
Level 1	Turkish Reading Comprehension, Written Expression, Oral Communication, Listening Comprehension [Level 1 (Grades 1, 2, 3, and 4) Textbook]-1	(Özmen, 2020)
	Turkish Reading Comprehension, Written Expression, Oral Communication, Listening Comprehension [Level 1 (Grades 1, 2, 3, and 4) Textbook]-2	(Özmen, 2020)
Level 2	Turkish Reading Comprehension, Written Expression, Oral Communication, Listening Comprehension [Level 2 (Grades 5, 6, 7, and 8) Textbook]-1	(Özmen et al., 2020)
	Turkish Reading Comprehension, Written Expression, Oral Communication, Listening Comprehension [Level 2 (Grades 5, 6, 7, and 8) Textbook]-2	(Özmen et al., 2020)
Level 2 Thematic Textbooks	Functional Literacy Textual Reader: *My Daily Life* [Level 2 (Grades 5, 6, 7, and 8)]	(Özmen, 2020)
	Functional Literacy Textual Reader: *My Social Life* [Level 2 (Grades 5, 6, 7, and 8)]	(Özmen, 2020)

Table 2

Descriptive Statistics and Lexical Diversity Measures of Textbooks for Students with Intellectual Disabilities

Textbook and the level	Total Words (N)	Unique Words (V)	TTR (%)	Yule's K
Level 1 Textbook 1	1,632	374	22.9	57.1
Level 1 Textbook 2	924	389	42.1	50.6
Level 1 Total	**2,556**	**624**	**24.4**	**46.2**
Level 2 Textbook 1	1,252	510	40.7	42.8
Level 2 Textbook 2	1,949	721	37	41.7
Level 2 Total	**3,201**	**1.008**	**31.5**	**38.1**
My Daily Life (*Günlük Yaşamım*)	2,289	641	28	48.4
My Social Life (*Sosyal Yaşamım*)	2,952	735	24.9	50.0
Thematic Textbooks Total	**5,241**	**1.151**	**22**	**41.1**
Grand Total (All Textbooks)	**10,998**	**1.861**	**16.9**	**35.4**

Note. N = Total word count (tokens); V = Unique words (types/headwords); TTR = Type-Token Ratio; Yule's K = Vocabulary repetition (lower values indicate lower repetition rate [thus higher lexical diversity]).

Table 3
Frequency Distributions of Unique (V) and Total Word Counts (N) Across Textbooks and Aggregated Levels

	Frequency Band	Word Count (V)	Percentage (%)
Level 1 Textbook 1	1-10	344	92,0
	11-20	20	5,3
	20+	10	2,7
	Total	374	
Level 1 Textbook 2	1-10	383	98,5
	11-20	4	1
	20+	2	0,5
	Total	389	
Level 1 Total	1-10	580	92,9
	11-20	25	4
	20+	19	3
	Total	624	
Level 2 Textbook 1	1-10	499	97,8
	11-20	8	1,6
	20+	3	0,6
	Total	510	
Level 2 Textbook 2	1-10	695	96,4
	11-20	20	2,8
	20+	6	0,8
	Total	721	
Level 2 Core Textbooks Total	1-10	951	94,3
	11-20	41	4,1
	20+	16	1,6
	Total	1008	
Level 2 *My Daily Life*	1-10	599	93,4
	11-20	32	5
	20+	10	1,6
	Total	641	
Level 2 *My Social Life*	1-10	680	92,5
	11-20	35	4,8
	20+	20	2,7
	Total	735	
Level 2 Thematic Textbooks Total	1-10	1041	90,4
	11-20	76	6,6
	20+	34	3
	Total	1151	
Grand Total (All Textbooks)	1-50	1835	98,6
	51-100	18	0,97
	101-150	5	0,27
	151+	3	0,16
	Total	1861	

Table 4

Distribution of the Top 99 Most Frequent Words (Frequency ≥ 21) Across Grammatical Categories

Word Category	Frequency (n)	Percentage (%)*
Noun	52	52,5
Verb	24	24,2
Adverb	11	11,1
Pronoun	10	10,1
Adjective	9	9,1
Conjunction	5	5,1
Preposition (Postposition)	4	4
Interjection	4	4
Total Target Words	99	

**Note.* A single word form may function under multiple grammatical categories depending on its contextual usage; consequently, the sum of cumulative percentages exceeds 100%. n = Number of distinct high-frequency unique word counts assigned to each category.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupllementalMaterialsCorpusWordFrequencyListsandDescriptiveStatistics.pdf