

דחיסות לקסיקלית בהוראת שפה ובהערכתה

איהאב ח'ליל אבו-רביעה**

מילות מפתח: רכישת שפה, הערכת שפה, אוצר מילים, מדדים לקסיקליים, מילות תוכן ומילות תפקיד

<https://doi.org/10.54301/PASZ3788>

תכונות של אוצר המילים כמדד מוביל בהערכת שפה

פיתוח מיומנות הכתיבה הוא פעילות בסיסית בתהליך ההוראה (East, 2006), וחיוברים הנכתבים בזמן מוגבל הם לרוב האמצעי להערכתה (Yoon, 2018). נמצא שמגוון המאפיינים הלשוניים שלפיהם אפשר להעריך את איכות הכתיבה, כגון מאפיינים לקסיקליים, תחביריים, פונולוגיים ומורפולוגיים (Kim & Crossley, 2018), תכונות של אוצר מילים הן אמצעי ניבוי חזק לאיכות הכתיבה יותר מאשר תכונות לשוניות אחרות (Kyle & Crossley, 2019; Vögelin et al., 2018), על כן פותחו מדדים רבים להערכת מיומנות לקסיקלית (Crossley et al., 2011). בבסיס ערך זה יוצג מדד לקסיקלי - מדד הדחיסות הלקסיקלית - ויוצעו לו יישומים להוראה ולהערכה מדויקת יותר של מיומנות הכתיבה בשפת אם (L1) ובשפה שנייה (L2) ולבחירה יעילה יותר של טקסטים במבחני L1 ו-L2.

דחיסות לקסיקלית

מדד הדחיסות הלקסיקלית (lexical density) הוא כלי לחישוב שיעור מילות התוכן מכלל המילים בטקסט כתוב או דבור (Ure, 1971; Halliday, 1989; Malvern et al., 2004; Nasser & Thompson, 2021). בדיקה לקסיקלית זו מושתתת על הבחנה בין מילות תוכן ומילות תפקיד. מילות התוכן (content words) הן מילים שיש להן ערך לקסיקלי או סמנטי (Demir-Vegter et

* ד"ר איהאב ח'ליל אבו-רביעה מלמד במרכז ללימודים קדם-אקדמיים באוניברסיטת בן-גוריון בנגב ובהתמחות בלשון העברית במכללה האקדמית לחינוך ע"ש קיי. את עבודת הדוקטור שלו כתב במחלקה ללשון העברית באוניברסיטת בן-גוריון בנגב בהנחייתם של פרופ' רוני הנקין-רויטפרב וד"ר רועי גפטר. ערך זה נכתב בתמיכת המלגה ע"ש צבי ינאי של משרד המדע והטכנולוגיה (מספר 16617-3) ומלגת נגב - רוח דרומית.

(al., 2014) שנושא את החלק העיקרי במשמעות של המשפט. הן נמצאות בלב המסר שהמשפט נועד להעביר והן כוללות: אובייקטים, אירועים, מאפיינים, מיקומים והוראות (Corver & Van Riemsdijk, 2013). מבחינת חלקי דיבר הן כוללות שמות עצם (דוגמת **מכונית**); פעלים (דוגמת **עמד**); שמות תואר (דוגמת **חרוצה**); ולפי חלק מהחוקרים גם תוארי פועל (דוגמת **לאט**).

מילות/מיליות התפקיד (function words) הן מילים שנדרשות לפי דקדוק השפה כדי לחבר מילות תוכן יחד לכדי משפט (Stubbs, 2002; Shi et al., 2006; Richels et al., 2010). הן משמשות בעיקר בתפקיד תחבירי ולא סמנטי, והמשמעות שלהן אינה ברורה כמו המשמעות של מילות התוכן (Hochmann et al., 2010). בקבוצה זו נכללות ה"א היידוע; מילות יחס (דוגמת **אצל**); מילות קישור (דוגמת **או**); מילית השעבוד (דוגמת **ש**); מילות קריאה (דוגמת **אוי!**); יסודות עזר (דוגמת **צריך**); כינויי רמז (דוגמת **זאת**); כינויי גוף (כגון **אתן**); כמתים (דוגמת **יותר**); מספרים (דוגמת **אחד**).

הדחיסות הלקסיקלית - בגרסתה השכיחה - מחושבת לפי מספר מילות התוכן בטקסט חלקי כלל המילים שבו (Halliday, 1989; O'Loughlin, 1995; Strömquist et al., 2002; Johansson, 2008). המדד מציע יחס ישיר בין מידת הדחיסות הלקסיקלית להיקף התוכן המובע (Read, 2000). לצורך הדגמה, הדחיסות הלקסיקלית במשפט "**נער צעיר דיבר עם אישה מבוגרת**" שיעור הדחיסות הלקסיקלית הוא 83% שכן יש 5 מילות תוכן (**נער, צעיר, דיבר, אישה, מבוגרת**) מתוך שש מילים בסך הכול (חמש מילות תוכן, ומילת תפקיד אחת: **עם**).

מדד הדחיסות הלקסיקלית מכמת את הידע ואת הרעיונות בטקסט (Michel et al., 2007) והוא נחשב למחוון לעושר טקסטואלי (Berman et al., 2011). עם העלייה בגיל וברמת הכשירות הלשונית עולה רמת הדחיסות הלקסיקלית (Ravid & Levie, 2010; Berman et al., 2011; Kim, 2014; Gregori-Signes & Clavel-Arroitia, 2015). לפיכך, בעזרת מדידת הדחיסות הלקסיקלית אפשר לעקוב אחר ההתפתחות הלקסיקלית של דוברי L1 וגם של דוברי L2 (Laurén, 2002; Johansson, 2008; Bulté & Housen, 2012).

שכיחות לקסיקלית

המאפיין העיקרי המבחין בין שתי קבוצות המילים - מילות

ניתן לנסות להעלות את רמת הדחיסות הלקסיקלית של אותו צירוף שמני במקומות שאפשר, למשל, להשתמש בצירופי סמיכות חבורה (**כוכבי השמיים**) במקום צירופי סמיכות פרודה (**הכוכבים של השמיים**).

• **בהערכת כתיבה:** מורים ומרצים שמעריכים כתיבה של תלמידים או סטודנטים, בעיקר בשאלות ההבעה בכתב שבהן הנבחנים כותבים חיבורים או סיכומים, יכולים לבדוק את רמת הדחיסות הלקסיקלית בכל חיבור בעזרת בדיקה של מספר מילות התוכן יחסית למילות התפקיד בשלוש שורות באמצע החיבור. בעזרת פעולה זו הם יכולים לקבל הערכה מספרית המכמתת את הידע שנמסר בחיבור (Nasseri & Thompson, 2021). כאמור, ככל שיש יותר מילות תוכן יחסית למילות תפקיד - במספר מוגדר של מילים - אזי המידע המבוטא בכתיבה רב יותר. עניין זה עשוי לשקף את מידת ההיכרות של הלומד עם עולם התוכן של נושא הכתיבה וגם את יכולתו של הכותב לשלב מידע רב יחסית לכותבים אחרים במספר מוגדר של מילים (Biber & Gray, 2010). נוסף לכך, שיטה זו מאפשרת למעריכים להעריך ביעילות הבדלים במיומנויות כתיבה של לומדים שונים באותה קבוצה ואף של התקדמות אישית של תלמיד בפרק זמן כלשהו.

• **בבחירת טקסטים למבחני שפה וטקסטים לתהליך**

ההוראה: באופן כללי טקסטים שהדחיסות הלקסיקלית בהם גבוהה הם טקסטים קריאים פחות. דהיינו, רמת הקלות שבה ניתן להבין אותם כאשר קוראים אותם למטרה ספציפית היא נמוכה יותר (Harrison & Bakker, 1998; Castello, 2008; Khamahani, 2015). כתוצאה מכך נדרש זמן רב יותר כדי לקרוא אותם ולהביןם. מחברי בחינות L1 ו-L2 ברמות ההשכלה השונות יכולים לבחור טקסטים למבחנים לפי רמות הדחיסות הלקסיקלית שלהם. ככל שעולה גיל הלומדים וגם רמת הכשירות הלשונית שלהם, אפשר לבחור טקסטים דחוסים לקסיקלית יותר/קריאים פחות. למשל טקסטים בבחינות רמת השפה באוניברסיטאות ובמכללות יהיו דחוסים לקסיקלית יותר מאשר הטקסטים שבבחינות הבגרות, שכן שפה אקדמית ידועה בדחיסות הלקסיקלית הגבוהה (Biber & Gray, 2016). שיטת דירוג זו תקפה גם בבחירת הטקסטים שהמורים מלמדים במהלך שנת הלימודים בגילים השונים. ניתן ללמד טקסטים בעלי רמות דחיסות לקסיקלית גבוהות יותר ככל שהכיתה גבוהה יותר, ואף ניתן להעלות את רמת הדחיסות הלקסיקלית של הטקסטים במהלך אותה שנת לימודים. שימוש נוסף למדד זה הוא בהשוואה בין טקסטים לצורך צמצום הבדלים ברמת הקושי של אותם טקסטים במועדים שונים של אותו מבחן. שכן לפעמים נבחנים טוענים להבדלים ברמת הקושי של הטקסטים

התוכן ומילות התפקיד - הוא השכיחות או התפוצה של מילים השייכות לקבוצות אלה. השכיחות ביחס בין שתי הקבוצות היא גם המפתח ליישום המדד בהוראת שפה ובהערכתה. קבוצת מילות התפקיד כוללת מעט מאוד מילים והיא קבוצת המילים השימושית ביותר בשפה (Shi et al., 2006). דהיינו, הדוברים או הכותבים משתמשים בהן יותר מאשר במילות תוכן. למשל, נמצא כי המילים *a, an, and, the* הן המילים השכיחות ביותר באנגלית ושיעורן מגיע לכ-20% מכל קורפוס (Milton, 2009), וכי המילה השכיחה ביותר בטקסטים דבורים וכתובים באנגלית היא תווית היידוע *the* (Schmitt & McCarthy, 1997). ממצאים דומים בעברית כ-L1 (Bolzky & Berman, 2020) ובעברית כ-L2 לערבית (אבו-רביעה, 2022) התקבלו בבדיקת שכיחותן של מילות/מיליות התפקיד; שני המחקרים הראו כי וי"ו החיבור, ה"א היידוע, מילית השעבוד ש, בְּ/בִּי היחס, לְ/לִי היחס הן המילים השכיחות ביותר בטקסטים שנבדקו. שכיחות זו היא פועל יוצא של תפקידן התחבירי. יתרה מזו, במחקר על ארבע שפות כולל עברית דווחו ערכים גבוהים מאוד של למות (ערכים מילוניים) של מילות תפקיד ביחס ללמות של מילות תוכן (Strömquist et al., 2002). מילים שכיחות הן לרוב קלות לרכישה ביחס למילים נדירות והן נגישות לרוב הלומדים. אך מילים נדירות יותר נגישות רק ללומדים ברמה מתקדמת (Tidball & Treffers-Daller, 2008), ועל כן שיעורן בכתיבה אינו סממן למיומנות לשונית גבוהה. כמובן, בכל אחת משתי קבוצות המילים יש מילים שכיחות לצד מילים נדירות.

יישומים בהוראת שפה ובהערכת הקניית השפה

אתמקד כאן בהשלכות יישומיות בהוראת הכתיבה ובהערכתה ובבחירת טקסטים להוראה ולהערכה.

• **בהוראת הכתיבה:** הוראת הכתיבה בבתי ספר, במכללות ובאוניברסיטאות מבקשת להעניק ללומד מיומנויות לכתיבה אקדמית וגם להבנת טקסטים אקדמיים. אפשר להשתמש במדד זה להערכת רכישת מיומנות כתיבה אקדמית (Linnarud, 1976; Snow & Uccelli, 2009; Lu, 2012), שכן שפה אוריינית ופורמלית, ובמיוחד שפה אקדמית, דחוסה לקסיקלית יותר מאשר שפה שאינה אוריינית ואינה פורמלית (Ure, 1971; Halliday, 1989; Biber, 2006). בהוראת הכתיבה ניתן ללמד את הכותבים דרכים שונות להגדלת הדחיסות הלקסיקלית כמו שימוש מוגבר בצירופים שמניים (Biber & Gray, 2016). המורים והמרצים יכולים ללמד את תלמידיהם שימוש בצירופי סמיכות (דוגמת **ספר המחר**); צירופים של שם עצם ושם תואר (דוגמת **משימה מאתגרת**); צירופים של שם עצם וצירוף יחס (דוגמת **הבית ליד המסילה**); צירופים של שם עצם ופסוקית (דוגמת **הספר שנמצא על המכתבה**). לאחר מכן

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. John Benjamins Publishing. <https://doi.org/10.1075/scl.23>

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9(1), 2–20. <https://doi.org/10.1016/j.jeap.2010.01.001>

Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511920776>

Bolozky, S., & Berman, R. A. (2020). Parts of speech categories in the lexicon of Modern Hebrew. In R. A. Berman (Ed.), *Usage-based studies in modern Hebrew: Background, morphology, and syntax* (pp. 265–330). John Benjamins Publishing Company. <https://doi.org/10.1075/slcs.210.10bul>

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In Housen, A., Kuiken, F., & Vedder, I. (Eds.), *Dimensions of L2 performance and proficiency* (pp. 21–46). Amsterdam, the Netherlands: Benjamins. <https://doi.org/10.1075/lllt.32.02bul>

Castello, E. (2008). *Text complexity and reading comprehension tests*. Peter Lang.

Corver, N., & Van Riemsdijk, H. (Eds.). (2013). *Semi-lexical categories: The function of content words and the content of function words*. Walter de Gruyter. <https://doi.org/10.1515/9783110874006>

Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011). What is lexical proficiency? Some answers from computational models of speech data. *TESOL Quarterly*, 45(1), 182–193. <https://doi.org/10.5054/tq.2010.244019>

Demir-Vegter, S., Aarts, R., & Kurvers, J. (2014). Lexical richness in maternal input and vocabulary development of Turkish preschoolers in the Netherlands. *Journal of psycholinguistic research*, 43(2), 149–165. <https://doi.org/10.1007/s10936-013-9245-7>

East, M. (2006). The impact of bilingual dictionaries on lexical sophistication and lexical accuracy in tests of L2 writing proficiency: A quantitative analysis. *Assessing Writing*, 11(3), 179–197. <https://doi.org/10.1016/j.asw.2006.11.001>

Gregori-Signes, C., & Clavel-Arroita, B. (2015). Analysing lexical density and lexical diversity in university students' written discourse. *Procedia-Social and Behavioral Sciences*, 198, 546–556. <https://doi.org/10.1016/j.sbspro.2015.07.477>

Halliday, M. A. K. (1989). *Spoken and written language*. Oxford: Oxford University Press.

Harrison, S., & Bakker, P. (1998). Two new readability predictors for the professional writer: pilot trials. *Journal of Research in Reading*, 21(2), 121–138. <http://dx.doi.org/10.1111/1467-9817.00049>

Hochmann, J. R., Endress, A. D., & Mehler, J. (2010). Word frequency as a cue for identifying function words in infancy. *Cognition*, 115(3), 444–457. <https://doi.org/10.1016/j.cognition.2010.03.006>

Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Lund Working Papers in Linguistics*, 53, 61–79.

באותו סוג של מבחן במועדים שונים (למשל מועד חורף לעומת מועד קיץ באותו שאלון בגרות).

מסקנות וסיכום

הבנת אופן הערכת רמת הדחיסות הלקסיקלית בחיבורים ובטקסטים, ההבדלים בין שתי קבוצות המילים שעליהן מתבססת בדיקה זו (מיליות תוכן ומילות תפקיד) ומאפייניהם והשפעת רמת הדחיסות הלקסיקלית על מידת קריאות הטקסט יכולים לשמש מורים, מרצים, בוחנים וכותבי תוכניות לימודים כדי לשפר, לדייק ולייעל את תהליך ההוראה וההערכה של השפה. זאת ועוד, הם יכולים להתאים במידת ביטחון גבוהה יותר טקסטים לרמת הלומדים או הנבחנים.

אין מדד אחד לבדו יכול לספק הערכה מדויקת לטיב הכתיבה של הלומדים או רמת הקלות/הקושי של טקסטים שעליהם הם נבחנים, ויש להשתמש במדדים שונים במידת האפשר כדי לקבל הערכה מדויקת יותר. למשל, נוסף על מדד הדחיסות הלקסיקלית ניתן להשתמש במדד הגיוון הלקסיקלי כדי להפוך את ההערכה למדויקת וליעילה יותר (אבו-רביעה, 2020 ב). בשלבים יישומיים מתקדמים יותר ניתן לראות שהיחס בין דחיסות ללקסיקליות לבין קריאות מורכב יותר (אבו-רביעה, 2017, 2020 א; Abu-Rabiah, 2020).

המקורות

אבו-רביעה, א' (2017). **הגיוון הלקסיקלי, הדחיסות הלקסיקלית ורמת המופשטות כמדדי התקדמות ברכישת העברית כ-L2 לערבית של צעירים בדואים בנגב**. [תזה]. אוניברסיטת בן-גוריון בנגב.

אבו-רביעה, א' (2020א). רמת המופשטות, הגיוון הלקסיקלי והדחיסות הלקסיקליות כמדדי התקדמות ברכישת העברית כשפה שנייה לערבית. **אורינות ושפה**, 8, 71–93.

אבו-רביעה, א' (2020ב). גיוון לקסיקלי: תיאוריה, פרקטיקה ומה שביניהן. **לקסי-קיי**, 13, 3–5. <https://doi.org/10.54301/xevu1230>

אבו-רביעה, א' (2022). **רכישת הלקסיקון העברי בקרב תלמידי תיכון דוברי ערבית בנגב: בחינה אמפירית**. [דיסרטציה]. אוניברסיטת בן-גוריון בנגב.

Abu-Rabiah, E. (2020). Lexical measures for testing progress in Hebrew as Arab students' L2. *Journal of Language and Linguistic Studies*, 16(3), 1096–1114. <https://doi.org/10.17263/jlls.803551>

Berman, R., Nayditz, R., & Ravid, D. (2011). Linguistic diagnostics of written texts in two school-age populations. *Written Language & Literacy*, 14(2), 161–187. <https://doi.org/10.1075/wll.14.2.01ber>

- Schmitt, N., & McCarthy, M. (1997). *Vocabulary: Description, acquisition and pedagogy*. Cambridge university press.
- Shi, R., Werker, J. F., & Cutler, A. (2006). Recognition and representation of function words in English-learning infants. *Infancy*, 10(2), 187–198. https://doi.org/10.1207/s15327078in1002_5
- Snow, C.E., & Uccelli, P. (2009). The challenge of academic language. In D.R. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp. 112–133). New York: Cambridge University Press. <https://psycnet.apa.org/doi/10.1017/CBO9780511609664.008>
- Strömqvist, S., Johansson, V., Kriz, S., Ragnarsdottir, H., Aisenman, R., & Ravid, D. (2002). Toward a cross-linguistic comparison of lexical quanta in speech and writing. *Written Language & Literacy*, 5(1), 45–67. <https://doi.org/10.1075/wll.5.1.03str>
- Stubbs, M. (2002). *Words and phrases: Corpus studies of lexical semantics*. Blackwell.
- Tidball, F. and Treffers-Daller, J. (2008). Analysing lexical richness in French learner language: What frequency lists and teacher judgement can tell us about basic and advanced words. *Journal of French language Studies*, 18(3), 299–313. <https://doi.org/10.1017/S0959269508003463>
- Ure, J. (1971). Lexical density and register differentiation. In G.E. Perren, & J. M. Trim (Eds.), *Applications of Linguistics* (pp. 443–452). Cambridge University Press.
- Vögelin, C., Jansen, T., Keller, S. D., Machts, N., & Möller, J. (2019). The influence of lexical features on teacher judgements of ESL argumentative essays. *Assessing Writing*, 39, 50–63. <https://doi.org/10.1016/j.asw.2018.12.003>
- Yoon, H. J. (2018). The development of ESL writing quality and lexical proficiency: Suggestions for assessing writing achievement. *Language Assessment Quarterly*, 15(4), 387–405. <https://doi.org/10.1080/15434303.2018.1536756>
- Khamahani, G. (2015). A corpus-based analysis of Tehran Times and Azeri News headlines: Focus on lexical density and readability. *International Journal of Humanities Social Sciences and Education*, 2(1), 12–16.
- Kim, J. Y. (2014). Predicting L2 writing proficiency using linguistic complexity measures: A corpus-based study. *English Teaching*, 69(4), 27–51. <https://doi.org/10.15858/engtea.69.4.201412.27>
- Kim, M., & Crossley, S. A. (2018). Modeling second language writing quality: A structural equation investigation of lexical, syntactic, and cohesive features in source-based and independent writing. *Assessing Writing*, 37, 39–56. <https://doi.org/10.1016/j.asw.2018.03.002>
- Kyle, K., & Crossley, S. A. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal*, 102(2), 333–349. <https://doi.org/10.1111/modl.12468>
- Laurén, U. (2002). Some lexical features of immersion pupils' oral and written narration. *Working Papers in Linguistics*, 50, 63–78.
- Linnarud, M. (1976). Lexical density and lexical variation: An analysis of the lexical texture of Swedish students' written work. *Studia Anglica Posnaniensia: An International Review of English Studies*, 7, 45–52.
- Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal*, 96(2), 190–208. https://doi.org/10.1111/j.1540-4781.2011.01232_1.x
- Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development*. New York: Palgrave Macmillan. <https://doi.org/10.1057/9780230511804>
- Michel, M. C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *International Review of Applied Linguistics in Language Teaching*, 45(3), 241–259. <https://doi.org/10.1515/iral.2007.011>
- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters. <https://doi.org/10.21832/9781847692092>
- Nasseri, M., & Thompson, P. (2021). Lexical density and diversity in dissertation abstracts: Revisiting English L1 vs. L2 text differences. *Assessing Writing*, 47, 100511. <https://doi.org/10.1016/j.asw.2020.100511>
- O'Loughlin, K. (1995). Lexical density in candidate output on direct and semi-direct versions of an oral proficiency test. *Language testing*, 12(2), 217–237. <https://doi.org/10.1177/2F026553229501200205>
- Ravid, D., & Levie, R. (2010). Hebrew adjectives in later language text production. *First Language*, 30(1), 27–55. <https://doi.org/10.1177/2F0142723709350529>
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511732942>
- Richels, C., Buhr, A., Conture, E., & Ntourou, K. (2010). Utterance complexity and stuttering on function words in preschool-age children who stutter. *Journal of fluency disorders*, 35(3), 314–331. <https://doi.org/10.1016/j.jfludis.2010.06.001>