

רמת המופשטות, הגיוון הלקסיקלי והדחיסות הלקסיקלית כמדדי התקדמות ברכישת העברית כשפה שנייה לערבית

איהאב ח'ליל אבו רביעה

החוג ללשון העברית, אוניברסיטת בן-גוריון

מדדי עושר לקסיקלי הם כלי חשוב להערכת רכישת הלקסיקון ומשמשים לעיתים קרובות במחקר התפתחות שפה שנייה (L2) (Johansson, 2008). מאז אור והאלידיי (Ure, 1971; Halliday, 1985) הוצעו בספרות המחקר מדדים שונים לבדיקת התפתחות הלקסיקון של לומדי L2. מאמר זה בוחן התקדמות ברכישת הלקסיקון של עברית כ-L2 ב-50 חיבורים טיעוניים שנכתבו על ידי תלמידי כיתות ט' ו-י"א בתיכון בדואי בנגב, לפי שלושה מדדים לקסיקליים: רמת מופשטות השפה, גיוון לקסיקלי ודחיסות לקסיקלית. מתודולוגיה: רמת מופשטות השפה חושבה לפי שיעור תמניות שמות העצם המופשטים מתוך כלל תמניות שמות העצם; גיוון לקסיקלי חושב לפי שיעור התבניות (types) -המילים השונות- חלקי כלל התמניות (tokens) -כל המילים; דחיסות לקסיקלית חושבה לפי שיעורן של תמניות מילות התוכן מכלל תמניות המילים. ממצאים: ראשית, נמצאה עלייה מובהקת ברמת מופשטות השפה עם הגיל; שנית, לא נצפתה עלייה מובהקת עם הגיל בדחיסות הלקסיקלית ולא בגיוון הלקסיקלי. תוכנית הלימודים בכיתה י"א מתמקדת בקישוריות כדי ליצור טקסטים קריאים. התמקדות זו מחייבת שימוש רב במילות התפקיד שמוביל לירידה בדחיסות. אי-העלייה בגיוון נובעת מהבדלי אורך החיבורים בשתי הקבוצות: תלמידי י"א כתבו טקסטים ארוכים בממוצע מאלה של תלמידי ט' ועל כן הגיוון אצלם ירד. מסקנות: רמת המופשטות יעילה כמדד להתקדמות ברכישת אוצר המילים גם בשלבים ראשוניים של הרכישה (כיתה ט') ואילו הגיוון והדחיסות מצריכים רמת כשירות גבוהה יותר (כיתה י"א); ניתן לבצע בדיקת רמת המופשטות בטקסטים שוני-אורך בשונה מבדיקת הגיוון; תוכנית הלימודים משפיעה על חלק מהמדדים הלקסיקליים; ולבסוף, יש לקחת בעירבון מוגבל את הנחת היסוד שעלייה במדדים האלה משקפת עלייה ברמת הכשירות הלקסיקלית.

מילות מפתח: רכישת שפה שנייה, התפתחות אוצר מילים, מדדים לקסיקליים, הערכת כתיבה, עברית לדוברי ערבית.

מבוא

רכישת הלקסיקון

החשיבות של חקירת תהליכי רכישת שפה הוכרה זה מכבר (Gardner & Lambert, 1972; Wode, 1981; Wilkins, 1985; Van Els et al., 1986; Spolsky, 1988, 1989). מחקר זה מתמקד בהיבט ספציפי של רכישת שפה שנייה שהוא רכישת הלקסיקון (lexicon) או אוצר המילים (vocabulary). המילים הן לבני היסוד של השפה, ומהן נוצרים מבנים גדולים יותר, כמו משפטים, פסקות וטקסטים. לומדי שפה רבים רואים ברכישה של אוצר מילים את עיקרה של רכישת השפה השנייה, ולכן הם מקדישים את רוב זמנם לשינון רשימות מילים (Read, 2000). וילקינס (Wilkins, 1972) אף ציין: "בלי דקדוק ניתן למסור מעט {ידע}, בלי אוצר מילים לא ניתן למסור לא כלום {של ידע}". נמצא כי ידיעת אוצר מילים מאפשרת שימוש יעיל בשפה (Nation, 1993) וכי

הלקסיקון הוא ההיבט החשוב ביותר בידע של L2 (Troike, 2006), כתוצאה מכך נחוצה מאוד בדיקת ידע אוצר המילים (vocabulary knowledge) ומדידתו.

כשירות לקסיקלית, מיומנות לקסיקלית, עומק לקסיקלי ועושר לקסיקלי הם שמות רבים לעניין אחד: שליטה בהיבטיו השונים של הלקסיקון, כמו שימוש במילה הנכונה בהקשר ושימוש במשמעויות השונות של המילה. מתוכם, המושג הנפוץ ביותר הוא עושר לקסיקלי (lexical richness) (Koizumi, 2012, p.555). ההגדרה הרחבה של כולם היא יכולת השימוש במילים באופן שונה ומלא. הווי אומר, כשירות לקסיקלית משלבת שני סוגים של כשירות: כשירות בין-לקסיקלית, כלומר היכולת לבחור מילה נכונה בין אחרות דומות לה; וכשירות תוך-לקסיקלית, כלומר היכולת להשתמש במילה אחת בכל ההקשרים הנכונים (Tanaka, 2012, p.1).

רכישת ערך לקסיקלי בשפת אם (L1) וב-L2 מסובכת יותר מזכירת מילה ופירושה: היא כוללת זיהוי של תבניות קוליות והתאמה לאורתוגרפיה (במקרה של רכישת השפה הכתובה), הגייה מדויקת, דיוק סמנטי, והתאמה לתוכן הרעיוני וההקשרי (Bogaards, 2001; Nation, 2005). מוסכם שרכישת ערכים לקסיקליים היא עירוב של אסטרטגיות של זיכרון ואסטרטגיות של קישור בין מילים עם זיכרון של הגדרות. כך הרוכש מצרף מילים חדשות בתוך קטגוריות סמנטיות שונות ובונה יחסים בין הקבוצות השונות עד שהלקסיקון מתפתח ומתבסס על קישורים סמנטיים פנימיים בתוך כל קבוצה בנפרד (Haastруп & Henriksen, 2000).

הרבה מחקרים ציינו שלרוכשי L2 יש פחות עומק של הידע הלקסיקלי (depth of lexical knowledge) מאשר לדוברי שפה ראשונה. כלומר, איכות השילוב של ההיבטים השונים של הלקסיקון אצל רוכשי L2 אינה זהה לזו של דוברי השפה הילידיים. דוברי L2 נוטים להעביר יחסים סמנטיים משפת האם שלהם לשפת היעד בתהליך של התערבות סמנטית משפת האם. מעבר זה משפיע על הבחירה הלקסיקלית הסמנטית המקובלת בשפת היעד ומוריד מהעומק הלקסיקלי שלהם (Ordóñez et al., 2002; Verhallen & Schoonen, 1993). ביטוי נוסף של התערבות נוסח על ידי בלום-קולקה ולוונסטון (Blum-Kulka & Levenston, 1983): רוכשי L2 בתחילת התהליך חושבים שיש תרגום אחד מקביל לכל מילה ב-L1, מה שמתברר להם מאוחר יותר כלא נכון. הם נעזרים במבנה של שפת אימם כדי ללמוד את L2 ובמקרי שוני בין השפות התערבות L1 גורמת לטעות ב-L2.

מחקרים רבים, למשל (Meara & Fitzpatrick, 2000; Laufer & Hulstijn, 2001; Vermeer, 2001; Jiang, 2007; Mondria & Wiersma, 2007; Nation, 2007; Qian, 2004; David, 2008; Fitzpatrick et al., 2008; Hilton, 2008) ניסו לכמת את התפתחות אוצר המילים על פי מדדים לקסיקליים, כגון רמת מופשטות, גיוון לקסיקלי, דחיסות לקסיקלית, דיוק לקסיקלי או תחכום לקסיקלי, מקוריות לקסיקלית, מספר הטעויות הלקסיקליות ועוד. נמצא שהתפתחות בלקסיקון יכולה לעודד התפתחות בדקדוק (Bell, 2009), שידיעת אוצר מילים מאפשרת שימוש יעיל בשפה (Nation, 1993) ושעושר לקסיקלי ב-L2 קשור להישגים אקדמיים באופן כללי (Polio, 2001; Daller et al., 2003). על כן בדיקת התפתחות לקסיקלית תורמת גם לניבוי הצלחה בתחומים אחרים, ולמעקב אחרי התפתחות קוגניטיבית כללית.

רוב המחקרים המשתמשים במדדים לקסיקליים לבדיקת התקדמות ברכישת אוצר המילים בוצעו על שפות שנכתבות באותיות לטיניות – בעיקר אנגלית – ואף לא אחד על העברית כשפה שנייה של דוברי הערבית בישראל. מחד גיסא, ישנה חשיבות גדולה לתחום רכישת הלקסיקון בספרות המחקר הבין-לאומית ומאידך

גיסא, ישנו מחסור במחקרים לאומיים המתחשבים בהקשר הישראלי הספציפי. על כן, ראיתי לנכון להקרין אור על לקונה מחקרית זו ולברר יעילותם של שלושה מדדים לקסיקליים להערכת התקדמות ברכישת העברית כ-L2 לדוברי הערבית – רמת מופשטות השפה, גיוון לקסיקלי ודחיסות לקסיקלית – תוך התייחסות לממצאים, אשר נובעים מסיטואציה חברתית-לשונית ספציפית.

המדדים הנבדקים להתפתחות לקסיקלית

רמת המופשטות

כבר לפני יותר משישים שנה הצביע פיאז'ה בתיאוריה המפורסמת שלו תיאוריית ההתפתחות הקוגניטיבית שבהתפתחות של הילדים קודמת החשיבה המוחשית והקונקרטי לחשיבה המופשטת (Piaget, 1959). בהגדרה פשוטה ופשטנית למדי, שם עצם מוחשי הוא שם שניתן לתפוס באחד מחמשת החושים, כמו ספר, ואילו שם עצם מופשט הוא שם שלא ניתן לתפוס באחד מחמשת החושים אלא הוא נתפס במחשבה, כמו שמחה. אך שמות עצם רבים אינם ניתנים לסיווג לפי הגדרה זו: יש בהם מאפיינים מוחשיים וגם מאפיינים מופשטים. למשל מכה אין לה ממדים משלה, אולם היא מורגשת באמצעות חושי האדם.

רביד מציגה סולם המורכב מעשר דרגות סמנטיות של שמות עצם ואינה מסתפקת בחלוקה דיכוטומית של מוחשי ומופשט: ראשית, שמות עצם מוחשיים, למשל שולחן; שנית, שמות פרטיים, למשל יוסף; שלישית, שמות עצם קיבוציים ושמות עצם המציינים מקום, למשל ספרייה ומשפחה; רביעית, שמות עצם המציינים תפקידים חברתיים, למשל דוד; חמישית, שמות עצם גנריים, למשל מישו; שישית, שמות עצם המציינים זמן, למשל דקה; שביעית, שמות עצם המציינים אירועים, למשל מסיבה; שמינית, שמות עצם מופשטים שאפשר לחוש (לשמוע, להרגיש ולראות) אותם, למשל מכות וצעקות, וגם שמות עצם בתוך צירופים כבולים, למשל ירד מהפסים ולשדר על אותו גל; תשיעית, שמות עצם מופשטים שלא ניתן לחוש אותם בחושים, למשל כבוד; עשירית, שמות עצם מופשטים גזורים, למשל ידידות (Ravid, 2006, p.796).

רביד מצאה שילדים משתמשים בעיקר בשמות עצם מוחשיים, פרוטוטיפים וחד-משמעיים וכשהם מתבגרים יותר, בשלב מתקדם יותר של רכישת שפה הם משתמשים יותר בשמות עצם מופשטים, מטפוריים ורב-משמעיים. היא מציינת גם שככל שעולה הגיל עולה השימוש בשמות עצם באופן כללי, עולה השימוש בשמות עצם מופשטים (Ravid, 2006, p.794) וגם עולה השימוש של שמות עצם מוחשיים במובנים מטפוריים (Ravid, 2005, p.337).

גם דוברי L2 רוכשים ביתר קלות מילים שכיחות ומוחשיות בשפת היעד מאשר מילים נדירות ומופשטות. ככל שזמן הרכישה של L2 מתארך, הם מתחילים להשתמש בשפה מופשטת יותר (Tanaka, 2012). לכן רמת המופשטות של L2, כמו גם של L1, בקורפוס נתון משמשת מדד חשוב להתפתחות לקסיקלית.

רמת מופשטות השפה (level of abstractness) הוא הממד הראשון שנבדק במחקר זה באמצעות ספירה של תמניות שמות העצם המופשטים יחסית לכלל תמניות שמות העצם בחיבור. עלייה בחלקם של שמות העצם המופשטים בטקסט נתון מעידה על שפה מופשטת יותר וכן על רמת כשירות לקסיקלית גבוהה יותר.

גיוון לקסיקלי

המדד השני הנבדק במחקר זה הוא מדד הגיוון הלקסיקלי (lexical diversity). אמנם יש חוקרים המזהים מושג זה עם עושר לקסיקלי (lexical richness), אך במחקר זה מקובלת ההנחה של חוקרים רבים אחרים שמבחינים בין שני המושגים ורואים בגיוון רק פן אחד של עושר לקסיקלי לצד, למשל, תחכום לקסיקלי, מקוריות לקסיקלית, ודחיסות לקסיקלית (Johansson, 2008, pp. 62-63). מדד זה כשמו כן הוא, מדד המתייחס למידת גיוון המילים בטקסט (Malvern et al., 2004).

מדד זה נבדק בעזרת נוסחאות רבות, שהשכיחה מביניהן היא נוסחת TTR (type-token ratio). וונדל ג'ונסון (Johnson, 1939, 1944) הוא הראשון שהציג את מדד ה-TTR. מדד הבודק את הגיוון הלקסיקלי בטקסט באמצעות ספירת התבניות (types) (המילים השונות) חלקי כלל התמניות (tokens) (כל המילים: הדומות וגם השונות) של המילים בטקסט. תבנית היא ערך מילוני, הכולל צורות נטייה שונות. למשל הצירוף *ילדים ילדות* נספר כשתי תמניות של תבנית אחת, זו של *ילד*. כך גם כל זמני הפועל נחשבים לתבנית אחת (*שמר, נשמר, ושמרי* הן שלוש תמניות של תבנית אחת: *שמר*), אך בניינים שונים מהווים תבניות עצמאיות (*כתב הכתיב* הן שתי תבניות שונות). ככל שמספר החזרות על התבניות קטן הגיוון הלקסיקלי עולה (Lieven, 1978; Bates et al., 1991).

מחקרים רבים הצביעו על חיסרון מתודי במדד ה-TTR: הוא מושפע מאורך הטקסט (Hess et al., 2007; Jarvis, 2002, 2012; Malvern et al., 2004; McCarthy & Jarvis, 2007). ההסבר לירידה זו בגיוון הלקסיקלי בגין אורך טקסט הוא שהתבניות של האדם מוגבלות ואילו תמניותיו אינן מוגבלות. כלומר, לאדם יש מספר מסוים של מילים שונות (תבניות) זמינות על נושא כלשהו אבל הוא יכול לכתוב או להגיד אינסוף מילים (תמניות) על אותו נושא. יתר על כן, לעיתים קרובות יש לכותב או לדובר מטרה לחזור על תבניות כלשהן לחיזוק דעתו או בשל צרכים פרגמטיים אחרים כמו הדגשת הרעיון. בעיה זו בולטת במיוחד בהשוואת טקסטים של בני גילאים שונים (כפי שנראה בהמשך, זה המצב במחקר זה). מחקרים הראו שעם העלייה בגיל הנבדקים, ה-TTR דווקא יורד (Johansson, 2008, p.63) בשל העובדה שאורכי הטקסטים עולים עם הגיל. לכן מומלץ לא להשתמש במדד TTR כמדד יחידאי לעושר לקסיקלי בטקסטים שאורכם שונה כי הוא מדד יעיל יותר כשמשווים טקסטים באורכים שווים.

גישת אוצר המילים התיאורטי (theoretical vocabulary) מנסה לנטרל את השפעת האורך על מדד ה-TTR. בגישה מוצעת בחירה רנדומאלית של מספר מילים מהטקסט ברצף או לא ברצף (למשל 100 תמניות) ואז מיונן לתבניות השונות וחישוב הגיוון הלקסיקלי. הגיוון הנמדד על קטע נתון נחשב למייצג את שאר הטקסט, בלי קשר לאורכו. כך ניתן למדוד טקסטים באורכים שונים בלי השפעה של האורך (Johansson, 2008, p.65). הפתרון שהציע ג'רוויס כדי להתגבר על השפעת האורך (Jarvis, 2013) הוא לחלק טקסטים ארוכים למדגמים של חמישים תמניות ולעשות TTR של כל מדגם ואז לחשב את הממוצע של ה-TTR בכל המדגמים בכל הטקסטים.

במחקר של קויזומי (Koizumi, 2012) ושל קויזומי ואינאמי (Koizumi & In'nami, 2012) על השפעת אורך הטקסט על כל מדדי הגיוון הלקסיקלי בטקסטים שנעים בין 50-200 תמניות נמצא כי כל המדדים של גיוון לקסיקלי מושפעים באופן זה או אחר מאורך הטקסט כשהטקסטים נעים בין 50-200 תמניות, אבל ההשפעה מצטמצמת יותר כשאורך הטקסטים נע בין 100-200 תמניות. ההמלצה שלהם היא להשתמש במדדי

הגיוון הלקסיקלי (לרבות מדד ה-TTR) בטקסטים שכוללים מאה תמניות עד מאתיים. ולמעשה הגיוון הלקסיקלי נבדק בעיקר בטקסטים שאורכם כ-100 תמניות (Malvern & Richards, 2002; Johansson, 2008). גיוון לקסיקלי ב-L1 וב-L2 נבדק גם בטקסטים ארוכים מ-100 תמניות, למשל במחקרים של Hess et al., 1986 ושל Choi & Jeong, 2016, אך למיטב ידיעתי לא ב-L2 עברית.

בדיקת מדד זה חשובה מכיוון שנמצא קשר חיובי בין מידת הגיוון הלקסיקלי לבין האיכות הכללית של טקסטים דבורים וכתובים (Yu, 2010) ומכיוון שהוא משפיע על מהימנות הדובר בעיני זולתו, על האופן שבו הנמענים יקבלו אותו ועל היעילות התקשורתית בינו לבין נמעניו, כי המאזינים מעדיפים מסר מורכב על פני מסר פשוט (Dillard & Pfau, 2002). אמנם לא תמיד יש קשר ישיר בין מסר מורכב לגיוון לקסיקלי: לפעמים מסר שטחי מובע בגיוון לקסיקלי גבוה ולהפך. אך עם זאת יש נטייה לשני הגורמים הללו (גיוון לקסיקלי גבוה ומסר מורכב) לבוא יחדיו. לקסיקון מורכב, בהנחה שהוא מוסר מסר מורכב, נתפס כמעניין ומושך בעיקר כי הוא מייצג בחירה לקסיקלית מורכבת, שמיוחסת לידע ואמינות. ומכאן החשיבות של מדידת יכולת זו וטיפול כמדד ליכולת לשונית ותקשורתית. מדד זה נחשב למנבא למיומנות הכללית בשפה של הלומדים (Zareva et al., 2005).

במחקר הנוכחי נבדק לראשונה הגיוון הלקסיקלי כאחד משלושה מדדים להתקדמות ברכישת העברית כ-L2 לערבית. לפי המלצות שנקבעו במחקר, נעשתה בדיקת הגיוון לפי מדד ה-TTR על טקסטים שכללו מאה תבניות במידת האפשר. לא השתמשתי בשיטות שהוצעו להשוואת הגיוון בין טקסטים שונים-אורך כי הם הוצעו להתמודד עם טקסטים ארוכים יותר (לפחות מאה מילים), אורך שלא התאפשר בכיתה ט'. על כך ארחיב בהמשך.

דחיסות לקסיקלית

המדד השלישי הנבדק במחקר זה הוא מדד הדחיסות הלקסיקלית (lexical density) שמתבסס על חלוקת המילים בשפה לשתי קטגוריות: מילות תוכן ומילות תפקיד (Fries, 1952). מילות תוכן הן מילים הכוללות ערך סמנטי או תוכן לקסיקלי (Ure, 1971). פעלים ושמות עצם הם מילות תוכן מובהקות, כמו סיפור הרעיש.

לגבי שמות תואר יש חילוקי דעות. לדעת רוב החוקרים, למשל האלידי וג'והנסון (Johansson, 2008, p.66; Halliday, 1985, p.64), הם כלולים בקבוצת מילות התוכן, אך למשל אור חולק על קביעה זו (Ure, 1971). תוארי הפועל שנויים עוד יותר במחלוקת לגבי הסיווג שלהם למילות תוכן או למילות תפקיד (Lyons, 1981; Halliday, 1989). יש המפרידים אותם לשני סוגים, אחד בעל ערך דקדוקי ונמנה עם מילות התפקיד והשני, למשל כל אלה שגזורים משמות תואר, בעל ערך לקסיקלי ונמנה עם מילות התוכן (Johansson, 2008, pp.66-67). בעברית הם מופיעים בצורות דקדוקיות רבות (Brandes & Ravid, 2017) לכן זוהי קטגוריה קשה במיוחד לסיווג. קדרי (תשמ"ה) בתוך (שלזינגר, תש"ס, עמ' 54) טוען שיש לחלק אותם לשתי קבוצות: הקבוצה הראשונה היא תוארי פועל מובהקים, כגון פתאום, ממש שהם מילות תפקיד כי אין הם עומדים בפני עצמם אלא משרתים מילים באחת משתי הקבוצות שבמילות התוכן המובהקות: אין להם מטען סמנטי מלא. המטען הסמנטי שלהם הוא חלק מהמטען הסמנטי של המגדיר שלהם. הקבוצה השנייה היא תוארי פועל הכוללים צורות מורכבות וצירופים המתפקדים תחבירית כתוארי פועל, שההשתייכות הקטגוריאליה המקורית של רכיביהם ניכרת בהם הם מילות תוכן, כגון במהירות. לצד מילות תוכן קטגוריה זו מכונה קבוצת המילים

הפתוחה, שכן היא מתרחבת כל הזמן (Gordon & Caramazza, 1982), או המילים הלקסיקליות (Shi & Werker, 2003).

מילות התפקיד הן מילים שנושאות תוכן דקדוקי והן נדרשות על ידי מבנה השטח כדי להדביק מילות תוכן יחדיו (Ure 1971; Degani et al., 2011). קטגוריית מילות התפקיד כוללת: מילות יחס (מעל), מילות קישור (לכן), כמתים (רוב), כינויים (אלה), מספרים (חמישה), הפועל היה (כמציין זמן), ה"א הידיעה וצורני זיקה (ש). לצד מילות תפקיד קטגוריה זו מכונה קבוצת המילים הסגורה (Gordon & Caramazza, 1982), או מילים דקדוקיות (Shi & Werker, 2003).

למרות הקושי הרב בהבחנה בין מילות תוכן למילות תפקיד הוקפד במחקר זה על עקביות במיון תוארי הפועל שבקורפוס. הקפדה זו נובעת ממסקנותיהם של האלידיי וג'והנסון (Halliday, 1985; Johansson, 2008) וחוקרים אחרים שלא משנה היכן יושם הקו המפריד בין שתי הקבוצות האלה כל עוד החוקרים נשארים עקביים.

רוב אוצר המילים בשפה הוא מילות תוכן, שכן קבוצת מילות התוכן היא קבוצה פתוחה שמתרחבת בדרכים שונות (Abdalla et al., 2010), למשל גזירה פנימית ושאיילה מלועזית. לעומת זאת, קבוצת מילות התפקיד היא קבוצה סגורה, קטנה יחסית ואינה מתרחבת, שיש בה כמה עשרות מילים (Juste et al., 2012). בהתפתחות של הפרט מילות התוכן קודמות למילות התפקיד. ילדים לעתים קרובות אומרים ואף כותבים משפטים בלי מילות תפקיד בכלל. אחרי שהם מתבגרים הם לומדים את מילות התפקיד ואת חשיבותן לקשר בין מילות התוכן ומתחילים להשתמש בהן (Johansson, 2008, p.66).

דחיסות לקסיקלית – שמכונה גם דחיסות מילונית – היא היחס בין מילות תוכן למילות תפקיד בטקסט נתון. טקסט דחוס לקסיקלית הוא טקסט שיש בו יותר מילות תוכן יחסית למילות תפקיד. הוא כולל יותר מידע מאשר טקסט שיש בו יותר מילות תפקיד (Johansson, 2008, p.65). ככל שיש יותר מילות תוכן, עולה חלקו של התוכן הלקסיקלי (הידע) הנמסר בטקסט. השיטה המקובלת ביותר במחקרים לחישוב דחיסות לקסיקלית בטקסט היא חישוב שיעור תמניות מילות התוכן מכלל תמניות הטקסט: מילות תוכן / כלל המילים X 100 (שראל, תשנ"ט, עמ' 406; Johansson, 2008, p.65).

באמצעות בדיקה זו אנו מקבלים רעיון על "דחיסת המידע" שבטקסט (information packaging) היות טקסט שבו יש אחוז גבוה של מילות תוכן מכיל יותר מידע מאשר טקסט שיש בו אחוז גבוה של מילות תפקיד (Johansson, 2008). יתרה מזו, בעזרת מדד זה ניתן לבדוק את "האוריינות" של המדגם הנבדק. בשפה פורמלית או בשפה אוריינית יש יותר מילות תוכן מאשר בשפה לא פורמלית ואילו בשפה לא פורמלית יש יותר מילות תפקיד מאשר בשפה פורמלית (Demir-Vegter et al., 2014). על כן היא אמצעי חשוב להערכת התקדמות ברכישת האוריינות, ובפרט ברכישת הלקסיקון.

שיטה

שאלת המחקר

שאלת המחקר היא: באיזו יעילות מודדים רמת המופשטות, הגיוון הלקסיקלי והדחיסות הלקסיקלית את ההתקדמות ברכישת הלקסיקון של העברית כשפה שנייה לדוברי ערבית בשתי קבוצות גיל: כיתות ט' ו-י"א?

משתתפים וקורפוס

מחקר זה נעשה על חיבוריהם של תלמידי תיכון ערבי בצפון הנגב. התלמידים היו בשתי קבוצות גיל: כיתה ט' וכיתה י"א. מכל כיתה השתתפו 25 תלמידים. כך נוצר קורפוס של 50 חיבורים. בכל קבוצת גיל וגם בקורפוס הכולל היה ייצוג כמעט זהה של שני המגדרים.

חומרים

המטלה

המשתתפים צפו בסרטון שבו נשים עובדות ועקרות בית דיברו על מידת האושר בחייהן שקשורה להיותן עובדות במשרה מלאה או עקרות בית. הסרטון ארך כחמש דקות ובמהלכו לא התנהל דיון בין המורה לבין התלמידים או בין התלמידים לעצמם. הסרטון היה בשפה הערבית כדי שלא תהיה השפעה ישירה של לשון הדמויות בסרטון על אוצר המילים של התלמידים בעת כתיבת החיבור. בעקבות זאת, כל משתתף כתב חיבור טיעוני ובו הביע את דעתו על נושא עבודת האישה. הכתיבה התבצעה כשהמורה משגיח על המשתתפים ומוודא שכל אחד כותב בעצמו בלי עזרה מאף אחד אחר. לא הייתה הגבלה של אורך החיבור וגם לא הגבלת זמן, אך כל התלמידים סיימו את הכתיבה שלהם בפחות משעה וחצי (שני שיעורים רצופים). המורה לא נתן להם פירושי מילים בעברית. הסרטון דן בנושא רלוונטי עבור החברה הבדואית והוא נבחר כדי שיעורר את התלמידים לכתוב. על החיבורים בוצעו שלוש הבדיקות הלקסיקליות.

חישוב המדדים הלקסיקליים

על כל חיבור נעשו שלוש הבדיקות: ראשית, רמת המופשטות. בעקבות מחקרה של רביד (Ravid, 2005) רמת מופשטות הטקסט הוגדרה כתמניות שמות העצם המופשטים חלקי כלל תמניות שמות העצם כפול 100. שם עצם מוחשי נספר רק אם ניתן לתפוס אותו באחד מחמשת החושים. למשל, בטקסט מסוים יש חמישה שמות עצם מופשטים מסך עשרה שמות עצם, אזי רמת המופשטות בטקסט היא 50%.

שנית, הדחיסות הלקסיקלית חושבה כמספר תמניות מילות התוכן חלקי כלל התמניות כפול 100. ננקטה שיטה אחת ועקבית לחלוקת תוארי פועל למילות תוכן ולמילות תפקיד. שלישית, הגיוון הלקסיקלי נבדק לפי הממד TTR (Johansson, 2008). גיוון לקסיקלי חושב כמספר התבניות השונות חלקי כלל התמניות כפול 100.

ספרות המחקר מכירה בהשפעה של אורך הטקסט על בדיקת הגיוון הלקסיקלי. כפי שצוין לעיל, מחקרים רבים ממליצים לערוך את בדיקת הגיוון הלקסיקלי על טקסטים של 100 תמניות עד 200 תמניות (Koizumi, 2012; Koizumi & In'nami, 2012). בשונה מהגיוון הלקסיקלי, לא נמצאה השפעה ישירה של אורך הטקסט על הדחיסות הלקסיקלית ועל רמת המופשטות. למרות זאת, החלטתי לבדוק גם את שתי ב-100 התמניות הראשונות בכל חיבור.

שיקולים בספירת המילים

שיקול עיקרי בספירת מילים בקורפוס מתייחס לייחוד האורתוגרפי של השפה העברית יחסית לשפות שבהן נערכו מחקרים מסוג זה. בעברית (ובערבית) רק חלק ממילות התפקיד הן תיבות נפרדות בכתב, כמו על ואילו השאר מצטרפות לתיבות של מילות תוכן, כמו בכפר. לכן יש מספר גדול של תיבות שכוללות גם מילת תפקיד

וגם מילת תוכן {ככפר : ב (מילת תפקיד); כפר (מילת תוכן)}. נשאלת השאלה כיצד להתייחס אליהן וכיצד להגיע לספירה נכונה של מילות התפקיד. במחקר זה, נבחרו שיטות שונות לפי צורכי הבדיקה, כי השאלה מהי מילה שונה מבדיקה לבדיקה.

למעשה, 100 תיבות בעברית כוללות הרבה יותר יסודות לשוניים ממאה מילים באנגלית. אדגים זאת באמצעות המשפט הבא: "הלכנו לביתנו שבכפר" כולל 3 מילים בעברית, אם סופרים תיבות, ואילו באנגלית המשפט המקביל "we went to our house (which is) in the village" כולל 8 או 10 מילים. המשפט באנגלית ארוך פי שלושה מהמשפט בעברית. סביר להניח שהפער באורך המשפטים בין שתי השפות נהיה ברור יותר ככל שהטקסטים מתארכים.

השאלה כעת היא האם תכונה זו של סינתטיות בעברית חשובה לצורך הבדיקות שבמחקר זה. בבדיקת דחיסות לקסיקלית שכל כולה מבוססת על ההבדל בין מילות תוכן למילות תפקיד חשוב להפריד תיבות המורכבות ממילות תוכן ותפקיד לרכיביהן ולספור כל יסוד לחוד. למשל, במשפט "והנערים הלכו" הכולל שתי תיבות בלבד נספרו 4 יסודות: שתי מילות תוכן: נערים; הלכו ושתי מילות תפקיד: ו-החיבור; ה"א הידיעה. לא הופרדו יסודות (צורני) נטייה וגזירה שאינן בגדר מילים, כך שנערים היא מילת תוכן אחת, וכך גם הלכו, כי אין זה מקובל בבדיקה זו. אף שבניתוח מורפולוגי ניתן לראות במילה נערים צירוף של נער + צורן ריבוי זכר, וגם במילה הלכו כלול צורן ריבוי זכר של הפועל בזמן עבר גוף שלישי.

לעומת זאת, בבדיקות רמת המופשטות והגיוון הלקסיקלי לא הופרדה תיבה שכללה מילת תוכן ומילת תפקיד לשתי מילים. לגבי רמת המופשטות, הסיבה היא שמילות התפקיד כלל אינן רלוונטיות בבדיקה שעוסקת רק בשמות עצם. באשר לבדיקת הגיוון הלקסיקלי, הסיבה היא שבבדיקה של ה-TTR מודדים את הגיוון הלקסיקלי לפי מספר התבניות חלקי כלל התמניות. בתיבה מורכבת כגון והנערים אין יתרון משמעותי בספירת שלושת היסודות לחוד (ה+נערים) שכן צפוי שמילות התפקיד החבורות יחזרו באותה תכיפות בממוצע בלי קשר לעושר הלקסיקלי של הכותב. לכן אפשר להתעלם מהן ולהתמקד ביסודות לקסיקליים ברורים כגון מילות תוכן ומילות תפקיד עצמאיות. תיבה שכללה מילת תוכן ומילת תפקיד נספרה רק כתבנית אחת: למשל בבתים נחשבה לתבנית של התבנית בית.

לאור החלוקות השונות של רצף צורנים למילים, יודגם להלן ניתוח של הפסוקית כשירד הגשם לפי בדיקת רמת המופשטות, הגיוון הלקסיקלי והדחיסות הלקסיקלית:

טבלה 1. בדיקת רמת המופשטות של הפסוקית כשירד הגשם

	2	1	
סה"כ	הגשם	כשירד	
היסוד	שם עצם	פועל: לא	
	מוחשי	נספר	

רמת המופשטות: $0\% = 100 \times 1/0$

התפתחות לקסיקלית בעברית של דוברי ערבית

טבלה 2. בדיקת הגיוון הלקסיקלי של הפסוקית כשירד הגשם

סה"כ	2	1
2	כשירד הגשם	תיבות =
		תמניות
2	גשם ירד	תבניות

הגיוון הלקסיקלי: 2 תבניות / 2 תמניות X 100 = 100%

טבלה 3. בדיקת הדחיסות הלקסיקלית של הפסוקית כשירד הגשם

סה"כ	4	3	2	1
4 צורנים	גשם	ה	כש ירד	צורנים
מילת תוכן	2 מילות תוכן	כן	לא	כן

הדחיסות הלקסיקלית: 4/2 X 100 = 50%

השערות

מצופה עלייה בביצועים בשלוש הקטגוריות הנבדקות בקורלציה ישירה עם הגיל וזמן החשיפה. דהיינו, רמות המופשטות, הגיוון הלקסיקלי והדחיסות הלקסיקלית יהיו גבוהים בחיבוריהם של תלמידי כיתה י"א יותר מאשר בחיבורי תלמידי כיתה ט'. אכן מחקרים רבים קישרו את העלייה בשלוש הקטגוריות עם התקדמות ברכישת שפה שנייה וגם עם עלייה בהישגים אקדמיים באופן כללי (Kroll & Tokowicz, 2001; Polio, 2001; Daller et al., 2003; Ravid, 2005).

תוצאות

להלן תוצאות הבדיקות לפי קבוצות:

טבלה 4. טבלה המשווה את הממוצע, את סטיית התקן ואת מבחן t של שלוש הבדיקות הלקסיקליות בשתי קבוצות הגיל

דחיסות לקסיקלית		גיוון לקסיקלי		רמת המופשטות		כיתה
SD	M	SD	M	SD	M	
.058001	41.72%	.081894	50.12%	.136566	35.76%	ט' (n=25)
.041569	39.60%	.089643	49.96%	.121500	46.24%	י"א (n=25)

t-Test	.0133*	.9487	.1520
--------	--------	-------	-------

*p < .05

רמת המופשטות

נתוני רמת המופשטות בכיתה ט' ובכיתה י"א, כפי שמוצגים בטבלה 4, מאששים את הציפיות הראשוניות: יש עלייה בממוצע שילוב שמות עצם מופשטים יחסית לכלל שמות העצם עם העלייה בגיל וזמן רכישת העברית כ-L2. התוצאות מוכיחות ששפתם של תלמידי כיתה י"א מופשטת יותר משפתם של תלמידי כיתה ט' (46.24% ו-35.76% בהתאם). ישנה עלייה של כעשרה אחוזים של מופשטות בשנתיים. לפי בדיקת t-test פער זה הינו פער מובהק סטטיסטית (0.0133; <math>p < .05</math>). מסתבר שבשנתיים האלה נרכשו שמות עצם מופשטים רבים שאכן נכתבים בחיבורים.

ספרות המחקר שנסקרה בתחום רמת מופשטות טקסטים של רוכשי L2 אינה מצביעה על מידת העלייה ברמת המופשטות באחוזים לפי זמן רכישה, אלא רק מציינת באופן כללי שיטתו של רב יותר של שפה מופשטת ככל שזוכים לזמן רב יותר ברכישת L2 (Ravid, 2005; Kroll & Tokowicz, 2001). נראה כי עמימות זו מוצדקת בשלב זה של המחקר כי תהליך הרכישה תלוי במספר גורמים מעבר לזמן החשיפה, כגון מידת הקשר בין שתי השפות, גיל הלומד, שיטות ההוראה (Filipović & Hawkins, 2013), המוטיבציה של הדובר ללמוד (Saranraj & Shahila, 2016) ועוד. לכן רק אחרי בידוד הגורמים ובדיקות כמותניות נפרדות ומשולבות ניתן יהיה לקבוע חד-משמעית את התלות בין מידת מופשטות לזמן רכישה לעומת גורמים אחרים.

כפי שהתבסס במחקר, רמת המופשטות של טקסט היא אחד המדדים לכשירות ב-L2. לפי מדד זה ניתן להסיק שרמת הכשירות בעברית כ-L2 בכיתה ט' נמוכה מאשר בכיתה י"א, כצפוי. על פי הממצאים בסעיף זה אני סבור שבדיקת רמת המופשטות כמדד להתפתחות העברית כ-L2 לערבית היא בדיקה יעילה גם בגילאים שנבדקו פה. ניתן לערוך אותה כבר בחטיבות ביניים ולקבל תמונת מצב מייצגת לרמת ההתקדמות ברכישת L2. היא אינה מצריכה גיל שבו רמת הכשירות גבוהה, כמו תיכונים, מכללות ואוניברסיטאות.

למרות ההבדל הברור באורך החיבורים של שתי קבוצות התלמידים (כפי שנראה בסעיף הבא), ההבדל ברמת המופשטות בין שתי הקבוצות מובהק סטטיסטית. בעקבות ממצאים אלה ניתן להסיק שבדיקה זו – בשונה מבדיקות לקסיקליות אחרות, בפרט גיוון לקסיקלי – אינה מושפעת מאורך הטקסט, לפחות לא בטקסטים קצרים עד כמאה תמניות.

גיוון לקסיקלי

לפני תחילת העבודה היה מצופה שהגיוון הלקסיקלי יעלה עם זמן הרכישה של L2. ציפייה אינטואיטיבית זו נתמכה גם בקריאה ראשונית, התרשמותית, של השורות הראשונות בחיבורים, שהעידה על לקסיקון מגוון יותר בכיתה י"א. אך טבלה 4 מראה ירידה קטנה ולא משמעותית בין כיתה ט' (50.12%) לכיתה י"א (49.96%). לפי בדיקת t-test פער זה אינו מובהק סטטיסטית (0.9487; $p > .05$). ממצאים אלה אינם תואמים את הציפיות שלנו המבוססות על מחקרים רבים כמו (Daller et al., 2003; Malvern et al., 2004). נשאלת השאלה מה ההסבר לאי-העלייה של הגיוון הלקסיקלי בין כיתות ט' ל-י"א. האומנם ייתכן מצב כזה אחרי שהתלמיד ללא ספק למד יותר לקסמות חדשות במהלך שנתיים?

הסבר אפשרי לאי-העלייה הוא האורך של החיבורים. ממוצע אורך חיבור בכיתה ט' הוא 76.28 תמניות ואילו בכיתה י"א הוא 98.88 תמניות. דהיינו, יש הבדל של כעשרים תמניות בממוצע בין שתי הקבוצות לטובת הקבוצה המבוגרת ביותר.

כאמור, אורך הטקסט ללא ספק משפיע על הגיוון הלקסיקלי שבו, כי מספר התבניות בדרך כלל מוגבל ומספר התמניות אינו מוגבל. ככל שהטקסט מתארך יותר אבל ממשיך לעסוק בנושא אחד מצומצם כך יש חזרה על תבניות שכבר הוזכרו כי התבניות מוגבלות וכתוצאה מזה הגיוון הלקסיקלי יורד. במקרה שהכותב מחליף נושא בתוך אותו חיבור, נפתח פתח לגיוון לקסיקלי נוסף ומדד הגיוון הלקסיקלי אינו יורד בהכרח, אך במקרה הזה מדובר בנושא אחד ומכאן הנטייה לחזור על מגוון מצומצם של לקסמות.

במספר רב של מחקרים שנעשו על מדד הגיוון הלקסיקלי נבדקו טקסטים שכללו 100-200 תמניות כי קורפוס קצר או ארוך מזה נתפס כלא מייצג למדדי העושר הלקסיקלי בצורה הטובה ביותר. בטקסט קצר ממאה מילים אין מספיק מילים שיעידו על העושר הלקסיקלי ואילו טקסט ארוך יותר ממאתיים מילים מושפע מגורמים כמו אורך שעלולים להשפיע באופנים שונים על התוצאות. לפיכך גם במחקר זה הועדפה בדיקת קורפוס בטווח ארוך זה, שנחשב אופטימלי, ובהתחשב ביכולות הכתיבה בתיכון ב-L2 הוחלט לספור 100 מילים. אך כפועל יוצא של היכולות הלשוניות של תלמידי כיתה ט' היו לי חיבורים שלא כללו שבעים תמניות. לא הייתה כוונה לקצר את הקורפוס האחד שייבדק לחמישים ושלוש תמניות, שזה מספר התמניות בחיבור הקצר ביותר בקורפוס, כי זה עלול לפגוע במידת המהימנות של מדדי העושר הלקסיקלי, לכן נבדקו כל התמניות בטקסטים שכללו פחות ממאה תמניות, ואת 100 התמניות הראשונות בטקסטים שכללו יותר ממאה.

סביר להניח שהגיוון הלקסיקלי של תלמידי כיתה י"א גבוה מזה של תלמידי כיתה ט' אבל בגלל שתלמידי כיתה י"א כתבו חיבורים ארוכים יותר היו אצלם יותר חזרות על תבניות מסוימות מאשר בקרב תלמידי כיתה ט' וכתוצאה מכך מדד הגיוון לא עלה. מצב זה אכן תואם את המחקרים הרבים שהצביעו על אי-עלייה ולפעמים ירידה במדד הגיוון הלקסיקלי כשמספר התמניות בטקסט עולה (Hess et al., 1986; Malvern et al., 2004; McCarthy & Jarvis, 2007, 2010).

סבורני שאילו תלמידי כיתה ט' היו כותבים חיבורים באורך כמאה תמניות והייתי משווה אותם עם מאה המילים הראשונות בחיבורי כיתה י"א, קרוב לוודאי שמדד הגיוון הלקסיקלי של תלמידי כיתה ט' היה נמוך מזה של תלמידי כיתה י"א. מכיוון שתלמידי כיתה ט' לא הגיעו לאורך החיבור הנחשב אופטימלי, אני סבור שמדד זה אינו מעשי בגילאים אלה ב-L2 עברית ל-L1 ערבית, לפחות בנגב.

על פי הממצאים בסעיף זה עדיף לא להסתמך על בדיקת הגיוון הלקסיקלי כמדד להתפתחות העברית כ-L2 לערבית בכיתה ט', ודאי לא כמדד יחיד וודאי לא לפני כן, כי רוב התלמידים בגילאים אלו לא מצליחים לכתוב חיבורים ארוכים מספיק על אותו נושא. מומלץ לערוך אותה בכיתות מבוגרות יותר: בתיכון או במוסדות להשכלה גבוהה מכיוון שבגילאים גבוהים יותר יכולים המשתתפים לכתוב יותר ממאה תבניות על אותו נושא.

הסתייגות נוספת מהקריטריון כולו: גיוון לקסיקלי יכול לפעמים להיות מדומה ולהסוות חסכים לשוניים: כתיבה יכולה לשלב תבניות רבות, בעושר לקסיקלי לכאורה, אך במבנים שגויים ומובנים שגויים. מבט על חיבוריהם של תלמידי כיתה ט' מראה ליקויים לקסיקליים גם כשהלקסיקון לכאורה מגוון.

דחיסות לקסיקלית

לפי אור (Ure, 1971) הדחיסות הלקסיקלית של רוב הטקסטים הדבורים ב-L1 נמוכה מ-40% ואילו הדחיסות הלקסיקלית בטקסטים כתובים גבוהה מ-40%. הסתייגות שהושמעה נגד קביעה זו מדגישה שיש להתחשב בשפה הנדונה (Johansson, 2008, pp.65-66): בשפות מורכבות מורפולוגית (כמו עברית וערבית) הדחיסות גבוהה יחסית בשל היותן של חלק ממילות התפקיד דבוקות למילת התוכן, וכך הן לא נספרות בספירה סיסטמטית של תוכנות ניתוח בלשני. עם זאת נמצאת הערכה זו תקפה לגבי עברית כ-L2 לערבית, אמנם לא בתחילת הרכישה אלא ברמות הגבוהות יחסית: הדחיסות בחיבורים הכתובים של כיתה ט' הייתה 41.72% ושל י"א 39.6% (ראו. טבלה 4).

לפני ביצוע המחקר היה מצופה שהדחיסות הלקסיקלית תעלה עם הזמן. מעיון בטבלה 4 ניתן לראות שממוצע הדחיסות הלקסיקלית אצל תלמידי כיתה ט' הוא (41.72%) ואצל תלמידי כיתה י"א הוא (39.6%), כלומר יש ירידה קטנה ולא משמעותית בין כיתה ט' לכיתה י"א. לפי בדיקת t-test פער זה אינו מובהק סטטיסטית (0.05 > p; 0.1520). ממצא זה אינו עולה בקנה אחד עם המחקרים הרבים שצוטטו קודם לכן (Daller et al., 2003; Malvern et al., 2004). אזי עולה השאלה איך ניתן להסביר את אי-העלייה של הדחיסות הלקסיקלית בין כיתות ט' ל-י"א. התשובה המוצעת כאן מסתמכת על תוכנית הלימודים של העברית כשפה שנייה לדוברי ערבית כשפת האם ובחינות הבגרות בשפה העברית כשפה שנייה.

שיקולי קישוריות ולכידות בהקשר של דחיסות לקסיקלית

תוכנית הלימודים של משרד החינוך (2011) מדגישה את חשיבות הבנתם של טקסטים ובשל כך מדגישה חשיבות יצירת טקסטים מקושרים ומלוכדים שקל יחסית לקרוא אותם ולהבנים. זאת בין היתר באמצעות שימוש בקשרים לוגיים – מילות קישור – ומאזכרים, שהם מילים שמאזכרות לנו מילים שהופיעו קודם לכן בטקסט (anaphora). הבנת הקשר הלוגי בין משפטים או בין פסקות באותו טקסט קשורה מאוד לשימוש הולם במילות הקישור ובהבנתן. יתרה מזו, המאזכרים מסייעים ביצירת הקישוריות והלכידות בתוך הטקסט. ככל שהטקסטים משלבים יותר קשרים לוגיים עולה רמת הקריאות שלהם בגלל שהם מפשטים את תהליך ההסקה של הקורא. לפי שלזינגר (תש"ס, עמ' 136) רמת הקריאות של טקסט, כלומר עד כמה קל לקרוא אותו ולהבינו, תלויה בין השאר בדחיסות: טקסט קריא יותר הוא טקסט דחוס לקסיקלית פחות, כלומר, יש בו יותר מילות תפקיד ופחות מילות תוכן. קשרים אלה הופכים את הקורא ליעיל יותר, דהיינו, לשותף פעיל בתוך הטקסט בגלל שהם מאפשרים תהליך חיזוי תקין במהלך הקריאה. הקשרים הלוגיים מציינים את הקשר בין המשפטים ולכן מקטינים את הצורך של הקורא בביצוע ניתוח קוגניטיבי מורכב של הטקסט. לכן אם הם חסרים הקורא נדרש לבצע ניתוח קוגניטיבי מורכב להבנת הקשרים בין המשפטים והפסקות. המודעות שהם מייצרים מנחה את הקורא לנוע קדימה ואחורה בתוך הטקסט בצורה נכונה (Olshtain & Cohen, 2005, p.317).

ככל שמתקרבים לבגרות בעברית (הבנה, הבעה וידע לשון) התלמידים נחשפים יותר לטקסטים שמשמשים בקשרים הלוגיים – מילות קישור – ובמאזכרים השונים וגם מתבקשים לכתוב חיבורים מקושרים ומלוכדים שיהיו ברורים והבנתם תהיה קלה. הם גם מקבלים הנחיות מטה-לשוניות מפורשות לגבי שימוש באמצעים אלה. כתוצאה מכך, ככל שמתקרבים לבגרות התלמידים מתחילים לכתוב חיבורים לפי אותם עקרונות של הטקסטים שהם קוראים כהכנה לבגרות. ניתן לצפות שתהיה עלייה בשימוש תקין ביסודות אלה ככל שמתקרבים יותר לבחינת המגן והבגרות, מה שגורר ירידה בדחיסות הלקסיקלית לצד עלייה

התפתחות לקסיקלית בעברית של דוברי ערבית

בקריאות של הטקסט כי מילות הקישור והמאזכרים הם חלק ממילות התפקיד שמסייעות ביצירת טקסטים קריאים.

כדי לבסס את ההסבר המוצע ייבחרו חיבורים משתי קבוצות התלמידים באופן אקראי ויושווה ביניהם השימוש במילות תפקיד שמטרתן ליצור קישוריות ולכידות בטקסט. נשווה בין החיבור הראשון והאחרון של כל קבוצת תלמידים.

טבלה 5. מילות קישור ומאזכרים בחיבור הראשון של כל קבוצה¹

כיתה	ט' 1	י"א 1
קשרים	ו' 4*, או 4*, אס, בסוף.	ו' 9*, או 2*, גם 2*, כי, אבל, בשביל.
מאזכרים	כ"ג 5*3, זה, ל 5*, אותנו.	ל, ב* 6*, כ"ג 5*3, של 7*.
סה"כ	22	35

התלמיד הראשון מכיתה ט' השתמש בעשרים ושתיים תמניות תפקיד ששייכות לתחום הקישוריות והלכידות ואליו התלמיד הראשון מכיתה י"א השתמש בשלושים וחמש תמניות תפקיד. יחסית לגודל הטקסט (90 תמנית). זהו הבדל ניכר.

התלמיד האחרון מכיתה ט' השתמש בעשרים ושלוש תמניות תפקיד ששייכות לתחום הקישוריות והלכידות ואליו התלמיד האחרון מכיתה י"א השתמש בעשרים ושבע תמניות תפקיד:

טבלה 6. מילות קישור ומאזכרים בחיבור האחרון של כל קבוצה

כיתה	ט' 20	י"א 20
קשרים	ו' 3*, אס, כי.	ו' 8*, גם.
מאזכרים	כ"ג 4*3, זה, ל 5*, ב* 3*, של 3* כ"ג 2*2.	ל 2*, ב* 5*, כ"ג 4*3, של 5*, את, זה.
סה"כ	23	27

בעת חיפוש מקבילות לתופעה זו במחקרים בין-לשוניים נחשפנו לממצא של בדיקות דחיסות לקסיקלית שערכה ג'והנסון לילדים בני 13 דוברי שוודית. בהשוואת טקסטים סיפוריים מול עיוניים בגיל זה נמצאה ירידה בדחיסות בסוגה העיונית, וזאת בניגוד לצפוי, שהרי הסוגה העיונית ידועה במחקר בלשני כדחוסה יותר מהסוגה הסיפורית. הסיבה שמציעה החוקרת ג'והנסון (Johansson, 2008, p.74) היא שבשלב זה של התפתחות האוריינות בשפת האם השוודית, הילדים גילו רגישות לדרישות הקישוריות והלכידות של הסוגה העיונית והרבו בקשרים לוגים וכינויים, כלומר מילות תפקיד, מה שהוריד את הדחיסות יחסית לסוגה הסיפורית. ביישום ההסבר על האוכלוסייה שלנו ניתן לצפות לתופעה זו בגיל יותר מאוחר שכן מדובר ברכישת L2. זה תואם את הממצאים בכיתה י"א, שם מופנית תשומת לב רבה ללכידות.

¹ * מסמנת מספר התמניות (מספר החזרות של אותה תבנית).

סביר להניח, עם זאת, שלא כל מילות התפקיד מעלות את לכידות הטקסט, יש כאלה שיוצרות מצב הפוך אם השימוש בהן לקוי. מילות תפקיד שגויות שמורידות את לכידות הטקסט מופיעות אצל תלמידי כיתה ט' פעמים רבות. למשל בחיבור הראשון בקבוצה זו מופיע המקטע: "לכל בעל יש לו...". פה יש שימוש יתר במילית הקניין, מה שדווקא מוריד מרמת הקישוריות של הטקסט לפי כללי העברית. מבחינת העברית יש פה כפילות כתוצאה מהתערבות L1 (ערבית) שבה מבנה כזה הוא בלתי מסומן ושגרת, ואילו בעברית הוא מסומן כמשפט ייחוד, לצורכי ניגוד למשל, כאשר פה אין צורך פרגמטי. אצל תלמידי כיתה י"א השימוש הלקוי הזה פוחת לפי התרשמותנו.

יתר על כן יש להבחין בין קשרים פרימיטיביים או ראשוניים לבין קשרים מתקדמים. הקשרים הראשוניים נרכשים בשלב מוקדם מאוד: ו, גם, אבל ועוד, ואילו הקשרים המתקדמים נרכשים בשלבים מתקדמים יותר של האוריינות: על כן, אילולא ועוד. כך שבדיקה כמותית של קשרים, ללא הבחנה לקסיקלית, לא בהכרח תעיד על התקדמות לקסיקלית. תלמידי כיתה ט' באופן כללי הרבו להשתמש בקשרים ראשוניים כגון ו-החיבור ואילו תלמידי כיתה י"א השתמשו רבות בקשרים מתקדמים כגון לפיכך, עם זאת, למרות זאת וכו'.

בחיבורי תלמידי כיתה ט' מופיעים בינומים (צמידים) של מילות תוכן כמו "טענה ודעה שלי", "לו דעה או טענה" ו"יש הרבה דעות וטענות" (חיבור ראשון כיתה ט'). אלה מעלים את הדחיסות לכאורה (ואת הגיוון הלקסיקלי בו בזמן) אך הם אינם מגבירים את הלכידות לפי נורמות של העברית אלא יוצרים טקסט מסורבל. סגנון בינומי זה מקובל בערבית, אך בעברית הוא חריג. הוא יוצר ציפייה למידע נוסף במילה השנייה, שאינו קיים למעשה, כך שהקורא עשוי להתבלבל.

שתי הטבלאות האחרונות (טבלאות 5 ו-6) מראות שתלמידי כיתה י"א משתמשים במילות קישור ומאזכרים (35 ו-27 תמניות בהתאמה), בעיקר שרמתם גבוהה, יותר מתלמידי כיתה ט' (22 ו-23 תמניות בהתאמה). למשל, באחד החיבורים של תלמידי כיתה י"א מופיע המשפט: "מצד אחד היא כמעט חסרת תועלת לחברה שלה, מצד שני היא מנקה את הבית עושה אוכל או עוזרת לילדיה". כאן ניכר שימוש במילות קישור גבוהות המציינות קשר לוגי של ניגוד. תלמידי כיתה ט', גם אם הם משלבים קשרים לוגים גבוהים, הם אינם בהכרח משתמשים בהם בצורה תקנית, על כן נוצרים אצלם הרבה משפטים לא ברורים. למשל תלמיד כיתה ט' כתב את המשפט: "היא {האישה} רוצה להיות אם מורה אם רופאה או עקרת בית" אמנם הוא השתמש במילת קישור של תנאי אם אך אין פה תנאי (אלא ברירה), כך שהשימוש במילת הקישור המתקדמת לקוי. ידוע שהשליטה בקשרים הלוגים בעברית היא אחת המשימות הקשות בהתפתחות האוריינות גם לדוברים ילידיים (Olshtain & Cohen, 2005, p.317).

כאמור לעיל, הסבר אפשרי לאי-העלייה בדחיסות הלקסיקלית אצל תלמידי י"א הוא התקרבותם לבחינת הבגרות בעברית והדגש המושם בתוכנית הלימודים שלהם על קישוריות ולכידות באמצעות שימוש במילות תפקיד נכונות.

אנו מניחים לפיכך כי אי-העלייה בדחיסות הלקסיקלית אינה מעידה על ירידה בכשירות הלקסיקלית אלא להיפך, היא מעוגנת בהקשר הלימודי, בהתערבות מטה-לשונית יזומה שמנוסחת בתוכנית הלימודים. כדי לתמוך בהסבר זה ניתן לראות שבשנים האחרונות בכל בחינות הבגרות בעברית לדוברי ערבית היו שאלות על קשרים לוגים ומאזכרים. בשאלון 014381 קיץ 2016 (שלוש היחידות הראשונות בבגרות "עברית לדוברי

התפתחות לקסיקלית בעברית של דוברי ערבית

ערבית") היו שש נקודות, מסך חמישים נקודות בשאלות על הטקסט, הקשורות בשאלות על מילות הקישור והמאזכרים. הנה שאלות מאותו שאלון:

}

1- לפניך שני משפטים.

-נהגים צריכים לבחור בסוג מוזיקה שלא תעורר בהם מחשבות, זיכרונות ורגשות.

-הם מסיחים את דעת הנהגים מן הכביש.

כתוב את שני המשפטים במשפט אחד. חבר את שני המשפטים במילת קישור בקשר לוגי של סיבתיות. (3 נקודות)

2- בשורות 19-20 כתוב: " {... } משום שהם עלולים להסיח את דעת הנהג מן הכביש".

למה מכוונת המילה "הם" במשפט זה? (3 נקודות)

{ זכות היוצרים שמורה למדינת ישראל, משרד חינוך

בשאלון 014381 חורף 2017 הוקדשו שבע נקודות לאותו עניין. זהו ניקוד גבוה שמדגיש את חשיבותם של נושאים אלה. הנה נוסח השאלות משאלון זה:

}

1- לפניך שני משפטים. קרא אותם, וענה על השאלות שאחריהם.

(1) כבר כיום ישראלים רבים רוכשים מוצרים באמצעות האינטרנט.

(2) יש תכנית לבנות עוד 123 קניונים בשנים הקרובות.

א. על פי הקטע { הקטע שמופיע לפני השאלות ובו דנים ביחס בין בניית קניונים לבין רכישה מוצרים באמצעות האינטרנט }, מהו הקשר הלוגי בין משפט (1) ובין משפט (2)?

הקף את התשובה הנכונה. (2 נקודות)

סיבה והתוצאה ויתור והסתייגות הכללה ופירוט בעיה ופתרון

ב. חבר את שני המשפטים (1) ו- (2) למשפט אחד. הוסף מילת קישור המתאימה לקשר הלוגי שציינת. (3 נקודות)

2- "בשנים שלאחר מכן הוקמו בארץ עוד עשרות קניונים כאלה". (שורה 17)

למה מכוונת המילה "כאלה" במשפט זה? (2 נקודות)

{ זכות היוצרים שמורה למדינת ישראל, משרד חינוך

יש הסתייגות נוספת לקישור בין דחיסות לקסיקלית גבוהה לכשירות לקסיקלית בקורלציה ישירה חד-משמעית עם הגיל: כאמור, ככל שמתקדמים בציר הרכישה משתמשים יותר במילות תוכן יחסית לכלל התמניות בטקסט הכתוב בנסיבות אקדמיות. אך ברור שהמצב הזה נכון עד לנקודה מסוימת ואופטימלית בציר הרכישה, ושאינן מצב שבו יהיה השיעור של מילות התוכן מכלל תמניות הטקסט גבוה מנקודה אופטימלית כי

מילות התפקיד משחקות תפקיד חשוב מאוד בהבנת הטקסט. בלעדיהן הטקסט לא יהיה כלל טבעי או קוהרנטי. מילות תפקיד הן חלק אינהרנטי בשפה טבעית. הוצאתן יוצרת טקסט ספציפי מאוד, למשל סגנון מברק או כותרות עיתון: "יום ב' 4:00 מגיע באר שבע" שמשמעותו ככל הנראה "ביום שני, בשעה ארבע אני מגיע לבאר-שבע". כמעט בלתי אפשרי לקיים סגנון זה ממעלה לשורה או שתיים. מכאן נגזר המיקוד בקישוריות בכיתה י"א. התלמידים הנבחרים בשפה אקדמית מתבקשים לכתוב טקסט שהוא אמנם דחוס לקסיקלית, שיש בו שימוש רב ונכון במילות תוכן להבעת הרעיונות, אך עם זאת קריא, מקושר ולכיד, שיש בו שימוש רב ונכון של מילות תפקיד. לצורך זה, חשוב האיזון הנכון בין מילות התוכן למילות התפקיד ולא עלייה במילות תוכן בלבד.

לסיכום, ישנו הסבר אחד משולב לממצאים שהוצעו בסעיף זה: ראשית, בכיתה ט' בדיקת דחיסות לקסיקלית כנראה אינה יעילה עדיין כמדד לכשירות לקסיקלית. אצל תלמידי כיתה ט' יש שימוש מוטעה בקשרים, לרבות השמטת מילות תוכן חיוניות, שיוצרת רושם של דחיסות כסימן לאוריינות. נוצרת דחיסות גבוהה אך לא תקנית. נוכל להסיק שבשלב זה של כשירות לשונית, מדד הדחיסות הלקסיקלית בצורתו הנוכחית אינו ישים כמדד להתפתחות לקסיקלית.

שנית, בכיתה י"א אחרי שהתלמידים עברו הכוונה מפורשת בתחום הקישוריות, מדד הדחיסות הלקסיקלית הוא נאמן יותר כסמן של כשירות לקסיקלית ולשונית-אקדמית. אך הכשרה מטה לשונית זו גורמת לירידה לכאורה בדחיסות הלקסיקלית בהשוואה לכיתה ט', במקום העלייה הצפויה.

נראה כי ישנו הסבר אפשרי נוסף להיעדר הבדל משמעותי בדחיסות הלקסיקלית אצל שתי הקבוצות והוא שפער הגילאים קטן יחסית ויש צורך בפער גדול יותר כדי לבדוק דחיסות לקסיקלית ב-L2. זאת ראינו במחקרה של ג'והנסון, שבין גילאי 10 ו 13 לא מצאה הבדלים בגיוון לקסיקלי ואף לא בדחיסות, ואף לא בין גילאי 13 ו 17. רק בבחירת טווחים גדולים נמצאו ההבדלים הצפויים (Johansson, 2008, p.77). אני מניח שגם במקרה הנידון העלייה בדחיסות הלקסיקלית, וייתכן גם בגיוון הלקסיקלי, תהיה מובהקת יותר בתנאים של הבדל גיל גבוה משנתיים בין הקבוצות הנבדקות.

המלצתנו אפוא היא לערוך בדיקה זו בכיתות מבוגרות יותר: בתיכון או במוסדות ללימודים גבוהים, אחרי שחלה העלייה הצפויה בשל החשיפה הנוספת. למסקנה דומה אבל בתחום המורפולוגיה מגיעים סאיג'י-חדאד וג'יוסי כשהם מציינים שחלק מהבדיקות הלשוניות (במקרה שלהם בדיקות מורפולוגיות) מצריכות רמת כשירות לשונית מסוימת על מנת לבצען ולקבל תוצאות מייצגות ומהימנות (Saiegh-Haddad & Jayusy, 2016, p.432).

סיכום

במחקר זה נבדקה התקדמות בכשירות לקסיקלית בעברית כ-L2 לערבית בנגב באמצעות שלושה מדדים: רמת מופשטות, גיוון לקסיקלי ודחיסות לקסיקלית. כאשר לוקחים בחשבון את תוכניות הלימודים בתחום עברית כ-L2 לערבית של מערכת החינוך המקומית, לא כל המדדים הלקסיקליים התגלו כיעילים לבדיקת התקדמות ברכישת אוצר המילים של L2 בכל הגילאים במידה שהיה ניתן לצפות מספרות המחקר: ראשית, צופתה עלייה בממוצע בשלוש הבדיקות הלקסיקליות עם העלייה בגיל. בנוגע לרמת מופשטות הטקסטים, שמשקפת התקדמות ברכישת L2, ציפיתנו התממשה: העלייה בשימוש בשמות עצם מופשטים יחסית לכלל שמות העצם

התפתחות לקסיקלית בעברית של דוברי ערבית

בחיבורים עלתה בקו ישר עם הגיל והחשיפה לעברית. למרות שהיה הבדל ברור באורך החיבורים של שתי קבוצות התלמידים, היה הבדל מובהק סטטיסטית ברמת המופשטות ביניהן. בעקבות ממצא זה הסקנו שבדיקה זו אינה מושפעת מאורך הטקסט, לפחות לא בטקסטים קצרים עד כמאה תמניות ושהיא תקפה כבר בכיתה ט'.

שנית, באשר לגיוון לקסיקלי, לא חלה עלייה מובהקת עם הגיל בעיקר בגלל שההשוואה הייתה בין טקסטים שוני אורך, שכן רוב תלמידי כיתה ט' לא כתבו טקסטים שכללו 100 מילים. על כן, אנו סבורים שאין לצפות לתוצאות מהימנות עד אחרי כיתה ט', כאשר התלמידים מגיעים לרמת כשירות לשונית אשר מאפשרת להם לחבר טקסטים מובנים וקריאים שכוללים 100 תמניות.

שלישית, בדומה לגיוון הלקסיקלי לא התגלתה עלייה מובהקת בדחיסות הלקסיקלית עם הגיל. ההסברים שהצענו לאי-העלייה כנגד המצופה הם: הדגשים בתוכנית הלימודים של כיתה י"א על כתיבת טקסטים מקושרים, מלוכדים וקוהרנטיים באמצעות שימוש רב יחסית במילות קישור ובמאזכרים, שהם מילות תפקיד, דבר שמוריד את הדחיסות הלקסיקלית. נוסף לכך, השימוש במילות התפקיד אצל תלמידי כיתה ט' עדיין אינו מגובש דיו (לפי נורמות של עברית) כדי לשמש מדד לכשירות לשונית: יש השמטת קישוריות מצד אחד ועודף קישוריות מצד שני, בין השאר בהשפעת הערבית. מסתבר שאין בהכרח קשר רציף וחד-ערכי בין דחיסות לקסיקלית לכשירות לקסיקלית, באוכלוסייה הנבדקת, ואולי באופן כללי יותר בבתי ספר ערבים בישראל. אי-עלייה בדחיסות מכיתה ט' לכיתה י"א עשויה להעיד דווקא על שיפור בקישוריות. אצל תלמידי כיתה י"א כבר יש שימוש נכון ותקני יותר בקשרים ובמאזכרים בעקבות ההתערבות המטה-לשונית שמתקיימת באמצעות תוכנית הלימודים. משלב זה ואילך ניתן לדעתנו לראות במדד הדחיסות הלקסיקלית סמן להתקדמות בכשירות לשונית-אקדמית ב-L2.

בעקבות הממצאים במחקר זה ניתן לחלק את המדדים לשתי קבוצות: *מדדים לקסיקליים התחלתיים* ו*מדדים לקסיקליים מתקדמים*. הקבוצה הראשונה כוללת רמת מופשטות ואילו השנייה כוללת גיוון לקסיקלי ודחיסות לקסיקלית. ההבדל בין שתי הקבוצות הוא שאת הראשונה אפשר לבצע בגילאים צעירים: כיתה ט' ומטה. כלומר, בשלבים לא מתקדמים של רכישת L2. לעומתה, את מבדקי הקבוצה השנייה כדאי לבצע בשלבים מתקדמים של רכישת L2 - כיתה י' ומעלה.

מחקר זה הוא מחקר חלוץ שבוחן שלושה מדדים להתקדמות לקסיקלית בעברית כשפה שנייה לדוברי ערבית. ממצאי מאמר זה מראים שיש עוד מקום רב למחקר נוסף בתחום אוצר המילים של העברית כ-L2 לערבית. ההמלצה היא לבחון את המדדים הלקסיקליים האלה בקורפוס גדול יותר שיכלול גילאים שונים (תלמידי חטיבת ביניים לעומת תלמידי תיכון; תלמידי תיכון לעומת סטודנטים) בקהילות שונות (הקהילה הבדואית בדרום לעומת קהילות דוברות ערבית אחרות) בטווחי זמן גדולים יותר (שנה וחצי, שנתיים ועוד). ישנו מקום גם לבדוק אם יש הבדלים בין שני המגדרים בהתפתחות הלקסיקון של העברית כשפה שנייה לערבית בעזרת מדדים לקסיקליים אלה. יש להתחשב בממצאי מחקר זה שחלק מהמדדים יעילים רק החל ברמת כשירות מסוימת. לגילאים נמוכים יותר יש צורך למצוא או לפתח מדדים אחרים ולבדוק את יעילותם בצורה מבוקרת.

הערת המחבר

תודתי הרבה נתונה למנחיי בעבודת הדוקטורט: פרופ' רוני הנקין-רויטפרב וד"ר רועי גפטר על הערותיהם החשובות במהלך שלבי הכתיבה השונים. המאמר נכתב בתמיכת מלגת הוועדה לתכנון ולתקצוב (ות"ת) במועצה להשכלה גבוהה (מל"ג) ומלגת נגב – רוח דרומית.

מקורות

משרד החינוך, (2011). עברית כשפה שנייה בבתי הספר הערביים לשלוש חטיבות הגיל: ג-י"ב. נדלה מתוך:

<http://cms.education.gov.il/NR/rdonlyres/202A56F5-EC9E-43CD-9592-0EAE76451DF2/212803/29916.pdf>

שלזינגר, יי (תש"ס). לשונות העיתון: מאפיינים סגנוניים בלשון העתונות היומית הכתובה על מדוריה השונים.

באר-שבע: הוצאת ספרים של אוניברסיטת בן-גוריון בנגב.

שאלון 014381, (קיץ 2016). נדלה מתוך:

<https://sites.google.com/a/etz.tzafonet.org.il/phenotpagrot/home/bhynwt-bgrwt-2016>

שאלון 014381, (חורף 2017). נדלה מתוך:

<https://sites.google.com/a/etz.tzafonet.org.il/phenotpagrot/home/bhynwt-bgrwt-2017>

שראל, צ' (תשנ"ט). הדחיסות המילונית בשיח הדבור ובשיח הכתוב (על מחקר חלוץ בתכונות מבחינה אחת).

בתוך ש' שרביט (עורך), מחקרים בלשון העברית העתיקה והחדשה (ספר היובל לפרופ' מ"צ קדרי) (עמ' 401–411).

רמת-גן: אוניברסיטת בר-אילן.

Abdalla, F., Robb, M. P., & Al-Shatti, T. (2010). Stuttering and lexical category in adult Arabic speakers. *Clinical Linguistics & Phonetics*, 24(1), 70–81.

Bates, E., Bretherton, I., & Snyder, L. S. (1991). *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge University Press.

Bell, H. (2009). The messy little details: a longitudinal case study of the emerging lexicon. *Lexical processing in second language learners*, 111–127.

Blum-Kulka, S., & Levenston, E.A. (1983). Universals of Lexical Simplification. In C. Faerch, & G. Kasper (Eds.), *Strategies in Interlanguage Communication* (pp.119–139). Longman, New York.

Bogaards, P. (2001). Lexical Units and the Learning of Foreign Language Vocabulary. *Studies in Second Language Acquisition*, 23, 321–343.

Brandes, G., & Ravid, D. (2017). A developmental study of prepositional phrases in Hebrew written text construction. *First Language*, 37(1), 83–105.

- Choi, W., & Jeong, H. (2016). Finding an Appropriate Lexical Diversity Measurement for a Small-Sized Corpus and its Application to a Comparative Study of L2 Learners' Writings. *Multimedia Tools and Applications*, 75(21), 13015–13022.
- Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied linguistics*, 24(2), 197–222.
- David, A. (2008). Vocabulary Breadth in French L2 Learners. *Language Learning Journal*, 36(2), 167–180.
- Degani, T., Prior, A., & Tokowicz, N. (2011). Bidirectional transfer: The effect of sharing a translation. *Journal of Cognitive Psychology*, 23(1), 18–28.
- 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.
- Dillard, J. P., & Pfau, M. (2002). *The persuasion handbook: Developments in theory and practice*. Sage Publications.
- Filipović, L., & Hawkins, J. A. (2013). Multiple factors in second language acquisition: The CASP model. *Linguistics*, 51(1), 145–176.
- Fitzpatrick, T., Al-Qarni, I., & Meara, P. (2008). Intensive Vocabulary Learning: A Case Study. *Language Learning Journal*, 36(2), 239–248.
- Fries, C. C. (1952). *The Structure of English*. Harcourt Brace, New York.
- Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and Motivation in Second Language Learning*. Rowley, MA: Newbury House Publishers.
- Gordon, B., & Caramazza, A. (1982). Lexical decision for open-and closed-class words: Failure to replicate differential frequency sensitivity. *Brain and Language*, 15(1), 143–160.
- Haastrup, K., & Henriksen, B. (2000). Vocabulary acquisition: Acquiring depth of knowledge through network building. *International Journal of Applied Linguistics*, 10(2), 221–240.
- Halliday, M.A.K. (1985). *Spoken and Written Language*. Victoria: Deakin University Press.
- Halliday, M.A.K. (1989). *Spoken and Written Language*. (2nd Edition). Oxford University Press, Oxford.
- Hess, C. W., Sefton, K. M., & Landry, R. G. (1986). Sample size and type-token ratios for oral language of preschool children. *Journal of Speech, Language, and Hearing Research*, 29(1), 129–134.

- Hilton, H. (2008). The Link Between Vocabulary Knowledge and Spoken L2 Proficiency. *Language Learning Journal*, 36(2), 153–166.
- Jarvis, S. (2002). Short Texts, Best-Fitting Curves and New Measures of Lexical Diversity. *Language Testing*, 19(1), 57–84.
- Jarvis, S. (2012). Lexical Challenges in the Intersection of Applied Linguistics and ANLP. In C. Boonthum-Denecke, P.M. McCarthy, & T.A. Lamkin (Eds.), *Cross-Disciplinary Advances in Applied Natural Language Processing: Issues and Approaches* (pp.50–72). IGI Publishing Hershey, PA.
- Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning*, 63, 87–106.
- Jiang, N. (2007). Semantic Transfer and Development in Adult L2 Vocabulary Acquisition. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge* (pp.101–126). Cambridge University Press, Cambridge.
- Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Lund Working Papers in Linguistics*, 53, 61–79.
- Johnson, W. (1939). *Language and Speech Hygiene: An Application of General Semantics*. Institute of General Semantics, Chicago.
- Johnson, W. (1944). Studies in language behavior: A program of research. *Psychological Monographs*, 56(2), 1–15.
- Juste, F. S., Sassi, F. C., & de Andrade, C. R. F. (2012). Exchange of disfluency with age from function to content words in Brazilian Portuguese speakers who do and do not stutter. *Clinical linguistics & phonetics*, 26(11-12), 946–961.
- Koizumi, R. (2012). Relationships between text length and lexical diversity measures: Can we use short texts of less than 100 tokens. *Vocabulary Learning and Instruction*, 1(1), 60–69.
- Koizumi, R., & In'nami, Y. (2012). WITHDRAWN: Effects of text length on lexical diversity measures: Using short texts with less than 200 tokens. *System*, 40(4), 522–532.
- Kroll, J., & Tokowicz, N. (2001). The Development of Conceptual Representation for Words in a Second Language. In J. Nicol (Ed.), *One Mind, Two Languages: Bilingual Language Processing* (pp.49–71). Blackwell Publishers, Malden, MA.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied linguistics*, 22(1), 1–26.

- Lieven, E.V.M. (1978). Conversations between Mothers and Young Children: Individual Differences and their Possible Implication for the Study of Child Language Learning. In N. Waterson, & C.E. Snow (Eds.), *The Development of Communication*. (pp.173–187). Wiley, Chichester.
- Lyons, J. (1981). *Language and linguistics*. Cambridge University Press.
- Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language testing*, 19(1), 85–104.
- Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development*. New York: Palgrave Macmillan.
- McCarthy, P. M., & Jarvis, S. (2007). vocd: A theoretical and empirical evaluation. *Language Testing*, 24(4), 459–488.
- McCarthy, P. M., & Jarvis, S. (2010). MTLT, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior research methods*, 42(2), 381–392.
- Meara, P., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System*, 28(1), 19–30.
- Mondria, J., & Wiersma, B. (2007). Receptive, Productive, and Receptive + Productive L2 Vocabulary Learning: What Difference Does It Make? In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge* (pp.79–100). Cambridge University Press, Cambridge.
- Nation, I. S. P. (1993). Vocabulary size, growth, and use. *The bilingual lexicon*, 6.
- Nation, I.S.P. (2005). Teaching and Learning Vocabulary. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp.581–595). Lawrence Erlbaum, Mahwah, NJ.
- Nation, I.S.P. (2007). Fundamental Issues in Modelling and Assessing Vocabulary Knowledge. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge* (pp.33–43). Cambridge University Press, Cambridge.
- Olshtain, E., & Cohen, E. (2005). Logical Connectors in Hebrew: How Well Do Eighth-Graders Master Them?. *Perspectives on Language and Language Development*. Springer, Boston, MA.

- Ordóñez, C. L., Carlo, M. S., Snow, C. E., & McLaughlin, B. (2002). Depth and breadth of vocabulary in two languages: Which vocabulary skills transfer?. *Journal of Educational Psychology, 94*(4), 719.
- Piaget, J. (1959). *Language and thought of the child*. Humanities Press, New York.
- Polio, C. (2001). Research Methodology in Second Language Writing Research: The Case of Text-Based Studies. In T. Silva, & P.K. Matsuda (Eds.), *On Second Language Writing* (pp.91–115). Lawrence Erlbaum, Mahwah, NJ.
- Qian, D. (2004). Second Language Lexical Inferencing: Preferences, Perceptions, and Practices. In P. Bogaards, & B. Laufer (Eds.), *Vocabulary in a second language: selection, acquisition, and testing* (pp.155–169). Amsterdam; Philadelphia: John Benjamins Pub.
- Ravid, D. (2005). Emergence of Linguistic Complexity in Later Language Development: Evidence from Expository Text Construction. In D. Ravid, & H. Bat-Zeev Shyldkrot (Eds.), *Perspectives on Language and Language Development* (pp.337–355). Kluwer, Dordrecht.
- Ravid, D. (2006). Semantic development in textual contexts during the school years: Noun scale analyses. *Journal of Child Language, 33*(4), 791–821.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Saiegh-Haddad, E., & Jayusy, A. (2016). Metalinguistic awareness in reading Hebrew L2. *Acquisition and Development of Hebrew: From Infancy to Adolescence, 19*, 353–386.
- Saranraj, L., & Shahila, Z. (2016). Motivation in Second Language Learning—A Retrospect. *International Interdisciplinary Research Journal, 4*(1), 7–13.
- Shi, R., & Werker, J. F. (2003). The basis of preference for lexical words in 6-month-old infants. *Developmental Science, 6*(5), 484–488.
- Spolsky, B. (1988). Bridging the gap: A general theory of second language learning. *Tesol Quarterly, 22*(3), 377–396.
- Spolsky, B. (1989). *Conditions for second language learning*. Oxford: Oxford University Press.
- Tanaka, S. (2012). New directions in L2 lexical development. *Vocabulary Learning and Instruction, 1*(1), 1–9.
- Troike, M. S. (2006). *Introducing second language acquisition*. Cambridge University Press. 2 edition.

- 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, London.
- Van Els, T., Bongaerts, T., Extra, G., Van Os, C. & Janssen-Van Dieten, A. (1986). *Applied Linguistics and the Learning and Teaching of Foreign Languages*. Edward Arnold, London.
- Verhallen, M., & Schoonen, R. (1993). Lexical knowledge of monolingual and bilingual children. *Applied linguistics*, 14(4), 344–363.
- Verhallen, M., & Schoonen, R. (1998). Lexical knowledge in L1 and L2 of third and fifth graders. *Applied Linguistics*, 19(4), 452–470.
- Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied psycholinguistics*, 22(2), 217–234.
- Wilkins, D.A. (1972) *Linguistics in Language Teaching*. London: Arnold.
- Wilkins, D.A. (1985). *Linguistics in Language Teaching*. Edward Arnold, London.
- Wode, H. (1981). *Learning a Second Language: an integrated view of language acquisition*. Tübingen: Gunter Narr Verlag.
- Yu, G. (2010). Lexical diversity in writing and speaking task performances. *Applied linguistics*, 31(2), 236–259.
- Zareva, A., Schwanenflugel, P., & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition*, 27, 567–595.