

לשוננו לעם

עורך: משה פלורנטין

מתוך

כרך נד, תשס"ג-תשס"ד



האקדמיה ללשון העברית
מוגן בזכויות יוצרים

מחקר העברית המדוברת – הצעד הראשון על רישום הדיבור לצורכי מחקר

במאמר זה אביא לפני הקורא הרהורים אחדים על דרכי רישום הדיבור לצורכי מחקר. עבודתי בתחום זה היא תוצאת הצורך לגבש כללים מנחים להעברת הטקסטים שעתידים להיכלל במאגר העברית המדוברת בישראל (מעמ"ד) אל ערוץ הכתב. מעמ"ד מתוכנן כך שיהיה בסיס למחקרים בתחומים רבים ומגוונים – לשוניים, חינוכיים, חברתיים, תרבותיים ותעשייתיים. מעמ"ד יכלול כחמישה מיליון מילים, שהוא מאגר גדול עבור לשון מדוברת. מעמ"ד יכונן כמאגר מייצג, ויכלול קטעי דיבור מאת דוברים בעלי רקע דמוגרפי שונה. הקטעים יוקלטו בנסיבות שיח מגוונות. ההקלטות תלווינה בתמליל סינכרוני. ההקלטות ותמליליהן יוצגו במקביל: התמליל יוצג בד בבד עם רצף הדיבור. קטעים נבחרים יתועתקו גם בתעתיק פונטי.¹

לשון מדוברת ולשון כתובה

בחברה אוריינית כמו החברה שלנו, שלה מסורת ארוכה של קריאה וכתובה, ההישענות המטה-לשונית על הלשון הכתובה רבה ועמוקת שורשים. בִּדְבָרְנו (או בְּכַתְבְּנו) על מרכיבים של לשון הדיבור אנו מדברים בעצם על מרכיבים של הלשון הכתובה. איננו מדברים על קולות או על הגאים כי אם על אותיות. אנו חושבים שאנו מדברים במשפטים. גם המסורת הבלשנית לוקה בטעות מושגית זו. כך נהוג בבלשנות השמית ובחקר העברית בכלל, הנשענות על מחקר דקדוקי בן אלף שנים ויותר, וכך נהוג בבלשנות הכללית. למשל, השורש, על פי מילון אבן-שושן, הוא "החלק היסודי של המילה בלא אותיות השמוש: השרש העברי הוא בִּרְבּוּ בן שלש אותיות" (ההדגשות שלי). את יחידת התחביר הבסיסי של הלשון המדוברת מכנים במונח **משפט**, אף כי המשפט אינו יחידה תקפה במבנה הלשון המדוברת, כפי שנתברר לאחרונה במחקר הכללי. בכלל, המודלים הבלשניים אשר אנו נוטים לכוון מבוססים על מסורת רבת שנים זו של בניית מודלים ללשון הכתובה. חקר הלשון המדוברת מבוסס, רובו ככולו, על תפיסות ומושגים שנקבעו ונתבססו בחקר הלשון הכתובה. הישענות זו על מסורת מחקרה של לשון הכתב נובעת,

1. על תכנון המאגר ועל תכנית העבודה ראה במאמרם של בנימין הרי, גיורא רהב וכותב דברים אלה בלשונו סד (תשס"ב), עמ' 265–287.

כנראה, מן העובדה הבסיסית, כי חקר הלשון יכול להתבצע אך ורק על בסיס נתונים כתובים. אני לא בדקתי את ההנחה הזאת, אולם חקר תולדות הבלשנות תומך בה תמיכה ממשית, שהרי את מבנה הלשון המדוברת אנו בודקים רק לאחר שיש בידינו העתקה של הלשון הזאת אל הערוץ החזותי, דהיינו אל הכתב. האזנה בלבד אינה מאפשרת התבוננות מעמיקה ברובדי הלשון השונים.

מן הראוי לזכור כי הלשון אינה עשויה מקשה אחת והיא איננה בעלת מבנה אחיד וקשור. המושג **לשון** הוא מושג־על לרצף רב־ממדי של לשונות הרבה, של דיבורם של אנשים ונשים בעלי נתונים אישיים מגוונים, המדברים וכותבים בנסיבות שונות, שעל פיהן משתנה מבנה הלשון שבפיהם. סוג הלשון הנפוץ ביותר מצוי, כמובן, בשיחת יום־יום, אשר בה נוקטים הדוברים לשון טבעית וספונטנית. הלשון הספונטנית היא זו המבוצעת בלא הכנה בתנאים רגילים של שיח, שיח שבמסגרתו אין הדובר פועל – במודע או שלא במודע – לכוון את צורת דיבורו או להתאימה לתנאים שאינו רגיל בהם (למשל בדברו אל מול מצלמות הטלוויזיה או אל מול קהל, או במעמד חשוב בעבורו כגון ראיון עבודה). חקר הלשון המדוברת, המתפתח לאחרונה במרכזי מחקר בעולם, רואה להבדיל בין הלשון הספונטנית לבין חלופות אחרות של הלשון המדוברת, שכן ללשון הספונטנית חוקיות שונה במובהק מלשונות הכתב ומן החלופות המדוברות הניזונות מלשון הכתב, עד כדי הצורך לתארה ולהסבירה בכלים שונים. לפיכך ראיתי לייחד לסוג הלשון הזה מונח שיבחינו משאר סוגי הלשון המדוברת, והצעתי לקרוא לו **לשון דבורה**. בספרות הבלשנית העברית משמשים שמות התואר **מדוברת** ו**דבורה** לחלופין בלא הבחנת משמעות ביניהם. ההבחנה נשענת על התחושה כי שם התואר **מדובר(ת)** נפוץ יותר בציבור, ומשום כך ראוי לו לשמש התואר הבלתי מסומן. שם התואר **דבור(ה)**, המשמש כמעט אך ורק בכתביהם של העוסקים בהתבוננות בלשון, אפשר לייחדו למשמעות מצומצמת. כך ייאמר, למשל, "הלשון המדוברת בישראל", אולם לא – לפחות לא בדרך טבעית וספונטנית – "הלשון הדבורה בפי העם". לפיכך שם התואר **מדוברת** בצירוף **לשון מדוברת** ישמש תואר בלתי מסומן לכל סוג של דיבור פה, אם טבעי וספונטני, ואם קרוי מן הכתב וכל מה שביניהם. במילים אחרות, הצירוף **לשון מדוברת** משמש לסימון ערוץ התקשורת שבדיבור על כל חלופותיו, והוא עומד לעומת ערוץ הלשון הכתובה על כל חלופותיו. הצירוף **לשון דבורה** ייועד, אפוא, למרחב חלופות השפה הספונטנית. הדוגמה שלהלן תמחיש את ייחודה של הלשון הדבורה:²

3 אלף שלוש מאות שקל ||

1 מ: [עומר]

4 ע: אז את

2 אתמול עשיתי טיפול לאוטו |

2. מרבית הדוגמאות מובאות מן הטקסטים שפורסמו בלשונו סד (תשס"ב), עמ' 289–314. סימונים: סוף שורה = גבול יחידת הנגנה; | = גבול יחידת הנגנה בהטעמת המשך (בדרך

5	הראשונה	24	מ: [בוא הנה {boəna}]
6	ששווה לך [לקחת אוטו]	25	אני כל יומיים --
7	מוזיקה מתחילה להתנגן >	26	כבר [המתדלק אומר
8	מ: [אלף] שלוש מאות שקל טיפול [לאוטו אתמול]	27	המתדלק
9	ס: [זה מה שאמרת] לה	28	כבר אומר לי
10	...3...	29	תגידי לי
11	ע: אין	30	מה
12	אני חושב שאני אגיע לקילומטר' הזה	31	יש לך עודף (כסף) \
13	זה הקטע	32	אומר לי
14	מ: למה \	33	את כל יומיים נמצאת פה
15	ע: ככה	34	.. [בצחקוק] מה עובר עליך
16	[מתחיל ללמוד יומיים בשבוע]	35	@@@@
17	מ: [אתה כשיהיה לך אוטו]	36	...4...
18	כשאתה לא תדפוק [חשבון לדלק]	37	מ: [בשקט, שרה עם התקליט <
19	אתה תעשה קילומטר'ים	38	(חל)פו הרבה שנים
20	אחי	39	2...3
21	חבל לך על הזמן	40	ע: אלף מאתיים שקל דלק
22	ע: בטח	41	יא ואראדי {ja wara'di}
23	ס: גם תחשוב על זה שאתה נוסע		
	לתלייביב [פעמיים בשבוע]		

רבות מחלופות הלשון המדוברת, ובמיוחד הלשון הדבורה, מתאפיינות בתכנון תוך כדי הדיבור, באי-בהירות, באי-התחייבות ובתלות חזקה בהקשרים לשוניים וחוץ-לשוניים. את הדיבור מאפיינת מהירות, ועמה הצורך בהתאמת ההאזנה והקליטה. יש בדיבור קטיעות, יחידות קצרות, חזרות מרובות, חפיפות והפסקות, הן הפסקות של ממש הן הפסקות ממולאות בהברות כגון e. אפשר שהייתה סטיות

כלל בעלייה קטנה או בירידה קטנה); || = גבול יחידת הנגנה בהטעמת סיום (בדרך כלל בירידה גדולה); \ = גבול יחידת הנגנה בהטעמת פנייה (הטעמה עולה; בדרך כלל לשאלה שאין בה מילת שאלה); -- = יחידת הנגנה קטועה; רווח = גבול בין מילים; - = מילה קטועה; .. = פסק קצר; ... = פסק בינוני; = פסק ארוך (עד שתי שניות); ...4... = פסק ארוך מאוד (בן שתי שניות ומעלה), בציון אורכו בשניות; ע: = ציון הדובר או הדוברת (=האות הראשונה של שמו או שמה; השמות בדויים כדי לשמור על האנונימיות של המוקלטים); (ע:) = אי-ביטחון בזהות הדובר; [= התחלת חפיפה בדיבור;] = סוף חפיפה בדיבור; [[]] = חפיפה בדיבור של יותר משני אנשים בו בזמן; {a}, {e}, אה {a} = היסוס בדיבור; <צחוק> = קולות שאינם קולות דיבור: צחוק, כחכות, אנחה, המהום, נשיפה, פתיחת דלת וכו'; <בלחש> = הערות המתמלל; (שלוס) = תמלול לא בטוח; @ = הברה לא מזוהה; @...@ = רצף לא מזוהה; {fte'ki} = תעתיק פונטי (IPA), הטעם מסומן במתג לפני ההברה המוטעמת; : = הארכה (בתעתיק פונטי); :: = הארכה ארוכה במיוחד (בתעתיק פונטי); $\bar{m}$ = טעם ראשי.

מרצף השיח וחזרה אליו להמשיכו. מרובים תיקוני נוסח, הופעות של צורות לא תקניות של הלשון המתוקנות רק לפעמים, תיקוני יתר, יסודות מטה-לשוניים ויחידות המשמשות בתפקוד הלשון והשיח עצמם ואינן נושאות משמעות תוכנית. בלשון הדבורה ישנם יסודות סלנג מרובים לצד יסודות לקסיקליים המיוחדים ללשון זו או מאפיינים אותה, צורות מילים ומבני תחביר שאינם מצויים – או אינם מצויים ככלל – בלשון הכתובה או בחלופות רבות אחרות של הלשון המדוברת, שימוש רב בכינויים, במבני ייחוד, במבנים פֶּרְטַקְטִיים (מבנים תחביריים נטולי מילות קישור) ועוד כהנה וכהנה. יסוד מוסד במבנה הלשון המדוברת הוא הרמה הסופר-סגמנטלית, דהיינו הפרוזודיה (טעם המילה, הגננה [=אינטונציה], הארכות והאצות, פסקים). הלשון הדבורה מתאפיינת גם בשוני בטעם המילה או במבני הגננה מחלופות אחרות של הלשון המדוברת. בהיעדר סימני פיסוק, הפרוזודיה היא המכתיבה את חלוקת השיח ליחידות ואת הקשר בין היחידות – בפי דובר יחיד או בחלופים בין הדוברים. כפי שנראה בבירור מן הקטע שלעיל, יחידות הלשון הדבורה שונות במהותן ממבני הלשון הכתובה או ממבני הלשון המדוברת בחלופותיה המושפעות מן הלשון הכתובה.

תעתיק ותמליל

כאמור, מחקר הלשון המדוברת קשה לו להיעשות במישרין בהאזנה לרצף הסימנים הקוליים. הדיבור מטבעו אקוסטי, קווי ומתמשך לאורך זמן, ומשום כך יש להעבירו אל הערוץ החזותי כדי לחקור את מבנה הלשון. יוצא מכלל זה מחקר פרטני על צלילי הלשון והפרוזודיה, שבעבורו יש צורך במגע בלתי אמצעי עם צלילי הדיבור (אולם גם אז יש לתרגם את התוצאות אל הערוץ החזותי כדי לפרסם את תוצאות המחקר). לשם חקירת הלשון יש אפוא להשתמש בטקסט כתוב. ההמרה מערוץ לערוץ אפשר שתיעשה בדרכים שונות, החל בשימוש בכתוב המקובל של הלשון ובסימני הפיסוק וכלה בהעברה מדוקדקת של צלילי הדיבור עם היסודות הפרוזודיים המתלווים אליו. אנו נכנה את העברת הדיבור אל הערוץ החזותי בכתוב המקובל במונח **תמליל**. המונח **תעתיק** ישמש לרישום הדיבור בסימנים פונטיים, דהיינו באותיות לטיניות או בסימנים מקובלים אחרים, כגון סימני התעתיק השמי או סימני האגודה הפונטית הבין-לאומית (IPA). להלן דוגמה לתמליל ולתעתיק מסוגים שונים:

תמליל	תעתיק לאותיות לטיניות	תעתיק צר למחצה	תעתיק צר
מה הקשר	ma hakešer	ma akefek	'maa'kefek
אף אחד לא בבית שלי	af exad lo babait šeli	af əxəð lo babait ʃəli	'afə,xaðlɔbɔɪt'baətʃə'li
הייתי נותנת לך את המפתח	hayiti notenet lexa et hamafteax	aiti notəd lexa tɪmafteax	'əjɪtəntə'deðle'xatamaf'teax
היית הולך	hayita holex	aita olex	a'itao'lex

הטור הימני מייצג תמליל בסיסי. בטור השמאלי ביותר תעתיק צר, שבו אין הפרדה בין מילים, ממש כפי שרצף הדיבור אינו מבחין בין מילים (ראה להלן). לידו תעתיק צר למחצה והפרדה בין המילים, ומימין תעתיק לאותיות לטיניות המשמש הן בידי הדוֹטוּת הן בידי בלשנים (בשינויים קלים, למשל $x \sim sh$, ξ).

תעתיק רחב יסודו בשיקולים פונולוגיים; תעתיק צר מנסה לקרב את הקורא לשונות הפונטית של הדיבור, אם ידועה המערכת הפונולוגית שלו ואם לאו. כל סוג של העברה אל הערוץ החזותי, תמליל כתעתיק, מעוגן במסד תאורטי, שהרי גם התעתיק הצר ביותר לא יוכל לתרגם את כל הצלילים המופקים בעת הדיבור. גם ברשימת IPA כל סימן פונטי – לתנועה, לעיצור או לתכונה פרוזודית – מתאים למגוון של סוגי צליל. לפיכך הוויתור על העברת מאפיינים או רכיבים כלשהם אל הערוץ החזותי תלוי במסד התאורטי שממנו אנו יוצאים ושעל פיו אנו עובדים. נקודה נוספת שיש לתת עליה את הדעת היא המתח המתמיד בין הצורך להעביר את הטקסט המדובר שיהיה קרוב ככל האפשר לדיבור עצמו, לבין הצורך שלא להכביר נתונים שיקשו על השימוש בחומר המחקרי. כפי שברור מן הדוגמה דלעיל, הקריאה בתעתיק הצר אינה פשוטה.

תעתיק פונטי רחב, או תעתיק פונמי, כפי שהוא נקרא לעתים, משקף תפיסה פונולוגית של הלשון. מטרה חשובה של התעתיק היא לתאר תיאור חדש את המערכת הפונטית-פונולוגית של הלשון, ותעתיק רחב בשלב ראשוני ימנע אפשרות לבסס תאוריה פונולוגית על בסיס ניסיוני לא מיוחד. בכל מקרה, כבר הזכרתי קודם כי העוסקים בתורת ההגה זקוקים לרצף הקולי עצמו לשם המחקר, ואמנם מעמ"ד יעמיד את הקלטות הדיבור עצמן למשתמשים. התעתיק הצר או הרחב ישמש את החוקרים לאחר שיבססו את התאוריה שיציעו בעבור הלשון הנחקרת. לחקר ענפי הלשון האחרים – הזקוקים ממילא לטקסטים בכמות רבה – תמליל נראה הפתרון המועדף מן השיקולים התאורטיים שנזכרו לעיל, וגם משום הקושי הטכני והעומס הרב בתעתוק לעומת תמלול, במיוחד כאשר כמות הטקסטים רבה. היתרון הגדול של התמליל הוא בשרירותיות שלו, בידעה שהתמליל אינו משקף את הרצף המדובר אלא את תרגומו השרירותי ואין הוא פשרה הנתונה באורח מהותי בתעתיק פונטי. המילים שיובאו בתמליל יהיו, אפוא, תרגום של יחידות הרצף הדיבורי ללשון הכתב והתאמתן לה. חיסרון ברור של התמלול הוא היותו סותר את התפיסה של הרצון לנתק את חקר לשון הדיבור מלשון הכתב.

שונות לשונית: קולות, פונמות וגרפמות

בטקסט המובא לעיל יש ארבע היקרויות של ההגה $[x]$, אשר שתיים מהן מיוצגות בתמליל ב־ח ושתיים ב־כ. בכמה אתנולקטים של העברית הישראלית שתי גרפמות אלו מבוטאות $[h]$ ו־ $[x]$ בהתאמה, וראוי לראות בהן שתי פונמות נפרדות. דעת אחדים היא כי שתי פונמות אלו קיימות אף בלשונם של דוברים אשר אינם

מבחינים בהגייה בין ח לבין כ. אחרים רואים הבדל במערכת הפונולוגית בנקודה זו בין הדיאלקטים השונים של העברית. הכרעה בין האפשרויות היא שאלה תאורטית, ותעתיק פונמי מן הדין שיעשה רק לאחר קביעה מבנית בשאלה זו. הפתרון לשאלה זו אפשר שיהיה במישור פונולוגי, מורפופונולוגי או מורפולוגי, והשיקולים התאורטיים שיקבעו בעניין זה יביאו בחשבון לא רק את דיבורו של אדם זה או אחר, אלא אף את הגישה העקרונית לשאלה אם המערכת הלשונית היא מערכת אחת בעלת שונות, או מערכות נבדלות שקשרי גומלין ביניהן.

דוגמה דומה היא היחס בין ע לבין א. תעתוק שתי המילים **על** ו**אל** בצורה זהה, למשל al, יש בו הצהרה פונולוגית בדבר הזהות הפונולוגית של שתי גרפמות אלו בעברית המדוברת, לפחות בצמד ההומונימי הזה ובדומיו, כגון or בעבור הצמד **עור** ו**אור**. השתקפות מעניינת של ההבדל והקשר שבין הלשון המדוברת ללשון הכתובה ניכרת במקרים שבהם ההקשר אינו מבהיר את משמעות המילה. אז יכריז הדובר, בביטוי של מטה-לשון, על האות שבה תיכתב המילה (למשל 'ain be'alef "אֵין באל"ף").³ הנה כי כן, בכמה מקרים ללשון הכתובה הבחנות שאינן קיימות בלשון המדוברת (אף כי, כאמור, ביצוען בדיבור של צמד הגרפמות ע-ו-א נתון לניתוחים פונולוגיים שונים, כמוהו כצמד ח-ו-כ). לשון הכתב תבחין בין ט לבין ת, אולם לא לשון הדיבור (לבד מזו של שרידי אתנולקטים מועטים). לשון הכתב לא תבחין בין v לבין o או u, שהרי שלוש פונמות אלה נכתבות כולן באות ו. לשון הדיבור תבחין בין p לבין f, אולם לשון הכתב לא תבחין בין העיצורים הללו. כמובן, כוונתי לכתב הלא מנוקד, שהוא דרך הכתיבה והקריאה הרגילה לנו. כשמערכת התנועות של העברית הכתובה מסומנת בניקוד, הריהי שונה שינוי מהותי ממערכת התנועות של העברית המדוברת. בלשון המדוברת אין, למשל, הבדל בין קמץ לבין פתח: שניהם מבוטאים a, אולם הקמץ אפשר שישמע כתנועת o. ההבדל היחיד בין **גִּדְּלָה** (=הגודל שלה) לבין **גְּדָלָה** (=היא נהייתה גדולה) בשפה הכתובה הוא בהימצאות המפיק בה"א. הלשון המדוברת מבחינה בין שתי המילים הללו באיכות התנועה הראשונה: "הגודל שלה" יבוטא godla; "היא נהייתה גדולה" יבוטא gadla.

דוגמאות אלה ואחרות מבהירות כי לשתי הלשוניות, הלשון המדוברת והלשון הכתובה, שתי מערכות פונולוגיות שונות. לאמתו של דבר, המונח פונולוגיה מטעה אם מכוונים אנו ללשון הכתובה בלבד, כי אין ללשון הכתובה צלילים. אם המונח פונמה יוגדר יחידת הלשון שמרכיביה צלילים, ומערכת פונמית כמערכת הגאי

3. ההכרזה על הכתיב לשם הבחנה בין הומונימים או הומופונים מהווה פן במכלול של תופעות שעניינן השפעת הכתיבה על השפה. תחום זה הוא נושא עבודת הדוקטור שכותב כעת ישי נוימן (אוניברסיטת פריז 3 Sorbonne Nouvelle), ואני מודה לו על שהפנה את תשומת לבי אל עבודתו.

הלשון, הרי שללשון הכתובה אין פונמות, יש לה אותיות, גרפמות. גם אם נקרא בקול מן הכתב, המרת הערוץ תיעשה בתרגום של היחידות המזעריות. מכאן שלא נוכל לדבר על פונולוגיה של לשון כתובה, אלא על מערכת גרפמית המקבילה לפונולוגיה של הלשון המדוברת, ויהיה עלינו לומר ששתי המערכות הללו אינן חופפות במלואן. כך, למשל, במערכות של ב-ו, ח-כ-ק כמו במערכות של היחס בין p לבין f, תעמודנה לעיון שאלות של שונות לשונית, ויהיו להן היבטים ופתרונות תאורטיים שונים.

לצד השונות הכללית המתוארת לעיל, זו שבין הלשון הכתובה למדוברת, הכוללת גם את היחס בין המערכות הפונולוגיות של שתי קבוצות דוברים עיקריות בחברה הישראלית, ניכרת גם שונות רבה בלשונם המדוברת של קבוצות דוברים מצומצמות ומסוימות, הן בשל מוצא עדתי שונה הן בשל רקע חברתי או חינוכי שונה הן בשל גיל או רקע דמוגרפי אחר. שונות, אם דמוגרפית ואם נסיבתית, דהיינו שונות התלויה בנסיבות הדיבור, היא מאפיין בסיסי וראשוני של הלשון. זוגות כגון xiba ~ kiba (כיבה) או yitpos ~ yitfos (יתפוס), אפשר למצוא בדיבורם של אנשים רבים, בלא קשר למוצאם. זוגות כגון bikef ~ vikef (ביקש) הם, ככל הנראה, משתנים המאופיינים דמוגרפית, אולם הנושא צריך בדיקה מעמיקה. שונות בדיבורו של יחיד ניכרת גם בדפוסי הצמצום האופייניים לדיבור מהיר. שתי דוגמאות מובהקות הן:

tsax ~ tsaxix -- tsaxim ~ tsixim (צריך - צריכה - צריכים)

ta- ~ et ha- (את ה')

בעוד שהאנגלית הכתובה הרגילה מאפשרת קיצורים המחקים צורות מדוברות (למשל: ain't בעבור is not, can't בעבור cannot ואפילו gonna בעבור going to), העברית עדיין לא יצרה תקנים כאלה, והצורות כגון **ת'ספר** במקום **את הספר** עדיין אינן נפוצות, וודאי שאינן תקניות. אפשר, לכאורה, להביע קיצורים כאלה גם בתמליל, אולם לא תמיד הקיצור קיצוני כרשום לעיל, ופעמים הרבה ההכרעה בשאלה איזו צורה תירשם אינה פשוטה. שאלה נוספת העומדת לפני המתמלל היא כיצד לאפשר חיפוש קל בתמליל ממוחשב של מאגר לשון גדול. עבודתו של חוקר לשון בעזרת מאגר כרוכה בחיפוש מתמיד של צורות ומבנים. חיפוש מבנים יתאפשר רק במאגר מתויג (כלומר, מאגר שכל אחת מן המילים שבו מנותחת ניתוח מורפולוגי או תחבירי מפורט). תמליל של מאגר כלשהו שאינו מתויג, מן הראוי שיפשט את דרכי חיפוש המילים ככל הניתן, על ידי היצמדות לצורה המלאה המקובלת בכתב. שרירותיות התמליל היא, כאמור, יתרונה הגדול. והרי שאלות של קיצור וצמצום צורות אף הן שאלות פונולוגיות בעיקרן שיש להן לעתים השלכות על המורפולוגיה, אך בדרך כלל לא על התחביר (וראה עוד להלן).

בעוד שהשונות המתוארת לעיל ניתנת לפתרון, בדרך כלל על ידי חוקי דיבור מהיר או חוקיות פונטית או פונולוגית אחרת, הרי על השונות המורפולוגית קשה יותר לגשר בעזרת בריחה אל הכתיב התקני. התמדה בעקרון הרישום על פי התקן תעלים מעיני המעיין והחוקר צורות כינויים וצורות פועל במ"ם סופית לנקבה ("אתם רוצות"), צורות פועל בתחילית י למדבר ("אני גם יעשה"), צורות מספרים שאינן תקניות ("שלוש תפוזים"), צורות נפוצות, כגון **אצלהם** (ולא **אצלם**) ודומותיהן. מן הראוי לחשוב על רשימת חלופות אפשריות, ולשלם את המחיר בחיפוש כפול. כאמור, תיוג המאגר יצמצם את הבעיה, כי התיוג מטבעו רושם את התפקיד המורפולוגי או התחבירי ומתעלם מחלופות מורפולוגיות או פונולוגיות.

הומוגרפים

חיפוש צורות יתאפשר אך ורק אם הצורות יובאו בתמליל בכתיבה אחידה ובלא הפרעות. ניקוד ממוחשב מתבצע על ידי הכנסת תווים בין האותיות, ומכאן שניקוד אינו רצוי בטקסט ממוחשב.

הבחנה בין הומוגרפים שתנועותיהם שונות צריכה להינתן מחוץ לגבולות המילה. הדבר ייעשה בתעתיק פונטי בסימון IPA וינתן בין סוגריים מסולסלים. דוגמאות:

ס: ההורים שלהם רוצים שהם יהיו בחגים בבית ||

ע: שמה {fe'ma}

ס: שהם יהיו בבית בחגים ||

אז שתקי {fte'ki} ||

בוא הנה {bona}

הדלק הזה |

קורע אותי |

עומר |

חבל על הזמן ||

בוא הנה {boana}

אני כל יומיים --

בדוגמה הראשונה אפשר לטעות בקריאה במילה **שמה** (**שמה** או **שמה**). בדוגמה השנייה (שתקי {fte'ki}) התעתיק נועד להסב את תשומת לב הקורא לצורות שאינן מקובלות כתקניות, אף כי אפשר שתהיינה מקובלות בלשון המדוברת בכלל או בלשון הדבורה בפרט.

שתי הדוגמאות האחרונות ("בוא הנה") ממחישות תפיסה שונה של הביטוי בלשון המדוברת מבלשון הכתובה, והעניין צריך עוד מחקר: נצטרך לבדוק את כל

ההופעות של הרצף הזה במאגר כדי להכריע בכמה שאלות, אם *boəna* היא "מילה" אחת או שתיים, אם יש הבדל במשמעות בין ביצועים שונים של הרצף, אם יש הבדל במין ובמספר בהופעות שונות של הביטוי, ואם נמצא כי אכן כך, צריך לבדוק אם ההבדל הזה קיים בכל ההופעות ויהיה צורך להסביר אותו.

בניגוד לסימון התנועות, בעת הצורך אפשר לסמן את ההברה המוטעמת גם בתוך המילה. התעתיק המקובל (IPA) מסמן את ההברה המוטעמת במתג לפנייה, אך הסימון הזה מפריע לרצף של האותיות. לעומת זאת, התמליל יוכל לסמנה על ידי שינוי בסגנון האות, למשל על ידי אות שמנה או על ידי תוספת קו עילי או תחתני על ההברה המוטעמת. מערכות המחשוב המודרניות מאפשרות, בדרך כלל, העברה מתכנה לתכנה או ממחשב למחשב.

כך טוב לעשות במקרים שמקום הטעם אינו תקני (אף אם מקומו זה של הטעם מקובל בלשון המדוברת), בשמות פרטיים, שבהם מקום הטעם עשוי להשתנות מדובר לדובר, או במקרים שמובאת מילה לא מוכרת או מילה שעשויה להיקרא בהטעמות שונות. דוגמאות:

תשעים ושמונה מהמקרים

היא לוקחת חופש אורית

מילים שאולות

כפי שממחישות שתי דוגמאות אחרונות אלה, התמליל אמנם מפשט את בעיותינו בהעברת הדיבור אל הערוץ החזותי, אולם ראינו כי אף התמליל אינו מונע חיבוטים. כיצד נתעתק את הרצף *ala'bokeʔ*? "עלא בוקר" או "על הבוקר"? ואת הרצף *ala'subax* – "עלא סובח" או "על הסובח"? ו-*ala'kefak* – "עלא כיפק" או "על הכיפק"? (והאם האיות הזה, ב־כ וב־ק, קביל?) במקרים כגון אלה, של מילים או ביטויים שאולים שכבר התאזרחו בלשון, על אף השרירותיות של התמליל, יש צורך בהכרעות תאורטיות לשם התמלול עצמו. בעברית יש מילים לועזיות רבות. האם נתמלל **רֹמְסְרִיִּסִים** או **רֹסְרִיִּסִים**? האם מילה אחת היא בעברית, כפי שמלמדת תוספת צורן הריבוי בסוף הביטוי או שניצמד לצורת המקור האנגלית ונתמלל כשתי מילים (עם מקף או בלעדיו)?

גם מבחינה טכנית ניתקל בקשיים:

ומה בדבר (ה)יִיאורֵאִילִים

מונחים כגון זה מופְּרָשׁ, בדרך כלל, בשימוש שבעל־פה או בכתבים השגור בלשון המקור (URL), ויש קושי לקרוא את המילה בתמלילה העברי. התאזרחותו של הביטוי, לפחות בקרב קבוצת דוברים מסוימת, מוכחת במקרה זה על ידי תוספת צורן הריבוי העברי. מילים שאולות שאינן מאזרחות או עגה מקצועית עשויים

להוות בעיה בפני עצמה. הדוגמה הבאה ממחישה את הבעייתיות שבהעתקת מילים לועזיות:

ס: [איזה יִנְדַּאי {jundaj} יש להם] |
 לִנְטֶרָה {lantʁa} |
 או אֶקְסֶנט {ek'sent} ||
 מ: (אין לי שמץ של מושג) ||
 ע: \{æksent} accent |
 ס: אה |
 יש להם אֶקְסֶנט {ak'sent} \

ככלל, מן הראוי שמילים לועזיות יתומללו בתעתיק עברי, שכן הן בדרך כלל חלק אינטגרלי מלשונו של הדובר, גם אם אינן נפוצות אצל כלל האוכלוסייה. תמלילה של מילה לועזית יובא בכתבה המקורי, אם ברור שהדובר התכוון לשרבב מילה לועזית בתוך דבריו. בדוגמה דלעיל הדובר עי' שינה את מבטאו הילידי כשביטא את שם המכונת, וגם הטעמתו את המילה הייתה שונה משל בת שיחו ס'.

המילה בלשון המדוברת

בעוד שרצף הדיבור אינו מבחין בין מילים, העברת הדיבור אל הערוץ החזותי יוצרת, ככלל, את ההפרדה המקובלת בין מילים. כך בוודאי בתמליל, אולם בדרך כלל גם תעתיקים פונטיים נוהגים את ההפרדה המקובלת למילים. על אף ההדגשה על השרירותיות של התמליל, מן הראוי להסב את תשומת הלב לשאלת המושג **מילה** בלשון המדוברת. המושג **מילה** אינו מוגדר דיו אפילו בתחום הלשון הכתובה, שבה תיראה המילה כרצף אותיות המוגבל בין רווחים או סימני פיסוק. בחקר הלשון המדוברת, בדרך כלל אנחנו מתייחסים למילה כפי שמסורת הכתיבה מתייחסת אליה, אף כי ברצף הדיבור אין רווחים שיוכלו להגדיר מילה. בחקר לשונות ללא מסורת כתיבה פותחו גישות שונות לקביעת החלוקה למילים. קריטריונים המשמשים להגדרת המילה במחקר הבלשני הכללי עשויים להיות פונטיים או מבניים, וכוללים, בין היתר, קיום סדק פוטנציאלי לפסק בדיבור, חוסר אפשרות לחלוקה של רצף, האפשרות לקיום עצמאי, גבולות להתפשטות יסודות פונטיים, שיקולים של הטעמה, חילופים של פרדיגמות ושיקולים סמנטיים. ואם בחברה אוריינית עסקין ובלשון שיש לה לצד הערוץ המדובר גם ערוץ כתוב, הרי בלשון המדוברת על גווניה השונים חלוקת המילים תהיה שונה מבלשון הכתובה ועשויה להתבסס על קריטריונים אחרים. מכל מקום עלינו לבדוק שאלות אלה מנקודת ראותו של הרצף הקולי ולא לצאת מן המקובל בחקר הלשון הכתובה. התבוננות ומסד תאורטי יכתיבו את הגדרת המילה ואת החלוקה למילים.

לא אדון פה בהרחבה באפשרויות השונות, אלא אדגים את השאלה התאורטית בעזרת אפשרות אחת. קריטריון מקובל להגדרת המילה בלשון המדוברת הוא הקריטריון הפרוזודי: מילה תוגדר יחידה אם היא בעלת טעם ראשי יחיד. על פי קריטריון זה, מספרים המסתיימים בַּעֲשָׂרָה ייחשבו בדרך כלל מילה יחידה:

e'xadeske 'fekel o ar'baeske 'fekel lekaʔtis
אחד עשרה שקל או ארבע עשרה שקל לכרטיס

כמובן, קריטריונים אחרים יגדירו את מקום הגבולות בין המילים, אולם גם בזאת לא אעסוק כאן. לענייננו בשאלת השלב הראשון בחקר הלשון המדוברת ובדבר הצורך להעבירה אל הערוץ החזותי, אעיר כי התמליל אינו מאפשר דיון בשאלה זו. התעתיק הפונטי, לעומת זאת, עשוי להבהיר את השיקולים שעמדו לפני המתעתק בחלוקת המילים.

גם סמיכויות אפשר שתיקלטנה כיחידת הנגנה אחת, אולם לא בהכרח. בשתי הדוגמאות להלן ברור כי הביטוי **סוף שבוע** הוא ביטוי שהתלכד למילה אחת, ואכן הלשון הכתובה מכירה את שני הכתיבים, **סוף שבוע** או **סופשבוע**.

העיקר שבוע (כחכו) --

העיקר שבוע הבא אני סופשבוע בבית מלון עומר |

ס: [נרד] --

מ: [סופ'] סופשבוע ל- לא [הקרוב]

|| הבא

אולם לא תמיד הקריטריון הפרוזודי יכריע בשאלה הכללית. לעתים יידרשו שיקולים צורניים נוספים, לעתים גם שיקולים סמנטיים. האם **בית ספר** היא מילה אחת או שתיים? מנקודת הראות הסמנטית הצירוף ראוי לו להיקרא כמילה אחת, וגם ביצועו כיחידת טעם אחת יתרום להגדרתו כמילה אחת. אולם שיקולים צורניים אחרים עשויים להעלות הרהור שני, במוצאו צורות מיוחדות שונות לצירוף: **בית הספר** לצד **הבית ספר**. אישית לא שמעתי **הבית הספר**, אולם שמעתי גם שמעתי **המזג האוויר** לצד **מזג האוויר** ולצד **המזג אוויר**. דוברים שונים מתייחסים לצירופים אלה בדרך שונה. אולם השונות אינה רק נחלתו של הכלל אלא גם נחלתו של היחיד. בטקסט שנקרא, ככל הנראה, מן הכתב, נמצאו כמה צורות של הצירוף **כתב יד**. על פי הקריטריון הפרוזודי, ברוב המקרים יוגדר הרצף הזה כשתי יחידות: 'ktav 'jad. היקרות אחת של הצירוף נשמעת מילה יחידה כשהטעם על ההברה הראשונה: 'ktavjad. במקרה אחר, בריבוי, לרצף כתבי היד טעם ראשי אחד: 'kitvea'jad.

ובכן, לא כל הדוברים זהים ולא כל ההקשרים זהים. לשונות תפקיד מכריע בחיי הלשון, ומכאן שעלינו להכיר בה כיסוד מוסד של מבנה הלשון ולהעניק לה תפקיד מכריע בהבנת הלשון ובניתוחה.

נושא אחר שבו אדגים את הצורך בבדיקה רעננה וללא משוא פנים של מבנה הלשון המדוברת הוא כינויי הדאטיב: האם כינויים עצמאיים הם או חבורים?

יש לך {yefxa} ישראלכרט \

אם אתה רוצה להקיא תגיד לי {ta'gidli}

אני אעצור לך {a,nia'fsoxlexa} ||

ולבסוף, נשווה לאלה את צירופי השייכות:

ע: [מקלחת שלי {mik'laxatfli}]

מ: [אמרתי לו שיבוא] לישון אצלי

|| הוא לא רצה ||

ע: [מקלחת שלי {mik'laxatfli}]

אני אוהב את שלי {etfe'li}

|| נו (מה) ||

ס: מה הקשר ||

אף אחד לא בבית שלי {ba'bait fe'li}

כפי שמלמדות הדוגמאות הללו, מן הראוי לבדוק את הגורמים לקליטיזציה (הידבקות הכינוי למילה הקודמת) וליצור מודל שיגדיר כהלכה ומתוך מתן תשומת לב לשונות את תפקוד היסודות הללו מכל הבחינות: פונולוגית, מורפולוגית, תחבירית וסמנטית.

לסיום, אני מביא דוגמה שתמחיש את מרחב העיון הדרוש לשם העברה מדוקדקת מערוץ לערוץ ואת חשיבות העיון הזה לחקר השפה. כאמור, הפרדה בין מילים אינה הליך דיבורי שכיח. כל מי שהאזין לדיבור בלשון זרה לו שמע למעשה רצף בלתי פוסק, וודאי שלא יכול להבחין בין מילים ברצף הנשמע. אם במילים עסקינן, החלוקה אינה פשוטה כלל וכלל. בני שיח מבינים, בדרך כלל, זה את זה, והחלוקה ליחידות הלשון אינה מהווה בעיה אצל השומע משום ידיעתו האינטואיטיבית את אוצר המילים ואת דקדוק הלשון. בכל זאת לעתים יקרה שהשומע אינו תופס כהלכה את דברי בן שיחו. אי-בהירות עשויה לנבוע מסיבות רבות, לשוניות וחוצי-לשוניות, ואם בני שיח עלולים לשנות בהבנת דברים, על אחת כמה וכמה מאזין השומע הקלטת שיחה שדובריה אינם בני שיחו. בהקלטותינו שמענו מפי דוברת יוצאת יוון בשנות השבעים לחייה:

a'marti'kax'po'kax'po

רצף זה אפשר לפרשו בשתי דרכים: (1) אמר תיקח פה קח פה; (2) אמרתי קח פה קח פה. דהיינו, "אמר: תיקח פה, קח פה!" או "אמרתי: קח פה, קח פה!". התמליל הראשון הוגש לי על ידי מתמללת מנוסה למדי, אשר היא עצמה השתתפה בשיחה זמן מה לפני התמלול. מבחינת רצף השיח, תמליל זה מייחס את

הדברים לדובר ממין זכר, ומעלה שתי שאלות דקדוקיות: (א) האם בלשונה של הדוברת בהקשר זה (שהוא הקשר דיווחי או סיפורי) אפשרי שיובא פועל זה בלא נושא מפורש? (ב) האם אפשרית השניות של צורת נטיית התחיליות ושל צורת הציווי ברצף יחיד? התמליל השני מייחס את הדברים לדוברת עצמה, ומראה שתי צורות ציווי עוקבות וחזרה מדויקת על הצירוף הפועלי. בדיקה מדוקדקת של הקשר הדברים הראתה שהפירוש השני עדיף. במקרים אחרים ייתכן שההכרעה לא תהיה אפשרית. כמובן, הסקת מסקנות על מבנה הלשון מן הדין שתבסס על יותר מהיקרות אחת, אולם (וודאי) מקרים כאלה אומרים דרשני. אין הם מועטים.

הערות אלו ודאי שאינן ממצות את הדיון בשאלת המילה בלשון המדוברת, אף לא בשאלת המרת הערוץ המדובר אל הערוץ החזותי. לא נגענו בשאלת צירופי המילים אל יחידות גדולות יותר, שהיא שאלה הנוגעת לתחבירה של הלשון המדוברת; לשאלות אלו הקדשתי מאמר מיוחד.⁴ אולם בכל תחום נותרו וייותרו שאלות רבות פתוחות, והכרעות חדות אינן אפשריות. הבעיות אינן פשוטות, וכל שיטה להמרת ערוץ יש בה חסרונות. הרקע התאורטי של החוקר, שאלות המחקר והשערות המחקר שהוא מביא אל בדיקת הטקסט המדובר הם שיכריעו בסופו של דבר מהי דרכו בהעברת הרצף הדיבורי אל הכתב. משום שהלשון המדוברת היא מהלך דינמי במהותו והלשון הכתובה היא סטטית, ההמרה אל ערוץ הכתב כמוה כהקפאת זרם החיים השוטפים בתצלום, שהוא מטבעו דו-ממדי, מקובע לזמן נתון ומשקיף אל המציאות המקובעת בו מנקודת מרחב יחידה. שום חוקר של הלשון המדוברת אינו יכול ואינו רשאי להסתפק בעיון בטקסט המתומלל או המתועתק, שכן ללשון המדוברת יסודות מרובים שאינם ניתנים להעברה אל ערוץ הכתב, יסודות המהווים חלק מהותי של מבנה הלשון.

4. המאמר "מדיבור לתחביר – מתאוריה לתמליל" יתפרסם בספר היובל לכבוד אהרן דותן.

מחקרים נבחרים על הלשון המדוברת בכלל ועל העברית המדוברת בפרט

הלשון המדוברת והלשון הדבורה - כללי

M. A. K. Halliday, *Spoken and Written Language*², Oxford 1989 • C. Blanche-Benveniste, *Approches de la langue parlée en français*, Paris 1997 • J. Miller and R. Weinert, *Spontaneous Spoken Language: Syntax and Discourse*, Oxford 1998 • D. Biber et al., *Longman Grammar of Spoken and Written English*, Harlow 1999

העברית המדוברת - כללי

ח' רוזן, "עברית בת ימינו כתובה ודבורה", דברי הקונגרס העולמי התשיעי למדעי היהדות, ישיבות מרכזיות: הלשון העברית והארמית, ירושלים תשמ"ח, עמ' 77-84; תגובות: מ"צ קדרי, עמ' 85-92; א' גלינרט, עמ' 93-99 • י' בן-טולילה, "העברית המדוברת", לשוננו לעם מ-מא (תש"ן), עמ' 266-278 • מדברים עברית: לחקר הלשון המדוברת והשונות הלשוניות בישראל, תעודה, יח, בעריכת ש' יזרעאל, תל-אביב תשס"ב

העברית המדוברת - פונטיקה ופונולוגיה

מ' חן, "מבטאה של העברית הישראלית", לשוננו לו (תשל"ב), עמ' 212-219, 287-300 • א' לאופר, "מחקרים בפונטיקה", דברי האקדמיה הלאומית הישראלית למדעים, 7/3, ירושלים תשמ"ו • י' בן-טולילה, מבטאי העברית המשמשת במושב של יוצאי מרוקו בנגב: פרק בפונולוגיה חברתית, עבודת דוקטור, האוניברסיטה העברית, ירושלים תשמ"ד • Sh. Bolozky, "Israeli Hebrew Phonology", *Phonologies of Asia and Africa (Including the Caucasus)*, ed. S. Kaye, 1, Indiana 1997, pp. 287-311

תעתיקים ותמלילים של עברית מדוברת ושל עברית דבורה

ח' בלנק, "קטע של דיבור עברי ישראלי", לשוננו כא (תשי"ז), עמ' 33-39 • י' בן-טולילה, מבטאי העברית המשמשת במושב של יוצאי מרוקו בנגב: פרק בפונולוגיה חברתית, עבודת דוקטור, האוניברסיטה העברית, ירושלים תשמ"ד, עמ' 232-349 • נ' שתיל, דגמי המשפט בעברית המדוברת, עבודת מוסמך, אוניברסיטת תל-אביב תשנ"ד, עמ' 119-172 (כולל תמלילי הטקסטים של בלנק) • נ' שתיל, קבוצות שמות עצם בעברית המדוברת, עבודת דוקטור, אוניברסיטת תל-אביב תשנ"ז, עמ' 303-338 • ש' יזרעאל, "מאגר העברית המדוברת בישראל (מעמ"ד): דוגמאות טקסטים", לשוננו סד (תשס"ב), עמ' 289-314 • H. Blanc, *Israeli Hebrew Texts: Studies in Egyptology and Linguistics in Honour of H. J. Polotsky*, Jerusalem 1964, pp. 132-152 • M. J. Chayen, *The Phonetics of Modern Hebrew*, Janua Linguarum, Series Practica, 162, The Hague 1973, pp. 39-45