

# הערות לאוצר המילים של הקורפוס: הצגת שכיחויות ובחינת מתודולוגיה לקסיקוגרפית לטיפול בטקסט דבור

דורון רובינשטיין

לזכר אבי דוד בנימין רובינשטיין, שנלב"ע בחודש  
אב תשע"ה, והוא בן שמונים שנה. שלי – שלו הוא.

## 1. פתיחה\*

במאמר זה אעיר הערות אחדות על אוצר המילים של הקורפוס ששימש בסיס למיזם. כאמור (יזרעאל תשע"ו, בספר זה), בקורפוס 21 קבצים פרי הקלטותיהם של תשעה מתנדבים. אורכו הכולל של הקורפוס הוא כ־200 דקות הקלטה, מתועד בו דיבורם של כ־40 אנשים לכל הפחות, ויש בו 26,474 תיבות, לרבות התחלות נפל, התחלות כפולות וקולות שאינם מילים.<sup>1</sup>

לצורך הבדיקה התקנתי לקורפוס מאגר במתכונת מסד הנתונים של מפעל המילון ההיסטורי ללשון העברית של האקדמיה ללשון העברית, ותיבות הקורפוס נותחו כאילו היו חלק ממסד הנתונים של המפעל. תוצאת עבודתי היא קונקורדנציה לקורפוס במתכונת מפעל המילון ההיסטורי ללשון העברית. את הקונקורדנציה התקנתי בעיקר על פי גרסאותיהם המתוכתבות של הדברים בקובצי תכנת ELAN (יזרעאל תשע"ו, בספר זה) תוך שנעזרתי בהקלטותיהם במקומות הקשים לפיענוח.

## 2. הערות לשיטה המילונית הנוהגת כאן

השיטה המילונית הנוהגת במפעל המילון ההיסטורי ללשון העברית – וכאמור, גם כאן – תוארה בהרחבה פעמים אחדות בעבר.<sup>2</sup> אחזור כאן על אחדים ממאפייניה,

\* ברצוני להודות לאורה שורצולד, לשלמה יזרעאל ולנועם אורדן על הערותיהם המאלפות והמועילות.

1 השוו לראו את שורצולד תשע"ו, בספר זה, וראו גם להלן, פרק 3.

2 תיאורה הטוב ביותר נרפס להערכתו בפרסום הזה: האקדמיה ללשון עברית תשכ"ט.

החשובים להבנת התיאור שבהמשך דבריי. אעיר כבר עתה כי על כמה מהם יצאו עוררין הן בתוך המפעל הן מחוץ לו, וארחיב על כך להלן בדברי הסיכום למאמר. ואולם עוד לפני כן ראוי להגדיר שלושה מונחים המשמשים בתיאור:<sup>3</sup>

תיבה – כל היקרות (או תמנית או תקרית) המצויה בקורפוס, לרבות התחלות הנפל, ההתחלות הכפולות והקולות שאינם מילים (token);

צורה – כל תבנית המצויה בקורפוס (type). הצורות כוללות בעיקרן את צורת היסוד ואת כל צורתיו הנטויות של הערך וכן את כל הצורות הללו כאשר קודמות להן מילות השימוש מש"ה וכל"ב. בהכרח, מספר הצורות שווה למספר התיבות או קטן ממנו, שכן צורות רבות נקרות בקורפוס פעמים רבות;

ערך – הערך המילוני או האנציקלופדי (entry). בהכרח, מספר הערכים שווה למספר הצורות או קטן ממנו, שכן ערכים רבים ממומשים בקורפוס בצורות שונות.

לדוגמה, במשפט "אה ... כלכ- כלכלה ו- כלכלה לבד" [C514\_2] יש שש תיבות (אה, כלכ-, כלכלה, ו-, כלכלה, לבד), חמש צורות (אה, כלכ-, כלכלה, ו-, לבד), וארבעה ערכים (אה |קול או מילה מלאכותית|, פִּלְפֶּלָה, וְ, בַד |במשמעות 'חלק, בדד'|).

הערכים עצמם מחולקים לשלוש קבוצות כדלקמן:<sup>4</sup>

א. פעלים – בקבוצה זו כללות כל המילים שצורתן צורת פועל, כלומר שורש הגזר בבניין מבנייני הפועל: הִיהַ (היי |קל|), אָמַר (אמר |קל|) וכיו"ב. בנייני הפועל הם אלו: קל, נפעל, פיעל, פועל, התפעל, הפעיל, הופעל, פוֹלֵל, פוֹלֵל, התפולל. ראוי להדגיש כבר כאן כי כל הפעלים שנמצאו בקורפוס נגזרו בבניינים הנ"ל. לשון אחר: לא נמצאה בקורפוס כל עדות לגזירת פועל שלא בבנייני הפועל המוכרים בדקדוק המסורתי.

כמו כן ראוי להעיר כי בשני עניינים הקשורים במערכת הפועל מפעל המילון ההיסטורי שונה מן הנוהג בדרך כלל במילונות העברית:

- 3 השוו לכאן את ההגדרות המתאימות אצל שורצולד וסוקולוף תשנ"ב, וכן את שורצולד תשס"ב: 94-95, וראו שם ביבליוגרפיה נוספת.
- 4 השוו לתיאור זה את תיאורה של שורצולד תשע"ו, בספר זה.
- 5 בניין זה לא נמצא בקורפוס הנדון כאן, וראו להלן, פרק 4.3.3.

(1) צורות הבינוני מתויגות כולן תחת מערכת הפועל ללא תלות במעמדן התחבירי.<sup>6</sup> לדוגמה, הצורה מְנַהֵל מקבלת את אותו ערך עצמו (נהל [פיעל], בינוני) בשתי המובאות האלה:

(א) אבל המצב הרוח הכללי שלי היום | לא נובע מאיך שרמי מנהל אותו בשלושה חודשים האחרונים || [OM; פועל];

(ב) כש | מושי כותב אה | המנהל שלי | לאיזה מנהל מבחינה סובייקטיבית | הוא התכוון || [שם; שם עצם].

בדומה לכך, גם הצורה נְכוֹן מקבלת את אותו הערך (כון [נפעל], בינוני) בשתי המובאות האלה:

(ג) אז נכון שיש ניואנסים || C612\_4; שם תואר במעמד תחבירי של חג"ם?<sup>7</sup>  
(ד) - יפה תתחדשי ||

- זה הבנים קנו לי ||

- יפה ||

- איזה חמודים ||

- נכון || Y311; תואר הפועל.<sup>8</sup>

וכן הדבר בכל צורות מְפַעֵל ומְפַעֵל, ששויכו ללא יוצאת מן הכלל לבינוניים של הבניינים פועל והופעל, בהתאמה.<sup>9</sup>

(2) צורות שם הפעולה במשקלים אחדים מתויגות כְּזֶמֶן הדקדוקי "שם פעולה" תחת מערכת הפועל, ואינן מעמידות ערך לעצמן. לדוגמה, במשפט "שנבע מהמ- ניהול הקלוקל של עינת" [OM] ערכה של הצורה נְהוּל הוא נהל [פיעל], שם פעולה. הדבר תקף גם כאשר שם הפעולה מציין שם עצם מובהק: "עוד מעט תהיה לי דירה נחמדה להשכיר" [Y111; דירה – דור [קל], שם פעולה].

ב. "שמות" – בקבוצה זו נכללים שמות העצם, שמות התואר, תוארי הפועל, כינויי הגוף, כינויי הרמז והמיליות.<sup>10</sup>

6 השוו לכאן את מישור תשמ"ז, את מישור תשנ"ז, וכן את ההקדמה לבהט ומישור תשנ"ה.

7 השוו לכאן את כוזר תשנ"ב.

8 ראו גם להלן, פרק 4.3.2.

9 93 מ'96 היקרויותיו של בניין פועל בקורפוס הן בצורת מְפַעֵל, ו-31 מ'32 היקרויותיו של בניין הופעל הן בצורת מְפַעֵל. על ארבע ההיקרויות הנותרות ראו להלן, פרק 4.3.3, והשוו לכאן גם את דבריה של שורצולד תשע"ו, בספר זה, על העניין (ושם ביבליוגרפיה נוספת).

10 לעתים (רחוקות מאוד!) נקבע הערך בניגוד מפורש לצורה המסורה בקורפוס. לדוגמה, לצורה מְזוֹרֵן נקבע הערך התקני מְזוֹרֵן, ועליו נוסף ציון לייחוד בצורתה בהשוואה לערך

ג. "רשימות" – בקבוצה זו נכללים ערכים אשר אינם משתלבים שילוב מלא במבנה הלשון העברית או שמקומם הרגיל אינו במילון או שמקומם בערכי המְשָׁנָה במילונים המקובלים. בקונקורדנצייה הנוכחית הוגדרו בעיקר הרשימות האלה: קולות או מילים מלאכותיות, שמות פרטיים ושמות משפחה של אנשים, שמות מקומות, שמות חברות ומוסדות,<sup>11</sup> שמות ייחוס (ישראלי, יפני וכיו"ב), שמות מספר,<sup>12</sup> ראשי תיבות,<sup>13</sup> מילים ארמיות הנתונות בהקשר עברי, מילים לועזיות שמקורן בלשונות הודו-אירופיות, ומילים ערביות. רק ערך אחד מקבוצה זו – אה [קול או מילה מלאכותית] – היה נפוץ די הצורך כדי להידון בהמשך המאמר (והוא גם מופיע בדוגמה שלעיל, בראשית הפרק).

עד כאן תיאור השיטה המילונית. הערות נוספות ליישומה בניתוח הקורפוס אביא להלן בפרק 3.

### 3. הערות לקורפוס וליישום שיטת מפעל המילון ההיסטורי בניתוח

כאמור לעיל, בקורפוס 26,474 תיבות. בררת המחדל בתיוגן הייתה לרשום להן ערכים "מפורשים" – ערכים שצוינו בהם אותיות הערך ושאר פרטי התיוג, כפי שהוצגו לעיל. ואולם הממצאים אילצו אותי לרשום לחלק מן התיבות ערכים "לא מפורשים" – ערכים שאותיותיהם לא פורשו, אלא הומרו בסימן שאלה: ? . תיבות אלו משתייכות בעיקרן לשתי קבוצות:

הקבוצה הראשונה כוללת את התחלות הנפל ואת ההתחלות הכפולות. את חלקן אפשר היה לפענח ולתייג ברמה גבוהה של ודאות, ואולם פיענוחן של כמה עשרות מהן נותר עלום, והן קיבלו ערכים לא מפורשים. במקרים המתאימים שויכו המילים הללו לקבוצת הערכים המתאימה מן הקבוצות המנויות לעיל: ? (התפעל), ? (שם פרטי) וכיו"ב: "כי אף אחד לא כורת <את העצים. ד"ר > | אז הם מזדקנים | מת- נחלשים | ונופלים ||" [Och; בתיבה מת- ניכרת תחילתה של צורת בינוני

התקני. השוו לכאן את קביעתה ההפוכה של שורצולד תשע"ו, בספר זה. כמו כן השוו לכאן את בהט תשנ"א ואת התשובה "מזרן או מזרון?" באתר האקדמיה במרשתת.

11 השמות הפרטיים, שמות המקומות ושמות החברות והמוסדות הם בעיקרם ערכים אנציקלופדיים ולא מילוניים.

12 רשימה זו משמשת במספרים 11-19, המורכבים כל אחד משתי תיבות, וכן לתיוג אותיות המשמשות בגימטרייה: א, ב וכיו"ב.

13 אלו נמצאו בקורפוס בעיקר בלשונם של חיילים: מש"ק, פק"ש וכיו"ב.

בכניין התפעל]. למילות שימוש (אחת או כמה) שהופיעו כתיבות לעצמן, נרשמו ערכיהן המפורשים: כָּ, הָ, וְ, כִּי, לְ, מִן, שֵׁ.<sup>14</sup>  
 הקבוצה השנייה כוללת מילים זרות שאינן מקובלות בדיבור העברי ה"רגיל" (קראש [crash], רקומנדיישן [recommendation] וכיו"ב) ומילים זרות שפיענוחן נותר עלום.<sup>15</sup> מילים אלו קיבלו את הערך הלא מפורש הזה: ? [מילה לועזית].  
 קבוצה נוספת של תיבות גרמה להגדלה מזערית של מספר התיבות המנותחות בקורפוס: ב־62 מקרים נרשמו לאותה התיבה שני ערכים זה לצד זה (לדוגמה: הבנאדם – בֵּן, אָךְ [OCh]; יאללה – יָא [ערבית], אֵלֵּלָה [ערבית] ×15), מפי שבעה דוברים בשישה קבצים<sup>16</sup>), ובמקרה אחד נוסף נרשמו לתיבה אחת שלושה ערכים מילוניים זה לצד זה (שכש – שֵׁ, כֵּ, שֵׁ). אותן תיבות נספרו פעמיים ושלוש פעמים בספירת תיבות הקורפוס, בהתאמה, וכתוצאה מכך גדל מספר תיבותיו ל־26,538 (עלייה בת כ־0.024%).

התיבות הקטועות, הן אלו שקיבלו ערכים מפורשים הן אלו שקיבלו ערכים לא מפורשים, התיבות הלועזיות שקיבלו ערכים לא מפורשים, התיבות שנרשמו להן שניים או שלושה ערכים, וכן התיבות שערכיהן אנציקלופדיים, מונות בסך הכול 1,284 תיבות, כלומר כ־4.8% מתיבות הקורפוס. בשל פיזורן הרב בין ערכי הקונקורדנצייה ובשל אי־השפעתן על התמונה הכללית המצטיירת ממנה הן לא הושמטו מן הספירה, והן נכללו בכל הספירות שלהן בפרק 4.

- 14 א. כנהוג במפעל המילון ההיסטורי, לא נרשמו ערכיהן של מילות השימוש בראשי מילות התוכן.  
 ב. השוו לפסקה זו את דבריה של שורצולד תשע"ו, בספר זה, על התחלות הנפל וההתחלות הכפולות, וראו את דוגמאותיה.
- 15 הקביעה אילו מילים תקבלנה ערך מפורש ואילו לא היא סובייקטיבית, וכדיקה על ידי אדם אחר או בזמן אחר יכולה להניב תוצאות שונות. מכל מקום, אין לקביעה הזאת השפעה על התמונה הכללית המצטיירת מן הקונקורדנצייה.
- 16 חשוב להעיר כי גם למילות הצירוף יא אללה (×6), מפי ארבעה דוברים בארבעה קבצים) נרשמו אותם הערכים בדיוק. במילון שייכתב על פי הרישום הזה תינתן כמוכח הרעת הן להברל בביצוע הן להברל במשמעות, וראו לדוגמה בן־אמוץ ובן־יהודה תשל"ב (הערכים: יֵה־אֵלֵּלָה, יֵלֵּלָה!) וכן רוזנטל 2005 (הערכים: יָא אֵלֵּלָה, יָאֵלֵּלָה, בהתאמה).

#### 4. ניתוח אוצר המילים בקורפוס

הקורפוס השלם מונה כאמור 26,538 תיבות. יש בו 5,552 צורות, שתויגו ל-2,236 ערכים. מן התיאור המובא לעיל ברור שאין כל טעם להביא את פירוטם המלא, ואולם יש מקום רב להביא סטטיסטיקה נבחרת מן הדברים. זאת אעשה בפרק זה.<sup>17</sup>

##### 4.1 תפוצת הצורות בקורפוס

5,552 הצורות שבקורפוס נבדלות הכול רב מאוד בתפוצתן. הדברים יראו היטב לעין אם נחלק את תיבות הקורפוס לשלישונים על פי תפוצת צורותיהן:

מספר הערכים	מספר הצורות	אחוז מן הצורות	מספר התיבות	אחוז מן התיבות	השלישון
35	36	0.65%	8,884	33.5%	השלישון השלישי
261	417	7.51%	8,778	33.1%	השלישון השני
2,199	5,099	91.84%	8,876	33.4%	השלישון הראשון
<sup>18</sup> 2,236	5,552	100.00%	26,538	100.0%	סך הכול

עם השלישון הראשון (התחתון) נמנות הצורות הנקרות בקורפוס 1-7 פעמים, עם השלישון השלישי (העליון) נמנות הצורות הנקרות בו למעלה מתשעים פעמים, ואילו כל שאר הצורות נמנות עם השלישון השני. השלישון הראשון מלמדנו כי כ-92% מן הצורות בקורפוס מייצגות רק שלישי מתבותיו, ואילו השלישון השלישי מלמדנו כי אחוז דומה מן התיבות מיוצג על ידי פחות מאחוז אחד מן הצורות! הדבר מעיד על משקלן הרב של צורות אלו בלשון הדיבור. להלן פירוטן ופירוט תפוצתן:<sup>19</sup>

17 לדברים המובאים כאן עשויה להיות השתמעות פדגוגית רבה על הוראת העברית הן כלשון אם הן כשפה זרה.

18 סיכום השלישונים נותן את המספר 2,495, הגבוה ממספר הערכים – 2,236. הסיבה לכך היא שצורות מן הערכים הנפוצים במאגר מופיעות ביותר משלישון אחד, ולפיכך ערכים אלו נספרים כאן פעמיים ושלוש. לדוגמה, הערך אָת נספר כאן שלוש פעמים: הצורה את נמנית עם השלישון העליון, הצורות אותו, אותה ועוד נמנות עם השלישון התיכון, ואילו הצורות באותו, באותה, אותנו ועוד נמנות עם השלישון התחתון.

19 צורות וערכים שתפוצתם שווה דורגו דירוג שווה הן בטבלה הזאת הן בטבלאות שלהלן. הצורות והערכים שאחריהם דורגו על פי מיקומם בסדר השוטף של הפריטים, כמודגם בטבלה.

הדירוג	הצורה	הערך	ספירת התיבות	אחוז מן התיבות
1	לא	לֹא	1,058	4.0%
2	זה	זֶה	998	3.8%
3	אני	אֲנִי	630	2.4%
4	אה <sup>20</sup>	אה [קול או מילה מלאכותית]	497	1.9%
5	מה	מָה	406	1.5%
6	את	אֶת	393	1.5%
7	כן	כֵּן	330	1.2%
8	לי	לִי	264	1.0%
9	אז	אֲז	259	1.0%
9	הוא	הוא	259	1.0%
11	יש	יֵשׁ	241	0.9%
12	אבל	אֲבָל	231	0.9%
13	אתה	אַתָּה	199	0.7%
14	לך	לְךָ	186	0.7%
15	היא	הִיא	184	0.7%
16	של	שֶׁל	182	0.7%
17	על	עַל	179	0.7%
18	גם	גַּם	167	0.6%
19	את	אֶת	164	0.6%
20	כל	כָּל	153	0.6%
21	אם	אִם	150	0.6%
22	ה	ה	149	0.6%
22	היה	היּי וקל, עבר (= הִיּה)	149	0.6%
24	עם	עִם	144	0.5%
25	פה	פֶּה	137	0.5%
26	או	אוּ	128	0.5%
27	איזה	אֵיזָה	122	0.5%
28	הם	הֵם	113	0.4%
29	ש	שֶׁ	109	0.4%
30	כי	כִּי	108	0.4%
31	ו	וּ	107	0.4%

20 הצורה אה מייצגת כאן למעשה שני קולות, מילת הקריאה אָה וקול ההיסוס אַה, שבמחשבה שלאחר מעשה ראוי היה להפרידם זה מזה.

הדירוג	הצורה	הערך	ספירת התיבות	אחוז מן התיבות
32	אֵין	אֵין	104	0.4%
33	עכשיו	עֶכְשָׁו	99	0.4%
34	למה	לָמָּה	98	0.4%
35	נכון	כּוֹן וּנְפַעֵל, בינוני (= נְכוֹן) <sup>21</sup>	95	0.4%
36	שם	שָׁם	92	0.3%
סך הכול	36	35	8,884	33.5%

#### 4.2 תפוצת הערכים בקורפוס

בדיקת תפוצת הערכים מעלה תמונה בעלת אופי דומה: 13,378 תיבות, שהן כמחצית מתיבות הקורפוס, ובהן 533 צורות, שהן פחות מ-10% מן הצורות, תויגו ל-52 ערכים, שהם כ-2.3% בלבד מכלל הערכים. ברומה לתפוצת הצורות, הדבר מעיד על משקלם הרב של ערכים אלו בלשון הדיבור. להלן אפוא פירוט הערכים הנפוצים בקורפוס ופירוט תפוצתם:

הדירוג	הערך	ספירת הצורות	אחוז מן הצורות	ספירת התיבות	אחוז מן התיבות
1	זֶה	14 <sup>22</sup>	0.25%	1,312	4.9%
2	לֹא	6 <sup>23</sup>	0.11%	1,118	4.2%
3	אֲנִי	10	0.18%	771	2.9%
4	לְ	12	0.22%	729	2.7%
5	אֵת	18	0.32%	580	2.2%
6	אה וקול או מילה מלאכותית]	1	0.02%	497	1.9%
7	היִי [קל] (= הִיָּה)	33	0.59%	444	1.7%
8	מָה	8	0.14%	432	1.6%
9	שָׁל	16	0.29%	374	1.4%
10	כֵּן	6	0.11%	345	1.3%
11	הוּא	7	0.13%	334	1.3%
12	אָז	5	0.09%	293	1.1%

21 ראו לעיל, פרק 2.

22 אלו חמש הצורות הנפוצות בערך: זֶה (998×; ראו גם לעיל, פרק 4.1), הִיָּה (80×), כֵּן (65×), וְזֶה (64×), שָׁזָה (52×).

23 אלו כל צורות הערך: לֹא (1,058×; ראו גם לעיל, פרק 4.1), שָׁלֹא (28×), וְלֹא (27×), הִלֹא (2×), לֹלֹא (2×), וְשָׁלֹא (1×).



הדירוג	הערך	ספירת הצורות	אחוז מן הצורות	ספירת התיבות	אחוז התיבות
13	כל	20	0.36%	292	1.1%
14	יש	3	0.05%	288	1.1%
15	אמר וקלן (= אָמַר)	29	0.52%	273	1.0%
16	אָתָּה	6	0.11%	254	1.0%
17	היא	6	0.11%	241	0.9%
18	אָבֵל	2	0.04%	232	0.9%
19	על	14	0.25%	224	0.8%
20	עשי וקלן (= עָשָׂה)	31	0.56%	205	0.8%
21	גם	4	0.07%	186	0.7%
21	ירע וקלן (= יָרַע)	16	0.29%	186	0.7%
23	אָתְּ	4	0.07%	184	0.7%
24	אָם	5	0.09%	178	0.7%
25	ה	6	0.11%	170	0.6%
26	פה	5	0.09%	168	0.6%
27	הם	5	0.09%	158	0.6%
28	עם	3	0.05%	146	0.6%
29	ראי וקלן (= רָאָה)	22	0.40%	142	0.5%
30	אִיזָה	6	0.11%	139	0.5%
30	בוא וקלן (= בָּא)	25	0.45%	139	0.5%
32	יום	15	0.27%	132	0.5%
32	כי	3	0.05%	132	0.5%
34	רצי וקלן (= רָצָה)	13	0.23%	131	0.5%
35	יכל וקלן (= יָכַל)	16	0.29%	129	0.5%
35	שָׁ	8	0.14%	129	0.5%
37	או	1	0.02%	128	0.5%
38	שָׁם	7	0.13%	123	0.5%
39	וְ	5	0.09%	117	0.4%
40	אֵין	5	0.09%	115	0.4%
41	דָּבַר	14	0.25%	114	0.4%
42	צָרִיךְ	6	0.11%	110	0.4%
43	עָכְשׁוּ	6	0.11%	107	0.4%
44	כּוֹן ונפעלן (= נָכוֹן)	7	0.13%	106	0.4%

הדירוג	הערך	ספירת הצורות	אחוז מן הצורות	ספירת התיבות	אחוז מן התיבות
45	הלך וקלן (= הֶלֶךְ)	25	0.45%	105	0.4%
46	לָמָּה	2	0.04%	99	0.4%
47	עוֹד	6	0.11%	97	0.4%
48	זאת	7	0.13%	96	0.4%
48	טוב	5	0.09%	96	0.4%
50	אַחַר	5	0.09%	94	0.4%
51	חשב וקלן (= חָשַׁב)	18	0.32%	93	0.4%
52	בְּ	11	0.20%	91	0.3%
<b>סך הכול</b>	<b>52</b>	<b>533</b>	<b>9.60%</b>	<b>13,378</b>	<b>50.4%</b>

### 4.3 מערכת הפועל בקורפוס

מערכת הפועל בקורפוס דורשת תיאור נפרד. פועלי הקורפוס מונים 5,904 תיבות, שהתפלגו בין 2,196 צורות ותויגו ל-699 ערכים. כאמור לעיל (פרק 2), צורות שם הפעולה תויגו אף הן תחת מערכת הפועל. צורות אלו מונות 428 תיבות, שהן כ-7.2% מכלל הפעלים, והן תויגו ל-150 ערכים. ב-72 מְעַרְכִים אלו תועדו שמות פעולה בלבד, ללא זמני פועל אחרים: דְּיוֹק (דוק [פיעל]  $\times 31$ ); רק בצורה בְּדִיוֹק); דִּירָה (דור וקלן  $\times 18$ ); בְּחִינָה (בחן וקלן  $\times 18$ ); בעיקר בצורה מְבַחֵינָה) ועוד. משקלם המועט יחסית של שמות הפעולה ופיזורם הלקסיקלי הרב אינו פוגע בתמונה הכללית המצטיירת ממערכת הפועל, ולפיכך לא אפריד אותם משאר חלקי המערכת בתיאורים שלהלן.

#### 4.3.1 תפוצת צורות הפועל בקורפוס

בדומה להבדלים בתפוצת הצורות הכללית בקורפוס, גם צורות הפועל נבדלות הבהל רב מאוד בתפוצתן. הדברים ייראו היטב לעין אם נחלק את תיבות הקורפוס פחות או יותר לרביעונים על פי תפוצת צורותיהן:

מספר הערכים הפועליים	אחוז מצורות הפועל	מספר הצורות	אחוז מתיבות הפועל	מספר התיבות	הרביעון
18	1.6%	35	23.7%	1,402	הרביעון הרביעי
103	9.2%	203	28.7%	1,692	הרביעון השלישי
279	25.2%	553	23.8%	1,405	הרביעון השני
613	64.0%	1,405	23.8%	1,405	הרביעון הראשון
<sup>24</sup> 699	100.0%	2,196	100.0%	5,904	סך הכול

עם הרביעון הראשון (התחתון) נמנות צורות הפועל הנקרות בקורפוס פעם אחת בלבד, עם הרביעון השני נמנות הצורות הנקרות בו 2-4 פעמים, עם הרביעון השלישי נמנות הצורות הנקרות בקורפוס 5-19 פעמים, ואילו הצורות הנקרות בו 20 פעמים ומעלה נמנות עם הרביעון הרביעי (העליון). הרביעון הראשון מלמדנו כי כ-64% מצורות הפועל בקורפוס (כלומר כמעט שני שלישים מהן!) מייצגות רק כרבע מתיבות הפועל שבו, ואילו הרביעון הרביעי מלמדנו כי אחוז דומה מתיבות הפועל מיוצג על ידי פחות משני אחוזים מצורותיו! גם כאן הדבר מעיד על משקלן הרב של צורות אלו בלשון הדיבור. להלן פירוטן ופירוט תפוצתן:

הדירוג	הצורה	הערך	הזמן הדקדוקי	ספירת התיבות	אחוז מתיבות הפועל
1	היה	היי וקלן (= הִיָּה)	עבר	149	2.5%
2	נכון	כון ונפעל (= נִכּוֹן)	בינוני	95	1.6%
3	רוצה	רצי וקלן (= רִצָּה)	בינוני	81	1.4%
4	יודע	ידע וקלן (= יָדַע)	בינוני	80	1.4%
5	יכול	יכל וקלן (= יָכַל)	בינוני פֶּעַל	71	1.2%
6	יודעת	ידע וקלן (= יָדַע)	בינוני	68	1.2%
7	אומרת	אמר וקלן (= אָמַר)	בינוני	60	1.0%
8	להיות	היי וקלן (= הִיָּה)	מקור	55	0.9%
9	אמרתי	אמר וקלן (= אָמַר)	עבר	49	0.8%
10	אומר	אמר וקלן (= אָמַר)	בינוני	47	0.8%
11	הייתי	היי וקלן (= הִיָּה)	עבר	46	0.8%

24 סיכום הרביעונים נותן את המספר 1,013, הגבוה ממספר הערכים – 699. גם כאן הסיבה לכך היא שצורות מערכי הפועל הנפוצים במאגר מופיעות ביותר מרביעון אחד, והשוו להערה 18 לעיל.

הדירוג	הצורה	הערך	הזמן הרדדוקי	ספירת התיבות	אחוז מתיבות הפועל
12	עושה	עשי וקל] (= עֲשֶׂה)	בינוני	44	0.7%
13	לעשות	עשי וקל] (= עֲשֶׂה)	מקור	38	0.6%
14	בדיוק	דוק ופיעל] (= דִּיק)	שם פעולה	31	0.5%
15	משנה	שני ופיעל] (= שָׁנָה)	בינוני	28	0.5%
16	נראה	ראי ונפעל] (= נִרְאָה)	בינוני	27	0.5%
17	בוא	בוא וקל] (= בָּא)	ציווי	26	0.4%
17	יהיה	היי וקל] (= הִיָּה)	עתיד	26	0.4%
17	חושב	חשב וקל] (= חָשַׁב)	בינוני	26	0.4%
17	פשוט	פשט וקל] (= פָּשַׁט)	בינוני פֶּעוּל	26	0.4%
21	היו	היי וקל] (= הִיָּה)	עבר	25	0.4%
21	תראי	ראי וקל] (= רְאִי)	עתיד	25	0.4%
23	הייתה	היי וקל] (= הִיָּה)	עבר	24	0.4%
23	יכולה	יכל וקל] (= יָכַל)	בינוני פֶּעַל	24	0.4%
23	נורא	ירא ונפעל] (= נִרְאָה)	בינוני	24	0.4%
26	להגיד	נגד והפעיל] (= הִגִּיד)	מקור	22	0.4%
26	רואה	ראי וקל] (= רְאִי)	בינוני	22	0.4%
28	מלא	מלא וקל] (= מָלֵא)	בינוני פֶּעַל	21	0.4%
28	עשיתי	עשי וקל] (= עֲשֶׂה)	עבר	21	0.4%
28	תראה	ראי וקל] (= רְאִי)	עתיד	21	0.4%
31	אמרה	אמר וקל] (= אָמַר)	עבר	20	0.3%
31	בא	בוא וקל] (= בָּא)	בינוני פֶּעַל	20	0.3%
31	מבין	בון והפעיל] (= הִבִּין)	בינוני	20	0.3%
31	היית	היי וקל] (= הִיָּה)	עבר	20	0.3%
31	רוצים	רצי וקל] (= רְצִי)	בינוני	20	0.3%
<b>סך הכול</b>	<b>35</b>	<b>18</b>	-	<b>1,402</b>	<b>23.7%</b>

#### 4.3.2 תפוצת ערכי הפועל בקורפוס

בדיקת תפוצת ערכי הפועל מעלה תמונה בעלת אופי דומה לתפוצת הערכים הכללית: 31 ערכים, שהם כ-4.4% בלבד מכלל ערכי הפועל, מונים 2,969 תיבות, שהן כמחצית מתיבות הפועל בקורפוס. את נתוני תפוצתם של ערכים אלה עֲמַתִּי עם נתוני תפוצתם של 498 פעלים, שמסר לי חבר הסדנה, הפרופ' שמואל בולוצקי,

על פי קורפוס שברשותו, ובו כ-5,300,000 תיבות מלשון העיתונות.<sup>25</sup> להלן אפוא פירוט ערכי הפועל הנפוצים בקורפוס ופירוט תפוצתם:

הדירוג	הערך	ספירת הצורות	אחוז מצורות הפועל	ספירת התיבות	אחוז מתיבות הפועל	הדירוג ברשימת בולוצקי
1	היי וקל' (= הִיָּה)	33	1.5%	444	7.5%	1
2	אמר וקל' (= אָמַר)	29	1.3%	273	4.6%	2
3	עשי וקל' (= עָשָׂה)	31	1.4%	205	3.5%	4
4	ידע וקל' (= יָדַע)	16	0.7%	186	3.2%	8
5	ראי וקל' (= רָאָה)	22	1.0%	142	2.4%	7
6	בוא וקל' (= בָּא)	25	1.1%	139	2.4%	6
7	רצי וקל' (= רִצָּה)	13	0.6%	131	2.2%	9
8	יכל וקל' (= יָכַל)	16	0.7%	129	2.2%	3
9	כון [נפעל] (= כִּוֵּן)	7	0.3%	106	1.8%	-
10	הלך וקל' (= הִלָּךְ)	25	1.1%	105	1.8%	17
11	חשב וקל' (= חָשַׁב)	18	0.8%	93	1.6%	16
12	נגד והפעיל' (= נִגַּד) <sup>26</sup>	14	0.6%	85	1.4%	106
13	לקח וקל' (= לָקַח)	19	0.9%	68	1.2%	43
14	נתן וקל' (= נָתַן)	25	1.1%	60	1.0%	18
15	בון והפעיל' (= הִבִּין)	14	0.6%	59	1.0%	21
16	נגע והפעיל' (= הִגִּיעַ)	18	0.8%	56	0.9%	5
17	נסע וקל' (= נָסַע)	21	1.0%	52	0.9%	111
17	קני וקל' (= קָנָה)	13	0.6%	52	0.9%	97
19	יצא וקל' (= יָצָא)	22	1.0%	51	0.9%	13
19	שום וקל' (= שָׁם)	14	0.6%	51	0.9%	188
19	שמע וקל' (= שָׁמַע)	12	0.5%	51	0.9%	50
22	דבר ופיעל' (= דִּבֶּר)	17	0.8%	49	0.8%	23
22	תחל והפעיל' (= הִתְחִיל)	22	1.0%	49	0.8%	24

25 להלן: רשימת בולוצקי. אני מנצל מקום זה כדי להביע את תודתי לפרופ' בולוצקי על העמדת ממצאיו לרשותי טרם פרסומם על ידו. עיבוד הקורפוס נעשה בשני שלבים: תחילה מיוני תיבותיו לצורותיהן מיון אוטומטי בתכנת הקונקורדנציה WordSmith (<http://www.lexically.net/wordsmith/>), גרסה 4; לאחר מכן עובדו צורות הפועל ומווגו לערכיהן בעיבוד ידני.

26 בפועל זה הצורות העיקריות הן צורות עתיד (אָגִיד וכיו"ב) ומקור (לְהִגִּיד).

הדירוג	הערך	ספירת הצורות	אחוז מצורות הפועל	ספירת התיבות	אחוז מתיבות הפועל	הדירוג ברשימת בולוצקי
24	חזר [קל] (= חִזַּר)	18	0.8%	48	0.8%	20
25	נסי [פיעל] (= נָסָה)	22	1.0%	44	0.7%	22
26	קרי [קל] (= קָרָה)	9	0.4%	43	0.7%	31
27	ראי [נפעל] (= נִרְאָה)	8	0.4%	41	0.7%	11
28	בוא [הפעיל] (= הִבִּיא)	16	0.7%	40	0.7%	36
29	אכל [קל] (= אָכַל)	15	0.7%	39	0.7%	137
29	עבר [קל] (= עָבַר)	17	0.8%	39	0.7%	12
29	שני [פיעל] (= שָׁנָה)	4	0.2%	39	0.7%	72
<b>סך הכול</b>	<b>31</b>	<b>555</b>	<b>25.3%</b>	<b>2,969</b>	<b>50.3%</b>	-

הדמיון במיקומם היחסי של הפעלים בשתי הרשימות ברבע העליון של הטבלה בולט לעין: שני הפעלים הראשונים – הָיָה וְאָמַר – שווים במקומותיהם בשתייהן, ואילו ששת הפעלים הבאים בקורפוס זהים לשישה מן הפעלים שבמקומות השלישי עד התשיעי ברשימת בולוצקי, אך לא באותו הסדר. הפועל החמישי בתפוצתו ברשימת בולוצקי – הִגִּיעַ – נמצא רק במקום השישה-עשר בקורפוס, ואילו הערך נִכּוּן, התשיעי בתפוצתו בקורפוס, אינו נמנה עם ערכי הפועל ברשימת בולוצקי. החל במקום העשירי הרשימות מתרחקות זו מזו ואינן בנות השוואה.

### 4.3.3 תפוצת בנייני הפועל בקורפוס

גם בדיקת תפוצת בנייני הפועל מעלה דמיון בין הקורפוס לבין רשימת בולוצקי, הן בתפוצת הערכים הן בתפוצת התיבות.

להלן טבלה המשווה את נתוני תפוצת הערכים בין בנייני הפועל בשני הקורפוסים:

הדירוג	הבניין	הקורפוס	אחוז מערכי הפועל בקורפוס	רשימת בולוצקי	אחוז מערכי הפועל ברשימת בולוצקי
1	קל	242	34.6%	166	33.3%
2	פיעל	161	23.0%	106	21.3%
3	הפעיל	110	15.7%	100	20.1%
4	התפעל	60	8.6%	41	8.2%
5	נפעל	58	8.3%	63	12.7%

הדירוג	הבניין	הקורפוס	אחוז מערכי הפועל בקורפוס	רשימת בולוצקי	אחוז מערכי הפועל ברשימת בולוצקי
6	פועל	42	6.0%	7	1.4%
7	הופעל	18	2.6%	10	2.0%
8	התפולל	6	0.9%	2	0.4%
9	פולל	2	0.3%	3	0.6%
סך הכול	9	699	100.0%	498	100.0%

סדר תפוצתם של שלושת הבניינים הראשונים – קל, פיעל, הפעיל – שווה בשני הקורפוסים, ואף התפלגויותיהם קרובות זו לזו. אף התפלגויותיהם של שאר הבניינים קרובות זו לזו, למעט התפלגותו של בניין פועל. את משקלם הרב יחסית בקורפוס של הבניינים פועל והופעל יש לייחס לתיוג צורות הבינוני שלהם כפעלים: בקורפוס תועדו רק ארבע צורות עבר ועתיד בבניינים אלו, ובשלושה ערכים בלבד: חָקַק, נָתַב, חָחַף.<sup>27</sup> בבניינים התפולל ופולל נרשמו רק הערכים האלה: התמוּדָד (×9), הסתוּבָב (×9), התמוּטָט (×3), התכוּנָן (×2), התהוּלָל (×1), התעוּרָר (×1), סוּבָב (×1), שוּחָח (×1).

להלן טבלה המשווה את נתוני תפוצת התיבות בין בנייני הפועל בשני הקורפוסים:

הדירוג	הבניין	הקורפוס	אחוז מתיבות הפועל בקורפוס	רשימת בולוצקי	אחוז מתיבות הפועל ברשימת בולוצקי
1	קל	3,733	63.23%	175,038	54.39%
2	הפעיל	789	13.36%	58,688	18.24%
3	פיעל	730	12.36%	42,939	13.34%
4	נפעל	356	6.03%	29,066	9.03%
5	התפעל	141	2.39%	10,591	3.29%
6	פועל	96	1.63%	3,289	1.02%
7	הופעל	32	0.54%	1,325	0.41%
8	התפולל	25	0.42%	133	0.04%
9	פולל	2	0.03%	745	0.23%
סך הכול	9	5,904	100.00%	321,814	100.00%

קל לראות כי סדר תפוצת הבניינים שווה ברובו בשתי הרשימות: קל, הפעיל, פיעל, נפעל, התפעל, פועל, הופעל. הבניינים התפולל ופולל מחליפים מקומות בסוף הרשימה, ואילו בניין פולל נעדר ממנה לחלוטין. כאמור (פרק 2), לא נמצאה בקורפוס כל עדות לגזירת פועל שלא בבנייני הפועל המוכרים בדקדוק המסורתי.<sup>28</sup>

#### 4.4 תפוצת כינויי הגוף הפרודים בקורפוס<sup>29</sup>

בקורפוס תועדו רוב כינויי הגוף הפרודים בלשון העברית. ארבעה כינויי גוף נעדרו ממנו: אָנוּכִי, אָנוּ, אֶתָּן, הֶן. במקרים אחדים השתמשו הדוברים בכינוי הֶם במקומות שעל פי הדקדוק הקלסי נדרש בהם התאם נקבה, כלומר הכינוי הֶן. לדוגמה:  
 א. הם <ההחלטות. ד"ר> לא נקבעות | לא על ידי לא על ידך | [P931];  
 ב. ברור שאת צריכה להרגיש בבית שלך || את לא רוצה את התמונות האלה \ הם יהיו במחסן || אין לי שום בעיה || [Y34].

להלן התפוצה המלאה של כינויי הגוף הפרודים בקורפוס:

הדירוג	הכינוי	הספירה	אחוז מכינויי הגוף הפרודים
1	אָנִי	771	37.2%
2	הוּא	334	16.1%
3	אֶתָּה	254	12.3%
4	הִיא	241	11.6%
5	אֶתָּ	184	8.9%
6	הֶם	158	7.6%
7	אֶנְחֵנוּ	82	4.0%
8	אֶתָּם	47	2.3%
סך הכול	8	2,071	100.0%

28 השוו את הממצאים בפרק זה לממצאיו של מירקין תשל"ט: 135-138. מירקין מחלק את שבעת בנייני הפועל הראשונים שברשימה לשלוש קבוצות לפי סדר תפוצתם: במקומות 1-3 הבניינים ה"פעילים" (קל, פיעל, הפעיל); אחריהם, במקומות 4-5, הבניינים ה"חזורים"-סבילים" (נפעל, התפעל); ולבסוף, במקומות 6-7, הבניינים ה"סבילים" (פועל, הופעל).

29 השוו לפרק זה את מאמרה של כהן תשע"ו, בספר זה.



## 5. סיכום

הבאתי כאן נתונים ראשוניים פרי ניתוח סטטיסטי של אוצר המילים של קורפוס הסדנה.

הבחירה במתכונת מפעל המילון ההיסטורי לניתוח הקורפוס עשויה להיראות תמוהה עד אבסורדית, שכן המפעל עוסק בניתוחם של טקסטים כתובים, חלקם בני יותר מאלפיים שנה, ואילו הקורפוס הוא קורפוס דבור שהוקלט בתחילת המאה העשרים ואחת: הריחוק בזמן והשינוי במדיום עשויים היו לכאורה לעשות את המלאכה בלתי אפשרית. ואולם עצם כתיבת המאמר והצגתו כאן מוכיחים כי הדבר אפשרי, וזאת בזכות גורמים אחדים.

הראשון שבהם הוא ללא ספק האחדות המורפולוגית של הלשון העברית, שאפשרה עד כה את הצגת כל רובדי הלשון שלה במילון אחד, דוגמת מילונייהם של בן-יהודה, גור ואבן-שושן, וכן מפעל המילון ההיסטורי ללשון העברית (בן-יהודה 1908-1959, גרזובסקי תרצ"ה-תרצ"ז, אבן-שושן תש"ח-תשי"ב, בן-חיים תשי"ט). ליד האחדות המורפולוגית של הלשון העברית אפשר להציב גם את כתיבה המקובל, שעל אף כל הגוונים שבו, שימר את אחדות ייצוגה בכתב.<sup>30</sup>

האחדות המורפולוגית והכתיב המקובל של הלשון העברית הם שאפשרו את תכתובן של הקלטות הקורפוס כפי שתוכתבו. אמנם התכתוב כרוך באבדנם של גוונים רבים בדיבור אשר אינם באים לידי ביטוי בכתב, ואולם אין מנוס מאבדן זה בכל דיון כתוב בלשון הדיבור. כשם שהנוסח המתוכתב הוא מעין תְּכָלִיל (פרטיטורה) המאפשרת לקוראיה לנסות לשחזר את ביצועם המקורי של הדברים, כן הוא כשר גם להיות בסיס לניתוח אוצר המילים שלהם, כפי שנעשה במאמר זה.<sup>31</sup>

לכל אלו סייע גורם נוסף, והוא תכונתו של מאגר המילון ההיסטורי, שאין בו תלות בין צורתה של התיבה לבין ערכה המילוני. תכונה חשובה זו של המאגר מאפשרת את קביעת ערכה התקני של כל תיבה המופיעה בו, תהא צורתה אשר תהא, ויהא מקורה אשר יהא: צורה משובשת בכתב היד, צורה משובשת בחיבור מודפס או צורה "משובשת" בשל ביצוע "שגוי" בדיבור (או בייצוגו הספרותי). כמו

30 כידוע, הסדרתו המלאה של הכתיב העברי לא באה לעולם אלא בתחילת המאה העשרים, עם פעולתו של דוד ילין (תרס"ה), ואחריו – עם קביעותיהם של ועד הלשון העברית והאקדמיה ללשון העברית.

31 השוו לפסקה זו את דבריו של יזרעאל תשס"ב: "המילים המופיעות בתמליל יהיו אפוא תרגום של יחידות הרצף הדיבורי ללשון הכתב והתאמתן לה" (ההדגשה שלי. ד"ר). השקפה דומה הציג יזרעאל במאמרו מתש"ע. כן ראו את דבריה של שורצולד תשע"ו, בספר זה, על בעיית הכתיב במילון.

כן המאגר מאפשר לציין את הערכת מנתח התיבה אם השיבוש הוא מקרי (כלומר ברמת ה־parole בלבד) או שהוא מעיד על שינוי מערכתי (כלומר ברמת ה־langue).

במאגר המילון ההיסטורי נמצא דוגמאות מעין אלו:<sup>32</sup>

א. כיון שראה הירשילי את משה־יוסי נצב עליו, מיד נצנצה בו רוח עצה ופתח ואמר לו בלשון קונד"ס זו, שהנערים השוכבים משתמשים בה בשעת שמחה והוללות.

– סושה־סוסי, משה־יוסי! סאתה סוצה, אתה רוצה, סהיות סולה, להיות חולה?  
– סוא, לא, סס וסלום, חס ושלום! – החזיר לו משה־יוסי אף הוא בלשון־שוכבים זו.

(מנדלי מוכר ספרים תרנ"ז: 309; ועוד פעם בסמוך למקום זה);

ב. ומאטי פונפי הופך פתאום פרצופו החרומפי לחתן־הסעודה, מאַנפֿך ושואל:

– מְבִנְקֶשׁ אַנְתָּה אַנְשָׁה, בחור?

(ביאליק 1909: 490 | = פרק ח); ועוד פעמיים בסמוך למקום זה);

ג. – 'מ'ז'תומרת! – התרגש נחום – תותחים או לא תותחים!

(ס' יזהר תשי"ח: 365; ועוד פעם אחת בעמ' 331 בפי אותו דובר).

התיבות המודגשות בשתי המובאות הראשונות קיבלו את ערכיהן התקניים בצירוף הציון לייחוד בהגיינתן (שאינו מעיד במקרים אלו על שינוי בתפיסת הצורה): משה [שם פרטי]; יוסי [שם פרטי]; אָתָּה; רצי [קל], בינוני; היי [קל], מקור; חלי [קל], בינוני וגו'. המובאה השלישית מכניסה ממד חדש להתמודדות: בניגוד למנדלי ולביאליק, שיצרו דיבור עברי בפי גיבוריהם, שלא דיברו עברית במציאות, ס' יזהר כבר ניסה לחקות דיבור עברי חי. הצורה מ'ז'תומרת תקבל בכוא היום את ערכיה על פי היחידות המילוניות שמהן היא מורכבת: מָה; זאת; אמר [קל], בינוני, וגם בה יצוין הייחוד בהגיינתה. הגייה דומה הוצגה בקורפוס הנוכחי בתכתוב "מה זאת אומרת" (5×, מפי ארבעה דוברים בארבעה קבצים; לשתי היקרויותיה בקובץ OCh הובא גם תעתיק פונטי: imazto meket), ותיבותיה תויגו לערכים הנ"ל. גם אילו הובאה ההגייה שבקורפוס בתכתוב דוגמת מ'ז'תומרת או mazto meket, לא היה הדבר מעלה או מוריד מן האפשרות להתמודד עמה בכליו של מפעל המילון ההיסטורי, והיא הייתה מתויגת באופן דומה.

נראה אפוא כי כשם שכליו של מפעל המילון ההיסטורי כשרים לניתוחם של כתבים עבריים עתיקים, כך הם כשרים לניתוחה של העברית המדוברת בתחילת

32 אני מבקש להביע כאן את תודתי לאקדמיה ללשון העברית על הרשאתה להשתמש בחומרם אלו במאמרי.

המאה העשרים ואחת. ואולם מן הדיון בפרקים הקודמים נראה גם כי לשם ניתוח מדויק יותר של הדברים נדרשות דרישות אחדות:

- א. עיון נוסף במעמדם של צורות הבינוני<sup>33</sup> ועיון במעמדם של שמות הפעולה;
- ב. הגדרה מדויקת יותר של היחס שבין המילה המדוברת למילה הכתובה, ובכלל זה גם הגדרת מעמדם של המיליות מש"ה וכל"ב;<sup>34</sup>
- ג. הגדרה מדויקת יותר של הטיפול בכל אותם מבעים שאינם מגיעים לכדי מילים שלמות (התחלות הנפל וההתחלות כפולות).

אחדות מבעיות אלו אינן ייחודיות לניתוחה של לשון הדיבור, ומכל מקום, הדברים צריכים עיון נוסף.<sup>35</sup>

ואולם למרות המגבלות שציינתי, עדיין אפשר ללמוד מן הניתוח כמה דברים על אוצר המילים של העברית:

- א. על תפוצתן הרבה מאוד של צורות מסוימות בהשוואה לרוב הצורות האחרות;
- ב. על תפוצתם הרבה מאוד של ערכים מסוימים בהשוואה לרוב הערכים האחרים;
- ג. על ההבדל בתפוצת הפעלים בין לשון הדיבור לבין לשון הכתב, שיוצגה כאן על ידי רשימת בולוצקי. נראה לי שההבדל הזה איננו רק תלוי מדיום (דיבור וכתב), אלא גם תלוי סוגה: מעניין לדעת אילו הבדלים יימצאו בין לשון הדיבור לבין סוגות שונות בלשון הכתב.

כפי שהראיתי בגוף המאמר, לחלק מן הממצאים עשויות להיות השתמעויות חשובות כבר עתה, בעיקר בתחום הפרדגוגי, והדברים צריכים עיון נוסף גם כאן.

## ביבליוגרפיה

- אבן-שושן א'. תש"ח-תשי"ב. מלון חדש. ירושלים: קרית-ספר.  
 בהט ש'. תשנ"א. "הסינר, המזרן והנמל". לשוננו לעם מב, ב. 71-72.  
 בהט ש' ומישור מ'. תשנ"ה. מילון ההווה – מילון שימושי לעברית התקנית. תל-אביב וירושלים: ספרית מעריב; איתאב.  
 בן-אמוץ ד' ובן-יהודה נ'. תשל"ב. מלון עולמי לעברית מדוברת. ירושלים: א. לוינ-אפשטיין.

33 וראו לעיל הערות 6 ו-9.

34 וראו לעיל הערה 14.

35 כן נציין כי גם מפעלים אחרים, דוגמת מיל"ה ו-NMELRC, התמודדו עם ניתוח לשון הדיבור. ואולם השוואת הממצאים המוצגים במאמר הזה עם מפעלים אלו כרוכה הן בהשוואת מבני מסדי הנתונים שלהם זה לזה הן בהשוואת שיטות העבודה בהם זו לזו, והדברים עומדים הרבה מעבר לגבולות המאמר.

בן-חיים ז'. תשי"ט. "לעשייתו של המילון ההיסטורי ללשון העברית של האקדמיה ללשון העברית". לשוננו כג. 102-123.

בן-יהודה א'. 1959-1908. מלון הלשון העברית הישנה והחדשה. ירושלים.

גרוזבסקי י'. תרצ"ה-תרצ"ז. מלון השפה העברית. תל-אביב: דביר.

האקדמיה ללשון עברית. תשכ"ט. המפעל ודרכי עשייתו. ירושלים.

יזרעאל ש'. תשס"ב. "מאגר העברית המדוברת בישראל (מעמ"ד): דוגמאות טקסטים". לשוננו סד. 289-314.

יזרעאל ש'. תש"ע. "מדיבור לתחביר – מתאוריה לתמליל". בתוך: מ' בראשר וח"א כהן (עורכים). משאת אהרן: מחקרים בלשון מוגשים לאהרן דותן. ירושלים: מוסד ביאליק. 590-610.

יזרעאל ש'. תשע"ו. "מסד הנתונים למיזם 'חוקרים עברית מדוברת'". בתוך: ע' גונן (עורכת). חוקרים עברית מדוברת (תעודה כז). תל-אביב: אוניברסיטת תל-אביב. 37-49.

ליין ד'. תרס"ה. המבטא והכתיב בעברית. ירושלים: לונץ.

כהן ס'. תשע"ו. "על ערפל גופיניו": היערכות צייני הגוף ביחידות פרוזודיות פועליות בעברית המדוברת – ממצאים ראשוניים". בתוך: ע' גונן (עורכת). חוקרים עברית מדוברת (תעודה כז). תל-אביב: אוניברסיטת תל-אביב. 317-339.

כוזר ר'. תשנ"ב. "החג"ג – חלק דיבור או עמדה תחבירית?". לשוננו נו. 241-248. מזרן או מזרון? תש"ע.

[http://hebrew-academy.huji.ac.il/sheelot\\_teshuvot/MivharTeshuvot/Pages/25071006.aspx](http://hebrew-academy.huji.ac.il/sheelot_teshuvot/MivharTeshuvot/Pages/25071006.aspx).

מיל"ה. מרכז ידע לתקשוב בשפה העברית.

<http://www.mila.cs.technion.ac.il/heb/index.html>

מירקין ר'. תשל"ט. אוצר המלים בחיבוריו העבריים של ש"י אברמוביץ (מנדלי מוכר ספרים): ניתוח הבחינה המילונית, הרקדוקית והסגנונית בסיוע מחשב. עבודת דוקטור. האוניברסיטה העברית, ירושלים.

מישור מ'. תשמ"ז. "מה בין מילון מדעי למילון שימושי". בלשונות עברית חפ"שית 24. 69-73.

מישור מ'. תשנ"ז. "סוגיות מדעיות בהכנת מילון שימושי". בתוך: מ' בראשר (עורך). פרקים בעברית לתקופותיה: אסופת זיכרון לשושנה בהט (אסופות ומבואות בלשון ב). ירושלים: האקדמיה ללשון העברית. 265-271.

רוזנטל ר'. 2005. מילון הסלנג המקיף. ירושלים: כתר.

שורצולד (רודריג) א' וסוקולוף מ'. תשנ"ב. מילון למונחי בלשונות ודקדוק. אבן-יהודה: רכס. שורצולד (רודריג) א'. תשס"ב. פרקים במורפולוגיה עברית. כרך א. תל-אביב: האוניברסיטה הפתוחה.

שורצולד (רודריג) א'. תשע"ו. "כתיבת מילון לעברית המדוברת". בתוך: ע' גונן (עורכת). חוקרים עברית מדוברת (תעודה כז). תל-אביב: אוניברסיטת תל-אביב. 107-127.

NMELRC. *National Middle East Language Resource Center*. Available at: <http://nmelrc.org/>, <http://hebrewcorpus.nmelrc.org/>.

**המקורות הספרותיים**

ביאליק ח"נ. 1909. מאחורי הגדר. **השלח** כא. 217-226, 325-332, 483-496.  
מנדלי מוכר ספרים. תרנ"ז. בעמק הבכא (פרולוג; חלק ראשון, פרקים א-יב). **השלח** א.  
23-7, 308-319, 509-523.  
ס' יזהר. תשי"ח. **ימי צקלג**. תל-אביב: עם עובד.