

לשוננו לעם

עורכים: ע' איתן ומ' מדן

כרך לז, תשמ"ו



האקדמיה ללשון העברית
מוגן בזכויות יוצרים

המילון ההיסטורי — פיצד הוא נעשה?

לכל מי שעובד במפעל המילון ההיסטורי של האקדמיה ללשון העברית מזדמן לשוחח עם מכריו על עבודתנו, וכל מי שאנו מספרים לו על המילון, שואל אותנו את השאלה הנצחית: "באיזו אות אתם עומדים?". אנחנו מסבירים, שפעת אנחנו רק אוספים חומר, ולא פותבים את המילון ממש. ואז שואלים אותנו: "אם כן, לאיזו אות אתם אוספים חומר עכשיו?"

ובכן, המילון ההיסטורי לא פכה עושים אותו. המילון ההיסטורי ללשון העברית נעשה ע"י עיבוד אוצר המלים של כל תקופה. מכאן, שאיסוף החומר המלא למילון היסטורי פירושו התקנת קונקורדנציות כוללות לכל תקופות הלשון.

הקונקורדנציות המצויות אינן מספקות את צורכנו. ראשית, מספר החיבורים שיש להם קונקורדנציות אינו מרובה — יחסית לכלל החומר הדרוש. יש קונקורדנציות רק למקרא ולמקצת הספרות התלמודית.

הפגם העיקרי של הקונקורדנציות המצויות הוא, שהן מבוססות על הדפוסים. אשר למקרא, מצב הדפוסים אמנם מניח את הדעת, ואין הבדלים מהותיים בין הדפוסים המודרניים ובין

פה"י הקדומים (חוץ מן המגילות המקראיות של מדבר יהודה, שהן בעיה לעצמה). אבל הדפוסים של התושבע"פ, ובכללם אלו שהקונקורדנציות המצויות מבוססות עליהם — אינם ראויים לטיפול מדעי.

אבל הטעם העיקרי, הפוסל את הקונקורדנציות המצויות, הרי הוא חוסר הגמישות לטיפול במחשב. שיתוף המחשב בעשיית המילון מחייב תכנון כולל ומיוחד, ועוד נחזור לעניין הזה אחר כך. יוצא מכאן, שנאלצנו לקבל עלינו את מלאכת הכנת הקונקורדנציות, שעל יסודן ייפתח המילון.

אלא שהכנת קונקורדנציות מחייבת בחירה של מהדורות מהימנות של החיבורים המעובדים. והרי כולנו יודעים, שאין מהדורות טובות. הספרות העברית הקדומה איננה בבחינת ממצא ארכיאולוגי, המשמר את פרצופו המקורי במשך הדורות. הספרות הקדומה לא נמצאה, אלא נמסרה, ובמסירה מיד ליד נטבעו טביעות אצבעותיהם של המווסרים.

נאלצנו אפוא לקבל עלינו גם את התקנת הטקסטים, שעליהם יתבססו הקונקורדנציות.

סיכומו של דבר: קודם שניגש לעצם פתיבת המילון, צריך להיעשות במפעלנו מבצע עצום: הכנת מהדורות והכנת קונקורדנציות. עבודה זו לבדה מתוכננת לפרק זמן של דור תמים.

- ובכן, את עבודתנו אפשר לחלק לשני שלבים:
- א. שלב איסוף החומר, ובו אנו עומדים עכשיו.
 - ב. שלב הפתיבה והעריכה של המילון עצמו.

השלב השני, שלב הפתיבה והעריכה, אינו יכול לצאת לפועל קודם שכל החומר פולו יהיה אגור וערוך בקונקורדנציות. (לאחרונה נעשתה עבודה של ניסיון; פורסמה ב"לשוננו" מו, התשמ"ב.)

השלב הראשון, של האיסוף, פולל עיבוד של חיבורים והכנת קונקורדנציות להם: בשלב הזה כל חיבור הוא יחידה לעצמה, ועם תום הכנת המהדורה של כל חיבור אפשר להכין את הקונקורדנציה שלו. שלב זה נתון עכשיו בעיצומו, בכך אנו עוסקים כבר כ-20 שנה, והפעולות חוזרות חלילה כל פעם שנבחר עוד טקסט לעיבוד:

א. עלינו לבחור את הקורפוס הספרותי שיציג את אוצר המלים של כל תקופה ותקופה. החיבורים של הספרות העתיקה נבחרים מתוך "ספר המקורות" (האקדמיה ללשון העברית, מפעל המילון ההיסטורי, התשפ"ג). הבחירה טעונה אישור של ועדה מיוחדת, הנקראת "מועצת המילון", המורכבת ממומחים בלשון ובתולדות הספרות העברית. הבחירה מתחשבת בגורמים שונים: צריך שיהיה ייצוג לכל סוגי הספרות (הספרות היפה, העיתונות, הספרות המקצועית, ההלכה ועוד), צריך שיהיה ייצוג לסופרים או לחיבורים שיש בהם משום ציון דרך בתולדות הלשון, צריך שיהיה ייצוג לגוּי העברית לפי ארצות ולפי עדות. מן התקופות הקדומות הגיעו לידינו חיבורים שהיקפם ניתן לעיבוד מְמָצָה: המקרא והפתיבות המיוחדות לתקופתו, המגילות הגנוזות, חיבורי התנאים. תקופות אחרות עשירות כל-כך, שאי-אפשר לעבד את כלל היצירה שנכתבה בהן, וגם אין צורך

בכך. העיתונות, למשל, לא תיוצג אלא באחוזו זעיר ביותר, יחסית לשפע המצוי בה.

ב. בחירת המסירה.

לתקופות שלאחר המצאת הדפוס אנחנו בוחרים במהדורה הראשונה של החיבור, שהיא, בד"כ, בת זמנו של המחבר, ושהמחבר עצמו סמך את ידו עליה.

לתקופות שלפני המצאת הדפוס אנו נוקקים לכה"י. אנחנו מזמינים תצלומים של כ"י מן האוספים שבספריות שונות בעולם.

מגמתנו היא להתשמש בכה"י המהימן ביותר לחיבור הנדון, שהוא עפ"ר הקדום ביותר שהגיע לידינו. לגבי חיבורים אחדים נמצאו כ"י משופחים ביותר, "אבות טקסטים". חיבורים אחרים צריכים בדיקה יסודית של כה"י שלהם. יש חיבורים שהועדף להתקנים לפי כ"י שאיננו הקדום ביותר, מפני שזה אינו שלם (כגון כ"י אַרפורט של התוספתא). מפל מקום קיבלנו את העיקרון, שדפוס טוב וכ"י גרוע — כ"י גרוע עדיף.

אבל אף המעולים שבכ"י אינם נקיים מליקויים — ליקויים פיסיים, כגון מקומות קרועים או מטושטשים, וליקויים ענייניים, כגון טעויות אנוש של סופרים (השמטות וכיו"ב). לכן, בצד כ"י המשמש בסיס למהדורה, אנו נעזרים בכ"י אחר (או בכ"י אחרים וכן בדפוסים ראשונים) של החיבור הנדון. אנחנו מוסיפים פיסוק, אבל לא מתערבים בנוסח אלא במקומות מעטים, ובכל מקרה שאנו משנים מנוסח המקור אנו מציינים בהתקנה עצמה מה היה הנוסח המקורי.

בספרות החדשה הבעיה חמורה הרבה פחות. כאן אנחנו

מקיימים את הפיסוק שבמקור, ומידת ההתערבות שלנו בנוסח מצטמצמת עפ"י ר' לכדי סימן קריאה בלבד, להבעת תמיהה על הנוסח, ולא יותר.

חשוב להדגיש, שפעבודתנו אנחנו מתארים את כל השינויים שחלו בכה"י במשך תולדותיו. דרפס של פ"י לשמש במשך תקופה ארוכה, ודורות מאוחרים התערבו בנוסח המקורי. אנו מתארים כל מחיקה בכה"י, כל תוספת בין השורות, כל תיקון, כל ספק בקריאה — עד כמה שעיננו מגעת. לא במקום אחד ולא בשני מקומות מצאנו גרסה טובה מתחת למחיקה. ואנו מביעים במפורש את ההבדל בין תיקונים שתיקנו הסופרים בגוף כה"י לבין תיקונים שאנחנו מתקנים בנוסח, במערכת מפורטת של סימנים, סוגי סוגריים שונים וכיו"ב, שלא יהיה שום ספק לקורא, מה מצוי במקור ומה אין מצוי בו.

החיבור המותקן מנוקב על סרט מגנטי, ומודפס הדפסה רגילה לשם הגהה.

ברגע זה, מספר התיבות בטקסטים המותקנים במפעל מסתכם בכשבעה מיליונים, מחציתן מן הספרות העתיקה, ומחציתן מן החדשה. ג. הכנת הקונקורדנציה.

משהוכן הטקסט לעיבוד, רושמים לכל תיבה ותיבה שבו את ערכה המילוני:

"מאימתי" ערכה "אימתי"

"קורין" ערכה "קרא, קל בינוני"

"את" ערכה "את"

"שָׁמַע" ערפּה "שָׁמַע, שֵׁם תפילה"
"בערבין" ערפּה "ערב",

וכן הלאה.

פעולה זו נעשית פשהטקסט מועתק במאונך, תיבה אחת בכל שורה, בטור בימין הדף. הערך נרשם משמאל לטור זה. בתום הרישום המחשב אוסף את המלים שבטור משמאל, מסדר אותן בסדר הא"ב, מצרף להן הַקָּשֶׁר, והרי קוֹנְקוֹרְדַנְצִיה.

ברישום הערך, המחשב ממלא תפקיד של שותף לכל דבר. פידוע, המחשב מסוגל לחזור באופן אוטומאטי על פעולות שנצטוה לבצע פעם אחת. זה ההבדל בין המחשב ובין סתם "גולם"... המחשב "זוכר", כביכול, אם נתקל בתיבה מסוימת או שלא נתקל בה, והוא "זוכר" — אם נתקל בה — מה הערך שנרשם. ובכן, בזמן שהמחשב נתקל באותה תיבה פעם שנייה, ושלישית, ורביעית... — הרי הוא רושם לה את הערך שרשם לה בראשונה. בתנאי — שהתיבה כתובה ממש באותיותיה: אם נתקל המחשב בתיבה "מאימתי", במ"ם השימוש, עדיין אינו יודע איזה ערך עליו לרשום לתיבה "אימתי" בלא מ"ם השימוש; ואם נתקל בה ביו"ד אחרי האל"ף, אינו יודע מה ייטלה בתיבה "אמתי" בלי יו"ד, וכן בשתי יו"דין בסופה, וכיו"ב. בכך המחשב דומה בהחלט ל"גולם" מצוי... רק אם התיבה שנה בפול — אז הוא רושם לה אותו ערך.

כמובן, יש לעבור על הפעולה של המחשב, ולבדוק אחריו, שכן, מלבד התיבות שהמחשב לא ידע לפתור מחוסר ניסיון קודם, יש שלא פינו המחשב לפירוש הנכון של התיבה. המחשב מסוגל לרשום לתיבה "שָׁמַר" את הערך "מַר" (=שָׁ+מר);

לתיבה "אָחר" את הערך "אָחר" וכדומה — הפול לפי מיטב זיכרונו (והדברים אמורים בטקסטים לא־מנוקדים!). במידה רבה צמצמנו את ההחטאות של המחשב ע"י כך, שפשהוא מציע שני ערכים או יותר, השכיח שבהם נכתב ראשון. סידור הערכים לפי שכיחותם נעשה באופן אוטומאטי, והסטטיסטיקה מתעדכנת עדכון שוטף במהלך העבודה.

מכאן יובן, שכל שמספר הטקסטים המעופדים מתרבה, כך הזיכרון של המחשב מעשיר. כיום המחשב מסוגל לפתור לבדו כ־98 אחוז מפלל התיבות של כל חיבור חדש שיועמד לפניו, ומהם כ־90 אחוז פתרון נכון. הרישום האוטומאטי פוטר אותנו מהעתקות והגהות, וגם זה רֶחַח לא־מבוטל.

אין זו הדרך היחידה של טיפול אלקטרוני באוצר מלים. למשל, במפעל השאלות והתשובות של אוניברסיטת בר־אילן פותחה שיטה, הנותנת תוצאות לגמרי לא־רעות, על יסוד טבלת "שורש/צורנים נוספים" (במלים אחרות: מילון ודקדוק), שהזינו בה את המחשב מתחילה, בלא לרשום ערכים. למעשה, רשום שם כל מה ש י י ת כ נ בלשון, והמחשב מציין מה מזה קיים במציאות, וכמובן, היכן.

שתי שיטות אלו למה הן דומות?

אנשי המחשב מספרים על ניסיון, שנעשה בראשית השימוש במחשבים, ללמד את המחשב לשחק שח ודמקה. המתכנתים לא התקשו להכין תכנית, שתלמד את המחשב את המהלכים ואת הפללים של משחק הדמקה, והמחשב מסוגל לנצח כל מתחרה, על־פי הוראות שנקבעו לו מראש. אבל לגבי

משחק השח, נוכחו המתכנתים, שהפללים מורכבים כל-כך, שעדיף, פשוט, לחבר את המחשב לשולחן שחקנים במשך זמן רב, והמחשב ילמד מניסיונם של השחקנים, בלא שיעמיסו עליו תכנית קשה ולא-מציאותית.

הלכנו אפוא בשיטת השימוש בזיכרון המחשב, שהיא המתאימה יותר לטיבם של הטקסטים המעובדים במפעל המילון ולדרישות שלנו, כגון ארגון החומר לפי ערכים, ולא רק לפי צורות התיבות.

המובאות מסודרות בקונקורדנציות שלנו בסדר פרונולוגי, שהוא הסדר המתבקש במילון היסטורי. אבל הזאיל והחומר מועתק על סרט מגנטי, אפשר לקבל אותו מאורגן בסדר אחר, בצד הסדר הפרונולוגי: לפי הפתיבים, הצורות, התיבה השכנה וכיו"ב.

מן הסתם ייערך המילון עצמו בדרך אלקטרונית, והמובאות יועתקו בו העתקה מכאנית, מן המחשב אל הדפוס, בלא צורך בהגהות.

הטקסטים והקונקורדנציות של מפעל המילון ההיסטורי יש להם ערך לעצמם, ואינם שלבי ביניים בלבד.

בידינו קונקורדנציה מודפסת לתקופה שלאחר חתימת המקרא ועד סוף תקופת התנאים. היא נדפסה בעותק אחד, בעצם לצורך הגהה ובדיקה, אבל היא עומדת לשימוש, והיא מסייעת לנו בתכנון התקופות הבאות. תלמידים מן החוג ללשון העברית ומחוגים אחרים, חוקרים ואנשי לשון, באים למפעלנו ומשתמשים בקונקורדנציה ובטקסטים שבידינו. אגב, כבר

אפשר לבשר, שבתכניתנו לפרסם בקרוב מאוד את הקונקורדנציה שלנו, בצורת "מיקרופיש". המעוניין בדוגמה מוחשית לדרך הטיפול בטקסט בעזרת המחשב — וגם היא אינה ממצה את כלל האפשרויות — ידפדף במהדורת משלי בן-סירא, הטקסט, קונקורדנציה וניתוח אוצר המלים (יצאה לאור בשנת התשל"ג).

הזכרתי, שנעשה ניסיון של עריכה בצורת חוברת לדוגמה. החוברת לדוגמה פוללת כמה עשרות ערכים, ערוכים עריכה סופית — כלומר, כפי שהיינו רוצים לראות אותם ערוכים במילון ההיסטורי, לכשייערך. אם נתחשב בדבר, שהחומר עדיין לא נאסף בכללותו — אנחנו עומדים אי-שהוא באמצע איסופו — חוברת כזו לא תוכל להיות אותנטית, כאילו לקוחה מן המילון ההיסטורי עצמו ממש. חוברת זו נועדה לעצב דגם של עריכה למילון, לפי התפיסה שלנו היום, בעיקר בהשראת הפרופ' זאב בן-חיים, העומד בראש המפעל.

בחרנו בשורש ער"ב על כל הסתעפויותיו. השורש ער"ב הוא מן הבעייתיים בעברית (בעיקר מן הפחינה האטימולוגית), ודוקא משום כך בחרנו בו לעשות את הניסיון הזה. לאחר שהבחננו בו עשר הוראות בסיסיות (מהן הפוללות שמות שאינם קשורים בפעלים, כגון עֹרֵב, עֶרְבָה), מנינו בו לא פחות מ-90 ערכים (כל בניין של הפועל הוא ערך לעצמו). לחומר זה הובאו כ-1400 מובאות, מתוך כלל של כ-10000 שבדינו.

המובאות סופקו לנו ע"י המחשב, פשהמלה המצוטטת נתונה בהקשר רחב למדיי (פ-30–60 תיבות); לצורך העריכה העמדנו את ההקשר על האורך שנראה לנו אופטימאלי.

על המלים המשוּיכות לשורש ער"ב, שפבר נרשמו להן ערכים, ויכולנו לקבל מהן מובאות בדרך הנהוגה אצלנו, הוספנו והפקנו בדרך אוטומאטית את המלים משורש זה, המצויות בטקסטים שלא נרשמו להם ערכים, אבל פבר הועתקו במפעל והועלו על סרטים מגנטיים. הדבר נעשה ע"י תכנית קלה יחסית, בשל עובדה מקרית: שורש זה ניתן לזיהוי כמעט חד-משמעי במחשב; שלוש האותיות ע-ר-ב, בסדר זה — גם פשמצטרפות אליהן אותיות אחרות, אבל פששלושתן בסדר זה — נתונות בתיבה אחת רק אם היא מתייחסת לשורש ער"ב; למען האמת נטפלו אליהן עוד מלים מעטות, שאינן שייכות — או שהחלטנו שלא לשייך אותן — לשורש ער"ב: השרשים המורחבים ערב"ב, ערב"ל, המלה עקרב, וכמה שמות פרטיים, כגון שם העיר פעטערבורג (לנינגראד של היום), בפתיב הנהוג אצל סופרים אחדים, עי"ן פנגד תנועת [e] על דרך הפתיב ביידיש. את המלים האלה סילקנו בפעולה ידנית.

מן התקופות שעדיין לא עובדו, או שעובדו שלא בשלמותן, נוספו השלמות מן המילונים וממקורות אחרים, אם נראה לנו שיש בהן צד של חידוש: ערך חדש, משמעות חדשה, תיעוד עתיק למלה המוחזקת צעירה. השלמות אלו נבדקו במקורן, כדרכנו.

התועלת שלנו מן החוברת לדוגמה, הרי היא שהשיטה

הנקוטה אצלנו תעמוד למבחן, ויתבררו הנקודות הטעונות שיפור.

הניסיון הזה יעלה את הבעיות המעשיות, העתידות לעמוד לפני הדור הבא, הדור שייפול בחלקו לערוך את המילון כולו. הדבר מחייב אותנו להפיק את הלקחים הדרושים ולספק ליורשינו את הפלים הנאותים, שלא יודקו לחזור ולהשלים עבודות תשתית המוטלות עלינו, דור המאספים.

חיים אהרנוביץ

הורתם בעברית ולדתם ביידיש

ה. חזק

מה שאירע ל"קונדס" אירע ל"חזק". השם הזה רווח ביידיש בכתיב עברי (במבטא יידי: חוֹזֶק) בכל תפוצות יהודי אשכנז, במשמע של לעג וצחוק (עושה את פלוני ללעג ולצחוק = ער מאכט חזק פון איהם), וגם בכינוי של תם (פניו פנים של חזק = ער האט א פנים פון חזק). והוא גיבור של אנקדוטות מצחיקות.

מוצאה של חזק במשמע הזה לוט בערפל. יש גורסים, שבאה מן התשובה "בחזק יד הוציאנו ה' ממצרים" על שאלת