

ממנוע חיפוש לשלטי דרכים

עוזי אורנן

ממנוע חיפוש לשלטי הדרכים

עוזי אורנן

במאמר זה נטען כי הפתרון הנכון לקשיים המיוחדים שמעמיד הכתב העברי לפני מי שמבקש ליצור כלים חישוביים לעיבוד העברית במחשב, ובמיוחד למי שמבקש לבנות מנוע חיפוש בטקסטים עבריים, טמון במציאת תעתיק נאות למילים העבריות שהמחשב צריך לעבד אותן. נפתח בתיאור תכונותיו של הכתב העברי ובתכונות הנדרשות לתעתיק נאות לעברית, נבדוק את עמדתה של האקדמיה ללשון העברית ואת החלטותיה בעניין התעתיק, ותוך הצגת אפשרויות התעתיק וטיפוסיו נצביע על חולשתה של ההחלטה החדשה על תעתיק שלטי הדרכים.

א. תכונות הכתב העברי

1א. הקשיים המיוחדים של עיבוד העברית במחשב נובעים מארבע תכונות המאפיינות את שיטת הכתיבה העברית:

1. רוב התנועות במילה אינן נכתבות.
 2. מילות השימוש הזעירות נכתבות בלא רווח בינן ובין המילה שאחריהן.
 3. מכפל עיצורים נכתב באות אחת ולא באות כפולה.
 4. האותיות א'ה'ו'י יש שהן משמשות כעיצור ויש שהן משמשות כתנועה.
- תכונות אלה גורמות לרב משמעות של המילה העברית הכתובה. במוצק כל מחרוזת אותיות בטקסט היא בעלת קרוב לשלוש משמעויות, וזה הקושי העיקרי שיש להתמודד איתו בפענוח טקסט בכתב עברי. קשיים אחרים בהקמת מנוע חיפוש, כגון פענוח המורפולוגיה הענפה, גילוי המבנה התחבירי וחיפוש המשמעות המיוחדת של מופעי מילים רבות משמעות, אינם נובעים מן התכונות המיוחדות של הכתב העברי, ולא יידונו כאן.

2א. את הבעיות שמציגה שיטת הכתב העברי לפני מפתחי מנוע חיפוש ניתן לפתור על ידי ייצוג המילים העבריות בשיטת כתיבה שלא יהיו בה התכונות שנמנו לעיל. לכאורה אפשר להשתמש באותיות העבריות עצמן ולהוסיף למילים תוספות שונות שיוכלו לבטל את החסרונות הנ"ל. למשל, להשתמש בכתב עברי מנוקד. אולם דרך זו לקויה מכמה טעמים:

1. טקסט מנוקד דורש כתיבה לא בערוץ אחד, כשכל סימן בא אחרי קודמו. הכתב המנוקד מוצג בשלושה או ארבעה ערוצים, ומערך כזה הרבה יותר מסובך מאשר מערך של ערוץ אחד, שהוא ייצוג נאמן של שטף הדיבור שהמילים אמורות לייצג.

2. הניקוד מכיל סימנים דו משמעיים, כגון נקודה בתוך האות, העשויה לסמן אחת מארבע משמעויות ("דגש קל", "דגש חזק", מפיק, וגם – מה שאינו מצוי בספרי לימוד - עיצור ל' יחיד אחרי תנועת חיריק, כגון במילה "נְקִיָּה" הנקודה איננה מסמנת "דגש חזק"); סימן שווא (בפתיחת הברה או בסיום הברה); הסימן קמץ, העשוי לסמן תנועת A או תנועת O; הסימן פתח שמציין תנועת A אחרי האות שמתחתיה הוא כתוב, אבל לפעמים הוא מסמן תנועת A שמקומה לפני האות שמתחתיה הוא כתוב (פח – חוץ).

3. הניקוד מכיל סימנים כפולים לייצוג אותה מהות לשונית: סימן קמץ או סימן סגול בסוף המילה שלפני האות ה' והאות ה' עצמה מייצגים שניהם את התנועה A או E בסוף המילה (מוֹרָה, מוֹרָה או מוֹרָה).

4. הניקוד אינו מוסר בבהירות את מבנה המילה, ורק התבוננות עמוקה יותר מגלה את המבנה שמאחורי המילה המנוקדת. למשל, לפי צורות נטויות של המילה עֵז ניכר שמילה זו מסיימת בשני עיצורים זהים הבאים יחד בסופה (העיצור ז'), שאם לא כן מה פתאום יש דגש באות ז' בצורת הריבוי עֵזִים, לעומת

הריבוי של עץ - עצים? ומדוע בצורת הנטייה של לב יש דגש חזק ב-ב' ואנו אומרים לבות, לעומת גב - גבים? לפי הניקוד והנטיית ניתן אפוא להסיק מה המבנה של המילה, אבל מבנה זה לא תמיד נמסר בגלוי בשיטת הניקוד, כגון אין סימן של דגש חזק באות האחרונה של המילה לב או עז, והן נכתבות בדיוק כמו גב או עץ. גם פשר הנקודה ב-י' אחרי חיריק אינו מוסבר על פי הניקוד (לעיל 2/2).

3א. לכאורה כבר ב-1957 קבעה האקדמיה "תעתיק מדויק" לעברית. התעתיק שהיא קבעה היה מכוון למילים עבריות מנוקדות, ולכל סימן ניקוד היה תחליף משלו באות לטינית. לפני כחמישים שנה ידעה הבלשנות רק שתי גישות לתעתיק: תעתיק אותיות או תעתיק הגאים. התעתיק של 1957 מיוסד על גישת תעתיק האותיות. התכלית שאליה מכוונים כללי התעתיק הייתה לתעתק טקסט עברי מנוקד. האקדמיה קבעה תחליף נפרד בתעתיק לטיני לכל סימן הקיים בטקסט העברי המנוקד. כך זכה, למשל, תעתיק התנועה פתח לסימן שונה מתעתיק התנועה קמץ, ועוד סימן אחר ניתן לחטף-פתח. וכן קבעו הבדל בין קובוץ לשורוק, שכן האחד "תנועה קטנה" (או "קצרה") והשני תנועה גדולה (או ארוכה). כמובן, כל ההבחנות האלה נעשו על ידי שימוש בסימנים דיאקריטיים. אבל האם "התעתיק המדויק" יוכל לשמש לצורכי העיבוד במחשב? לעיל ראינו כיצד הטקסט העברי המנוקד אינו מוסר בבהירות את מבנה המילה. לא זו בלבד, ההבחנה בין "תנועה קטנה" ו"תנועה גדולה" נעשית מיותרת כשנשתמש בתעתיק המוסר בבהירות את מבנה המילה. משיקולים אלה היה ברור שתהיה זו החלטה שגויה להשתמש ב"תעתיק המדויק" כבסיס לעיבוד טקסטים עבריים במחשב. להשלמת התמונה ראוי להוסיף כאן כי מלכתחילה היה ברור ששום אדם שאינו עוסק במיוחד בניקוד לא יוכל להשתמש בתעתיק מדויק זה. לכן החליטה האקדמיה כבר אז גם על נוסח "פשוט" שבו ויתרו על רוב סימני הניקוד. "התעתיק הפשוט", בנוי אפוא גם הוא על הניקוד, אבל מוותרים בו על סימני הבחן דקים במיוחד בין סימני הניקוד המתועתקים.

4א. האפשרות האחרת, להוסיף סימני עזר אחרים לאותיות הכתב העברי, תכביד מאוד על המפתח ועל הקורא. לא זו בלבד שיהיה קשה מאוד להתרגל לשיטה חדשה, אלא שהכנסת סימנים נוספים עלולה לפגוע בתמונת הכתב העברי, שהוא החומר שלנו המצטבר כבר אלפי שנים. לכן מוטב שנשתמש לצורכי העיבוד בכתב הלטיני, המוכר לכל אדם שלמד אפילו רק כמה כיתות בבית הספר, והוא משמש בהצלחה מאות רבות של שפות בעלות הגיות שונות ומבנים שונים.

5א. לא זו בלבד. זה כמאתיים שנה קיים תעתיק עברי אחד פחות או יותר לאותיות לטיניות, והוא שימש ומשמש את כל מי שנזקק ללימודי עברית בארצות אירופה. כל אחד מן העיצורים העבריים מסומן בו בסימן נפרד משלו. מכיוון שבציבור שלנו כל הבלשנים, המפתחים, מכיני האלגוריתמים, המתכנתים והמשתמשים מכירים כולם את הכתב הזה, ומכיוון שהמחשבים שלנו מכילים את מערכת הסימנים המגוונת של הכתב הלטיני, אין ספק שבחירה זו היא הסבירה ביותר.

לשם עיבוד במחשב ניתן לתעתק את רוב האותיות העבריות ללא היסוסים, כגון $g = \text{ג}$, $d = \text{ד}$, $h = \text{ה}$ וכיוצא בהן. אבל איך לתעתק את האותיות א' ה' ט' ע' צ' ש' ? אמנם לכל אחת מהן יש תעתיק מקובל בספרי הלימוד העבריים: א מתועתקת

$$א' \text{ או } \text{?} , h = \text{ח} \text{ או } x , \text{ט} = \text{ט} , \text{ע} = \text{או} \text{ '}, \text{צ} = \text{ז} \text{ או } c \text{ או } \zeta , \text{ש} = \text{ש}$$

אבל את הסימנים המיוחדים האלה קשה לעבד בפעולה פשוטה במחשב, וכאשר ביקשתי לפתח מערכת של "מכונת קריאה" ושל מנוע חיפוש בעברית, לא יכולתי להשקיע עבודה בבניית אותיות מיוחדות שקל יהיה להגיע אליהן בשימוש פשוט במחשב. לכן החלטתי להשתמש לצורך העבודה - באופן זמני -, רק בקלידים פנויים הקיימים ממילא בכל מחשב, רובם בשורה העליונה. ואלה הם: $\text{א} = \text{^}$, $\text{ח} = \text{X}$, $\text{ט} = \text{@}$, $\text{ע} = \text{\&}$, $\text{צ} = \text{C}$, $\text{ש} = \text{\$}$. בחירה זו ברובה לא הייתה מקרית. רוב הסימנים האלה כבר שימשו חוקרים שונים או מערכות תעתיק אחרות. אמנם הייתה זו סטייה ניכרת מן התעתיק המקובל בעולם, אבל שימוש זה, כאמור, נועד להיות שימוש זמני בלבד.

1. שאלת התעתיק הראוי לעברית עלתה כעת בציבור על הפרק היות וועדת הדקדוק של האקדמיה ללשון העברית החליטה לאחרונה ליצור תעתיק מיוחד לשמות היישובים בכבישי ישראל. והתעתיק המיוחד הזה הופך את כל המקובל בתעתיק העברית על פיו. לפני כל ביקורת מהותית על ההחלטה הזאת יש לומר שאין שום צורך ליצור תעתיק מיוחד לשימוש מיוחד לשלטי הדרכים או לשימוש מיוחד אחר. כאשר יש תעתיק לטיני טוב לשפה הנכתבת בכתב לא לטיני – תעתיק זה צריך לשמש בכל מקום ולכל שימוש.

ההחלטה הזאת היא: לקבל נוסח אנגלי של הגיית האותיות הלטיניות, ולכתוב באותה אות לטינית את העיצורים העבריים הנבדלים כאשר ההיגוי שלהם נשמע באותו אופן בפי דוברי עברית או בפי חלק גדול או חלק ניכר מדוברי העברית. כך למשל, לא להשתמש כלל באות Q כדי לתעתק בה את האות העברית ק, כמקובל, ולתעתק אותה כאילו היא פ. כמו כן – להימנע מכל סימן דיאקריטי מקובל.

2. הנימוק המובא כהסבר לסילוק ה Q מתעתיק העברית הוא שבאנגלית לא מקובל לכתוב Q בלי שאחריה תבוא U. נגד נימוק זה יש לומר בפשטות, התעתיק של המילה העברית איננו הופך אותה למילה אנגלית. גם כשהיא כתובה בתעתיק לטיני מילה זו נשארת מילה עברית במהותה (ראו להלן הערה 1). מילה איננה נהפכת למילה בשפה אחרת אם כותבים אותה באותיות אחרות. הנה יהודים כותבים באותיות עבריות כבר כמה מאות שנים את השפה הגרמנית הישנה (הנקראת אידיש). זה לא הופך את השפה הזאת לעברית. גם לאדינו כותבים באותיות עבריות, וגם ערבית כתבו יהודים במשך שנים רבות באותיות עבריות. זה לא הפך את הטקסטים האלה לטקסטים עבריים.

כמו כן לא נכון שזה המנהג באנגלית. שמות זרים כותבים גם באנגלית ב Q- בלי U אחריה. למשל Iraq, Qalqilia, או Aqaba, והרי באנגלית יש שלוש אותיות שוות בהיגוי C, K, Q - מה פשוט מלהשתמש בשתיים מהן כדי לסמן 'כ' ו-'ק' עבריות בתעתיק שונה? כל תייר שיראה את השם Qedma יידע לבטא אותו בדיוק כאילו ראה את השם הזה כתוב Kedma. לתייר אין שום הפסד. אולי "לא נוח לו" מראה המילה... אבל לעומת זאת תעתיק שאינו מבחין ולא מבדיל בין הפונמות העבריות יש בו נזק עצום לדוברי עברית ולאלה שיש להם ידיעות מסוימות בעברית.

3. כי זאת לדעת, התעתיק לאות לטינית בשמות לרחובות ובשלטי הדרכים אינו משרת תיירים זרים בלבד. מאות אלפים מתושבי ישראל, כשהשלט מופיע מול עיניהם, רואים בו קודם כול את האותיות הלטיניות. אלה הם עולים חדשים, עולים ותיקים שבאו בבגרותם ועובדים זרים. אפילו אלה שלמדו קצת עברית משיחות ברחוב יכולים לזהות מילים עבריות על-פי התעתיק הלטיני. בשביל כל אלה השילוט באות לטינית בצד המילה בכתב עברי הוא ההזדמנות לקרוא גם עברית. גם לחלק מן התיירים יש ידע מסויים בכתב העברי, וחשוב להם – כמו לנו – שהאות הלטינית בשלטי הדרכים תאפשר את זיהוי האות העברית המקורית. שלא תיכתב למשל K גם בתעתיק של שם היישוב כסלון (Ksalon) וגם בתעתיק של השם קורנית (Koranit), אלא Qoranit, Ksalon. אם לא נעשה כן, יאמר אחר כך מישו: "איזו שפה משונה היא העברית, K כותבים בעברית פעם 'כ' ופעם 'ק'... שחזור לאותיות עבריות מתעתיק מעוות כזה הוא בלתי אפשרי. מי שראה תעתיק שגוי כזה של "קדמה" לא יידע לעולם אם באותיות עבריות כותבים 'ק' או 'כ'".¹

¹ ההחלטה של ועדת הדקדוק היא מהפך גמור בעמדת האקדמיה עד היום. הקביעה הברורה של עמדת האקדמיה ניכרת בספר מהודר שהוציאה האקדמיה לפני שבועות אחדים לקראת ביקורו של ראש הממשלה מר אהוד אולמרט בישיבה חגיגית של מליאת האקדמיה שנערכה ב-19.6.06 בבית האקדמיה בירושלים. הספר נקרא "האקדמיה ללשון העברית בת חמישים". העריכה נעשתה בידי המזכיר המדעי של האקדמיה, היא הגב' רונית גדיש. בפרק מיוחד על התעתיק כן נאמר בעמ' 143:
"שאלת ה-Q -

לפי כללי התעתיק האות ק מתועתקת באות Q. לדוגמה: SHEQEL (שקל). כלל זה, המעמיד את האות Q ללא תנועת U אחריה, עורר ומעורר ביקורת ציבורית לצד כללים אחרים מכללי התעתיק המרגיזים את הציבור (בעיקר בתעתיק שמות יישובים המופיע על שלטי הדרכים; יוער כאן כי רוב המבקרים יוצא מתוך הנחה מוטעית שמדובר בתעתיק לאנגלית). החלטת האקדמיה לתעתק ק ב-Q נבעה משני נימוקים: האחד – האות Q היא מקבילתה של הקו"ף באלף-בית הלטיני מבחינת מוצאה הטיפוגרפי. השני – מכיוון שהכתב הלטיני מעמיד לרשותנו את האפשרות להבחין בין שתי האותיות פ ו-ק, אין סיבה לוותר על הבחנה זו. יצוין שה-Q משמשת לתעתיק מערבית, כגון בשמות Iraq (עִירַק), Qalqilya (קלקיליה), ואם כן התופעה של Q

4. לא זו בלבד. האותיות א' ו-ע', שמאות שנים מקובלות בעברית שיטת תעתיק המבחינה ביניהן, יסולקו מן המילים ולא יועתקו כלל. המילים "בו" ו"בוא" יתועתקו שתיהן ל- bo, ובמקום שיש הכרה לסמן א' או ע' באמצע מילה לא יהיה הבדל בסימון בין א (') ובין ע ('). והרי גם אלה שאינם מבדילים בין א' ו-ע' בדיבור שומרים עליהן באותיות נפרדות כשהם כותבים. תלמיד שלא יבחין ביניהן ייכשל בבחינה, ומי שאינו יודע להחליט אם יש א' או ע' במילה לא יתקבל לעבודה בשום מקום המצריך כתיבה. איך אפשר להעלים בתעתיק העברית עובדות מוצקות אלה של השפה העברית? הוועדה החליטה גם שהאות ו' והאות ב' רפה ייכתבו שתיהן באותה אות V, ש-ת' ו-ט' ייכתבו שתיהן ב T - בעוד שתעתיק ט' צריך להיות T בהוספת נקודה מתחתיה. בכתיבה בעברית שלנו אנו מקפידים על הבחנה מובהקת בין אותיות של עיצורים שונים, גם אם בדיבור עיצורים אלה נהגים באותו אופן. את אותן ההבחנות יש לשמור גם בתעתיק, כי כפי שציינה האקדמיה - ראו בהערה 1 לעיל - התעתיק אינו הופך את השמות האלה למילים אנגליות. הן נשארות עבריות גם בתעתיק, אבל בעברית שגויה ומעוותת. לפי התעתיק המעוות הזה אם תרצה לכתוב "צ׳ייר קוף צעיר וחברו מטיילים ליד היער", תקבל "תסייר כוף תסאיר בחברו מטיילים ליד היאר...".

2. כאמור, הכתיבה באותיות לטיניות של שמות היישובים אינה הופכת את השמות למילים אנגליות. אבל ברוב הארצות שיבוא אליהן התייר הוא ימצא שלטי דרכים לא באנגלית. כך בצרפת, באיטליה, בנורבגיה או בצ'כיה וברוב המדינות בעולם. שלטי הדרכים המודיעים על שמות היישובים אינם כתובים באנגלית. התייר האנגלי או האמריקני המבקר בארצות האלה אינו יודע לבטא אותן. באיטליה הוא אינו יודע איך מבטאים chicco ((kee'-ko=), uscita ((oo-shee'-tar=), או egli (=el'-yee). בשבדיה הוא טועה בקריאת djur (= yoor), או skida (= shee'-da). גם בישראל, כמו בכל העולם, הפְּתִיב של שם היישוב אינו יכול לכלול הדרכה פונטית לדוברי אנגלית

6. התעתיק לאותיות לטיניות הקיים כמאתיים שנה היה מקובל גם באנגליה בתור התעתיק הנכון לעברית, והוא עומד ביסוד העקרונות שקבעה הוועדה החשובה -

Committee on Geographical Names for British Official Use.

בתעתיק כלל עולמי זה הוקפד כאמור על הבחנה ברורה בין העיצורים, והוא כלל סימנים דיאקריטיים (כמו ברוב שפות אירופה שיש בהן אקסנטים וסימני עזר למיניהם). התעתיק המדויק הזה היה מיטב ההבנה הבלשנית באותו זמן (לא היום). אבל כבר לפני עשרות שנים מרדו בו האנגלים לצרכים פנימיים שלהם. למשל, בספר בשם Palestine & Transjordan שהוציאה Naval Intelligence Division ב-1943, כך נאמר בעמ' 7:

"All-diacritical-signs-have-been-omitted,...-the letter-k,-except-in-'Iraq',-has-been-preferred-to-q;-the-terminal-h has-been-appended-after-e-(e.g.-Huleh,-not-Hule-or-Hula)"

לאינטליג'נס הימי הבריטי לא הייתה שום מחוייבות לשפה העברית. מה איכפת להם אם התעתיק שלהם ניתן לשחזור לכתב העברי או לא; מה איכפת להם מה הן הפונמות העבריות; מה איכפת להם אם הן מתערבבות; מה מפריע להם אם אותה אות עברית מועתקת בצירוף אותיות "אנגליות" או לא. אבל האקדמיה ללשון העברית מחוייבת גם על-פי החוק וגם על-פי התקנון שלה, לשמור על הלשון העברית גם בתעתיק, ובהחלט ועדת הדקדוק, אם אכן תהיה סופית, הפרה האקדמיה גם את החוק וגם את התקנון.

ג. לחצים על האקדמיה

ללא U אינה המצאה של התעתיק מעברית. גם בדיונים החוזרים עמדה האקדמיה על דעתה בעניין זה ואף לא העמידה הצעה לשנות את הכלל."

טקסט מאלף זה, המאשר את כל דברינו בגוף המאמר, התפרסם בהוצאה המכובדת של האקדמיה. ישיבת ועדת הדקדוק, שבה נקבעו הכללים החדשים הופכים את כל עמדות האקדמיה על פיהן, נערכה ב-5.6.06, כשבועיים לפני שהספר נמסר לחברים והתפרסם.

החלטה זו של האקדמיה היא תוצאה של לחצים שהופעלו עליה לקבוע "כתיב אנגלי" בתמרורים לשמות היישובים. לחץ זה בא בעיקר מצד משרדי התיירות והתחבורה, אולי נבהלה האקדמיה גם מהחלטת הכנסת לחייב תעתיק אחיד בכל השלטים בארץ תוך חמש שנים. בישיבת האקדמיה מיום 22.5.06, הזהירה הגב' רונית גדיש, המזכיר האקדמי, שאם האקדמיה לא תשנה מייד את עמדתה בעניין תעתיק שלטי הדרכים, תיטול ממנה הכנסת את זכות ההחלטה. וכך האקדמיה, מטעמים פוליטיים שאינם קשורים כלל במהותה של הלשון העברית, טעמים שאין להם שום הצדקה בלשנית או מדעית משום סוג שהוא, נכנעה ללחץ זה.

ההחלטה הנדרשת הזאת לא יכלה להתקבל במליאת האקדמיה. זו הייתה החלטה של ועדה. חברי המליאה (גוף בן שלושים חברים ויותר) חששו באותה ישיבה, כמו גם בהרכב הזדמנויות קודמות, להיכנס לעומק שאלת התעתיק, ובמקום זה "הסמיכו את הוועדה" בצורה מעורפלת ושנויה במחלוקת לקבל החלטה מחייבת. מתוך שלושים ויותר חברי המליאה שנתבקשו להביע דעתם בכתב לוועדה, רק חמישה (!) טרחו לכתוב את דעתם על העניין... ההחלטה על "כתיב אנגלי" בשלטי הדרכים שנתקבלה כך בחופזה היא החלטה שראוי לבטלה בכל דרך. אם חלילה לא תתנער האקדמיה מכניעתה לתכתיב האנגלי, יתמלאו שדותינו וחוצותינו בתעתיקים של מילים עבריות לפי המבטא הדמוי-אשכנזי ולפי כתיב מעוות המנסה לחקות מנהגי שפה זרה.

ד. עברית אחת לאורך כל שנותיה

1ד. האקדמיה על כל חבריה, וכן רוב כמעט מלא של הציבור דובר העברית, מתייחסים אל השפה העברית כאל שפה אחת הקיימת מאז נכתבו הכתובות העבריות הראשונות בראשית האלף הראשון לפני הספירה או אולי גם קודם לכן. גם את העברית שבפינו אנו חושבים כמובן לחלק פנימי ומהותי של השפה העברית. האקדמיה עצמה איננה מסכימה עם אלה הרואים בעברית שלנו איזו שפה חדשה הנקראת "ישראלית" (שאמנם התומכים במציאותה מודים בשפה רפה שהיא נובעת מאיזו שפה עתיקה שקוראים לה "עברית"...). גם הציבור המשכיל רואה ב"ישראלית" רק את אחד המבעים של העברית. וכשאנו מכינים מגוון חיפוש כוונתנו שנוכל להפעיל אותו על כל טקסט עברי שהוא, בין שהוא חדש בין שהוא מתקופות קודמות של העברית.

והנה באה האקדמיה ובכניעתה לדרישה של "כתיב אנגלי" למילים העבריות, בתעתיק המשקף כביכול מבטא "ישראלי", דמוי אשכנזי, היא מכירה למעשה, וגם להלכה, בקיומה של שפה נפרדת, מודרנית, ששמה כביכול "ישראלית". הכניעה לכתיבה שאיננה מפרידה בין פונמות עבריות שונות המובעות תמיד בכתב סותרת את המחויבות של האקדמיה, שהיא -

"לכוון את דרכי התפתחותה של הלשון העברית לפי טבעה, לפי צרכיה ואפשרויותיה בכל תחומי העיון והמעשה... בכתב, בכתיב ובתעתיק."

2ד. כפי שניכר ממובאה זו, הלקוחה מתקנון האקדמיה ללשון העברית, התעתיק הוא חלק מן העברית, וחובת האקדמיה היא להגן גם על תעתיק נכון של המילים העבריות. שום עם המכבד את עצמו ובוודאי לא האקדמיה ללשון שלו, לא ישנה את מטבע הלשון שלו כדי "להקל", כביכול, על תיירים לבטא את המילים שבשפתו. מה גם שכל "הקלה" כזאת כלל אינה מקילה. חשבו לרגע לדוגמה על רחוב "התור" בחיפה. שם קצר ופשוט. בשלט המוצב ברחוב זה כתוב: hator. אבל כשתייר אמריקני יפנה אל עובר אורח לשאול אותו איפה הרחוב הזה, הוא יבטא שם זה לפי מנהגי ההגייה של האנגלית, וישמיע משהו כמו "הייתר"... מי אותו עובר אורח שיבין את כוונתו של התייר? בירושלים שאל אותי פעם תייר אמריקני איפה הֶמָג סטריט. לא ידעתי. אילו היה אותו תייר פוגש לראשונה בכתב את השם Tel Aviv היה שואל על אֵיבב. כאמור, התעתיק אינו יכול לכלול הדרכה פונטית אנגלית למילים בשפה אחרת. איך יקרא התייר האמריקני את שמות היישובים Shaked, Even Sapir, Magen או katif? התייר, בכל ארץ שהוא מתייר בה, יודע שקריאת המילה בשפה שהיא זרה לו אינה כדרך שהוא קורא את שפתו האנגלית. את זאת יבינו התיירים דווקא כשלא נתאמן להמציא בשבילם כתיב "אנגלי" למילים שאינן בשפתם.

1א. התעתיק חייב להיות בר שחזור. ארגון התקינה העולמי, ISO, המאגד את מכוני התקנים של כל העולם, עסק הרבה בסוגיית תעתיקים של שפות שיש להן כתב לאומי שאינו לטיני. ארגון ISO קבע את עקרון השחזור כתנאי ראשון במעלה, עיקרון שאסור להפר אותו - כל שיטת תעתיק חייבת להיות כזאת שתאפשר שחזור אוטומטי, הן על-ידי אדם הן על-ידי מכוונה: מן הנוסח המתועתק חזרה אל הנוסח המקורי של הכתב הלאומי.

2א. כל עיצור יתועתק באות לטינית אחת ולא בצירוף של שתי אותיות. תכונה זו חשובה למבנה העברית שרוב המילים בה בנויות כצירוף של שורש ומשקל. הרוב המכריע של המילים בעברית בנוי על יסוד צירוף של שורש ומשקל. מילה כמו מְשַׁמֵּר מורכבת מן השורש ש'מ'ר' וממשקל *mi_a* כשכל אחד מעיצורי השורש נכנס לפי הסדר לאחד המקומות הריקים של המשקל, ואין לשים שני עיצורים באותו מקום ריק (המסומן כאן בקו). אם תשתמש בצירוף של שתי אותיות לסימון פונמה אחת - התוצאה תהיה שהאות השנייה של הצירוף תתקבל כמסמנת את העיצור השני של השורש! מכאן נובע במיוחד שאסור לחקות את הנעשה בשפות אחרות שבהן האות *h* משמשת בתור "סימן עזר" הבא בצירוף לאות אחרת, כמו למשל, *kh, ch, sh, th, zh* וכדומה. באיטלקית אכן משמשת האות *h* כסימן עזר. באיטלקית זה אפשרי כי בשפה זו אין פונמה */h/* וגם לא הגה *[h]*. גם באנגלית זה אפשרי, כי *h* כפונמה באנגלית יכולה לבוא רק בראש מילה או כשלפניה ואחריה תנועה. לכן באנגלית במילה שיש בה *h* אחרי עיצור אין אפשרות לטעות.

אבל בעברית האות ה' מסמנת עיצור מוגדר השווה בערכו לכל עיצור עברי אחר. רק בסוף מילה עברית יכולה האות ה' לסמן או עיצור או תנועה. לכן, אי אפשר להציע בתעתיק של מילים עבריות להשתמש באות *h* כ"סימן עזר" שיצביע על הגה אחר. הפתרון צריך להיות - שימוש בסימני הבחן קלים (דיאקריטיים), כמו שמקובל ברוב השפות האירופאיות. כאשר יש צורך בסימון של הגה השונה במידת מה מן המשמעות הפונטית המקובלת של אות לטינית - משתמשים בסימני הבחן, כגון נקודה או קו מתחת לאות או מעליה, סימן *v* קטן מעל האות (למשל סימון ההגה ש' בצ'כית ובעוד שפות הוא על-ידי *v* קטנה מעל האות S).²

אכן באנגלית לא מקובלים סימנים דיאקריטיים. אבל שמות היישובים בתעתיק אינם מילים אנגליות, וסימנים דיאקריטיים מצויים ברוב השפות שתיירים דוברים בהן או עשויים להכיר אותן, בין היתר למשל, צרפתית, גרמנית, ספרדית, איטלקית, שוודית, פינית ועוד לרוב. פרט מעניין הוא שגם בכללי התעתיק מעברית ללועזית שקבעה האקדמיה בשנת 1957 כלולים סימנים דיאקריטיים. קודם כול אלה הם התעתיקים של האותיות ג', ז', צ', שכולם תועתקו כש *v* - קטנה כתובה מעל האות *g, z, c*. יתר על כן, לפי כללי "התעתיק הפשוט" של האקדמיה מותר להשמיט את האות א' כאשר לפניה ואחריה יש תנועה, אבל אז, אם זה גורם לאי הבנות, מותר לשים *tréma* מעל התנועה השנייה. הכוונה למילים כגון בריאות, *bri'ut* אם הושמט סימן ה-א' (כלומר כתבת *briut*) קובעת האקדמיה, רצוי לכתוב זאת כך *briüt*: על יסוד מה מהססת כעת האקדמיה לקבוע הבחנות בין הפונמות על-ידי סימני הבחן דיאקריטיים?

3א. פונמה עברית שיש לה שני מבטאים יש לתעתק באות אחת קבועה, ולא באותיות שונות לפי המבטא. עקרון בלשני זה משמש בשפות רבות, למשל באנגלית במילה *devote* ובמילה *devotion* נהגה ה-*t* פעם כ *[t]* ופעם כ *[sh]*, אבל הכתיב לא משתנה, כי לדוברי אנגלית יש עניין לשמור על הקשר בין שתי צורות אלה השייכות ל"מילה אחת". כך גם בעברית: הפונמות ב'כ'פ' יש לכל אחת מהן שתי הגיות, אבל לדוברי עברית יש עניין לשמור על הקשר בין, למשל, שתי הצורות בדק - יבדוק. לכן גם בתעתיק יש לשמור על קשר זה: *badaq - yibdoq*. וכך האותיות *v, f* אינן משמשות אותנו בתעתיק. כמו כן ראו נא את האותיות *G C* שברוב השפות הן משמשות כל אחת לסימון שני הגאים, וההיגוי נקבע לפי מעמד האות במילה.

² אף כי ועדת הדקדוק החליטה להימנע מסימנים דיאקריטיים, שכן האנגלית אינה סובלת אותם, היא לא יכלה להימנע כליל משימוש בגרש (אותו האנגלית סובלת). הנה בהודעה הרשמית שהתפרסמה על התעתיק הוזה מתועתקת כ רפה בצירוף *kh*, אבל מה תעשה אם תצטרך לתעתק מילה כגון הקהל? לזה התקינו תקנת "גרש מפריד". התעתיק של המילה הקהל המובאת כדוגמה בהחלטה הוא: *hak'hel*, שהרי אם תכתוב *hakhel*, בלי הגרש המפריד, יהיה זה התעתיק של המילה הכל...

1. "תעתיק המבנה"

1.1. העקרונות האלה נשמרים בהקפדה בתעתיק שעמד כמצע לוועדה שהקים מכון התקנים הישראלי בשנות התשעים של המאה העשרים. הוועדה פעלה במסגרת ארגון ISO והתעתיק עבר את כל שלבי התקן של ISO עד האחרון. תעתיק זה נותן רישום אחיד לכל מילה עברית מכל תקופה שהיא, וכך הוא מבטיח את הקביעה החשובה, הנכספת, שהשפה העברית היא שפה אחת לאורך כל תולדותיה, וה"ישראלית" היא כולה בגדר העברית, ואין צריך לומר שהתעתיק הוא בר שחזור לטקסט העברי המקורי.

בניגוד לשתי שיטות התעתיק המקובלות, "תעתיק האותיות" (טרנסליטרציה) ו"תעתיק ההגאים" (טרנסקריפציה), שהנחו את האקדמיה בכללי התעתיק של 1957 (על כך להלן), תעתיק מכון התקנים הוא "תעתיק המבנה". אין הוא מתיימר לייצג את ההגאים הנשמעים של העברית - ייצוג כזה היה יוצר נוסחי תעתיק שונים לפי מבטאו של כל דובר, או של כל עדה או של כל תקופה היסטורית. הוא גם אינו מבקש לסמן את כל האותיות או את כל הנקודות שבכתב המנוקד של העברית - כי הכתב המנוקד, כאמור, אינו מגלה בתצוגתו את המבנה של המילה כראוי, וכן כי יש אותיות הממלאות תפקידים שונים במילים ולא ראוי לתעתקן תמיד באותו סימן. וכן, חלק ניכר מסימני הניקוד מיותר לגמרי לגבי כל תקופה של העברית, ובוודאי לגבי תקופתנו

לעומת זאת, אם אתה בונה את התעתיק על יסוד מבנה המילה - ולא לפי סימני האותיות והנקודות, וגם אינו מתכוון כלל לתת לפני הקורא הנחיה איך לבטא את המילים העבריות (יש הרבה דרכי הגייה, והן אינן יכולות להיות חלק מן התעתיק האחיד) - אין אתה צריך לשים לב לדקויות של סימני התנועות. מבנה המילה בתעתיק זה יבהיר ממילא, בעצם קיומו הגלוי, לאילו סימני ניקוד מקוריים הכוונה.

לא כאן המקום להראות בפרוטרוט איך על יסוד תעתיק המבנה ניתן לזהות את הניקוד המדויק, אבל ראוי לספר כאן כי כבר שנים רבות יש בידי תוכנית מחשב מלאה הערוכה לפי אלגוריתם בלשני בלבד (ולא לפי שכיחויות או הסתברויות) העושה זאת. היא מקבלת טקסט בתעתיק המבנה כקלט ומוציאה כפלט טקסט בכתב עברי מנוקד, מדויק וברור.

2.1. לפי בקשת מכון התקנים הישראלי אמורה האקדמיה לדון בקרוב בבדיקת התעתיק הזה לפי צרכיה ואפשרויותיה של העברית, כדי להבטיח שהוא איננו פוגע במבנה השפה העברית בשום צורה. אבל אף על-פי שהדיון בתעתיק זה לכאורה כבר התחיל באקדמיה לפני כשנתיים, מיהרה ועדת הדקדוק של האקדמיה והחליטה בטרם עת, לפני סיום הדיון הזה - לאשר לצורכי השלטים בדרכים את "התעתיק האנגלי" שביקשו ממנה אנשי מע"ץ, משרד התיירות ומשרד התחבורה. זהו תעתיק חדש ההופך על פיה את עמדת האקדמיה, מנוגד למקובל בעולם המדעי, מעולם הוא לא נדון במליאת האקדמיה ומעולם לא הובא לידיעה ולשמיעת חוות דעת של הציבור. אם נשב בחיבוק ידיים אנו עתידים לראות שמות אנגליים ולא עברית ברחובותינו ובשדותינו...

3.1. סגולותיו אלה של תעתיק המבנה המוצע מטעם מכון התקנים - (1) שהוא התעתיק היחיד המאפשר לתעתק לפיו כל טקסט עברי שהוא, בין שהוא טקסט לא מנוקד בן ימינו המכיל גם מילים שמקורן בלועזית, בין שהוא שיר של משורר ספרדי מן המאה השתים עשרה, בין שהוא טקסט של לשון חכמים, משנה או מדרש, ובין שהוא טקסט מקראי או כתובת עברית עתיקה, וזאת מבלי לסמן בו את הניקוד המקורי או הצפוי; (2) שהוא מאפשר שחזור מלא ומדויק של הטקסט בתעתיק לכתב עברי לא מנוקד מקורי, ואף (3) יכולתו להיכתב לפי טקסט ללא ניקוד ולהפוך אוטומטית לטקסט מנוקד - כל אלה עושים את התעתיק הזה למכשיר מצוין לכל "צרכיה ואפשרויותיה של הלשון העברית בכל תחומי העיון והמעשה, באוצר המילים, בדקדוק, בכתב, בכתב ובתעתיק".

4. לא נראית שום סיבה מדוע לא יוכרז תעתיק זה כמתקבל לצורכי תעתיק של טקסטים עבריים כל שהם, ובמיוחד טקסטים לא מנוקדים. תעתיק זה עשוי לשמש בצד "התעתיק המדויק" משנת 1957, שנועד לשמש כתעתיק לטקסטים מנוקדים. "התעתיק המדויק" מכוון לצורכי תעתיק בסימון מלא של טקסטים מנוקדים, ואילו התעתיק המוצע הוא כללי, אך מתאים במיוחד לתעתיק טקסט בכתב לא מנוקד. האקדמיה אולי תציע שינוי בסימן זה או זה, אבל העקרונות ודאי לא ישתנו. אולם, כאמור לעיל, רק חמישה מחברי האקדמיה (מתוך שלושים ומעלה) טרחו להביע את דעתם על הנושא הזה כשנתבקשו להביע אותה בכתב ולשלוח את פרי מחשבתם אל הוועדה לדקדוק. אם הוא תעתיק נאות, אין סיבה לנסח תעתיק מיוחד לשלטי הדרכים.

5. הרבה יותר פשוט, כנראה, לעסוק בדיון על מונחים בודדים, אבל כשעומדת על הפרק באמת "שאלה לזהות" (או "שאלה נכבדה"), כמו השאלה איך ייראו רחובותינו ושדותינו, איך ייכתבו שמות היישובים בכתובות של מכתבים - שם מושלך אלם. צר לי לספר שגם המכתב המובא להלן ששלחתי אל חברי האקדמיה לא הועיל:

"אנא אל תהססו. תפקידנו כחברי האקדמיה על אחריותה הרבה ללשונו
מחייבת כל אחד מאיתנו לא להימנע מהבעת עמדה ברורה ונחרצת.
בידינו החובה והזכות להחליט היום והדורות הבאים ידונו את מעשינו."

ז. סיום

בלחץ המוסדות הממשלתיים מיהרה ועדת הדקדוק להחליט בעניין שלטי הדרכים ולא דחתה את הקביעה עד לאחר שיוחלט על תעתיק המבנה, ואולי ניתן היה אז לקבוע שלטי הדרכים ייכתבו לפיו, אלא מיהרה לקבל מה שקיבלה כמתואר לעיל. הנה כך עומד הציבור לראות בקרוב את שלטי הדרכים כתובים בראש חוצות בתעתיק זר, מטעה, שאינו מוסר כראוי את תמונת המילים העבריות. אכן, בכי תמרורים...

האם ממלאת האקדמיה את חובתה המדעית, הבלשנית והלאומית המוטלת עליה מטעם החוק ומטעם הציבור? דומה יותר שהיא נכנעת לתביעת פוליטיקאים המבקשים להשליט תכתיבים רוחניים או גם גשמיים של מדינת חסות אמריקנית, תוך הסתגלות מהירה של שפת הילידים למנהגי השפה של בעלי הבית, כשאלה באים לסייר באחוזתם.... אם נשב בחיבוך ידיים נראה בקרוב בראש כל חוצות מילים דמויות אשכנזית בכתב אנגלי אמריקני. זה יהיה מעוות שלא יוכל לתקון.

לבסוף, אנא ראו את הדוגמות. האם תייר שאינו יודע עברית יאבד את דרכו אם יופיעו לפניו הכתובות הבאות בשלטי הדרכים?

Ge`ulim, Gadot, Ğat, Har Tab̄or,
Vered ha-Galil, Zek̄arya, H̄ulda,
Ṭirat Karmell, Ṭp̄aḥot, Ya`ala,
Yeruh̄ḥam, Kb̄iš Allon, Leb̄ona,
Miçpe, Nehora, Neve Šalom,
`ein Vered, Paṭṭiš, Šde Boqer,
Šmurat H̄ula, Reh̄ob̄ ha-Tor.