

לשוננו לעם

עורך: מ' בר־אשר

מתוך

כרך מ-מא, תש"ן

קובץ לשנת הלשון



האקדמיה ללשון העברית
מוגן בזכויות יוצרים

תרומת המחשב לבלשנות העברית

בדורנו השתלט המחשב על חיי היום-יום: בשירותי הבנק האוטומטיים למיניהם, במשחקי מחשב, באבחונים רפואיים ובהזמנת נופש. המחשב עתה אופף את כל פינות החיים, וזה פן אחד של הדמיון בין המחשב ללשון, שהרי גם הלשון אופפת כול כבר מראשית התרבות. הפן האחר הוא תרומת חקר מדעי המחשב לבלשנות. כפי שנראה, קשרים אלה ותרומות אלה מתרחשים ומתגשמים גם בלשון העברית.

א. תרומת המחשב ככלי – ללשון ככלי

דיברנו בלשון, כתבנו בה, האזננו וקראנו דורות רבים לפני בוא המחשב. ואכן לאנשים הרבים שניחנו בכשרים התקשורתיים הללו אין המחשב ככלי תקשורת מעלה או מוריד, ואולם פיתוח תוכנות מחשב לדיבור מלאכותי (Dectalk); לקריאה מן הכתב לעיוורים (מכונת קורצוויל) וכו', הועיל מאוד להעביר את העיוורים, החירשים, האילמים ומשותקי המוחין ממעמד של נדחים לשותפים, פעילים ושווים בחברה. היות שתוכנות אלה יכלו לנבוע רק הודות למאמץ מחקרי אדיר, נשוב להזכירן בהמשך.

אך גם לבני המזל שבנו, הרואים, המדברים והנעים בכוחות עצמם,

משמש המחשב כלי רב-עצמה לקידום מיומנויות לשוניות שונות.

תוכנות שירות המופצות עתה לרשות "צרכני הלשון" הן המילונים האלקטרוניים. בתוכנות אלו המחשב מוזן ברשימות מלים עבריות ומקבילותיהן מצד המשמעות בשפת היעד. הקשת מלה בשפת המוצא תגרום לשליפה (טכנית) של המלה משפת היעד (עלידי מצביע). כאן נעשה שימוש במחשב ככלי – אין כאן שום תפיסה בלשנית או כושר מחשבי מיוחדים.

מגוון רב של שימושי מחשב נעשים בתחום קרוב: הן התוכנות להדרכה בכתיבה ושיפורה. כאן יש תוכנות פשוטות (טכניות) המאתרות שגיאות כתיב, ושגיאות הדפסה כאותן מחרוזות שאינן נמצאות ברשימות המלים.

מורכבות יותר הן התוכנות המספקות למשתמש הצעות תיקון. בדרך-כלל משלבות הצעות התיקון גם את המלים הנרדפות למלה המוצעת, למשל:

"הגיה פתאום מן **הברב** **מרצח־חרש**" (כהן יעקב, שירים עג).

מארב, מסתר פזמק, כסות רגל התקן לזרימת מים קול, (בדקדוק) מבנה עיצוריים) ותנועה	הארב הגרב הברז הברה
--	------------------------------

בדוגמה זו מוצעים כמה תיקונים אפשריים למלה השגויה **הברב**. תוכנות מסובכות יותר מסייעות לביקורת סגנון הכתיבה ושיפורה. תוכנות אלה בודקות את הטקסט הכתוב לפי התאמתו למשלב מסוים (כתיבה מדעית, כתיבה טכנית, כתיבה ברמת חטיבת ביניים, כתיבת פרסומת ועוד). במקרים כאלה תנתח התוכנה את הטקסט ותצביע לפני המשתמש על מקומות שבהן אותרה סטייה מן הסגנון הנהוג במשלב המסוים (אוצר מלים, סיבוכיות המשפטים ואורכם, לקסיקון לא מקצועי בטקסט מקצועי ועוד). נעיר כי לפי הידוע לנו לא פותחה עדיין תוכנה כזו עבור הלשון העברית. באנגלית קיימות למשל תוכנות Proof ו-Epistle של יבמ.

תחום חשוב נוסף בתוכנות השירות הוא הלומדות. לומדה במובן האידאלי היא תוכנה המציגה לתלמיד גוף ידע והסברים (הוראה) ומזמנת לו מטלות לתרגול גוף הידע. כל אלה מותאמים לקצב האישי של כל תלמיד ותלמיד, תוך מעקב נבון אחר התקדמותו, הבנתו את החומר, ואיתור קשיי התלמיד (הלומדה יוצרת "מודל תלמיד"). לומדה טובה צריכה לנצל את מדיום המחשב באופן שהיא תציג איכויות מיוחדות השונות מאלה המוצגות בהוראה פרונטלית. לומדות מושלמות כאלה יכולות למעשה להחליף את ההוראה הפרונטלית (בתחום הידע המסוים). בתי-תוכנה רבים כותבים לומדות ומשווקים אותן. ואולם רובן ובכלל זה הלומדות שבתחום הלשון העברית, אינן מושלמות: בדרך-כלל הן מתרכזות בתרגול ולא בהוראה, וגם זה אינו מתאים עצמו ללומד המסוים אלא מהווה למעשה מעין חוברת תרגילים אלקטרונית. חלק מלומדות אלה לא שינה במאום את תפיסת ספרי הלימוד אולם הוא מגיש אותן לתלמיד ב"עטיפה" מפתה יותר דוגמת משחק מחשב. למשל,

תרגול הבחנה בין כתיב מלא לכתיב חסר על-ידי לכידת בעלי חיים ביער: היער מוצג בצבע על הצג ובו בעלי חיים. על התלמיד לנקוב בשמותיהם (דב-דוב, חלד-חולד, נמיה-נמייה, עיט-עייט, תכי-תוכי ועוד), אם איית התלמיד נכון את השם יוכנס בעל-החיים לכלוב גן-החיות.

לע"ד, יש לייחד את השם לומדה להוראה המתוכנתת, ואילו תוכנות תרגול על-ידי משחק ראוי לכוונתן בשם "שוחקה" (שחקה). בשיטות אלה נעשה שימוש גם להוראת שפות זרות ובכללן עברית. סדרה כזו פיתחה למשל, היחידה להוראת עברית באוניברסיטה העברית. יש אף ששולב כאן מרכיב קולי. ניצול דומה נעשה גם בתוכנות ללימוד הקריאה. נעיר כאן על הצד הפסיכולוגי בשימוש הקול: לומד מבוגר (דוגמת עולה חדש) אינו "מעוכב" בדיבורו אל המחשב כשם שהיה אולי מעוכב בדיבורו אל המורה.

התחום המפורסם ביותר הוא התרגום על-ידי מחשב, שהרי הניסיונות לתרגם טקסטים על-ידי מחשב הם בני דור ומעלה. בדרך-כלל מוגבלות התוכנות לתרגום של סוג טקסטים קבוע באוצר מלים מוגבל; רובן משלבות עם התרגום האוטומטי גם עבודה ידנית: במקום שיש כמה אפשרויות לתרגום מעבירה התוכנה את ההכרעה הסופית לידי המשתמש; התוכנות שונות גם במבנן, ובעיקר ביחס שבין החלק המילוני שבהן ובין התחביר והמורפולוגיה.

בזמננו חודשו שני פרויקטים גדולים בתחום זה: (א) פרויקט התרגום של השוק האירופי המשותף (Eurota), (ב) פרויקט התרגום של המרכזים המדעיים האירופיים של יבמ. בדרך-כלל מספר דוברי השפות הלאומיות הקטנות מגיע רק לכמה מיליונים בכל אחת, אך מובן שהם מבקשים לקבל כל חומר כתוב בלשונם שלהם. המרכזים רוצים לספק את החומר הטכני של יבמ בשפת הארץ. ונציין בסיפוק כי עברית היא אחת משפות היעד לתרגום, בצד השפות הסקנדינביות, הפינית, הפורטוגזית והצרפתית.

תרגום כשלעצמו הוא משימה קשה, אם כי בתחומים מוגדרים זוהי משימה אפשרית. ואולם, תרגום מעברית לשפה אחרת היא משימה קשה פי כמה בגלל מגבלות הכתב העברי. אפשרויות הקריאה הרבות של משפט ממוצע מחייבות את המשתמש להקדים הכנה ידנית של הטקסט לקראת התרגום ולהשקיע עבודת עריכה ידנית רבה של פלט תכנית התרגום.

לשם הבהרה נתבונן נא במשפט העברי הבא: הפרות בעלות הבוקר חוזרות ורבות. קל להראות שמשום שאין הכתב העברי מפריד אותיות שימוש ממלת הערך, ואין הכתיב המקובל (בלתי מנוקד) מציין דיגוש, סדרת המלים הזאת היא רב-משמעית, ואלה יוכלו להיות תרגומיה לאנגלית:

1. When the sun rises violations repeat and increase/
2. When the sun rises the cows return and reproduce/
3. When the sun rises the cows fight again/
4. The cows who own the cowboy return and fight/
5. The cows who own the cowboy fight again/
6. The cows who own the cowboy return and reproduce/

אדם הקורא את המשפט העברי יוכל לקבוע על-פי ידיעותיו הכלליות ועל-פי ההקשר מהו התרגום הנכון. אבל איך יעשה זאת המחשב? נעיר, כי הוספת סימני ניקוד לקלט הניתן למחשב עשויה להקל מאוד על האדם – קורא הפלט, אך היא בעייתית מאוד בשלבי העיבוד: תחילה, נדרשת הכנה של הטקסט הבלתי מנוקד על-ידי נקדן מיומן; לצד זאת, נדרשת עבודה טכנית ותכנותית של הקצאת סימני ניקוד מסימני המחשב (עתים על חשבון סימנים נדרשים כאותיות לטיניות נמוכות). דרישה זו עלולה גם להגביל את העבודה למחשב ומדפסת מסוימים. משנפתרו בעיות חיצוניות אלה, נדרש פיתוח של תוכנה כבדה שתבחין (לפי מטלת התכנית) בין סימני ניקוד רלוונטיים לשאינם רלוונטיים (כגון חטפים במיון משקלים ודיגוש וריפוי במיון שורשים).

הכתב הפונמי שפיתח פרופ' עוזי אורנן (מתן קלט באותיות לטיניות ושיקוף המבנה התאורטי של המלה) עוקף את בעיות הניקוד שנמנו לעיל. כתב זה הוכיח עצמו כפרשת דרכים בעיבוד עברית במחשב, בפרויקטים שונים של אורנן: תכנית לניקוד אוטומטי; תכנית להפקה של פלט פונטי (כללי קריאה); תכנית לקבלה של כתב עברי בלתי מנוקד (לפי כללי האקדמיה ללשון העברית); תכנית אינדקס וקונקורדנציה של פסיקת בית המשפט העליון ותכניות ניתוח ויצירה של פעלים ושמות.

נציין בסיפוק כי למרות הקשיים הרבים שמעמיד הכתב העברי, כבר מצויים ניצנים ראשונים של שפות המחשב הנכתבות בעברית (בייסיק ופרולוג עבריים). חשיבותן בכך שהן מאפשרות למתכנת העברי להתקשר

אל המחשב בשפתו שלו – עברית, ומעלות את העברית (כלפי המתכנת העברי) לשפה בת תחרות לאנגלית.

ב. תרומת המחשב לחקר הלשון

עד כאן הזכרנו תוכנות הניתנות למשתמש כשהן מוכנות. בעיקר היו אלה חבילות תוכנה. המשתמש אינו מתערב בתכנית, עליו רק לספק לתכנית את הנתונים (קלט) שברצונו לעבד, ואז יקבל את התוצאות (הפלט) הרצויות לו. כל עניינו של המשתמש מכון כלפי התוצאות ולא כלפי התהליך. גם בהוראה מתוכנתת, מעוניין הלומד בידע שהתוכנה מקנה לו, בהפניות שהיא מפנה אותו או בהערכה שהיא נותנת לו, ואין הוא מתעניין בתהליך ההכרעה של התכנית.

שונה מאלה תרומת המחשב למחקר הלשון. היבט זה של המחשב נכון לחלקו לשני חלקים: (א) תרומת המחשב ככלי למחקר הלשון (בלשנות); (ב) תרומת מחקר המחשב למחקר הלשון.

1. המחשב ככלי למחקר הלשון

המחשב העומד לרשות כל חוקר הוא כלי אמין ביותר. כלי מהיר שאין לו מתחרה, בעל זיכרון עצום, שאינו מתעייף, אינו נח בחגים ואינו שובת. חוקרים שיש בידיהם נתונים הדורשים ניתוח, מיון, מנייה וכד' יכולים לנצל את תכונותיו אלה של המחשב. שימוש זה משחרר את החוקר מן הפעולות הטכניות הרבות הנדרשות לביצוע מחקר, ומפנה את מחשבתו, זמנו ומרצו לניסוח התאוריה, לקביעת המתודה וכד'. גם כאן מעוניין המשתמש בתוצאות – בפלט ולא בתהליך. משום כך מנוצלות גם כאן חבילות תוכנה קיימות המיועדות למיון נתונים (לאחר שהמשתמש מספק למחשב את הנתונים ואת קנייהמידה למיון), או לשליפה ואחזור. אם אין בנמצא תוכנה המתאימה לצרכיו של החוקר, יפקיד החוקר את מלאכת התכנות בידי מתכנת או בית-תוכנה ויזמין אצלם תוכנה שתתאים לצרכיו. כך למשל, המילון ההיסטורי של האקדמיה ללשון העברית מיוסד על תוכנות שנכתבו במיוחד לצורכי המילון. תחילה היו אלה תוכנות לצורך מיון בלבד, בהמשך העבודה נכתבו תוכנות המנסות לייחס כל מחרוזת של אותיות (מלה) לאיזה ערך שכבר נבדק במילון. ייחוס מלים לא-מנוקדות מן הטקסט לערך שכבר

קיים במילון עובר גם בדיקה של בלשן מומחה, אך לפי הדיווחים, תוכנות אלה, בשילוב ניקוד חלקי, מצליחות עד שכיום למעלה מ-95% של "ניחוש" המחשב אכן קולעים לערך הנכון.

אופן השימוש הזה במחשב משותף לחוקרים בכל תחומי המדע. השוני שבתכנים אינו "מעניין" את התכנית. תכנית מיון למשל, אדישה לשאלה אם נתבקשה למצוא ולמיין תווים המתפרשים אצלנו כסוג דם של הורים וילדם (A, O וכד'), או שנתבקשה למצוא ולמיין תווים המתפרשים אצלנו כצורני יחסה (A, O וכד'). יתר על כן, אפילו מקור הנתונים, אם באו כאותות אורכי-גל וניתרגמו לאותות חשמליים (סורק-על-קולי), או אם באו דרך הקשה של כתב-יד עתיק על לוח המקשים של המחשב ומומרים גם הם לאותות חשמליים – חד הם עבור התכנית.

מובן מאליו כי תנאי הכרחי לשימוש בתוכנות אלה הוא נגישות המחשב לנתונים. ברוב המחקרים הלשוניים מדובר בנגישות המחשב לטקסטים. הדרך המקובלת להזין את זיכרון המחשב בטקסטים היא באמצעות אדם המקיש אותם על לוח המקשים. בשיטה זו משתמשים עד היום להכנסת טקסטים עבריים למחשב, כך בעבודת המילון ההיסטורי, כך בפרויקט השו"ת באוניברסיטת בר-אילן, וכך למעשה גם במשרד המשפטים, אלא שכאן רוב החומר שמבקשים למיין מוקלד ממילא במחשב, וזאת על-ידי הקלדנים של בתי המשפט או אפילו על-ידי השופטים עצמם. דרכים שכיחות פחות הן הזנה על-ידי סורק אופטי או על-ידי קריאה בקול. שיטות אלה כמעט שאינן בשימוש בלשון העברית. גם קלטים אלה מיתרגמים לאותות חשמליים. מובן מאליו כי אם בתקופה שלפני המחשב די היה בניתוח כמה עשרות דוגמות כדי להגיש עבודה סמינריונית או כדי לפרסם מאמר בכתב-עת, הרי היום נפוצות המחשב והעברת הדגש מאגירה ומיון למתודולוגיה ותאוריה הגדילו את היקף הדוגמות הנדרשות עשרות מונים. אולם, יכולת זו מותנית בזמינות של טקסטים גדולים על-גבי סרט מגנטי או דיסקה. בתחום הלשון העברית עדיין עניים אנו בכך. עם זאת, יש לציין כי מאגרי טקסטים מתרבים והולכים עם ריבוי השימוש במחשב למיון ולשליפה בעבודות מוסמך ודוקטור בכל הלימודים הטקסטואליים.

משהוכנסו לזיכרון המחשב הנתונים הדרושים, אפשר להפעיל את התוכנה הרצויה: מיון, שליפה, אחזור וכד'. פעולות אלה מספקות ללשונאי, לחוקר הספרות, לעורך-הדין, לספרן וכד' קונקורדנציות אלקטרוניות, חיתוך וארגון

של הנתונים לפי מפתחות שונים וכד'. בהינף יד יכול הלשונאי לקבל, למשל, את כל הקטעים שבהם במשפט אחד ובמרחק שאינו עולה על שלוש מלים באות שתי המלים **מחשב** ו**עברית**.

נפוצותו הרבה של המחשב במרכזי המחקר ובקרב כלל הציבור משנה בהדרגה את אופיים של ספרי העזר והופכת אותם לכלי עזר ממוחשבים. הדבר ניכר כבר בקטלוגים הממוחשבים המצויים ברוב הספריות האוניברסיטאיות. אך גם בספרי העזר הלשוניים מתרחש השינוי: אף-על-פי שהמחשב מסוגל להפיק מילון שלם או קונקורדנציה שלמה ולהדפיס בדפוס מרשים בזמן קצר ביותר, זמינות החומר והמחשב מאפשרים למשתמש לקבל בכל פנייה רק את המלה או המלים המסוימות הנדרשות לו באותה עת בלא צורך להציג או להדפיס את כל יתר המלים. לחלופין, מאחר שאין המשתמש מוגבל להדפסה יחידה וחד-פעמית של כל החומר, הוא יכול לבצע שאילתות מורכבות כצירופי מלים כאמור לעיל. המשתמש יכול גם לשאול על היעדר מלה, כגון: היכן מופיעה המלה **אוני** במקרא שלא בצירוף המלה **בן**?

בהינף יד יכול הלשונאי להשוות טקסטים: המחשב יציג לו את השינויים, התוספות והגריעות בצורה גרפית ברורה ובהירה. תכונות המחשב כפי שנמנו כאן הפכוהו כלי אהוד מאוד על לשונאים, אנשי ספרות ואף מוסיקולוגים המתעניינים באפיון סגנונו של מחבר. הלשונאים מחפשים אוצר מלים מסוים, אפיונים תחביריים מסוימים וכד'. יתר על כן, תוכנות שונות שפותחו לאפיונים אלה משמשות את החוקרים לגילוי מחברים של יצירות אנונימיות. מאחר שאין לנו עדיין טקסטים גדולים של ספרות עברית כתובים במחשב, תחום זה עדיין אינו מפותח בעברית.

לעומת זאת, בארצות רבות, ובעיקר בארצות דוברות אנגלית, כבר נתבססו מאגרי ידע ומאגרי נתונים והם נמכרים לכל דורש. כשם שכבר הורגלנו אנו או הורגלו ילדינו לקנות דיסקות של משחקי מחשב, כך ניתן לקנות בארצות האלה את אחד ממילוני אוקספורד, תיזאורוס וכד'. וכפי שהבהרנו, תנאי הכרחי להפעלת התוכנות הוא קיום הטקסטים בזיכרון המחשב. לעומת זאת, חוקר הלשון העברית עומד מול חסר יסודי במאגרי נתונים. אין בנמצא מאגרים מסחריים, ומוסדות מחקר שבידיהם טקסטים גדולים או מילונים, משתמשים בהם בעיקר לצורך מחקריהם שלהם. החסר הזה מעכב מאוד את חוקרי העברית והסטודנטים. סטודנט הלומד אנגלית

יכול, למשל, לרכוש דיסקה ובה כל שירי בלייק, לעבדם במחשב כרצונו ולהגיש עבודה סמינריונית המסתמכת על כל החומר. בעברית אנחנו רחוקים ממצב כזה. מאגרי הטקסטים הגדולים בעברית מרוכזים במילון ההיסטורי של האקדמיה ללשון העברית, במפעל השו"ת ובמכון לאחזור מידע המתנהלים באוניברסיטת בר-אילן (בראשות פרופ' יעקב שויקה) ובפרויקט תרגום השבעים (בראשות פרופ' עמנואל טוב) החוג למקרא באוניברסיטה העברית. אך עדיין אין החומר הזה עומד חופשי כחומר גלם לרשות הציבור. במיוחד מורגש חסר במילון ובתיזאורוס עבריים ממוחשבים. מובן שכאן, כבבעיית התרגום שנדונה לעיל, אין להתעלם מן הקושי שמעמיד הכתב העברי בכל עיבוד של טקסט בלתי מנוקד. רב-המשמעות שבכתב מונעת אפשרות טובה של שליפה אוטומטית (בלא עריכת מומחה) של ערכי קונקורדנציה מטקסט. ברוב המקרים גם ניסוח שאילתה המורכבת משתי מלים אינו מסלק את ריבוי הקריאות. בדוגמה שהבאנו לעיל ניכר כי בבקשנו את שילוב המלים **בִּקֵּר וְהָפְרוֹת** נקבל גם את המלים **בִּקֵּר וְהָפְרוֹת**. כפי שהוזכר לעיל, גם כוחו האדיר של המחשב וגם והתוכנות המחוכמות שנכתבו לפעולות אלה עדיין לא פתרו את הקשיים שמעמיד הכתב העברי.

2. תרומת מדעי המחשב לבלשנות

מדעי המחשב והבינה המלאכותית המסונפת אליהם נתנו דחיפה אדירה לחקר הלשון, לא רק על-ידי כוח האדם והמשאבים הרבים שהם מייחדים לנושא זה: חברות לתעשייה עתירת ידע ויחידות מקבילות בצבאות העולם מבצעות מחקרים לשוניים בשיטות של מדעי המחשב והבינה המלאכותית, ועל-ידי חוקרים שהוכשרו אצלם. החברות אף תומכות במחקרים לשוניים הנעשים בחוגים למדעי המחשב ולבינה מלאכותית באוניברסיטאות ובמכללות טכנולוגיות. יש מעריכים שמספר חוקרי הלשון בחוגים לשוניים שקול למספרם בחוגי מדעי המחשב והתעשיות עתירות הידע. יש מעריכים כי יחס זה עתיד להשתנות לטובת מדעי המחשב והתעשיות עתירות הידע. משום כך קורסים, סמינרים ועבודות דוקטור שעניינם עיבוד שפות טבעיות במחשב הם כיום מרכיב קבוע ברוב תכניות הלימוד של החוגים למדעי המחשב. יתר על כן, פיתוח שיטות פורמליות לחקר השפות הטבעיות העמיד גם ספרי לימוד לבלשנות ומקראות, המותאמים לאנשי מדעי המחשב.

הקשר שבין מדעי המחשב והבינה המלאכותית לבין חקר הלשון בא משלושה כיוונים: האחד – תאוריה ושיטות של מדעי המחשב; השני – הדמיה; והשלישי – "מכשור לשוני".

תאוריה ושיטות של מדעי המחשב

מאמץ מחקרי רב מושקע במחקר ובפיתוח של שפות המחשב, ושל כלים להערכת שפות אלו, היינו הוכחת תכונותיהן המתמטיות והוכחת דקדוקיותן. מאמץ זה בתחום מדעי המחשב הצמיח תאוריות על מבנה של שפות וכלים להוכחת תכונותיהן. תאוריות וכלים אלה עוסקים בשפות ובמובן זה הם בלשניים, ואולם היות שהם פותחו ברובם על-ידי מתמטיקאים חסרי השכלה בלשנית מסורתית, נוצרו כאן תאוריות וכלים בלשניים חדשים שונים מן המקובל בבלשנות המסורתית. הערת אגב – הקורא יתעניין ודאי לדעת, כי בקרב שפות המחשב הנפוצות כבר צמחו להגים שונים. למשל, לשפה לעיבוד אותות ורשימות Lisp נקבע נוסח אחד כ-Lisp (Commonlisp Standard) ויש גם Franzlisp, Interlisp ועוד.

המהפכה שהביאה ליצירת בלשנות חדשה של השפות הטבעיות ינקה רבות מניצול המודלים והכלים המתמטיים שפותחו ונוצלו במדעי המחשב. נקודת הציון המרכזית שבה הוא ניסיונו של נועם חומסקי, בלשן ומתמטיקאי, לתאר את הלשון הטבעית לפי מודלים של השפות הפורמליות (הכוללות את שפות המחשב). מיד בראשית דרכו זו, ניסח חומסקי את היררכיית השפות הקרויה על שמו Chomsky Hierarchy (שפות תלויות הקשר וחסרות הקשר על סוגיהן), והיא משמשת להבחנה יסודית בכל עשייה ומחקר בבלשנות המודרנית. גם רעיון ה'רשת הסמנטית', היינו גרף שצמתיו מייצגים עצמים וקשתותיו מייצגות יחסים, בא מתחום ייצוג הזיכרון במדעי המחשב. גישת מדעי המחשב והמהפכה שבאה עימה יצרו בלשנות השואלת שאלות וחוקרת נושאים שלא העסיקו את אנשי המדע באסכולה הקודמת.

הדמיה

המפגש השני של מדעי המחשב והבינה המלאכותית עם הבלשנות הוא הניסיון לפתח ולהדמות (סימולציה) במחשב מודלים של פעולת המוח. הכוונה לפיתוח מחשבים (חומרה ותוכנה) המחקים את אופן פעולת המוח האנושי (מחשבים מקביליים, רשתות-ניורונים ועוד). מודלים אלה עוסקים

בתפקוד המוח בכלל, אך חקר ההתנהגות הלשונית תופס מקום נכבד בפיתוח המודלים והדגמתם בהדמיה.

"מכשור לשוני"

הכיוון השלישי הוא מחקר ויישום של "מכשור לשוני". תחת כותרת זו ניתן לכנס שני סוגים של מוצרים ומחקרים: (א) מכשור עתיר ידע שנועד להשלים עבור האדם המוגבל תפקוד לשוני לקוי; (ב) מרכיב השיח של מערכות מומחה.

א) הזכרנו בפתח דברינו מכונות מפיקות דיבור מלאכותי. מכונות אלה יכולות לקבל קלטים בצורות שונות, למשל, הן יכולות לקבל קלט מקורא אופטי במכונות קריאה בשביל עיוורים, או מהקשה שמקיש על לוח המקשים אילם המבקש לשוחח בטלפון (או אנשים אחרים לצרכים אחרים). המחקר כאן נועד ליצור מוצר ולא לבנות מודל המחקר תהליך אנושי (כפי שהוזכר לעיל). מכשור כזה, לאחר שפענח את סימני הכתב, מעבדם לתעתיק פונטי או לדרך ייצוג אחרת, שעל פיה תתחברנה יחידות השמע למלים.

חלק זה של מעבר מסימני כתב אל התעתיק הפונטי הוא חלק בעל אופי מחקרי לשוני מובהק. היקפו משתנה משפה לשפה: מובן שיש הבדל ניכר מבחינה זו בין איטלקית לאנגלית. ואולם בעברית כאמור, הבעיה חמורה פי כמה. כפי שראינו במשפט הפרות בעלות הבוקר חוזרות ורבות, לשם הפקה של מכשור כזה בעברית לא די בחוקי ביצוע פונולוגיים. יש להוסיף למערכת גם חוקי תחביר, סמנטיקה ופרגמטיקה שיצמצמו את אפשרויות הקריאה. פתרון חלקי זה יבוא על חשבון פשטות המערכת, ולעתים תאבד גם נייחותה, שהיא תכונה הכרחית לשם ניצולה המלא על-ידי האדם המוגבל. מחקר בסיסי בנושא זה בעברית נערך בשנת תשמ"ו במרכז המדעי של יבמ ישראל.

ב) מרכיב תוכנה מיוחד, המנהל שיחה בשפה טבעית עם משתמש אנושי, הוא הממשק הלשוני של חלק ממערכות המומחה, היינו הקישור על-ידי הידברות בשפה טבעית בין המשתמש למערכת המומחה.

מערכת מומחה היא תוכנה אינטליגנטית המיוסדת על אוסף מקיף של עובדות שנאספו מפי מומחים בתחום מסוים, ויודעת להסיק מסקנות וללמוד בתחום המומחיות הזה. מפורסמת, למשל, מערכת Mycin היודעת לערוך אבחון רפואי של מחלות זיהומיות בדם.

המשתמש מפנה למערכת שאלות, והיא עונה לו. יש מערכות המצמצמות את התקשורת המשתמש עם התכנית למתכונת של תפריט, היינו על המשתמש לבחור שאלה או תשובה מתוך רשימה סגורה שהתכנית מציגה. ואולם דרך זו מגבילה את האדם הנועץ בתכנית ואף מרחיקה אותו. משום כך, חברות עתירות ידע המפתחות מערכות מומחה שתשמשה אנשי מקצוע (כגון רופאים או גאולוגים הרחוקים מעולם המחשבים) מקצות משאבים רבים לחקר הבנה ויצירה של שפה טבעית, והן מיישמות את תוצאות המחקרים האלה בממשקים המנהלים שיחה חופשית וטבעית עם המשתמש, למשל, Lunar, מערכת מומחה המסייעת לגאולוגים לנתח ממצאים מהירח. מאחר שמדובר כאן במערכת מסחרית המתכוונת לפעול במקום המומחה, עליה להיות מושלמת: מבחינת ענייננו כאן, נדרשת (מצד המערכת) הבנה מושלמת של שפה טבעית (בתחום אוצר המלים של נושא המומחיות, כמובן). כלומר, בניגוד לדוגמות קודמות של מחקר אקדמי גרידא, כאן אין כל ערך לתוכנה תחבירית מצוינת אם התוכנה הסמנטית לקויה. כאן על כל הרכיבים להתחבר יחד לחלק אחד מושלם. עוד ברור כי שלמות הביצוע מחייבת טיפול בתופעות שאינן דווקא בגבולות משפט אחד, כמו למשל אנפורה או כינויים. עדיין אין בעברית מערכת מומחה המשוחרת שיחה חופשית, אם כי נעשית עבודה בכיוון זה.

מחקר בהיקף כזה ובשלמות כזו, הנעשה בחברות תעשיות, מפתח גישות חדשות של חקר הלשון ומודלים חדשים של הבנת שפה טבעית. כמו במכשור הלשוני שלעיל, מודלים אלה אינם בהכרח מודלים של הדמיית המוח: עליהם לתת תוצאות אנושיות אך אין הם חייבים להגיע אליהן באותה דרך.

המרכיב הלשוני וחקר הלשון תופסים מקום מרכזי בתחום הבינה המלאכותית שבמדעי המחשב. אין פלא אפוא, שהמתמטיקאי ואיש המחשבים אלן טיורינג ניסח את המבחן הבא כפתרון לשאלה "האם המחשב חושב?" זהו מבחן שיחה: אדם יושב מול מסוף, ומקושר באמצעותו עם שני בני שיח אחרים: האחד מהם אדם ואילו השני מחשב. האדם מנהל שיחה עם שני בני שיח אלה. אם אין האדם יכול לקבוע מי מבני שיחו הוא האדם ומי המחשב – אפשר לומר כי המחשב הזה הוא בעל בינה.