תרי דבריא (2020)

עדן אזולאי

eden1998e@gmail.com

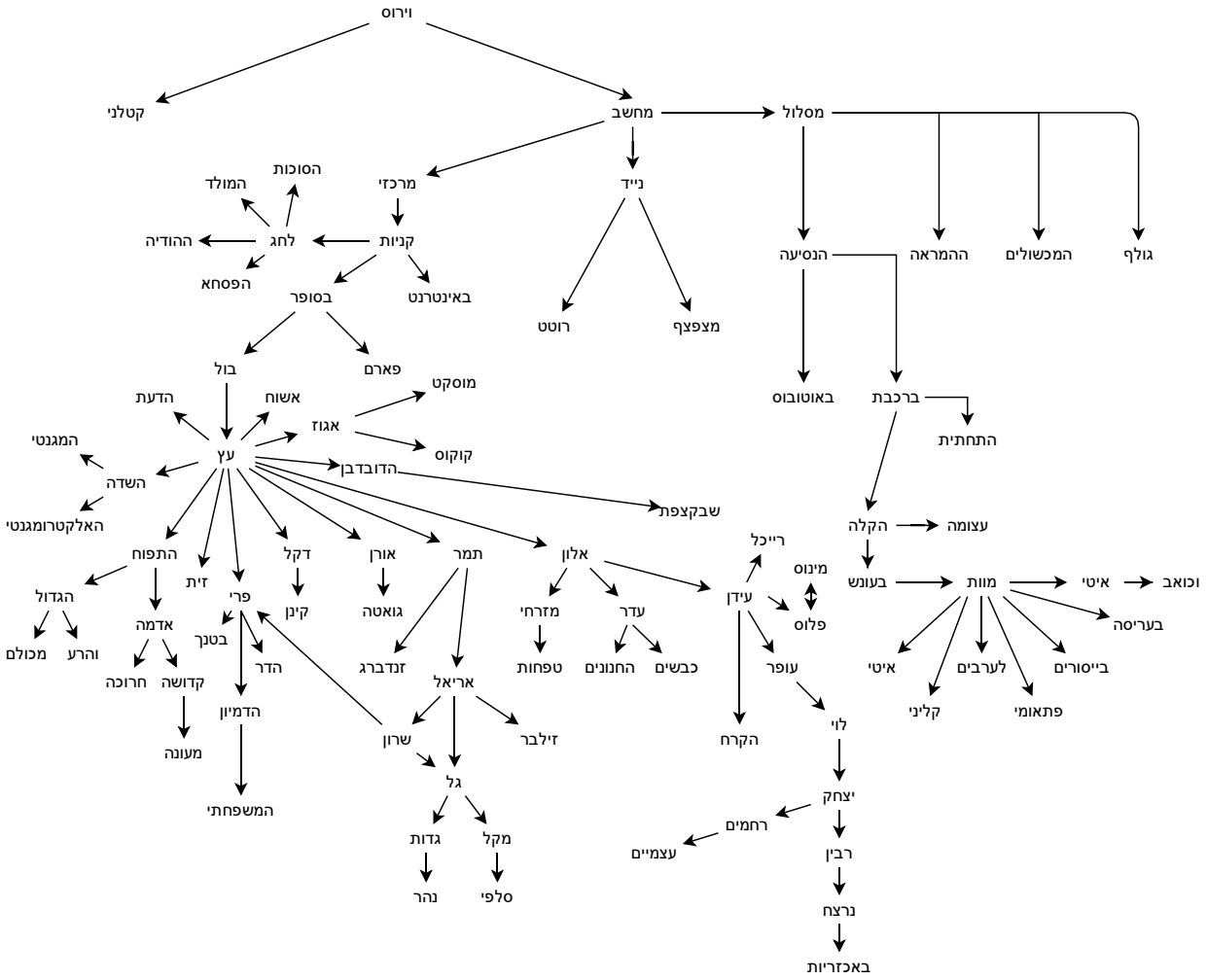
העבודה נעשתה במסגרת שהות-אומן "הפלוגר!", מבית "השוט!" - ניוזלטר ניו-מדיה, בהנחיית איל גרוס.

צירופי מילים בעברית

חילצנו זוגות מילים ממאגר גדול של טקסטים - ציוצים בטוויטר, מילות שירים, וכתוביות מסרטים וסדרות. עיבוד הטקסט וחילוץ רצפי המילים נעשה עם טוקנייזר שפיתחנו. אין זה מספיק למצוא צירופים שחוזרים על עצמם הרבה פעמים, שכן אנו מעוניינים בזוג מילים שהוא "צירוף מעניין", בין השאר בכך שיהווה יחידה סמנטית ורעיונית העומדת בפני עצמה. אחת הבעיות היא סינון של זוגות שהם חלק מביטויים ארוכים יותר. לדוגמה, צירוף שחוזר הרבה הוא "ראש הממשלה בנימין נתניהו", וממנו אנחנו רוצים לקחת את "ראש הממשלה" ו(אולי) את "בנימין נתניהו", ולא רוצים את "הממשלה בנימין". בהקשר זה חקרנו מדדים סטטיסטיים שונים על מנת לכמת את עצמאות הביטוי מן המילים הסובבות אותו. בפרט המצאנו וריאציות של אנטרופיה שמשקללות את האנטרופיה של הסביבה עצמה. כך, ביטוי שתמיד מופיע אחריו סימן מסויים ולכאורה תלוי בו, יכול שייחשב לעצמאי מאותו הסימן אם מדובר בסימן גנרי כמו סימן פיסוק. לאחר סינון של צירופים המכילים מילים מאד שכיחות, צירופים המופיעים מעט פעמים, צירופים בעלי מתאם סטטיסטי נמוך בין חלקיהם, וצירופים שבאופן מובהק אינם עצמאים, נותרנו עם רשימה של כ-45,000 זוגות מילים ומדדיהם, אותם ניתן למצוא כאן: bit.ly/hebig. ניסינו לאזן בין איכות לכמות, ואם חטאנו לצד הכמות, ניתן להיעזר במדדים השונים במקום שנדרשים פחות צירופים שהם יותר איכותיים. למיטב ידיעתנו זהו המאגר המקוון החופשי הגדול ביותר של זוגות מילים בעברית, ואנו תקווה כי ישמש את החוקרים, המפתחים והיוצרים בתחום העיבוד הממוחשב של השפה העברית.

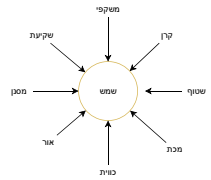
אמנות פלואוץ' (Flowch-Art)

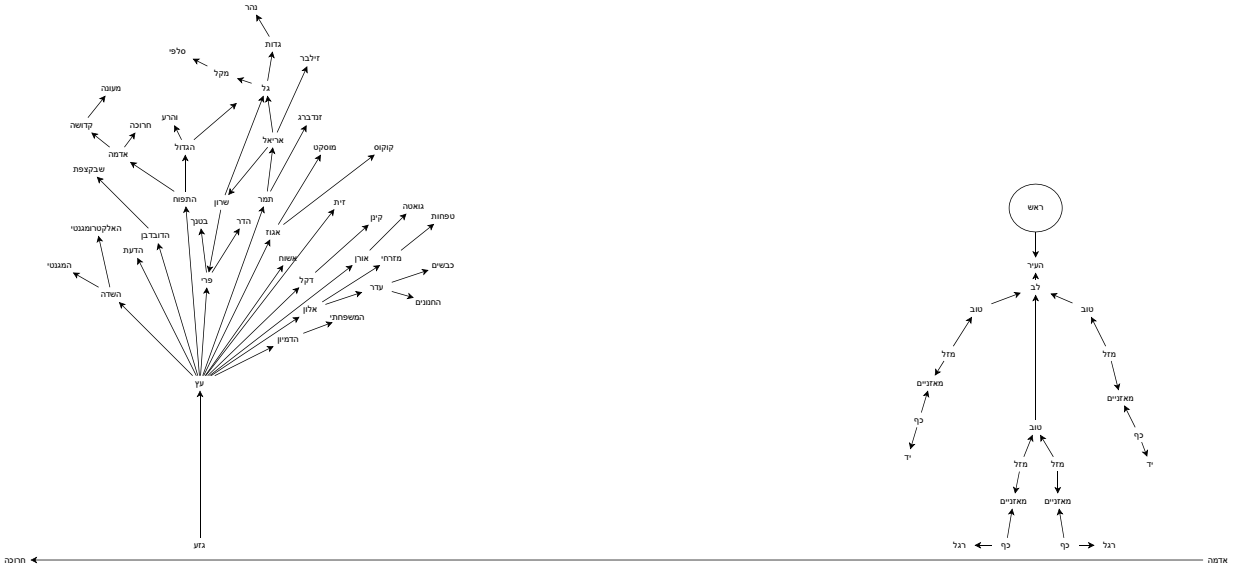
אם ניקח זוג מילים כמו "שומר-ראש", וזוג מילים אחר שמתחיל במילה האחרונה כמו "ראש-העיר", ובן הלאה, נוכל לבנות שרשראות של מילים שמכסות שדות סמנטיים מגוונים ויוצרות אסוציאציות מעניינות. בדוגמה הבאה התחלנו מהמילה "וירוס" והתחקנו אחר "ההתפשטות הסמנטית" שלה דרך שרשראות ביטויים שונות. כך קיבלנו גרף ברוח החקירה האפידמיולוגית המתחקה אחר מגעים של מקרי ההדבקה בוירוס הקורונה.



ניתן גם לצייר גרפים "אוטולוגיים", במובן שהטופולוגיה שלהם משקפת את הרעיון הסמנטי הגלום במילים ובביטויים שמרכיבים אותם. למשל:

- שמש עם קרניים שהן ביטויים שנגמרים במילה שמש.
- גרף-עץ שמתאר את הענפים שיוצאים מהמילה עץ.
- מסלול מ-"כף-רגל" ועד "ראש", שעובר דרך "לב", ויוצר צורה של אדם.





הלחמי-ביטויים

באמצעות אותה טכניקה של חיבור ביטויים בעלי מילה משותפת, אפשר ליצור גם הלחמי-ביטויים. כדי לקבל הלחמים מפתיעים ולא טריוויאלים, דרשנו שהמילים או הביטויים מהם הם מורכבים יהיו בעלי משמעות סמנטית רחוקה (לצורך זה השתמשנו בייצוג מתמטי שלהם שנקרא שיכוני-מילים; word embeddings). הנה כמה דוגמאות שאהבנו:

חכמת רחוב סומסום

תפוחי אדמה חרוכה

דמות אב קדמון

פורצי דרך אגב

אסיר תודה רבה

יחסינו לאן נעלמת

השבח לאל קעידה

נחלת בנימין נתניהו

דב חנין זועבי

במקרים בו אחד הביטויים אינו "עצמאי" כפי שהגדרנו לעיל, אלא תלוי במילה שתמיד מופיעה אחריו (קרי אנטרופיה אפס), אפשר לקבל ביטויים מפתיעים במיוחד כמו:

מוש בן זונה

אפס ביחסי מין

בעבודה עלו עוד רעיונות וכיוונים רבים - מחקריים, שימושיים ואומנותיים, ותוצרים נוספים, כמו גם הקוד, יפורסמו בהמשך.