

מדע בשפה שלנו: מדד ז'רגון להערכת הבהירות בתקשורת המדע

אילת ברעם-צברי
המחלקה לחינוך למדע וטכנולוגיה, הטכניון
ayelet@technion.ac.il

אביב שרון
המחלקה לחינוך למדע וטכנולוגיה, הטכניון
aviv.sharon@gmail.com

Measuring Mumbo-Jumbo: A Preliminary Quantification of the Use of Jargon in Science Communication

Aviv J. Sharon
Department of Education in Science and
Technology, Technion – Israel Institute of
Technology

Ayelet Baram-Tsabari
Department of Education in Science and
Technology, Technion – Israel Institute of
Technology

Abstract

A worldwide educational effort, supported by leaders of the scientific community, aims to teach scientists effective science communication, including honing the skill to discuss science with less professional jargon. However, avoiding jargon is not trivial for scientists for several reasons, and this demands special attention in teaching and evaluation. Despite this, no standard measurement for the use of scientific jargon in speech has been developed to date. Here we propose a computer-assisted method to assess the use of scientific jargon in spoken texts, using freely available online corpora (collections of texts) and freely downloadable software. Analyzed transcripts included academic speech, scientific TEDTalks, and communication about the discovery of a Higgs-like boson at CERN. Findings suggest that scientists use less jargon in communication with a general audience than in communication with peers, but not always less obscure jargon. These findings may lay the groundwork for (self-)evaluating the use of jargon through technology.

Keywords: media and science, science communication, media training, jargon, interaction between experts and publics.

תקציר

מפעל חינוכי חובק-עולם, הנתמך על ידי מנהיגי הקהילה המדעית, שואף ללמד מדענים לתקשר ביעילות עם הציבור הרחב, ובכלל זה, לשפר את מיומנות הדיבור על נושאים מדעיים תוך מיעוט בשימוש בז'רגון מקצועי. עם זאת, זוהי מיומנות קשה ללמידה עבור מדענים, והיא דורשת התייחסות ראויה בהערכה במסגרות להוראת תקשורת המדע. חרף זאת, טרם פותח מדד אחיד להערכת השימוש בז'רגון מדעי בדיבור. כאן מוצעת שיטת הערכה ממוחשבת לשימוש בז'רגון, תוך שימוש באוספי טקסטים (קורפוסים) ותוכנות שכולם זמינים באינטרנט בחינם. לצורך פיתוח המדד, נותחו תמלילי שיח אקדמי, הרצאות מדע פופולארי מתוך כנסי "טד" (TED) ותקשורת בנושא גילוי חלקיק דמוי-היגס במאיץ החלקיקים ב-CERN. ממצאי הניתוח עולה כי מדענים משתמשים בפחות ז'רגון כשהם מתקשרים עם קהל רחב מאשר בתקשורת עם עמיתיהם, אך לא תמיד בז'רגון בהיר יותר. ממצאים אלו עשויים להניח תשתית להערכת השימוש בז'רגון במסגרות להכשרה לתקשורת המדע, ובכלל זה הערכה עצמית, בעזרת הטכנולוגיה.

מילות מפתח: מדע בתקשורת, תקשורת המדע, מיומנויות תקשורת, ז'רגון, אינטראקציה בין מומחים לציבור.

מבוא

הקהילה המדעית מכירה יותר ויותר בחשיבות של שיפור דפוסי התקשורת שלה עם הציבור הרחב (להלן "תקשורת המדע"), ומקצה משאבים הולכים וגדלים להוראת מיומנויות תקשורת מתאימות לכך בקרב מדענים (Nisbet & Scheufele, 2009). בין היתר, הספרות בנושא מנחה את המדענים להמעיט בשימוש בז'רגון מדעי כאשר הם פונים לקהל רחב, כדי שדבריהם יהיו מובנים וכדי שהם יתקבלו באהדה רבה יותר (לדוגמה: Meredith, 2010; Dean, 2009; Baron, 2010).

למרות זאת, עד כה כמעט לא ניתנה הדעת לפיתוח שיטות הערכה ומדידה לתוצאות ההכשרה בתחום זה בכלל, ולשימוש הלומדים בז'רגון בפרט (Author, 2012). מיחשוב ההערכה עשוי להקל על הערכה אחידה, אוטומטית וחדשנית של תוצרי הלמידה (Conole & Warburton, 2005). כך למשל, כדי לקבוע באופן מהימן עד כמה מילה בה משתמש מדען היא "ז'רגונית", מוצע להשוות מילים וטקסטים עם דגימות של תמלילי תקשורת המדע שכבר נותחו באופן ממוחשב, על מנת לפתח "מדד ז'רגון" אחיד לשימוש במסגרות להכשרה לתקשורת המדע.

סקר ספרות

למידה והערכה של תקשורת המדע

קיימות בעולם מסגרות הכשרה רבות לתקשורת המדע, במסגרות אקדמיות ולא-אקדמיות כאחד, לאורך תקופות הכשרה שונות. תכני הליבה של תוכניות ההכשרה כוללים מיומנויות תקשורת הנדרשות בין מדענים (הכנת מאמרים, מצגות, פוסטרים) וכן התמודדות עם תקשורת ההמונים (ניסוח הודעות לעיתונות, עריכת ראיונות ברדיו ובטלוויזיה וכדומה). (Mulder, Longnecker, & Davis, 2008). הערכת תוכניות תקשורת המדע נעשית בדרך כלל בשיטות המתמקדות בעמדות ותחושות הלומדים על החוויה, ללא הערכת המיומנויות שנרכשו (Gascoigne & Metcalfe, 1997; Miller, Fahy, & The ESConet Team, 2009).

ז'רגון מדעי: הגדרה, איפיון והערכת השימוש

משלב (register) הוא אוסף מאפיינים לשוניים שונים, כגון הגייה, מורפולוגיה, אוצר מילים, תחביר ומאפייני שיח, המשתנה בתלות בנסיבות חיצוניות כגון תפקידי המשתתפים בשיחה והיחסים ביניהם, מטרת התקשורת ביניהם ועוד (Biber, 1988, 1995). כך, ניתן להגדיר "משלב מדעי" של שפה כלשהי כמשלב בו משתמשים מדענים כאשר הם מדברים על מדע עם עמיתיהם ותלמידיהם. למשל המדעי מאפיינים לשוניים שונים, כולל אוצר מילים ייחודי, הוא הז'רגון המדעי.

אנו טובעים את המושג "ז'רגוניות" לצורך אפיון הדרגה בה השימוש במילה מסוימת מוגבל למשל המדעי. כלומר, ככל שמילה ז'רגונית יותר, כך היא מצויה פחות בשפה הכללית ומוכרת פחות למי שאינם מדענים.

הז'רגון המדעי הוא כלי מנטלי חשוב למדענים, והם נעזרים בו במהלך עבודתם לייצוג סכמות מנטליות, להמשגה של עובדות ותגליות חדשות ולתקשורת יעילה עם עמיתים (Grupp & Heider, 1975; Jucks, Schulte-Löbber, & Bromme, 2007).

ההימנעות מז'רגון היא משימה מאתגרת עבור מדענים, שהורגלו בשימוש בשפה מקצועית, כחלק מהכשרתם ותהליך החיברות¹ שלהם כאנשי מדע (Lemke, 1990). בנוסף, אנשים נוטים להניח בטעות כי גם אחרים יודעים מה שידוע להם, ו"קללת ידע" זו תקפה גם לגבי ז'רגון (Hayes & Bajzek, 2008). מכאן, ההימנעות מז'רגון דורשת מאמץ מודע ומכוון לתקשר בצורה בהירה, וזוהי מיומנות נרכשת הדורשת ידע וניסיון (Stableford & Mettger, 2007).

ז'רגון מדעי נמדד בעבודות קודמות בעזרת כלים כגון משוואות קריאות (readability formulas) וניתוח אוצר המילים בעזרת רשימות מילים סגורות וקצרות יחסית (Ley & Florio, 1996), אך כלים אלה אינם יכולים לשקף את היקף השימוש במילות ז'רגון רבות.

¹ סוציאליזציה, הקניית דפוסי וערכי התרבות של החברה

לפתרון בעיה זו, אנו מציעים להיעזר בשיטה אחרת מתחום הבלשנות החישובית להערכת השימוש בז'רגון מדעי, בעזרת קורפוסים של אנגלית כללית ואנגלית מדעית. **קורפוס** הוא אוסף טקסטים טבעיים, של שפה דבורה או כתובה, המייצג שפה כלשהי, (או להג או משלב של שפה) (McEnergy, 2006; Xiao, & Tono, 2006). נהוג להשוות בין אוצרות המילים בקורפוסים שונים באמצעות מדידת הבדלים בתדירויות המילים ביניהם. **תדירות מילים** היא מספר המופעים של מילה בטקסט או בקורפוס (Paquot & Bestgen, 2008); תדירות מנורמלת היא מספר מופעי המילה למיליון מילים בקורפוס (McEnergy et al., 2006).

לשם "כיוול" כלי הערכה לז'רגון בשפה האנגלית, ביקשנו למדוד תדירויות מילים באוספי תמלילים שונים על מנת לתאר הבדלים באופן השימוש בז'רגון באנגלית בין: (1) תקשורת מדעית בין מדענים, לבין (2) דוגמאות מוצלחות לתקשורת המדע.

שאלות מחקר

(1) כמה ז'רגון יש (= מה שיעור מילות הז'רגון מכלל המילים) בדוגמאות מוצלחות של תקשורת המדע, והאם יש הבדל בין שכחות הז'רגון בדוגמאות אלו לשכחות הז'רגון בתקשורת בין מדענים?
 (2) מה שיעור הז'רגוניות של מילות הז'רגון (= מה השכיחות היחסית של כל מילת ז'רגון בשפה מדעית לעומת השפה הכללית) בדוגמאות מוצלחות של תקשורת המדע, והאם יש הבדל בין שיעור הז'רגוניות של מילות הז'רגון שבדוגמאות אלו לשיעור הז'רגוניות של מילות הז'רגון שבתקשורת בין מדענים?

השערות מחקר

(1) דוגמאות מוצלחות לתקשורת המדע כוללות שיעור קטן יותר של מילות ז'רגון מכפי שכוללת תקשורת בין מדענים;
 (2) מילות הז'רגון שמופיעות בדוגמאות מוצלחות של תקשורת המדע הן ז'רגוניות פחות (שכיחות יותר בשפה הכללית) ממילות הז'רגון שמופיעות בתקשורת בין מדענים.

מתודולוגיה

איסוף דגימות וניתוחן

שלושה אוספי תמלילים של שפה דבורה באנגלית שמשו במחקר זה, וכולם נאספו ממקורות הזמינים לציבור בחינם באינטרנט:

(1) תמלילי מפגשים מדעיים אקדמיים מאוניברסיטת מישגן שבארה"ב, כולל הרצאות, דיונים, פגישות קבוצות מחקר ועוד ("אקדמיה", 43 תמלילים, 487,671 מילים במצטבר) (Simpson, Briggs, 2002; Ovens, & Swales, 2002). אירועים אלו מהווים חלק מעבודת המדענים או מההכשרה המקצועית של מדענים לעתיד, ולכן שימשו כדוגמאות לתקשורת בין מדענים.

(2) תמלילי הרצאות מדע פופולארי מתוך כנסי "טד" (TED) בשנים 2010 ו-2011 ("טד מדע", 31 תמלילים, 69,290 מילים במצטבר). "טד" היא עמותה העורכת שני כנסים בשנה בקליפורניה ובסקוטלנד בנושא "רעיונות שראוי להפיץ", בתחומים כגון טכנולוגיה, עיצוב, כלכלה, מדע, חינוך ועוד. מעל 1,200 סרטוני הרצאות מהכנסים הועלו לאתר הכנס ומתוכם שיעור סרטוני המדע עולה על רבע (TED Conferences, LLC, 2012). הסרטונים מושכים מיליוני גולשים מהציבור לצפות בהרצאות על מדע מרצונם החופשי (Kessler, 2011), ולכן תמלילי הרצאות אלו נבחרו לשמש כדוגמאות מוצלחות לתקשורת המדע בין המדענים לציבור.

(3) תמלילי הרצאות פופולאריות, שאינן מדעיות ועוסקות בעיצוב, מתוך כנסי "טד" באותן השנים שמהן נאספו תמלילי טד מדע ("טד עיצוב", 28 תמלילים, 53,780 מילים במצטבר). קבוצה זו של טקסטים שימשה כבקרה.

בקרה חיצונית

כבקרה חיצונית, בוצעה חזרה נוספת על השיטה על מדגמים מתוך שני תמלילים מאירועים בהם נמסר מידע על גילוי חלקיק דמוי-היגס ביום 4 ביולי, 2012, במרכז המחקר CERN, אשר הוקלטו והועלו לאתר האינטרנט של המרכז:

(1) שני סמינרים מדעיים על התגלית, שניתנו על ידי פרופ' ג'ו אינקנדלה (Incandela) וד"ר פביולה גיאנוטי (Gianotti), ונמענו לקהל פיזיקאי חלקיקים (דוגמה לתקשורת בין מדענים);

(2) התבטאויות שני הדוברים הללו במסיבת עיתונאים מיד לאחר הסמינרים, שנמענו לקהל רחב (דוגמה לתקשורת המדע).

שלב ראשון: סינון על פי שכיחות בשפה

בתחילה, אוחרו מתוך הטקסטים מילים שנחשדו כז'רגון, משום שלא השתייכו לרשימת 9,000 משפחות המילים השכיחות ביותר בשפה האנגלית (לדוגמה, משפחת care, הכוללת את cared, carer, carers, careful, carelessly, uncared, uncaring). מספר משפחות זה נבחר משום שידוע כי דרושות 8,000 עד 9,000 משפחות מילים באנגלית כדי להבין טקסטים כתובים כגון עיתונים, תמלילי סרטים ורומנים (Nation, 2006). מילים שלא נכללו ברשימות אלו, ולכן סווגו כ"חשודות", כללו מגוון רחב של מילים נדירות יחסית בשפה האנגלית. חלקן מדעיות, אך לא כולן.

הסינון בוצע בעזרת התוכנה AntWordProfiler (Anthony, 2009) ורשימות המילים המצורפות לתוכנה Range (Heatley & Nation, 1994), שכולן זמינות להורדה חינם באינטרנט. בנוסף, התוכנה AntWordProfiler מייצרת גם סטטיסטיקות על התפלגות המילים בטקסט בתבניות (types; סוגי מילים; לא כולל חזרות) ובתקריות (tokens; מופעי מילים, כולל חזרות). כך למשל, בעזרת ההבחנה בין תבנית לתקרית ניתן לומר כי במשפט "ורד הוא ורד הוא ורד" יש שש תקריות אך רק שתי תבניות ("ורד" ו-"הוא").

שלב שני: מיון על פי שכיחות יחסית בטקסטים מדעיים וכלליים

כל המילים החשודות שנמצאו בשלב הראשון נבדקו בכדי להפריד את המילים המקצועיות מתחום המדע משאר המילים הנדירות. לצורך המיון, חקרנו את שכיחותה של כל אחת מתבניות המילים ה"חשודות" בשני קורפוסים: (1) מאמרי מחקר מדעיים באנגלית (מאגר PERC², "הקורפוס המדעי"), ו-(2) מאגר כללי של השפה האנגלית הבריטית על כל רבדיה (BNC³, "הקורפוס הכללי").

אם תבנית מילה הופיעה בשני הקורפוסים, איחזרנו את יחס תדירויות השימוש בה בכל אחד מהם. כך, למשל, המילה thermocline הופיעה בשני הקורפוסים, אך שכיחות מופיעה גדולה יותר בקורפוס המדעי מאשר בקורפוס הכללי.

חילקנו את תבניות המילים לחמש קטגוריות לפי דפוס הופעתן בשני הקורפוסים:

- (A) מילים שהופיעו רק בקורפוס המדעי (למשל metalloproteases) – הוכרו כמילות ז'רגון,
- (B) מילים שהופיעו בשני הקורפוסים, אך בתדירות מנורמלת גבוהה יותר בקורפוס המדעי (לדוגמה, thermocline),
- (C) מילים שהופיעו בשני הקורפוסים, אך בתדירות מנורמלת גבוהה יותר בקורפוס הכללי (לדוגמה, honeycombs),
- (D) מילים שהופיעו רק בקורפוס הכללי (כגון whimsical),
- (E) המילים שלא נמצאו באף מאגר (כגון facebook, infographics, polytheism, osteocytes) אך גם.

לאחר מכן, תבניות המילים בקטגוריה B מוינו פעם נוספת, לפי תוצאת מבחן מובהקות להבדל בין תדירויות כל מילה בשני הקורפוסים. מבחן המובהקות שבו נעזרנו היה Log-Likelihood, מבחן

² <http://scn.jkn21.com/~percinfo/index.html>

³ <http://www.natcorp.ox.ac.uk>

מקובל להשוואה בין תדירויות מילים בבלשנות חישובית (Dunning, 1993). תבניות מילים שנמצאו בתדירות גבוהה יותר בקורפוס המדעי ועבורן $p < 0.05$, הוכרו כמילות ז'רגון (תת-קטגוריה B1) ושאר המילים בקטגוריה B – לא (תת-קטגוריה B2).

לסיכום, תבנית מילה שעמדה באחד משני הקריטריונים הבאים הוגדרה כמילת ז'רגון: (1) הופיעה בקורפוס המדעי אך לא בקורפוס הכללי (קטגוריה A), או (2) הופיעה בשני הקורפוסים אך בתדירות גדולה יותר **במובהק** בקרב מאמרים מדעיים (תת-קטגוריה B1).

שלב שלישי: הקצאת ציוני ז'רגוניות למילות הז'רגון

לבסוף, לכל תבנית מילת ז'רגון (בקטגוריות A או B1) ניתן ציון ז'רגוניות. ציון הז'רגוניות למילים בכל אחת מהקטגוריות חושב בדרך שונה:

(1) אם המילה הופיעה לפחות פעם אחת בקורפוס הכללי (קטגוריה B1), הז'רגוניות חושבה כלוגריתם (בסיס 10) של יחס התדירויות של המילה בקורפוס המדעי ובקורפוס הכללי. פונקציית הלוגריתם נבחרה משום שהיא משקפת את סדר הגודל של יחס התדירויות: אחדות (10^0), עשרות (10^1), מאות (10^2) או אלפים (10^3). כך, למשל, המילה thermocline שהופיעה בשני הקורפוסים, אך בשכיחות גדולה פי 444 בקורפוס המדעי מאשר בקורפוס הכללי, הופיעה **בשני** סדרי גודל יותר (10^2 , פי מאות פעמים) בקורפוס המדעי מאשר בקורפוס הכללי. לכן הציון של המילה הוא מעל **שתיים**: $\log_{10}(444) \approx 2.65$. מילים אחרות, ז'רגוניות יותר, הגיעו אף עד לציונים מעט מעל שלוש.

(2) אם מילת הז'רגון נמצאה רק בקורפוס המדעי ולא בקורפוס הכללי (קטגוריה A), הוענק לה הציון שלוש, עקב הנחה שמרנית שהמילה נפוצה במשלב המדעי בשלושה סדרי גודל יותר (כלומר, פי אלף) מאשר במשלב הכללי. ערך זה נבחר משום שהוא מעט נמוך מציון הז'רגוניות המרבי שהתקבל למילה כלשהי במחקר (ראה פרק "תוצאות").

הנוסחה הבאה מסכמת את החישוב:

$$\text{Jargonness} = \begin{cases} \log_{10} \left(\frac{\text{frequency}_{\text{scientific}}}{\text{frequency}_{\text{general}}} \right) & (\text{frequency}_{\text{general}} > 0) \quad \text{Category B1} \\ 3 & (\text{frequency}_{\text{general}} = 0) \quad \text{Category A} \end{cases}$$

תוצאות

שלב ראשון: סינון על פי שכיחות בשפה

גם שיעור **תבניות** המילים החשודות וגם שיעור **תקריות** המילים החשודות (מספר הפעמים שהופיעו, כולל חזרות) השתנו בין קבוצות הטקסטים שנותחו (שני מבחני פרופורציה תלת-מדגמיים, בכל אחד $p < 0.001$). בתמלילי האקדמיה נמצא שיעור המילים החשודות הגדול ביותר, גם בתבניות וגם בתקריות (טבלה 1).

טבלה 1. שיעורי מילים חשודות (נדירות) בשלושה אוספי טקסטים באנגלית

טד עיצוב (בקרה)	טד מדע (תקשורת המדע)	אקדמיה (תקשורת בקרב מדענים)	
53,780	69,290	487,671	מספר תקריות מילים התחלתי (tokens)
995 (1.85%)	1,439 (2.08%)	12,909 (2.65%)	תקריות מילים חשודות (tokens) (מספר מוחלט ואחוז מתוך המספר ההתחלתי)
5,936	6,578	14,088	מספר תבניות מילים התחלתי (types)
663 (11.17%)	841 (12.79%)	3,636 ^(a) (25.81%)	תבניות מילים חשודות (types) (מספר מוחלט ואחוז מתוך המספר ההתחלתי)

^(a) 65 מילים הושמטו מקבוצה זו כי בוטאו ותומללו רק באופן חלקי.

שלב שני: מיון על פי שכיחות יחסית בטקסטים מדעיים וכלליים
 בתמלילים שנבדקו נמצאו פרופורציות שונות של תבניות מילים ז'רגוניות בקרב המילים החשודות (טבלה 2 ואיור 1). באקדמיה, השיעור הגדול ביותר של תבניות מילים ז'רגוניות (43.3%), בהשוואה ל-37.5% מתבניות המילים החשודות ב"טד מדע" ול-19.2% מתבניות המילים החשודות ב"טד עיצוב" (מבחן פרופורציה תלת-מדגמי, $p < 0.001$).

טבלה 2. התפלגות תבניות מילים חשודות (נדירות) על פי שכיחותן היחסית בקורפוסים (מאגרי טקסטים) מדעיים וכלליים

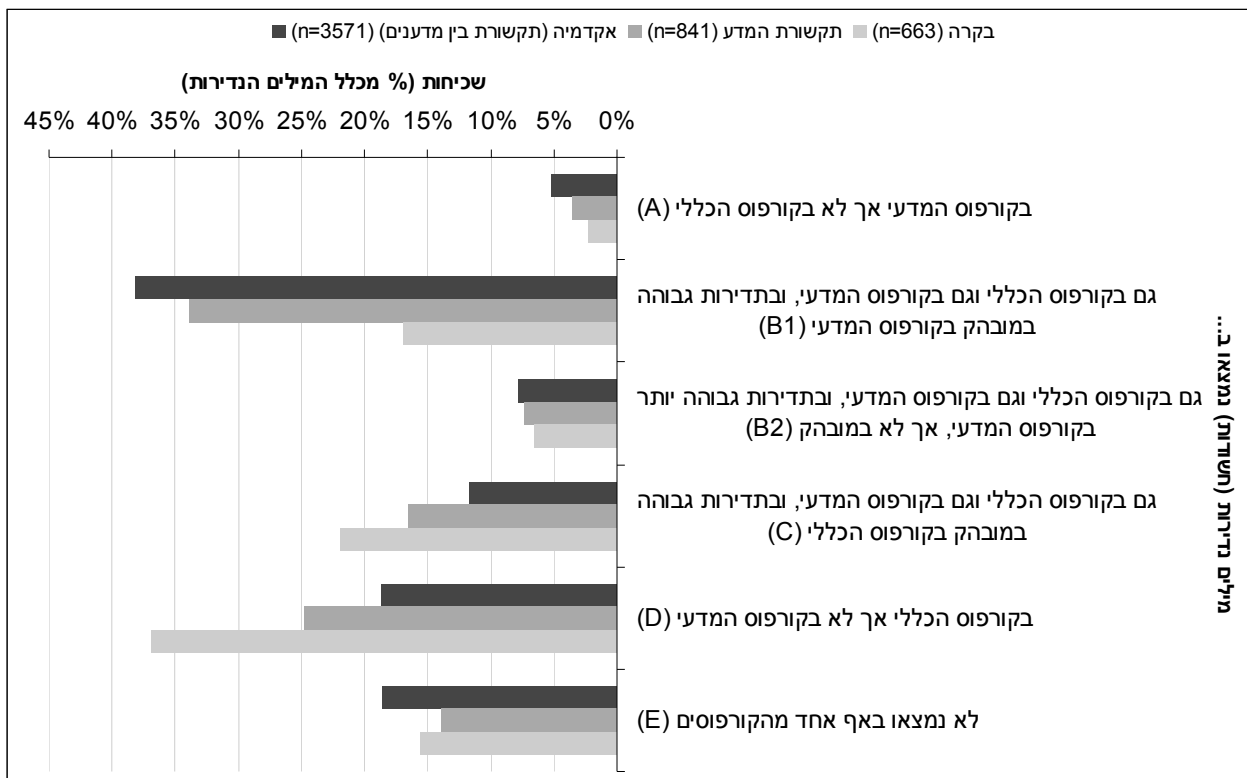
טד עיצוב (בקרה), (n = 663)	טד מדע (תקשורת המדע), (n = 841)	אקדמיה (תקשורת בקרב מדענים), (n = 3571)	דוגמאות	ז'רגון מדעי?	תבניות מילים חשודות שנמצאו...	קטגוריית	טקסטי
15 (2.26%)	30 (3.57%)	184 (5.15%)	Allergenicity, Postsynaptically	כן	בקורפוס המדעי ^(א) אך לא בקורפוס הכללי ^(ב)	-	A
155 (23.38%)	347 (41.26%)	1640 (45.93%)			גם בקורפוס המדעי וגם בקורפוס הכללי, אך בתדירות גדולה יותר בקורפוס המדעי (סה"כ)	-	B
112 (16.89%)	285 (33.89%)	1362 (38.14%)	Ethanol, Photoreceptor	כן	מתוכן במובהק ^(ג) בתדירות גדולה יותר בקורפוס המדעי	B1	
43 (6.49%)	62 (7.37%)	278 (7.78%)	Hallucinogen, Prerecorded	לא	מתוכן בתדירות גדולה יותר בקורפוס המדעי, אך לא במובהק	B2	
145 (21.87%)	139 (16.53%)	419 (11.73%)	Hyperactive, Decaffeinated	לא	גם בקורפוס המדעי וגם בקורפוס הכללי, אך בתדירות גדולה יותר בקורפוס הכללי	-	C
245 (36.95%)	208 (24.73%)	667 (18.68%)	Brunch, Choreography	לא	בקורפוס הכללי אך לא בקורפוס המדעי	-	D
103 (15.54%)	117 (13.91%)	661 (18.51%)	Essentialistic, Velociraptor	-	לא בקורפוס הכללי ולא בקורפוס המדעי	-	E
127 (19.2%)	315 (37.5%)	1,546 (43.3%)		כן	סה"כ ז'רגון מדעי ^(ד)		
433 (65.31%)	409 (48.63%)	1,364 (38.20%)		לא	סך כל המילים שאינן ז'רגון מדעי ^(ה)		

^(א) PERC (Professional English Research Consortium) Corpus

^(ב) BNC (British National Corpus)

^(ג) Log-likelihood, $p < 0.05$.

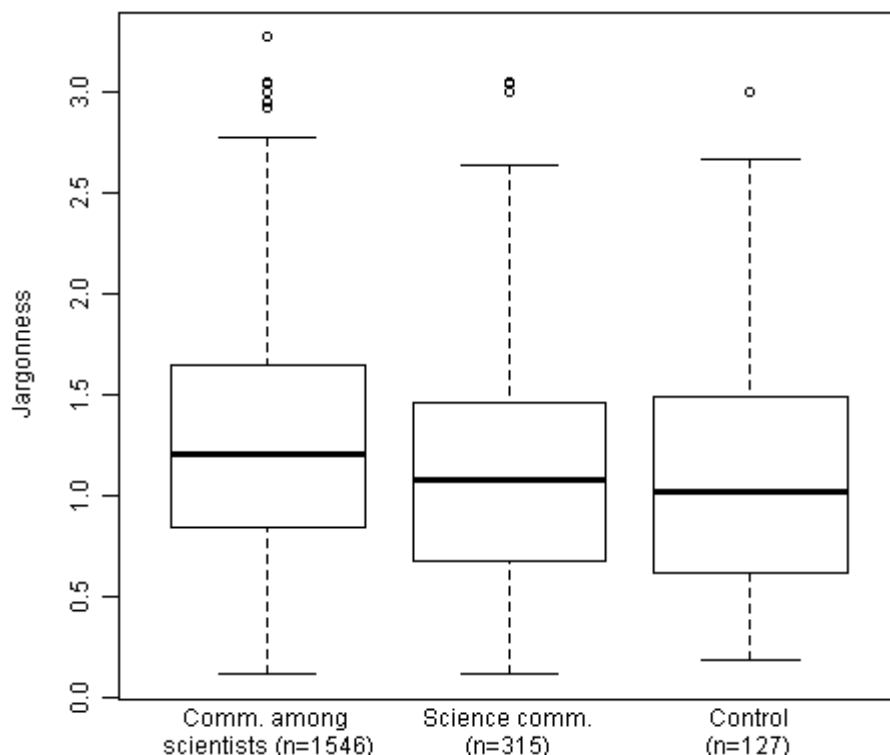
^(ד) סך המילים בקטגוריות A ו-B1.
^(ה) סך המילים בקטגוריות B2, C ו-D.



איור 1. התפלגות תבניות מילים חשודות (נדירות) על פי שכיחותן היחסית בקורפוסים (מאגרי טקסטים) מדעיים וכלליים. תבניות מילים בקטגוריות A ו-B1 נחשבות לז'ירגון מדעי.

שלב שלישי: הקצאת ציוני ז'ירגוניות לתבניות מילות הז'ירגון

תבניות מילות הז'ירגון בתמלילים מהאקדמיה (n = 1546) במובהק "ז'ירגוניות" יותר מתבניות מילות הז'ירגון ב"טד מדעי" (n = 315) (Wilcoxon-Mann-Whitney (WMW) test, p < 0.001), מה שמאשש את השערה מספר (2) (ראה "השערות המחקר"). עם זאת, לא נמצא הבדל סטטיסטי בז'ירגוניות של תבניות מילות הז'ירגון ב"טד מדעי" וב"טד עיצוב" (קבוצת הבקרה) (WMW test, ns). (איור 2).



איור 2. ז'רגוניות של תבניות מילות ז'רגון (קטגוריות A ו-B1). לתבניות המילים בתקשורת בין מדענים ("Comm. among scientists", אקדמיה) יש ז'רגוניות גבוהה יותר מאשר לתבניות המילים בתקשורת המדע ("Science comm.", טד מדע). (Wilcoxon-Mann-Whitney Test, $p < 0.001$)

בקרה חיצונית

בוצעה חזרה על שיטת הניתוח בתמלילי סמינרים מדעיים (1572 תקריות מילים, 473 תבניות) ומסיבת עיתונאים (1,645 תקריות מילים, 501 תבניות) מפי שני דוברים זהים בנושא גילוי חלקיק דמוי-היגס ב-CERN. הסמינרים המדעיים הכילו שיעור גבוה של תבניות מילים חשודות ביחס למסיבת העיתונאים (5.92% לעומת 2.59%; מבחן פרופורציה דו-מדגמי, $p < 0.01$). בשני המקרים היו רוב תבניות המילים החשודות ז'רגון מדעי (82% בסמינרים ו-77% במסיבת העיתונאים). עם זאת, בניגוד למשוער, הז'רגוניות של תבניות מילות הז'רגון הייתה גבוהה יותר במסיבת העיתונאים מאשר בסמינרים (WMW test, $p < 0.05$).

מגבלות המחקר

מהימנות המחקר מוגבלת עקב השימוש באוספי תמלילים קטנים יחסית, שהיו הזמינים ביותר לחקר המשלבים שנבדקו. בנוסף, בעוד תמלילי "טד" מתעדים דיבור שתוכנן ותורגל מראש, בתמלילי האקדמיה יש לצד ההרצאות האקדמיות גם שיחות ספונטניות. לכן נמצאו מילים שבוטאו ותומללו חלקית בתמלילי האקדמיה, וכן עשוי היה להיגרם הבדל בשכיחות מילות הז'רגון.

זאת ועוד, השיטה המוצעת מתעלמת ממאפיינים רבים שעשויים להשפיע על בהירות טקסט מדעי, ובהן הבהרות לז'רגון שהוזכר, צירופי מילים בעלי מובנים ייחודיים במדע (כגון "המפץ הגדול"), ועוד. השיטה גם מעניקה ציוני ז'רגוניות שונים למילים בעלות משמעות זהה אך כתיב שונה (לדוגמה, vapor ו-vapour) ולהטיות שונות של אותה המילה (למשל, algorithm ו-algorithms).

דיון ומסקנות

בקרב תבניות המילים החשודות, נמצא שיעור גבוה יותר של ז'רגון מדעי בתקשורת אקדמית מאשר בתקשורת המדע, מה שמאשש את ההשערה הראשונה.

באשר למידת הז'רגוניות של מילות הז'רגון, התקבלה מהמחקר תמונה מורכבת יותר. במקרה אחד (סרטוני "טד" מול אקדמיה) נמצא גם כי הז'רגוניות נמוכה יותר בתקשורת המדע מאשר באקדמיה, מה שמאשש את ההשערה השנייה, ואילו במקרה אחר (סמינרים על חלקיק דמוי-היגס מול מסיבת עיתונאים באותו הנושא) הנתונים סותרים את ההשערה. אנו מניחים שיש השפעה למידת התכנון והתרגול מראש של ההתבטאויות על הז'רגוניות שלהם, ומציעים להמשיך ולחקור בכיוון זה.

אם מדענים ישתמשו במילים שמוכרות לציבור, תוך היעזרות באופן שהציבור מבין ותופס אותן, תהיה בכך ודאי תרומה להבנה ההדדית בין המדענים לשאר הקבוצות בחברה. ככל הידוע, זהו היישום הראשון של שיטות הערכה ממוחשבות להוראת תקשורת המדע, מה שעשוי לקדם מטרה זו. חרף מגבלותיה, השיטה המוצעת רגישה הן לשיעור המונחים המקצועיים בטקסט באנגלית והן למידת "ז'רגוניותם" של כל אחד מהמונחים. שיטת ההערכה הממוחשבת עשויה לאפשר לעוסקים בתחום תקשורת המדע להעריך עד כמה מונחים מדעיים מוכרים לציבור, וזאת בהתבסס על מאגרי טקסטים אותנטיים הזמינים באינטרנט.

מקורות

- Anthony, L. (2009). *AntWordProfiler*. Tokyo, Japan: Waseda University. Retrieved from <http://www.antlab.sci.waseda.ac.jp/>
- Author. (2012). *Science Communication*.
- Baron, N. (2010). *Escape from the ivory tower: a guide to making your science matter*. Washington: Island Press.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D. (1995). *Dimensions of register variation*. Cambridge: Cambridge University Press.
- Conole, G., & Warburton, B. (2005). A review of computer - assisted assessment. *ALT-J*, 13(1), 17–31. doi:10.1080/0968776042000339772
- Dean, C. (2009). *Am I making myself clear? A scientist's guide to talking to the public*. Cambridge, Mass.: Harvard University Press.
- Dunning, T. (1993). Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics*, 19(1), 61–74.
- Gascoigne, T., & Metcalfe, J. (1997). Incentives and Impediments to Scientists Communicating Through the Media. *Science Communication*, 18(3), 265–282. doi:10.1177/1075547097018003005
- Grupp, G., & Heider, M. (1975). Non-Overlapping Disciplinary Vocabularies. In S. B. Day (Ed.), *Communication of scientific information*. (pp. 28–36). Basel: Karger.
- Hayes, J. R., & Bajzek, D. (2008). Understanding and Reducing the Knowledge Effect: Implications for Writers. *Written Communication*, 25(1), 104–118. doi:10.1177/0741088307311209
- Heatley, A., & Nation, I. S. P. (1994). *Range*. New Zealand: Victoria University of Wellington. Retrieved from <http://www.victoria.ac.nz/lals/about/staff/paul-nation>
- Jucks, R., Schulte-Löbber, P., & Bromme, R. (2007). Supporting Experts' Written Knowledge Communication Through Reflective Prompts on the Use of Specialist Concepts. *Zeitschrift für Psychologie / Journal of Psychology*, 215(4), 235–245.

- Kessler, S. (2011, June 27). With 500 Million Views, TED Talks Provide Hope for Intelligent Internet Video. *Mashable*. Retrieved May 20, 2012, from <http://mashable.com/2011/06/27/ted-anniversary/>
- Lemke, J. L. (1990). *Talking science: language, learning, and values*. Norwood, N.J.: Ablex Pub. Corp.
- Ley, P., & Florio, T. (1996). The use of readability formulas in health care. *Psychology, Health & Medicine, 1*(1), 7–28. doi:10.1080/13548509608400003
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: an advanced resource book*. London [etc.]: Routledge.
- Meredith, D. (2010). *Explaining research: How to reach key audiences to advance your work*. New York, N.Y.: Oxford University Press.
- Miller, S., Fahy, D., & The ESConet Team. (2009). Can Science Communication Workshops Train Scientists for Reflexive Public Engagement?: The ESConet Experience. *Science Communication, 31*(1), 116–126. doi:10.1177/1075547009339048
- Mulder, H. A. J., Longnecker, N., & Davis, L. S. (2008). The State of Science Communication Programs at Universities Around the World. *Science Communication, 30*(2), 277–287. doi:10.1177/1075547008324878
- Nation, I. S. P. (2006). How Large a Vocabulary is Needed For Reading and Listening? *Canadian Modern Language Review/ La Revue canadienne des langues vivantes, 63*(1), 59–82. doi:10.3138/cmlr.63.1.59
- Nisbet, M. C., & Scheufele, D. A. (2009). What's next for science communication? Promising directions and lingering distractions. *American Journal of Botany, 96*(10), 1767–1778. doi:10.3732/ajb.0900041
- Paquot, M., & Bestgen, Y. (2008). Distinctive words in academic writing: A comparison of three statistical tests for keyword extraction. In A. H. Jucker, D. Schreier, & M. Hundt (Eds.), *Corpora: Pragmatics and Discourse* (pp. 247–269). Presented at the 29th International Conference on English Language Research on Computerized Corpora (ICAME 29), Ascona, Switzerland.
- Simpson, R. C., Briggs, S. L., Ovens, J., & Swales, J. M. (2002). *The Michigan Corpus of Academic Spoken English*. Ann Arbor, MI: The Regents of the University of Michigan. Retrieved from <http://micase.elicorpora.info/>
- Stableford, S., & Mettger, W. (2007). Plain Language: A Strategic Response to the Health Literacy Challenge. *Journal of Public Health Policy, 28*(1), 71–93. doi:10.1057/palgrave.jphp.3200102
- TED Conferences, LLC. (2012, June 11). TED Talks. Retrieved June 11, 2012, from <http://www.ted.com/talks>