

פיתוח מדדים לאיתור
חיבורים משורבטים (חיבורי ג'יבריש)
ותיקוף המדדים מול חיבורי "פסיכורשת"

יעל שפרן
ענת בן-סימון

מאי 2014



מרכז ארצי לבחינות ולהערכה (ע"ר)
NATIONAL INSTITUTE FOR TESTING & EVALUATION
المركز القطري للامتحانات والتقييم
מיסודן של האוניברסיטאות בישראל

דוח מרכז 401
ISBN:978-965-502-182-0

**פיתוח מדדים לאיתור חיבורים משורבטים
(חיבורי ג'יבריש) ותיקוף המדדים מול
חיבורי "פסיכורשת"**

פרויקט השפה העברית (HLP)

יעל שפרן וענת בן-סימון

מאי, 2014

תוכן עניינים

4.....	תמצית.....
5.....	1. רקע.....
5.....	מטלת החיבור בבחינה הפסיכומטרית.....
5.....	מטלת החיבור בפסיכורשת.....
6.....	2. קריטריונים לאיתור חיבורי ג'יבריש.....
6.....	קריטריונים לאיתור חיבורי ג'יבריש בשפה העברית.....
7.....	קריטריונים לאיתור חיבורי ג'יבריש בשפה הערבית.....
7.....	מטרת המחקר.....
7.....	שיטה.....
8.....	3. תוצאות בדיקת דיוק האיתור של חיבורי ג'יבריש בשפה העברית.....
8.....	בדיקת דיוק הסיווג של החיבורים כחיבורי ג'יבריש או חיבורים תקינים.....
8.....	מדד ה- False Alarm.....
8.....	מדד ה- Miss.....
9.....	אחוז הצלחה באיתור חיבורים לא לנושא.....
9.....	התפלגות חיבורי הג'יבריש ומאפייניהם.....
12.....	4. תוצאות בדיקת דיוק האיתור של חיבורי ג'יבריש בשפה הערבית.....
12.....	בדיקת דיוק הסיווג של החיבורים כחיבורי ג'יבריש או חיבורים תקינים.....
12.....	מדד ה- False Alarm.....
12.....	מדד ה- Miss.....
13.....	אחוז הצלחה באיתור חיבורים לא לנושא.....
13.....	התפלגות חיבורי הג'יבריש ומאפייניהם.....
15.....	5. סיכום מסקנות והמלצות.....
16.....	הערות, דילמות והצעות נוספות לשיפור המדדים לאיתור חיבורי ג'יבריש.....
17.....	הצעות לצוות CAT לשיפור פסיכורשת.....
17.....	תכניות נוספות למחקר.....

רשימת טבלאות

8.....	טבלה 1: התפלגות החיבורים במדגם לפי קטגוריית התקינות.....
9.....	טבלה 2: התפלגות החיבורים לפי דיוק הזיהוי.....
10.....	טבלה 3: התפלגות חיבורי הג'יבריש לפי הציון הכולל במדד הג'יבריש.....
10.....	טבלה 4: התפלגות חיבורי הג'יבריש לפי הציון בכל אחד מהקריטריונים שהוגדרו.....
11.....	טבלה 5: חמשת הפרופילים הנפוצים ביותר של חיבורי ג'יבריש.....
11.....	טבלה 6: מתאמים בין הציונים שהתקבלו בקריטריונים המשמשים לאיתור חיבורי ג'יבריש.....
12.....	טבלה 7: התפלגות החיבורים במדגם לפי קטגוריית התקינות.....
13.....	טבלה 8: התפלגות החיבורים לפי דיוק הזיהוי (N=270).....
13.....	טבלה 9: התפלגות חיבורי הג'יבריש לפי הציון הכולל במדד הג'יבריש.....
14.....	טבלה 10: התפלגות חיבורי הג'יבריש לפי הציון בכל אחד מהקריטריונים שהוגדרו.....
14.....	טבלה 11: חמשת הפרופילים הנפוצים ביותר של חיבורי ג'יבריש.....
15.....	טבלה 12: מתאמים בין הציונים שהתקבלו בקריטריונים המשמשים לאיתור חיבורי ג'יבריש.....

רשימת נספחים

- נספח 1: דוגמאות לחיבורים בשפה העברית מקטגוריות סיווג שונות.....18
- נספח 2: דוגמאות לחיבורים בשפה הערבית מקטגוריות סיווג שונות.....22
- נספח 3: התפלגות פרופילים של חיבורי הג'בריש בשפה העברית.....25
- נספח 4: התפלגות פרופילים של חיבורי הג'בריש בשפה הערבית.....26
- נספח 5: טבלת הקריטריונים לזיהוי חיבורי ג'בריש בשפה העברית.....27

תמצית

מאז הטמעת החיבור בבחינה הפסיכומטרית ושילובו ב'פסיכורשת' (בחינה פסיכומטרית ממוחשבת ואדפטיבית להתנסות עצמית) נאספו במסגרת זו מעל 5,000 חיבורים. אוסף חיבורים זה כולל מגוון רחב של חיבורים שחלקם משקפים ניסיון לכתיבה מיטבית וחלקם נראים כשרבוט אקראי של מילים ומשפטים. קורפוס חיבורים זה סיפק הזדמנות לפיתוח מדדים לאיתור טקסטים שנראה בעליל שנכתבו כלאחר יד ואינם עומדים בשום קריטריון של כתיבה תקנית או חיבורים שאינם לנושא, להלן, חיבורי ג'יבריש. מטרתו של המחקר הנוכחי הייתה לפתח מדדים תקפים לאיתור חיבורי ג'יבריש.

לצורך זיהוי ממוחשב של חיבורי ג'יבריש גובשו עשרה קריטריונים (מדדים) עבור חיבורים בשפה העברית ושמונה קריטריונים (מדדים) עבור חיבורים בשפה הערבית ובוצעו ניתוחים שונים לחקירת הקריטריונים ומאפייניהם ולבדיקת יעילותם בזיהוי חיבורי ג'יבריש. כל המדדים מבוססים על מאפייני טקסט כמותיים המופקים באופן ממוחשב מהטקסט הכתוב.

לצורך בדיקת הרגישות של המדדים לעיל בזיהוי חיבורי ג'יבריש פותח אלגוריתם המפיק מדד ג'יבריש לכל חיבור ומסווג אותו כחיבור תקין או כחיבור ג'יבריש. דיוק הזיהוי נבדק על 4,108 חיבורים בשפה העברית ו-1,088 חיבורים בשפה הערבית שנכתבו במסגרת פסיכורשת, תוך שימוש בסיווג של מעריך אנושי כקריטריון. הבדיקה יושמה רק עבור חיבורים שאורכם עולה על 100 מילה.

התוצאות מלמדות כי רק 66.2% מ-4,108 החיבורים שנכתבו בשפה העברית ו-60.4% מ-1,088 החיבורים שנכתבו בשפה הערבית נמצאו על ידי המערכת הממוחשבת כחיבורים תקינים. מרביתם של החיבורים הנותרים היו קצרים מהנדרש (28% ו-34.3% בהתאמה), ומיעוטם (6.1% ו-5.3% בהתאמה) חיבורי ג'יבריש.

בבדיקת דיוק הסיווג של החיבורים כחיבורי ג'יבריש או חיבורים תקינים נמצא כי 99.4% מהחיבורים בשפה העברית ו-96.7% מהחיבורים בשפה הערבית שאורכם עמד בדרישות סווגו באופן מדויק כחיבורי ג'יבריש או חיבורים תקינים.

בהתייחס לטעויות הסיווג נמצא כי בהתייחס לחיבורים שפה העברית אף חיבור תקין לא סווג כחיבור ג'יבריש ורק 0.4% מכלל החיבורים שסווגו על ידי האלגוריתם כחיבורים תקינים הם למעשה חיבורי ג'יבריש. בהתייחס לחיבורים בשפה הערבית נמצא כי 2% מהחיבורים שסווגו על ידי המחשב כחיבורי ג'יבריש הם למעשה חיבורים תקינים ו-4% מהחיבורים שסווגו על ידי המחשב כחיבורים תקינים הם למעשה חיבורי ג'יבריש.

תוצאות המחקר מלמדים כי המדד הכולל לזיהוי חיבורי ג'יבריש מדויק דיו וניתן להשתמש בו בגרסה התפעולית של פסיכורשת ובהקשרים נוספים לפי הצורך.

1. רקע

כחלק מהליך ההכנה לבחינה הפסיכומטרית ולהתנסות בבחינה מציע מאל"ו לנרשמים לבחינה (ולקהל הרחב) אפשרות להיבחן בגרסה ממוחשבת של הבחינה הפסיכומטרית (להלן, פסיכורשת). גרסה זו כוללת שני פרקים מכל אחד משלושת התחומים הנבדקים בבחינה הפסיכומטרית (חשיבה מילולית, חשיבה כמותית ואנגלית ובכלל זה מטלת כתיבה המהווה חלק מהתחום המילולי).

מטלת החיבור בבחינה הפסיכומטרית

הוראות לנבחן – במטלת הכתיבה הכלולה בבחינה הפסיכומטרית מתבקשים הנבחנים לכתוב חיבור בנושא נתון. ההוראות לכתיבת החיבור תואמות את ההוראות הניתנות בבחינה הפסיכומטרית ובכלל זה מנחות את הנבחן לכתוב חיבור במסגרת השורות המוקצבות (50 שורות) ובאורך של 25 שורות לכל הפחות. בהינתן שבחיבורים הנכתבים בכתב יד מס' המילים לשורה נע בין 8 ל-14, אורך החיבורים צפוי לנוע בפועל מ-200-350 מילים לחיבור עד 400-700 מילים. הציון על החיבור ניתן על שני ממדים (סולם 1-6). כל חיבור מוערך על ידי שני מעריכים. ועל כן הציון הכללי בחיבור נע בין 4 ל-24 נק'.

הנחיות להערכת חיבורים חריגים – ההנחיות הניתנות למעריכים של חיבורים 'חריגים' שנכתבו במסגרת הבחינה הפסיכומטרית הן: (1) חיבור שאורכו פחות מ-10 שורות מזכה את החיבור בציון 4 (ציון מתוקנן=50) שהוא הציון המינימלי המקובל בהערכת חיבורים; (2) חיבור שאורכו 10-23 שורות קונס את החיבור בנקודה אחת בציון הכללי של כל מעריך; (3) טקסט שנכתב מעבר למסגרת השורות המוקצבות לא נכלל בהערכת החיבור; ו-(4) חיבור שנכתב לא לנושא מזכה את החיבור בציון 4.

מטלת החיבור בפסיכורשת

הוראות לנבחן - ההוראות לכתיבת החיבור בפסיכורשת תואמות באופן כללי את ההוראות הניתנות בבחינה הפסיכומטרית. במערכת הפסיכורשת אין הגבלה על אורך החיבור, עם זאת, המתנסים מתבקשים לכתוב חיבור באורך של 250-500 מילה. כדי להקל על הנבחנים להעריך את אורך הטקסט מוצג על המסך, במהלך כתיבת החיבור, מונה מילים, המחליף את צבעו מאדום כאשר החיבור קצר מ-250 מילה, לשחור כאשר החיבור באורך תקין, ושוב לאדום כאשר החיבור עולה על האורך הנדרש של עד 500 מילה. הציון על החיבור הנכתב במסגרת פסיכורשת מחושב באופן אוטומטי באמצעות מערכת NiteRater שפותחה במאל"ו ומשמשת להערכה ממוחשבת של חיבורים (NiteRater, 2007)¹.

הנחיות להערכת חיבורים חריגים – בהתאמה להנחיות הנהוגות להערכת חיבורים חריגים בבחינה הפסיכומטרית הוגדרו הכללים הבאים להערכה של חיבורים חריגים הנכתבים במסגרת פסיכורשת (הכללים מבוססים על ההנחה שמס' המילים לשורה בחיבור הנכתב בכתב יד הוא 10 מילים): (1) חיבור שאורכו פחות מ-100 מילים מזכה את החיבור בציון 4; (2) טקסט שנכתב מעבר

¹ NiteRater [Computer software]. (2007). Jerusalem: National Institute for Testing & Evaluation.

למסגרת השורות המוקצבות (מעל 500 מילה) לא נכלל בהערכת החיבור ; ו-(3) חיבור שאינו עומד בכללים מקובלים של כתיבה תקנית וחיבור שנכתב לא לנושא מזכים את החיבור בציון 4 ; במהלך השנה האחרונה, מאז הטמעת החיבור בבחינה הפסיכומטרית, נבחנו בפסיכורשת מעל 5,000 איש. אוסף החיבורים שהתקבלו כולל מגוון רחב של חיבורים שחלקם משקפים ניסיון לכתיבה מיטבית וחלקם נראים כשרבוט אקראי של מילים ומשפטים. קורפוס חיבורים זה מספק הזדמנות לפיתוח מדדים לאיתור חיבורי ג'בריש ("קשקשת לפי הצעת האקדמיה) - טקסטים שנראה בעליל שנכתבו כלאחר יד ואינם עומדים בשום קריטריון של כתיבה תקנית או חיבורים שאינם לנושא. מטרתו של המחקר הנוכחי היא לפתח מדדים תקפים לאיתור חיבורי ג'בריש.

2. קריטריונים לאיתור חיבורי ג'בריש

קריטריונים לאיתור חיבורי ג'בריש בשפה העברית

הוגדרו עשרה קריטריונים לאיתור חיבורי ג'בריש שנכתבו בשפה העברית. חיבור ייחשב כג'בריש אם מתקיימים עבורו אחד או יותר מהתנאים האלה :

1. **TTR** - יחס $TTR(type/token)$ נמוך : $0.30 >$
 2. **משפטים חוזרים** - 3 משפטים שונים החוזרים על עצמם לפחות פעמיים כל אחד במקום כלשהו בטקסט.
 3. **פסקאות קצרות** - אחוז גבוה של פסקאות קצרות מאד (פחות מ-5 מילה) : $0.90 <$
 4. **טקסט חופף גריין** (העתקת הגריין) - אחוז גבוה של משפטים חופפים גריין או מחרוזות חופפות גריין : $0.80 <$
 5. **אורך מילה** - אחוז גבוה של מילים באורך אות אחת ובאורך 10 ומעלה : $0.30 <$
 6. **מילים זרות** - אחוז גבוה של מילים שכתובות באותיות שאינן עבריות : $0.50 <$
 7. **רצפים לא חוקיים** - שלושה רצפים של 4 מילים לא חוקיות (מילים באורך 10 תווים או יותר, מילים שאינן קיימות במילון, מילים לא חוקיות בשפה ומילים החוזרות על עצמן)
 8. **טקסט שאינו חופף גריין** (לא לנושא) - אחוז מחרוזות וגם אחוז לקסמות חופפות גריין : $0.05 >$ או אחוז מזערי של מחרוזות חופפות גריין : $0.015 >$
 9. **מילים לא חוקיות בשפה** - אחוז מילים לא חוקיות (אות סופית באמצע מילה, אות תחילית בסוף מילה) : $0.05 <$
 10. **מילים לא במילון** - אחוז גבוה של מילים שאינן קיימות במילון : $0.30 <$
- היעדר כל פיסוק מזכה את החיבור ב- 0.5 נק'. כלומר אינו תנאי מספיק לתיוג החיבור כג'בריש אך מוסיף למדד הג'בריש.

קריטריונים לאיתור חיבורי ג'יבריש בשפה הערבית

הוגדרו שמונה קריטריונים לאיתור חיבורי ג'יבריש שנכתבו בשפה הערבית. ששת הקריטריונים הראשונים זהים לששת הקריטריונים הראשונים שהוגדרו עבור חיבורי עברית. בקריטריונים 7 ו-8 הוכנסו שינויים קלים.

חיבור ייחשב כג'יבריש אם מתקיימים עבורו אחד או יותר מהתנאים האלה:

1. **TTR** - יחס $TTR(\text{type/token})$ נמוך: $0.30 >$
 2. **משפטים חוזרים** - 3 משפטים שונים החוזרים על עצמם לפחות פעמיים כל אחד במקום כלשהו בטקסט.
 3. **פסקאות קצרות** - אחוז גבוה של פסקאות קצרות מאד (פחות מ-5 מילה): $0.90 <$
 4. **טקסט חופף גריין** (העתקת הגריין) - אחוז גבוה של משפטים חופפים גריין או מחרוזות חופפות גריין: $0.80 <$
 5. **אורך מילה** - אחוז גבוה של מילים באורך אות אחת ובאורך 10 ומעלה: $0.30 <$
 6. **מילים זרות** - אחוז גבוה של מילים שכתובות באותיות שאינן בערבית: $0.50 <$
 7. **רצפים לא חוקיים** - שלושה רצפים של 4 מילים לא חוקיות (מילים באורך 10 תווים או יותר, ומילים החוזרות על עצמן)
 8. **טקסט שאינו חופף גריין** (לא לנושא) - אחוז מחרוזות וגם אחוז לקסמות חופפות גריין: $0.03 >$
- היעדר כל פיסוק מזכה את החיבור ב-0.5 נק'. כלומר אינו תנאי מספיק לתיג החיבור כג'יבריש אך מוסיף למדד הג'יבריש.

מטרת המחקר

מטרתו של המחקר הנוכחי היא לבדוק את דיוק הזיהוי של חיבורים שנכתבו בשפה העברית או בשפה הערבית כחיבורי ג'יבריש תוך שימוש בכללים (מדדים) שהוגדרו לעיל לאיתור חיבורי ג'יבריש.

שיטה

לצורך בדיקת דיוק הזיהוי של חיבורי ג'יבריש בוצעה סדרה של בדיקות על חיבורים שנכתבו במסגרת ההתנסות בבחינה הפסיכומטרית בפסיכורשת.

לצורך בדיקת הרגישות של המדדים לעיל בזיהוי חיבורי ג'יבריש הוגדרו שלוש רמות סיווג של חיבורים: (1) חיבור תקין; (2) חיבור ג'יבריש ו- (3) חיבור קצר מהנדרש (פחות מ-100 מילה). דוגמאות לחיבורים בשפה העברית מקטגוריות סיווג שונות מופיעות בנספח 1, ודוגמאות לחיבורים בשפה הערבית מקטגוריות סיווג שונות מופיעות בנספח 2.

האלגוריתם שפותח לבדיקת תקינות החיבורים סיווג כל חיבור לאחת משלושת הקטגוריות לעיל.

לא נבדקה תקינותם של חיבורים הקצרים מ-100 מילה.

3. תוצאות בדיקת דיוק האיתור של חיבורי ג'יבריש בשפה העברית

בטבלה 1 מופיעה התפלגות של 4,108 חיבורים שנכתבו עד אוקטובר 2013 (במשך כשנה) במסגרת פסיכורשת כפי שמוינו לשלוש קטגוריות התקינות על-ידי האלגוריתם. 66% מהחיבורים נמצאו תקינים, 28.4% מהחיבורים היו קצרים מדי ו- 6.1% מהחיבורים סווגו כחיבורי ג'יבריש.

האורך הממוצע של החיבורים הקצרים הוא 25 מילים וסטיית התקן 29. מתוך 1,143 החיבורים הקצרים (בני פחות מ-100 מילה), 16% היו ריקים, 50% כללו פחות מ-10 מילים ו- 7% כללו 80 מילים ומעלה. 1,143 החיבורים הקצרים הוצאו מהמשך הניתוח.

טבלה 1: התפלגות החיבורים במדגם לפי קטגוריית התקינות

	חיבורים תקינים	חיבורי ג'יבריש	חיבורים קצרים	סך הכול
N	2,719	246	1,143	4,108
%	66.2	6.1	28.4	100

בדיקת דיוק הסיווג של החיבורים כחיבורי ג'יבריש או חיבורים תקינים

בטבלה 2 מופיעה התפלגות החיבורים לפי דיוק הסיווג שלהם כחיבורי ג'יבריש או חיבורים תקינים.

מדד ה- False Alarm

לצורך בדיקת מדד ה- False Alarm (זיהוי חיבורים תקינים כחיבורי ג'יבריש) נבדק באופן ידני דיוק הסיווג של 246 החיבורים שסווגו על ידי האלגוריתם כחיבורי ג'יבריש. תוצאות הבדיקה הצביעו על 0% False Alarm.

בדיקה נוספת של מדד ה- False Alarm בוצעה על ידי חישוב מדד הג'יבריש של 700 חיבורים שנכתבו במסגרת הבחינה הפסיכומטרית במועד אוקטובר 2012 והוקלדו למטרות מחקר. היות והחיבורים נכתבו במסגרת העברה תפעולית של המבחן ההנחה היא שכל החיבורים הם תקינים. בבדיקה של מדגם חיבורים זה לא זוהה אף חיבור כחיבור ג'יבריש, כלומר תוצאות הבדיקה הצביעו על 0% False Alarm.

מדד ה- Miss

לצורך בדיקת מדד ה- Miss (סיווג חיבורי ג'יבריש כחיבורים תקינים) נבדק באופן ידני דיוק הסיווג של 2,719 החיבורים שזוהו על ידי האלגוריתם כתקינים. רק 11 (0.45%) חיבורים מתוך 2,719 סווגו על ידי מעריך אנושי כחיבורי ג'יבריש. אפיון איכותני של חיבורים אלו מלמד שרובם חיבורים שכתובים באופן תקין אך לא לנושא, או חיבורים שנכתבו לנושא באופן חלקי בלבד (חיבורים שהפסקאות הראשונות שלהם כתובות לנושא והפסקאות הבאות כתובות לא לנושא).

טבלה 2: התפלגות החיבורים לפי דיוק הזיהוי

חיבורים שסווגו על ידי המחשב כחיבורי ג'יבריש (N=246)		חיבורים שסווגו על ידי המחשב כחיבורים תקינים (N=2719)		
חיבורים תקינים (False alarm)	חיבורי ג'יבריש (Hit)	חיבורי ג'יבריש (Miss)	חיבורים תקינים (Correct rejection)	סיווג מעריך אנושי
0	246	11	2,708	N
0	100%	0.4%	99.6%	%

אחוז הצלחה באיתור חיבורים לא לנושא

לצורך בדיקת דיוק הזיהוי של חיבורים שנכתבו לא לנושא נדגמו 100 טקסטים העוסקים בנושאים שונים מנושא החיבור בפסיכורשת. הטקסטים נדגמו מתוך המקורות הבאים: עיתון הארץ, ויקיפדיה, דה-מרקר, חיבורים בנושא 13T, (אוקטובר 2012). היות והטקסטים נלקחו מתוך קורפוסים תקינים, יש להניח שמדובר בטקסט תקני שהמאפיין היחיד ההופך אותו לטקסט חריג הוא שהטקסט נכתב לנושא שונה מנושא החיבור המופיע בפסיכורשת (18T). מדד הג'יבריש² איתר 78% מהחיבורים כחיבורי ג'יבריש-שאינם לנושא.

התפלגות חיבורי הג'יבריש ומאפייניהם

בטבלה 3 מופיעה התפלגות חיבורי הג'יבריש לפי הציון הכולל במדד הג'יבריש (מס' קריטריונים שזוהו). 24% מהחיבורים הוגדרו כחיבורי ג'יבריש על פי קריטריון אחד בלבד ואף לא אחד מהחיבורים הוגדר ככזה על פי כל הקריטריונים.

בטבלה 4 מופיע אחוז החיבורים שענו על כל אחד מהקריטריונים שהוגדרו לאיתור חיבורי ג'יבריש. ה'ליקויים' השכיחים ביותר היו: טקסט שאינו חופף גריין (75%), רצפים לא חוקיים (60%) ויחס TTR נמוך (53%).

² הקריטריונים: "טקסט שאינו חופף גריין" ו- "טקסט חופף גריין" המרכיבים את מדד הג'יבריש חושבו על הנושא 18T של חיבורי פסיכורשת

טבלה 3: התפלגות חיבורי הג'יבריש לפי הציון הכולל במדד הג'יבריש

הציון הכולל במדד הג'יבריש	שכיחות	אחוז
1	61	24.8
2	19	7.72
3	32	13.01
4	44	17.89
5	43	17.48
6	36	14.63
7	11	4.47

טבלה 4: התפלגות חיבורי הג'יבריש לפי הציון בכל אחד מהקריטריונים שהוגדרו

	קריטריון	אחוז
1.	TTR	52.85
2.	משפטים חוזרים	12.60
3.	פסקאות קצרות	7.32
4.	טקסט חופף גריין במרביתו	10.98
5.	אורך מילה	27.64
6.	מילים זרות	0
7.	רצפים לא חוקיים	60.16
8.	טקסט שאינו חופף גריין	75.20
9.	מילים לא חוקיות בשפה	44.31
10.	מילים לא במילון	34.15

בטבלה 5 מופיעים חמשת הפרופילים הנפוצים ביותר של חיבורי ג'יבריש (ההתפלגות המלאה של הפרופילים מופיעה בנספח 3), הפרופיל הנפוץ ביותר הוא טקסט שאינו חופף את הגריין אך תקין בכל שאר הקריטריונים (48%), והפרופיל השני בשכיחותו הוא טקסט שאינו חופף גריין, כולל רצפים לא חוקיים, כולל מילים לא חוקיות בשפה, וכולל מילים שאינן במילון (34%).

בטבלה 6 מופיעים המתאמים בין הציונים שהתקבלו בקריטריונים המשמשים לאיתור חיבורי ג'יבריש המלמדים על הזיקה ביניהם.

לבסוף נבדק גם המתאם בין מדד הג'יבריש הכולל לציון המנובא בפסיכורשת על 2,880 חיבורים. הציון המנובא הופק באופן ממוחשב והתבסס על משוואת ניבוי שהופקה מ- 112 החיבורים שהועברו במסגרת העברה ניסיונית (נושא18). של הנושא והייתה עבורם הערכה אנושית של שני מעריכים.

המתאם שהתקבל הוא -0.37. מתוך 225 חיבורי פסיכורשת שזוהו כחיבורי ג'יבריש, 29% מהחיבורים קיבלו ציון ממוחשב של 10 ומעלה ו- 14% קיבלו ציון ממוחשב של 16 ומעלה.

טבלה 5: חמשת הפרופילים הנפוצים ביותר של חיבורי ג'יבריש

קריטריונים המרכיבים את הפרופיל	אחוז	שכיחות	פרופיל 12345678910
טקסט שאינו חופף גריין	19.51	48	0000000100
רצפים לא חוקיים, טקסט שאינו חופף גריין, מילים לא חוקיות בשפה, מילים לא במילון	13.82	34	0000001111
רצפים לא חוקיים, טקסט שאינו חופף גריין, מילים TTR לא חוקיות בשפה, מילים לא במילון	8.54	21	1000001111
משפטים חוזרים, טקסט חופף גריין במרביתו TTR	5.28	13	1101000000
אורך מילה, רצפים לא חוקיים, טקסט שאינו חופף TTR גריין, מילים לא חוקיות בשפה	4.88	12	1000101110

טבלה 6: מתאמים בין הציונים שהתקבלו בקריטריונים המשמשים לאיתור חיבורי ג'יבריש

	TTR	משפטים חוזרים	פסקאות קצרות	טקסט חופף גריין במרביתו	אורך מילה	רצפים לא חוקיים	טקסט שאינו חופף גריין	מילים לא חוקיות בשפה
משפטים חוזרים	.16							
פסקאות קצרות	.14	-.11						
טקסט חופף גריין במרביתו	.12	.45	-.10					
אורך מילה	.33	-.23	.42	-.19				
רצפים לא חוקיים	.26	-.47	.20	-.30	.45			
טקסט שאינו חופף גריין	-.20	-.49	.13	-.46	.06	.22		
מילים לא חוקיות בשפה	.01	-.34	.10	-.26	.18	.71	.30	
מילים לא במילון	-.02	-.27	.13	-.25	-.02	.57	.25	.62

* הערכים המודגשים מציינים מתאמים גבוהים מ- 0.4 או נמוכים מ- 0.4-

4. תוצאות בדיקת דיוק האיתור של חיבורי ג'יבריש בשפה הערבית

בטבלה 7 מופיעה התפלגות של 1,088 חיבורים שנכתבו עד אוקטובר 2013 (במשך כשנה) במסגרת פסיכורשת לפי שלוש קטגוריות התקינות. 60.4% מהחיבורים נמצאו תקינים, 34.3% מהחיבורים היו קצרים מדי ו- 5.3% מהחיבורים סווגו כחיבורי ג'יבריש. התפלגות זו דומה מאוד להתפלגות שנמצאה עבור חיבורי פסיכורשת שנכתבו בשפה העברית.

האורך הממוצע של החיבורים הקצרים הוא 26 מילים וסטיית התקן 28. מתוך 373 החיבורים הקצרים (בני פחות מ-100 מילה), ב-18% מהחיבורים לא נמצאה אף מילה בשפה הערבית, 40% כללו פחות מ-10 מילים ו- 7% כללו 80 מילים ומעלה. כל החיבורים הקצרים הוצאו מהמשך הניתוח.

טבלה 7: התפלגות החיבורים במדגם לפי קטגוריית התקינות

	חיבורים תקינים	חיבורי ג'יבריש	חיבורים קצרים	סך הכול
N	657	58	373	1,088
%	60.4	5.3	34.3	100

בדיקת דיוק הסיווג של החיבורים כחיבורי ג'יבריש או חיבורים תקינים

בטבלה 8 מופיעה התפלגות החיבורים לפי דיוק הסיווג שלהם כחיבורי ג'יבריש או חיבורים תקינים.

מדד ה- False Alarm

לצורך בדיקת מדד ה- False Alarm (זיהוי חיבורים תקינים כחיבורי ג'יבריש) נבדק באופן ידני דיוק הסיווג של 58 החיבורים שסווגו על ידי האלגוריתם כחיבורי ג'יבריש. חיבור אחד מתוך ה-58 (1.7%) זוהה באופן שגוי כחיבור ג'יבריש למרות שהיה תקין.

בדיקה נוספת של מדד ה- False Alarm בוצעה על ידי חישוב מדד הג'יבריש של 1,000 חיבורים שנכתבו במסגרת הבחינה הפסיכומטרית במועד אוקטובר 2012 והוקלדו למטרות מחקר. היות והחיבורים נכתבו במסגרת העברה תפעולית של המבחן ההנחה היא שכל החיבורים הם תקינים. בבדיקה של מדגם חיבורים זה לא זוהה אף חיבור כחיבור ג'יבריש, כלומר תוצאות הבדיקה הצביעו על False Alarm 0%.

מדד ה- Miss

לצורך בדיקת מדד ה- Miss (סיווג חיבורים תקינים כחיבורי ג'יבריש) נבדק באופן ידני דיוק הסיווג של 212 חיבורים מתוך 657 שזוהו על ידי האלגוריתם כתקינים. 8 חיבורים מתוך 212 (9%) סווגו על ידי מעריך אנושי כחיבורי ג'יבריש. אפיון איכותני של חיבורים אלו מלמד שרובם חיבורים שכתובים באופן תקין אך לא לנושא, או חיבורים שנכתבו לנושא באופן חלקי בלבד (חיבורים שהפסקאות הראשונות שלהם כתובות לנושא והפסקאות הבאות כתובות לא לנושא).

טבלה 8: התפלגות החיבורים לפי דיוק הזיהוי (N=270)

חיבורים שסווגו על ידי המחשב כחיבורי ג'יבריש (N=58)		חיבורים שסווגו על ידי המחשב כחיבורים תקינים (N=212)		
חיבורים תקינים (False alarm)	חיבורי ג'יבריש (Hit)	חיבורי ג'יבריש (Miss)	חיבורים תקינים (Correct rejection)	סיווג מעריך אנושי
1	57	8	204	N
2%	98%	4%	96%	%

אחוז הצלחה באיתור חיבורים לא לנושא

לצורך בדיקת דיוק הזיהוי של חיבורים שנכתבו לא לנושא חושב מדד הגיבריש³ על 1000 חיבורים שנכתבו במסגרת הבחינה הפסיכומטרית (מועד אוקטובר 2012) והוקלדו לצרכי מחקר. חיבורים אלה נכתבו לנושא שונה מנושא החיבור בפסיכורשת אך כיוון שהם נכתבו במסגרת המבחן הפסיכומטרי יש להניח שמדובר בטקסט תקני שהמאפיין היחיד ההופך אותו לטקסט חריג הוא שהטקסט נכתב לנושא שונה מנושא החיבור המופיע בפסיכורשת בשפה הערבית (15T). מדד הגיבריש איתר 99% מהחיבורים כחיבורים שאינם לנושא.

התפלגות חיבורי הג'יבריש ומאפייניהם

בטבלה 9 מופיעה התפלגות חיבורי הגיבריש לפי הציון הכולל במדד הגיבריש (מס' קריטריונים שזוהו). 40% מהחיבורים הוגדרו כחיבורי ג'יבריש על פי קריטריון אחד בלבד ואף לא אחד מהחיבורים הוגדר ככזה על פי יותר מארבעה קריטריונים.

בטבלה 10 מופיע אחוז החיבורים שענו על כל אחד מהקריטריונים שהוגדרו לאיתור חיבורי גיבריש. ה'ליקוים' השכיחים ביותר היו: טקסט שאינו חופף גריין (67%), רצפים לא חוקיים (52%), יחס TTR נמוך (43%) ואורך מילה (31%).

טבלה 9: התפלגות חיבורי הג'יבריש לפי הציון הכולל במדד הג'יבריש

הציון הכולל במדד הג'יבריש	שכיחות	אחוז
1	23	39.66
2	14	24.14
3	10	17.24
4	11	18.97

³ הקריטריונים: "טקסט שאינו חופף גריין" ו- "טקסט חופף גריין" המרכיבים את מדד הגיבריש חושבו ביחס לנושא של חיבורי פסיכורשת 15T

טבלה 10: התפלגות חיבורי הג'בריש לפי הציון בכל אחד מהקריטריונים שהוגדרו

	קריטריון	אחוז
1.	TTR	43.10
2.	משפטים חוזרים	10.35
3.	פסקאות קצרות	1.72
4.	טקסט חופף גריין במרביתו	8.62
5.	אורך מילה	31.03
6.	מילים זרות	1.72
7.	רצפים לא חוקיים	51.72
8.	טקסט שאינו חופף גריין	67.24

בטבלה 11 מופיעים חמשת הפרופילים הנפוצים ביותר של חיבורי ג'בריש (ההתפלגות המלאה של הפרופילים מופיעה בנספח 4), הפרופיל הנפוץ ביותר הוא טקסט שאינו חופף את הגריין אך תקין בכל שאר הקריטריונים (27.6%), והפרופיל השני בשכיחותו הוא טקסט שאינו חופף גריין, שכולל רצפים לא חוקיים, מילים קצרות או ארוכות מדי (אורך מילה) ויחס TTR נמוך (17%).

בטבלה 12 מופיעים המתאמים בין הציונים שהתקבלו בקריטריונים המשמשים לאיתור חיבורי ג'בריש המלמדים על הזיקה ביניהם.

לבסוף נבדק גם המתאם בין מדד הג'בריש הכולל לציון המנובא שניתן ל-715 החיבורים בפסיכורשת. המתאם שהתקבל הוא -0.38. מתוך 58 חיבורי הג'בריש, 29% קיבלו ציון ממוחשב 10 ומעלה, ורק 5% קיבלו ציון ממוחשב 16 ומעלה.

טבלה 11: חמשת הפרופילים הנפוצים ביותר של חיבורי ג'בריש

קריטריונים המרכיבים את הפרופיל	אחוז	שכיחות	פרופיל 12345678
טקסט שאינו חופף גריין	27.6	16	00000001
טקסט שאינו חופף גריין, רצפים לא חוקיים, אורך מילה, TTR	17.2	10	10001011
טקסט שאינו חופף גריין, רצפים לא חוקיים	8.6	5	00000011
רצפים לא חוקיים	6.9	4	00000010
טקסט חופף גריין במרביתו, משפטים חוזרים, TTR	6.9	4	11010000

טבלה 12: מתאמים בין הציונים שהתקבלו בקריטריונים המשמשים לאיתור חיבורי ג'יבריש

	TTR	משפטים חוזרים	פסקאות קצרות	טקסט חופף גריין במרביתו	אורך מילה	מילים זרות	רצפים לא חוקיים
משפטים חוזרים	.16						
פסקאות קצרות	.15	-.04					
טקסט חופף גריין במרביתו	.35	.70	-.04				
אורך מילה	.32	-.23	-.09	-.21			
מילים זרות	.15	-.04	-.02	-.04	-.09		
רצפים לא חוקיים	.07	-.24	.13	-.32	.65	-.14	
טקסט שאינו חופף גריין	-.13	-.49	.09	-.44	.07	-.19	-.01

* הערכים המופיעים בהבלטה מציינים מתאמים גבוהים מ- 0.4 או נמוכים מ- 0.4-

5. סיכום מסקנות והמלצות

מאז הטמעת החיבור בבחינה הפסיכומטרית ושילובו פסיכורשת (בחינה פסיכומטרית ממוחשבת ואדפטיבית להתנסות עצמית) ועד ביצוע המחקר נאספו במסגרת זו כ-4000 חיבורים. אוסף חיבורים זה כולל מגוון רחב של חיבורים שחלקם משקפים ניסיון לכתיבה מיטבית וחלקם נראים כשרבוט אקראי של מילים ומשפטים. קורפוס חיבורים זה סיפק הזדמנות לפיתוח מדדים לאיתור טקסטים שנראה בעליל שנכתבו כלאחר יד ואינם עומדים בשום קריטריון של כתיבה תקנית או חיבורים שאינם לנושא, להלן, חיבורי ג'יבריש. מטרתו של המחקר הנוכחי היתה לפתח מדדים תקפים לאיתור חיבורי ג'יבריש.

לצורך זיהוי ממוחשב של חיבורי ג'יבריש גובשו עשרה קריטריונים (מדדים) עבור חיבורים בשפה העברית ושמונה קריטריונים (מדדים) עבור חיבורים בשפה הערבית. כל המדדים מבוססים על מאפייני טקסט כמותיים המופקים באופן ממוחשב מהטקסט הכתוב. בהמשך, בוצעו ניתוחים שונים לחקירת הקריטריונים ומאפייניהם ולבדיקת יעילותם בזיהוי חיבורי ג'יבריש שנכתבו בשפה העברית ובשפה הערבית.

לצורך בדיקת הרגישות של המדדים לעיל בזיהוי חיבורי ג'יבריש הוגדרו שלוש קטגוריות סיווג של חיבורים: (1) חיבור תקין; (2) חיבור ג'יבריש ו- (3) חיבור קצר מהנדרש (פחות מ-100 מילה). האלגוריתם שפותח לבדיקת תקינות החיבורים סיווג כל חיבור לאחת משלושת הקטגוריות לעיל. דיוק הזיהוי נבדק עבור שתי הקטגוריות הראשונות בלבד, תוך שימוש בסיווג של מעריך אנושי כקריטריון.

התוצאות מלמדות כי רק 66.2% מ- 4,108 החיבורים שנכתבו בשפה העברית ו- 60.4% מ-1088 החיבורים שנכתבו בשפה הערבית נמצאו על ידי המערכת הממוחשבת כחיבורים תקינים.

מרביתם של החיבורים הנותרים היו קצרים מהנדרש (28% ו- 34.3% בהתאמה), ומיעוטם (6.1% ו- 5.3% בהתאמה) חיבורי גיבריש.

בבדיקת דיוק הסיווג של החיבורים כחיבורי גיבריש או חיבורים תקינים נמצא כי 99.4% מהחיבורים בשפה העברית ו-96.7% מהחיבורים בשפה הערבית שאורכם עמד בדרישות (מעל 100 מילה) סווגו באופן מדויק כחיבורי גיבריש או חיבורים תקינים. נתונים אלה מלמדים כי המדד הכולל לזיהוי חיבורי גיבריש מדויק דיו וניתן להשתמש בו בגרסה התפעולית של פסיכורשת.

בהתייחס לטעויות הסיווג נמצא כי אף חיבור תקין לא סווג כחיבור גיבריש ורק 0.4% מכלל החיבורים שסווגו על ידי האלגוריתם כחיבורים תקינים הם למעשה חיבורי גיבריש. לעומת זאת בחיבורים בשפה הערבית נמצא 2% מהחיבורים שסווגו על ידי המחשב כחיבורי גיבריש הם למעשה חיבורים תקינים ו-4% מהחיבורים שסווגו על ידי המחשב כחיבורים תקינים הם למעשה חיבורי גיבריש.

הערות, דילמות והצעות נוספות לשיפור המדדים לאיתור חיבורי גיבריש:

- יש להחליט כיצד להתייחס לחיבור שחלקו תקין והמשכו גיבריש.
- קיים קושי בזיהוי חיבור הכולל טקסט שחלקו תקין וכתוב לנושא וחלקו כולל רשימה של מילים גבוהות ולא קשורות.
- יש לשקול את האפשרות להחליף את המדד הבודק 'חזרה על משפטים' במדד המבוסס על חלון של חמש מילים חוזרות החוזרות על עצמן בסדר קבוע. עם זאת, מדד זה צפוי להעמיס על המערכת.
- יש לשקול לבנות Stop-list של קללות כדי לזהות שימוש בקללות בחיבור. לצורך כך ניתן להיעזר ברשימת הסלנג של פרויקט HLP או להוציא את כל המחרוזות הטיפוסיות בחיבורים (types) בחיבורים שאינן מילים ולהפיק מהן רשימה של קללות או מופעי גיבריש טיפוסיים.
- משיקולי יעילות אפשר לשקול לבצע בשלב ראשון ניתוח מורפולוגי ללא הפגת עמימות מורפולוגית, לסנן את חיבורי הגיבריש וליישם הפגת עמימות רק על החיבורים התקינים.
- מומלץ לבדוק את האפשרות להיעזר במשתנים סמנטיים נוספים המופקים על ידי NiteRater שלא נכללים במדד הגיבריש, לאיתור חיבורים שאינם לנושא. בבדיקה ראשונית שנעשתה בהקשר זה נמצא כי שלושה מתוך ארבעת המדדים הסמנטיים המופקים על ידי NiteRater אינם יעילים מספיק באיתור חיבורים שנכתבו לא לנושא. עם זאת, המדד הרביעי - d_semantic (דרוג סמנטי מבוסס PCA) איתר בהצלחה 100 טקסטים (100%) שנלקחו ממקורות שונים ונכתבו לא לנושא⁴. מדד זה יושם גם על כל 2,719 חיבורי עברית שנכתבו במסגרת פסיכורשת ונמצאו תקינים וכתובים לנושא. בבדיקה זו נמצא שיעור מזערי (0.009) של false-alarm. דרוש מחקר נוסף כדי לבחון את תרומתו השולית של משתנה זה לקריטריונים שהוגדרו לאיתור חיבורי גיבריש שאינם לנושא. אפשרות נוספת היא להשתמש במודל שפה על החיבורים התקינים.

⁴ ע"פ הגדרת הקריטריון D_SEMANTIC > 10 (ממוצע+ס.ת בקורפוס פסיכורשת 18T : -9.74+0.12)

הצעות לצוות CAT לשיפור פסיכורשת

- יש לשקול את האפשרות להגביל את אורך החיבור (מס' המילים) לאורך הלגיטימי (השקול למס' המילים שניתן לכתוב בעמוד אחד), או לחילופין ליידע את הכותב שהערכת החיבור אינה מתייחסת לטקסט הנכתב בשורות העולות על המותר.
- רצוי להטמיע בפסיכורשת משוב לנבחן שכותב חיבור לא לנושא או חיבור גיבריש כדי להעביר את המסר שהמערכת מזהה חיבורים מסוג זה.

תכניות נוספות למחקר

- מומלץ להפיק הערכה אנושית למדגם חיבורי פסיכורשת לצורך בדיקת טיב הניבוי של הציון הממוחשב.
- מוצע לבצע ניתוח ממוחשב של קורפוס החיבורים הדיגיטלי של מבחן מפע"ם הן לצורך ישימות השימוש בהערכה ממוחשבת והן לצורך בדיקה נוספת של המדדים לאיתור חיבורי גיבריש.

כשיש יותר איכות, פוחת הצורך בכמות
ככל שמצליחים להופיק יותר הנאה ממאכל, כך אנו נזקקים לפחות ממנו,
וקל יותר לשמור על המידה הנכונה באכילתו.
לדוגמה, כאשר מתעוררת תשוקה לשוקולד, מיקוד הקשב בשוקולד ובהנאה מאכילתו,
עשוי להביא לסיפוקה של התשוקה כתוצאה מאכילת קוביית שוקולד אחת בלבד.
ללא מיקוד כזה, יתכן שיידרשו שלוש קוביות ואף יותר על מנת לספק תשוקה דומה.
בעקבות הגברת ההנאה מהכמות המופחתת, תבוא הנפש על סיפוקה ותרגיש בטוב,
בעוד רכיבי המזון המזיקים יפגעו בבריאות הגוף במידה המינימאלית האפשרית בלבד.
בדרך זו, ניתן לצמצם את צריכתם של רכיבי מזון מזיקים לגוף,
המצויים פעמים רבות במאכלים שנבחרים בלא-מודע כדי לספק צרכים נפשיים,
ואשר כלפיהם מתעוררת התשוקה.
כך מביאה הגברת ההנאה מהאכילה לשיפור הבריאות.
אם ברצונכם לנסות זאת באופן מעשי
בסדנאות שאני מעביר, המשתתפים בוחרים מאכלים שונים,
ואני מנחה אותם לאכול באופן שיעצים את הרגשת ההנאה מהאכילה.
קצרה כאן היריעה לפרט את התהליך כולו, אך הוא מתחיל בהשקטת הראש ככל הניתן ממחשבות
ומהסחות-דעת.
כשבא לכם לתרגל את הגברת ההנאה מאכילה, ביחרו מאכל טעים שאתם אוהבים,
מיצאו מקום שקט, והירגעו.
כל הבעיות והטרדות יכולות לקבל את מלוא תשומת לבכם אחר כך, ואם לא - נסו פעילות זו במועד אחר.
איכלו כמות קטנה מהמאכל. בנגיסה הראשונה,
נסו להגביר את תשומת הלב לטעם, כך שעוצמתו תתחזק.
איכלו עוד, והפעם הקצו תשומת לב גם להרגשת ההנאה עצמה.
ועכשיו - איכלו עוד ובלי לחשוב על כך, פשוט תיהנו...
בניגוד להדרכות הקודמות, אני ממליץ לעשות זאת פעם ביום לפחות (זה כיף!).
פעילות נוספת שאפשר ליהנות ממנה מאוד או פשוט לעשות אותה מהר זה להתקלח.
בזמן המקלחת אפשר לחשוב על דברים, להסתבן, להתגלח וכדומה... והופ - המקלחת נגמרה.
אם לא הספקתם לנסות את הגברת ההנאה מאכילה,
נסו להקדיש דקה מזמן המקלחת להנאה בלבד (ואז - מי צריך משהו מתוק?).

דוגמא 3: חיבור לא לנושא (לא אותר)

בישראל תלמידי התיכון נבחנו בבחינות בגרות העוזרות להם להתקבל לאוניברסיטה. כאשר מחשבים את ממוצע של בחינות אלו מקבלים על כל מקצוע שנלמד ברמת חמש יחידות 25 נקודות בונים. מדובר על יותר ממאה מקצועות שונים, עולה השאלה אם יש צורך להוריד את מספר המקצועות בעלי הבונוס לשמונה מקצועות ליבה. יש שיאמרו שרמת התלמידים נמוכה ולכן יש צורך לעודד אותם ללמוד את המקצועות הבסיסיים. אחרים יאמרו שבית הספר הוא המקום שבו הם מרחיבים את האופקים ולכן צריך להיות מבחר רחב, וכך כל תמיד ייחשף למקצוע שיעניין אותו, והרמה תישאר תוכל להיות גבוהה אם יעסיקו מורים מקצועיים.

תקופת בית הספר היא התקופה בה כל צעיר מתחיל להכיר את עצמו, מהם הדברים שמעניינים אותו, במה הוא טוב, ובמה פחות טוב. על בית הספר לעזור לאותו תלמיד להיחשף למידע מגוון בנושאים שונים. כאן נכנסות לתמונה מקצועות ההרחבה. בעזרת המגוון הרחב כל תלמיד יכול למצוא מקצוע אשר יעניין אותו, שייתכן שבדרך אחרת לא היה נחשף אליו. זהו השלב היחידי שאותו התלמיד יכול ללמוד כל מקצוע תוך התעלמות מוחלטת מהעתיד, לעומת האוניברסיטה שבחירת המקצוע שם נובעת ממספר רב של שיקולים.

אותם אנשים אשר רוצים לצמצם את מספר המקצועות בעלי הבונוסים יאמרו שהעובדה בשטח היא שרמת התלמידים נמוכה ועל כן יש למצוא פתרון לבעיה. פתרון אפשרי הוא צמצום הבונוס לשמונה מקצועות בסיסיים וכך התלמידים יהיו בעלי מוטיבציה רבה ללמוד ולהעמיק בהם. תלמידים שירחיבו מקצועות כמו אנגלית, מתמטיקה, היסטוריה ופיזיקה יהיו בעלי כלים רבים ויתכן שבעלי הצלחה גבוהה יותר באוניברסיטה. אך קודם יש לשאול את השאלה האם אנחנו פותרים את הבעיה או רק סימפטום שלה?

על פי האוניברסיטאות הסטודנטים נכנסים ברמה נמוכה יחסית לשאר מדינות העולם, אך האם זה נובע מריבוי מקצועות או אולי מרמת לימוד המקצועות. אם יעסיקו מורים מקצועיים ותכנית הלימוד של כל מקצוע תהיה ברמה גבוהה יכול להיות שהתלמידים יגיעו עם כלים מתאימים להתחיל לימודים באוניברסיטה.

לכן סוגיית המקצועות בעלי הבונוס אינה צריכה לעמוד על הפרק, אלא השאלה איך מעלים את רמת הלימודים מבלי לצמצם את הבחירה. כידוע לא כל תלמיד יכול להצליח בכל המקצוע וכל על הבחירה להישאר. יש לטפל בבעיה העיקרית – הרמה בה המקצועות נלמדים.

דוגמא 4: חיבור תקין

במדינות רבות, בהן ישראל, להורדת גיל הבחירה מ-21 ל-18 במספר מדינות, ובישראל מגיל 18 לגיל 17 או אף 16.

רבים תומכים בהצעה, אך אני סבור שזהו פתח לאסון.

בתור בן 18, אני מרגיש שאני מכיר את בני ה-16 וה-17, חלקם חברי. ומהיכרותי את הגיל הזה, אני מאמין שהורדת גיל הבחירות המדוברת הינה טעות מובהקת ממספר סיבות. הסיבה הראשונה, היא שבני נוער רבים אינם מתעניינים בפוליטיקה, ואם יבוקשו להצביע, לא ידעו למה להצביע, ואולי יצביעו מתוך לחץ חברתי למפלגה שאינם מכירים.

סיבה נוספת לכך שאני חושב שהורדת גיל הבחירות היא טעות, היא שבני נוער ידועים בחוסר אחריות ונסיון, ולכן יש אפשרות שבני נוער יצביעו בצורה שתפגע בהם מתוך יצרים רגשיים זמניים או מסיבות אחרות שאינן מנמקות את בחירתם בצורה טובה. כמובן שהאפשרות קיימת אף אצל בני ה-18 שזכאים לבחור כבר היום, אבל אני חושב שאף זוהי טעות, ואין סיבה להמשיך ולמנף את הבעיה הטמונה בדבר.

לסיכום, בתור בן 18, אני לא רואה בי ובבני גילי בוחרים לגיטימיים, וכך כמובן בבני ה-16 וה-17, כיוון שאין להם פנאי לתהיות פוליטיות, חלקם המכריע אינו מתעניין בפוליטיקה ועלולים בסבירות גדולה מאוד לדעת להצביע מתוך לחץ חברתי מצד חבריהם, ומשפחתם, או מסיבות בלתי לגיטימיות או אנטי-דמוקרטיות אחרות. לא להורדת גיל הבחירה!.

דוגמא 5: חיבור שנכתב לא לגמרי לעניין (לא אותר)

השאלה האם להוריד את גיל הבחירות הנוכחי, היא שאלה שמתחבטים בה כבר הרבה שנים.

ברוב המדינות גיל הבחירות הינו 18. עם זאת קיימות עדיין מדינות לא מעטות בהן גיל ההצבעה הוא 21. השאלה האמיתי העומדת מאחורי הדיו היא - מאיזה גיל אדם זכאי להיות מבוגר אחראי.

בכמה מדינות, ובהן ישראל, הוגשו הצעות חוק להורדת גיל הבחירות מ-18 ל-17 ואף ל-16,

ויש מדינות שכבר עשו כן. בדיון הציבורי על הורדת גיל הבחירה נשקלות שאלות כגון: מאיזה גיל אדם זכאי להשפיע על הרכב הפרלמנט, המקבל החלטות המשפיעות ישירה על חייו? האם בני הנוער בשלים דיים להחליט החלטות המשפיעות על הציבור כולו? מה יהיו ההשלכות של הכללת בני הנוער בקבוצת הבוחרים על המעורבות של בני הנוער בנעשה במדינה ועל המערכת הפוליטית בכללה?

למעשה, אחרי שקראתי שבדק את הבחינה מחשב, אני מעדיף למלא את החיבור בדברים שלא קשורים אליו בעליל. הרי בסופו של יום מה שיבדק הוא נכונות תכבירית של מילות קישור, הימצאות פיסוק שנראה מספק, ונוכחות ביטויים רבי נוכחות. כגון: "מרוב עצים לא רואים את היער", "תהום הנשיה", "בור סוד", "אוצר בלום" וכיוצא בזה. כמובן שגם מילים כגון- מרית, מריונטה, מקוף, מק והמקראמה מתאר את המק המקאברי עם המקוב הענק מקדמת דנא, גרם לרעד וחלחלה גם בלב בעלי העוז וחסרי המורא.

ניתן לראות שהמבנה שרשום בחיבורי הינו מבנה תקני אך, הנושא לא משתייך. הקונוטציה שעולה לי בקונטקסט זה, הינה קופסא שנשפטת על פי צורת העטיפה ולא על פי תוכנה. על זה נאמר -"אל תסתכל על הקנקן אלא בתוכו". אני מניח שבסופו של דבר, אם תעבירו את כל הבדיקות לממוחשבות - כולן יוכלו להיראות כך, מפוברקות עם מילים שלא רודפות אחת את השניה בקשר לוגי אותנטי.

לסיכום, אני מניח שאראה בקבלת ציוני האם נקודתי הוכחה או לאו. האיון של טענתכם לבדיקה נכונה על ידי מחשב גרמה לי עונג רב. אני מניח שאם החיבור יקבל ציון טוב זה יוכיח מעבר לכל ספק, כי לפעמים אכן כדאי לשמור על הסטאטוס קו. כמה מילות סיום - סטגנציה סורר ומורה נתן בפומבי בכף המקלעת, סיטרא אחרה, סיגים סטואי לסבר את האוזן.

פיתוח מדדים לאיתור חיבורי ג'יבריש ותיקוף המדדים מול חיבורי פסיכורשת

דוגמא 2: חיבור לא לנושא (אותר בהצלחה)

אנ מרגע חצארטנא ומבאדנא תעוד אל תאריר, אלזי סגלח אנאס וכטבח אשכאס קד געל מנחמ המל או זכרב בהמ המל. فالانسان بدون תאריר לא יוגד לה חلفיה תעכס صورته وبلورته لنفسه فبات التارير لى البعض ليس ذو اهميه على الرغم ان الحية تثبت عكس هذا.

الانسان كانسان ولد حر ويعيش حيات باستقلاليه بحاجه الى ان يصل الادراك والوعي لكي يحدد شخصيته وسلوكه ويعتبر التارير جزء لا يتجزء في المساهمه لكي يصل الى هذا الادراك فان التارير يعرض ويكشف لنا احداث وقصص وروايات واقعيه وحقيقيه قد عاشها اسلافنا وقد مرت با حضارتنا ويكتبها ويرويها لنا بكل مصداقيه موضوعيه خاليه من اي اعتبارات سياسيه ايديولوجيه او دينيه فلا يميز بين الجنس والعرق فتراه يعرض جميع الصور على حقيقتها مبين لنا السلبيات والايجابيات ومنها يستخلص العبر والعبارات, التي تعطينا خطوه في تطوير كياننا وذاتنا لتحسين معيشتنا والتعلم مما مارسه اجدادنا سواء كان ايجابيا او سلبيا.

على الرغم ان الكثير من البشر يعارضون فكرت التارير وتعليمه الا انهم غير مدركين بأنه حاكم لا محكوم فبمضي الوقت وعبور الزمن يكتب التارير او على يد عظماء او يروي عن حمقاء ولن ننسى انه المرجع الاساسي لبدء التكوين وولادة الانسان منذ ولادة ابونا ادم ولا اتوقع انه سيتوقف في الكتابه عن وعلى جميع البشر فعلينا استغلال هذا الشيء لعلنا نكون ممن كتبو التارير وممن كتب بهم التارير في سيطرتنا وسلوكنا الواعي , وفي انجازاتنا الشرفيه الرفيعه في تحقيق اهدافنا ومجارات الرياح في الوصل الى القمه فصعب ان تصل اليها ولكن الاصعب ان تحافظ عليها فيا ساده ان التارير لا يعترف في الفاشلين فهو حاد كسيف لا يميز بين مسؤول وعامل فقير بين ملك ورعيته فمن اراد البقاء في القاع المزدهم لن يصل الى التارير ولن يجد مكان في القاع بات لا مكان من التراحم.

عمليا وعلميا ان جميع البشر لديهم قوه باطنيه يستطيعون ان يبلغوا بها اعلا المناطق والمراتب والقوه وعليهم النظر الايجابي والتمحور حول قدرتهم والايمان بها فتاريخنا قد بدأ منذ ملايين السنين وهو مستمر واكبر اثبات على القوه البشريه لمن ليس له وعي او غاضا للنظر تاريخنا يسمو بنا ونحن قصه جاريه مستمره لا محاله .

דוגמא 3: חיבור תקין (אותר)

בחסב ראי תלמ התריח פי המדסה הו אמר מם מן נחיה וגיר מם מן נחיה ארי, אנה מם לانه מוסוע מן המוסיע الاساسية والاجبارية פי המדסה ולانه عن طريقه يمكننا ان نعلم احداثاً وقصصاً عن الماضي, ومهم ايضاً للاشخاص الذين يريدون ان يصبحوا معلموا تاريخ, وغير مם لان الناس لا يعطوه اهمية ولانه لا يفيدنا في حياتنا اليومية أي انه شيء لا يستعمل بالنسبة لنا.

בחסב ראי תדריס התריח פי המסע العربي מן המمكن ان يساهم في المجتمع ولكنه لا يساهم بالانسان ككل لان ليس الجميع لديه اهمية فائقة تجاه هذا الموضوع.

الحروب والاتفاقيات والمعاهدات, كلها كانت في زمن الماضي وكان لها اهمية اكثر من زمن الحاضر, او بالاحرى قلت نسبتها, ممكن ان يكون السبب هو انه لكل دولة يوجد نظام وحدود معينة وتقاليد وعادات, أي انه من غير الممكن ان تتعدى دولة على دولة اخرى, او ان تخترق حدودها.

وايضا تدریس התריח يساعد الانسان في حياته اليومية, مثال على هذا: بدون موضوع التاريخ لا يتم قبول الطالب لدخول الجامعة فيجب ان يكون بحوزته وحدتين في موضوع التاريخ على الاقل.

يساعد تدریس התריח, المجتمع وكما ذكر سابقاً لكي يكون الاشخاص على علم بما يدور من حولهم او ما كان يدور بالماضي من احداث ومجريات حصلت, وكما تعلم ان موضوع التاريخ وتدریسه هو امر مهم للجميع ولانه يوجد بعض المعلومات او بعض الموضوعات منه تتعلق بموضوع المدنيات والعلوم السياسية لذا يساعد على معرفة الاشياء وفهمها اكثر.

في باقي المجتمعات مثلاً في المجتمع اليهودي تدریس התריח مختلف عن تدریس התריח في المجتمع العربي, أي ان المادة تختلف كلياً وكما اعتقد ان وضع الاهتمامات هو نفسه وان استعماله للحياة اليومية قليل.

נספח 3: התפלגות פרופילים של חיבורי הג'בריש בשפה העברית

פרופיל 12345678910	שכיחות	אחוז	פרופיל 12345678910	שכיחות	אחוז
0000000100	48	19.51	1010101100	3	1.22
0000001111	34	13.82	1010101101	3	1.22
1000001111	21	8.54	1010101110	3	1.22
1101000000	13	5.28	1010101111	3	1.22
1000101110	12	4.88	1100000100	3	1.22
0000101110	7	2.85	0000001000	2	0.81
1000101100	7	2.85	0000001011	2	0.81
1100000000	7	2.85	0000101111	2	0.81
0001000000	6	2.44	0101000000	2	0.81
1000101000	6	2.44	1000001000	2	0.81
1000000100	5	2.03	1000001101	2	0.81
1000001100	5	2.03	1000100100	2	0.81
1000000000	4	1.63	0000001010	1	0.41
1000001110	4	1.63	0000001110	1	0.41
1000101111	4	1.63	0000101010	1	0.41
0010101111	3	1.22	0010101000	1	0.41
0100000000	3	1.22	1000001010	1	0.41
0100000100	3	1.22	1001000000	1	0.41
1000001011	3	1.22	1001001110	1	0.41
1000101001	3	1.22	1001101110	1	0.41
1000101010	3	1.22	1010001110	1	0.41
1000101101	3	1.22	1010100111	1	0.41
1001001100	3	1.22			

נספח 4: התפלגות פרופילים של חיבורי הג'בריש בשפה הערבית

פרופיל 12345678	שכיחות	אחוז
00000001	16	27.59
10001011	10	17.24
00000011	5	8.62
00000010	4	6.9
11010000	4	6.9
00001010	3	5.17
00001011	3	5.17
10000001	3	5.17
10000000	2	3.45
10001010	2	3.45
01000000	1	1.72
01000010	1	1.72
10000011	1	1.72
10000100	1	1.72
10010000	1	1.72
10100011	1	1.72

נספח 5: טבלת הקריטריונים לזיהוי חיבורי ג'יבריש בשפה העברית

שם הקריטריון	הסבר וערך סף	שמות המשתנים בתוכנה
1. TTR	$TTR > 0.30$	S_LETTER_STRING_TTR
2. משפטים חוזרים	3 משפטים שונים החוזרים על עצמם לפחות פעמיים כל אחד במקום כלשהו בטקסט.	S_SNTC_REPEATING_CNT
3. פסקאות קצרות	אחוז גבוה של פסקאות קצרות מאד (פחות מ-5 מילה): $0.90 <$	S_EXTREMELY_SHORT_NL_P
4. טקסט חופף גריין במרביתו	(העתקת הגריין) - אחוז גבוה של משפטים חופפים גריין או מחרוזות חופפות גריין: $0.80 <$	C_TEASER_SNTC_CONGRUENCE_P or C_TEASER_CONTENT_TYP_STRNG_P
5. אורך מילה	אחוז מילים באורך 10 ומעלה < 0.30	S_LETTER_STRING_LENGTH_P_1 + S_LETTER_STRING_LENGTH_P_10
6. מילים זרות	אחוז גבוה של מילים שכתובות באותיות שאינן עבריות: $0.50 <$	S_FOREIGN_WORD_P
7. רצפים לא חוקיים	שלושה רצפים של 4 מילים לא חוקיות (מילים באורך 10 תווים או יותר, מילים שאינן קיימות במילון, מילים לא חוקיות ומילים החוזרות על עצמן)	M_ILLEGAL_SEQUENCE_CNT
8. טקסט שאינו חופף גריין	(לא לנושא) - אחוז מחרוזות וגם אחוז לקסמות חופפות גריין: > 0.05 או אחוז מזערי של מחרוזות חופפות גריין: > 0.015	C_TEASER_CONTENT_TYP_STRNG_P and C_TEASER_CONTENT_TYP_LEX_P
9. מילים לא חוקיות בשפה	אחוז מילים לא חוקיות (אות סופית באמצע מילה, אות תחילית בסוף מילה): $0.05 <$	S_NON_LEGAL_END_LETTERS_P
10. מילים לא במילון	אחוז גבוה של מילים שאינן קיימות במילון: $0.30 <$	M_MORPH_NON_DICT_P