

סיגנאל של מערכת צפני ELS: ניסויים חדשים ב"תחום המורחב"

מאת דורון ויצטום

תקציר

בעבודת WRR [1] הועלתה טענה לגבי מציאותה של מערכת צפני ELS בספר בראשית. לשם כך נעשה ניסוי מדעי (שבו נמדדה רשימת נתונים בספר בראשית) שתוצאותיו תמכו בטענה זו. לפי ניסיון שהצטבר [2] מאז פורסמה עבודת WRR נראה הדבר, כי התופעה לה טענו WRR אינה מוגבלת לספר בראשית, אלא מצויה בתורה כולה. אנו רואים את כל התורה כחטיבה אחת לעניין התופעה שלפנינו [3], כאשר נושאים שונים מוצפנים באזורים שונים בספר. בבחינה זו, ספר בראשית בו נמדדה רשימת הנתונים, אינו אלא הקטע [1, 78064] מספר התורה. במאמרנו סיגנאל של מערכת צפני ELS: דעיכה מול הגברה [4] הסקנו, כי תחום ההצפנה של רשימת הנתונים הנ"ל קרוב יותר לתחום המורחב, שהוא הקטע [1, 156128] מן התורה, מאשר לתחום המקורי, הקטע [1, 78064] מספר התורה. לכן, יש עניין רב לערוך ניסויים נוספים בנושא רשימת נתונים זו בתחום המורחב. כאן אנו מדווחים על ניסויים נוספים בנושא רשימת הנתונים הנ"ל בתחום המורחב. בכולם נתקבלה מובהקות חזקה.

מבוא

טענתם המרכזית של MBBK [5] היא, כי נעשתה מניפולציה ברשימת הנתונים על מנת לקבל מובהקות חזקה בניסוי WRR [1] בספר בראשית. טענת MBBK מתמקדת בתת-רשימה של הנתונים, עבודה נתקבלה המובהקות החזקה ביותר בניסוי WRR. נכנה תת-רשימה זו L2. (ראה פרטים בנספח של [4], סעיף א). לגופה של טענה זו כבר ענינו באופן פרטני במקום אחר (ראה [6]).

במאמרנו סיגנאל של מערכת צפני ELS: דעיכה מול הגברה [4], הצגנו ניסוי אשר בודק, אף מבלי להיכנס לגוף הטענה, האם L2 מתנהגת כרשימת נתונים "תפורה". הרעיון מאחורי הניסוי הוא, שסיגנאל הנוצר באופן מלאכותי באמצעות "תפירה" של מערכת צפני ELS בקטע מוגדר מטקסט, צפוי לדעוך כאשר המדידות לזיהוי נעשות בקטע ארוך יותר, המכיל את הקטע המקורי. אנו מראים שם, בניסויי הביקורת, כי אכן סיגנאלים מלאכותיים מסוג זה דועכים (ובאופן חד) כאשר המדידות נעשות בקטע טקסט הכפול באורכו מן הטקסט המקורי.

לעומת זאת, בדיקה דומה בספר התורה עבור L2, העלתה תוצאה הפוכה. תחום המדידה המקורי היה קטע בן 78064 אותיות מתחילת ספר התורה (כלומר, כל ספר בראשית). תחום המדידה החדש הינו קטע מתחילת ספר התורה, הכפול באורכו מתחום המדידה המקורי. התברר, כי בתחום המדידה החדש הסיגנאל מתחזק באופן בולט במקום לדעוך.

במאמרנו זה אנו מוסיפים ובודקים באופן דומה, האם קיימת התגברות הסיגנאל גם עבור שני וריאנטים של L2, וזאת בעקבות רעיון של MBBK.

א. הטענה של חוסר סימטריה:

1. נזכיר, כי רשימות הנתונים ששימשו לניסויי WRR הן רשימות של זוגות ביטויים. בכל זוג, אחד מבני הזוג הוא שם או כינוי של אישיות מסוימת, ובן זוגו הוא תאריך הלידה או הפטירה של אישיות זו.

כל תאריך הוצג בשלוש תצורות. למשל היום הראשון בחדש תשרי הוצג כך:

A - "א' תשרי",

B - "בא' תשרי",

C - "א' בתשרי".

הדבר נקבע באופן אפריורי לפני הכנת רשימת הנתונים הראשונה (ראה על כך בנספח).

2. לטענת *MBBK* חסרה כאן תצורה רביעית: D – "בא' בתשרי".

הרציונל העומד מאחורי טענת "חסר" זו, מצריך סימטריה: D מתייחס ל-C כמו B ל-A.

נסביר: B נוצר מ-A באמצעות תוספת "ב" השימוש – ב"א' תשרי".

לכן יוצרים את D מ-C באמצעות תוספת "ב" השימוש – ב"א' בתשרי".

טענה זו הועלתה ממש מתחילת הוויכוח, והוצגה בהבלטה רבה בכל פרסומי *MBBK* כראיה למניפולציה. (ראה על כך בנספח. מבואר שם, כי טענה זו הופרכה כבר על ידינו).

3. אם שלישיית התצורות ABC נחשבת ללא-סימטרית בגלל שתצורה C חסרה "בן זוג", אפשר לתקן זאת בשתי דרכים:

(א) הוספת תצורה D – כפי שהציעו *MBBK* – וכך מקבלים את רביעיית התצורות ABCD.

(ב) הורדת תצורה C – וכך מקבלים את צמד התצורות AB.

נסמן ב- $L2(ABCD)$ את רשימה $L2$ כאשר התאריכים בה ניתנים ברביעיית התצורות ABCD.

נסמן ב- $L2(AB)$ את רשימה $L2$ כאשר התאריכים בה ניתנים בצמד התצורות AB.

אנו נעמיד את שתי האפשרויות במבחן. להלן נמדוד את הסיגנאלים של $L2(ABCD)$ ושל $L2(AB)$ בספר בראשית ובקטע המורחב.

ב. הגדרות:

אנו משתמשים באותן הגדרות שבמאמרנו סיגנאל של מערכת צפני *ELS*: דעיכה מול הגברה [4], ונביאן כאן בקיצור.

1. יהא T טקסט, נגדיר קטע בן n אותיות D , בטקסט T כך:

$$D \equiv ([t_{d_1}, t_{d_n}], T)$$

כאשר t_{d_1} הוא המספר הסידורי ב- T של d_1 , האות הראשונה בקטע D , ו- t_{d_n} הוא המספר הסידורי ב- T של d_n , האות האחרונה בקטע D .

2. יהא T טקסט, D קטע ממנו, $LIST$ רשימה של זוגות ביטויים, ו- $ELS(LIST, D)$ קבוצת *ELSs* המייצגים את הביטויים ב- $LIST$ והמוגדרים ב- D .

נגדיר סיגנאלים $S_i(LIST, D)$ ($i = 1, 2$) כך, שהסיגנאל יקבל ערך הגדל ככל שהמפגשים בין צפני ה-*ELS* יותר "מוצלחים". (למעשה, $S_i(LIST, D) \equiv 1 / P_i(LIST, D)$, כאשר P_1 ו- P_2 הוגדרו ב[1]. הפרטים נמצאים ב[4].)

ניתן לקבוע את מובהקות הסיגנאל, כלומר את הסיכוי "לקבל סיגנאל כה גבוה במקרה", באמצעות "מבחן טקסטים", כלומר באמצעות השוואה לקבוצה גדולה של טקסטים דומים (ראה ב[4] בנספח, סעיף ג).

3. יהא D' קטע ב- T , ארוך מ- D והמכיל אותו. הסיגנאלים המתקבלים עבור $LIST$ ב- D' יסומנו Q_i ($i = 1, 2$). נגדיר את מידת ההגברה Q_i :

$$Q_i \equiv S_i(LIST, D') / S_i(LIST, D)$$

$Q_i < 1$ מצביע על דעיכה, היחלשות S_i , ואילו $Q_i > 1$ מצביע על הגברת S_i .

4. יתכן כי עבור i שונים תתקבל מידת הגברה שונה. אף ייתכן ש- Q_1 יצביע על הגברה, בעוד Q_2 מצביע על דעיכה, או להיפך. אנו מעוניינים במידת ההגברה המירבית Q :

$$Q \equiv \max\{Q_1, Q_2\}$$

המדידות

חלק ראשון

מדידה 1: בדיקת $L2(ABCD)$ בקטע גדול מספר התורה.

- הקטע המקורי D , הוא "הקטע $[1, 78064]$ מן התורה" (1 הוא המספר הסדורי של האות הראשונה בתורה, ו-78064 הוא המספר הסדורי של האות האחרונה בספר בראשית), והוא ספר בראשית (G) : $D \equiv ([1, 78064], TORAH) = G$
- הקטע D' שהוא הקטע $[1, 156128]$ מן התורה: $D' \equiv ([1, 156128], TORAH)$

א. ערכנו שתי מדידות:

- בספר בראשית (G) .
- בקטע D' .

1. התוצאות בספר בראשית:

$$S_1(L2(ABCD), G) = 4.5979E+6, S_2(L2(ABCD), G) = 1.7889E+8.$$

2. התוצאות בקטע D' :

$$S_1(L2(ABCD), D') = 3.9210E+11, S_2(L2(ABCD), D') = 2.0161E+8.$$

לכן:

$$Q_1 = 8.5278E+4, Q_2 = 1.1270, Q = \max\{Q_1, Q_2\} = 8.5278E+4.$$

ב. צריך להעריך מהי המובהקות של התוצאה הנ"ל, דהיינו: "מה הסיכוי שיתקבל במקרה Q כה גדול?"

1. לשם כך בנינו אוסף גדול של N טקסטים, T_j , "דומים" לקטע D' מספר התורה, המוגדר לעיל. הקטע D' בנוי משני חלקים:

- מחלקו הראשון, $D \equiv ([1, 78064], TORAH) = G$
- ומתוספת – הקטע D'' מן התורה: $D'' \equiv D' - D = ([78065, 156128], TORAH)$

כל טקסט T_j אף הוא בנוי משני חלקים:

- מחלקו הראשון, $D \equiv ([1, 78064], TORAH) = G$
 - ומתוספת – הטקסט T''_j - שהוא ערבוב אקראי של הקטע D'' (ערבוב אקראי של מלים בתוך פסוקים. הדבר נעשה בדיוק כפי שנעשה ב[3] לגבי טקסט U שם).
- טקסטים אלה הם הטקסטים שהוכנו בעבודתנו [4].

2. כאשר מודדים את $L2$ בטקסט T_i , התוצאה הסופית "נהנית" מכל היתרונות (והחסרונות) שהיו במדידה המקורית של $L2$ בספר בראשית. במלים אחרות: כל רץ T_j נהנה בדיוק מאותה "רוח גבית". בודקים עבור N טקסטים T_j בכמה טקסטים n מתקבל ערך Q גדול או שווה ל- $8.5278E+4$. הסיכוי לקבל כזו תוצאה במקרה הוא

$$p = n/N$$

3. תוצאות:

מתוך 50,000 טקסטים T_j היו רק 4 טקסטים בהם נתקבל Q כה גדול. לפיכך:

$$p = 0.00008.$$

מדידה 2: בדיקת $L2(AB)$ בקטע גדול מספר התורה.

א. נתוני הטקסטים כנ"ל. ערכנו שתי מדידות:

- בספר בראשית (G) .
- בקטע D' .

1. התוצאות בספר בראשית:

$$S_1(L2(AB), G) = 3.0998E+4, \quad S_2(L2(AB), G) = 8.4746E+6.$$

2. התוצאות בקטע D' :

$$S_1(L2(AB), D') = 2.9279E+10, \quad S_2(L2(AB), D') = 4.5249E+7.$$

לכן:

$$Q_1 = 9.4454E+5, \quad Q_2 = 5.3394, \quad Q = \max\{Q_1, Q_2\} = 9.4454E+5.$$

ב. מובהקות התוצאה:

בהשוואה לטקסטים כנ"ל, נמצא כי מתוך 50,000 טקסטים T_j היו רק 19 טקסטים בהם נתקבל Q כה גדול. לפיכך:

$$p = 0.00038.$$

חלק שני

במאמרם [5] העלו *MBBK* טענות אחדות בנוגע לתאריכים שננקטו ברשימות הנתונים של *WRR* [1]. מטענות אלה רלוונטית רק טענה אחת ויחידה לגבי תת הרשימה $L2$: כי עבור 4 אישים יש להוסיף או לשנות תאריך (הפרטים ב[4] בנספח, חלק ה). במאמרנו [4] הגדרנו את הרשימה $L2M$ כתת הרשימה $L2$ עם השינויים הנ"ל בתאריכים (כל שאר התאריכים והפרטים, בפרט השמות והכינויים – נשארו כפי שהם, ללא שינוי).

מדידה 1: בדיקת $L2M(ABCD)$ בקטע גדול מספר התורה.

א. נתוני הטקסטים כנ"ל. ערכנו שתי מדידות:

- בספר בראשית (G).
- בקטע D' .

1. התוצאות בספר בראשית:

$$S_1(L2(ABCD), G) = 1.1832E+8, \quad S_2(L2(ABCD), G) = 8.2651E+9.$$

2. התוצאות בקטע D' :

$$S_1(L2(ABCD), D') = 4.2589E+14, \quad S_2(L2(ABCD), D') = 1.3228E+10.$$

לכן:

$$Q_1 = 3.5994E+6, \quad Q_2 = 1.6005, \quad Q = \max\{Q_1, Q_2\} = 3.5994E+6.$$

ב. מובהקות התוצאה:

בהשוואה לטקסטים כנ"ל, נמצא כי מתוך 50,000 טקסטים T_j היו רק 3 טקסטים בהם נתקבל Q כה גדול. לפיכך:

$$p = 0.00006.$$

מדידה 2: בדיקת $L2M(AB)$ בקטע גדול מספר התורה.

א. נתוני הטקסטים כנ"ל. ערכנו שתי מדידות:

- בספר בראשית (G).
- בקטע D' .

1. התוצאות בספר בראשית:

$$S_1(L2(AB), G) = 9.0106E+4, \quad S_2(L2(AB), G) = 4.2194E+7.$$

2. התוצאות בקטע D' :

$$S_1(L2(AB), D') = 1.8211E+13, \quad S_2(L2(AB), D') = 3.4723E+9.$$

לכן:

$$Q_1 = 2.0211E+8, \quad Q_2 = 8.2294E+1, \quad Q = \max\{Q_1, Q_2\} = 2.0211E+8.$$

ב. מובהקות התוצאה:

בהשוואה לטקסטים כנ"ל, נמצא כי מתוך 50,000 טקסטים T_j היו רק 5 טקסטים בהם נתקבל Q כה גדול. לפיכך:

$$p = 0.00010.$$

מסקנות

- א. בכל ארבע המדידות מדד ההגברה המירבית הצביע על התחזקות ברורה וחדה ובמובהקות חזקה.
- ב. העובדה כי רביעיית תצורות התאריך ABCD, ואף צמד התצורות AB, מצליחים יותר מאשר שלישיית התצורות ABC של WRR [1] – עשויה להצביע על האופן בו מקודד המידע בספר התורה. למשל, במקרה זה מתקבל הרושם, כי עיקר המידע מקודד באמצעות צמד התצורות AB.
- ג. האמור בסעיף הקודם הוא אינדיקציה בלבד וכדי לאששה יש צורך במחקרים נוספים.

נספח

א. רשימות הנתונים ששימשו לניסויי WRR הן רשימות של זוגות ביטויים. בכל זוג, אחד מבני הזוג הוא שם או כינוי של אישיות מסוימת, ובן זוגו הוא תאריך הלידה או הפטירה של אישיות זו. כל תאריך הוצג בשלוש תצורות. למשל היום הראשון בחדש תשרי הוצג כך:

A - "א' תשרי",

B - "בא' תשרי",

C - "א' בתשרי".

הדבר נקבע באופן אפריורי לפני הכנת רשימת הנתונים הראשונה, ע"י הבלשן יעקב אורבך ז"ל ששימש יועץ ל-WRR בעניינים הלשוניים. לפי הנחיה זו הוכנה רשימת הנתונים הראשונה ובוצע עליה ניסוי פירוט הרשימה הראשונה ותוצאות הניסוי פורסמו לפני ש-WRR נדרשו להכין רשימת נתונים נוספת. רשימת נתונים נוספת זו, הידועה כ-LIST 2, הוכנה אף היא באותו אופן, והיא הרשימה אשר לגביה נערך מבחן הפרמוטציות שתוצאותיו פורסמו ב[1].

ב. טענת MBBK, שהועלתה ממש בתחילת הוויכוח, והיתה השאלה הראשונה הנוגעת לגוף הנתונים [7], היא: "מדוע ברישום התאריכים השתמשם רק ב-3 צורות מ-4 (א תשרי, א בתשרי, בא תשרי, אך לא בא בתשרי)?" הם הוסיפו כי באנציקלופדיה מרגליות, שהיתה נקודת המוצא לרשימת האישים, נעשה שימוש בארבע התצורות.

טענה זו חזרה בהבלטה בכל פרסומיהם, כולל [5]. למעשה, הם טענו כי נבחרו רק שלוש הצורות ABC ולא D, כי כך השיגו WRR תוצאה טובה יותר.

ג. ובכן, איננו יודעים את שיקוליו של יעקב אורבך ז"ל, בפרט איננו יודעים אם בדק או התחשב בכתוב באנציקלופדיה מרגליות. אבל אנו יודעים שלושה דברים בבירור:

1. כי זו בחירה ראויה: למשל, כאשר האנציקלופדיה העברית השתמשה בתצורות סטנדרטיות בערכי האישים שבמדגם, היא השתמשה בדיוק בשלוש התצורות ABC, ולא בתצורה D.
2. כי אי אפשר להכחיש שזו בחירה אפריורית לגבי הרשימה השניה (ראה א לעיל).
3. כי לפי אמות המידה הסטטיסטיות ששימשו למדידת המובהקות בזמן יצירת הרשימה הראשונה – שימוש בתצורה D היה דווקא משפך את המובהקות! (ראה [6]).

7. השאלה לגבי תצורות התאריך מעניינת לכשעצמה (ראה מסקנות ב-ג).

הכרת תודה

החשובים נעשו באמצעות תוכנה של יעקב רוזנברג. רוברט האראליק עורר אותי לחשב את המובהקות באמצעות השוואה לטקסטים "דומים", ושולם סרברניק העלה, כדרכו, רעיונות מועילים. תודה ל-MBBK על שהסבו את תשומת לבנו לעניין תצורות התאריך.

מקורות והערות

1. עבודת WRR:

Witztum, D., Rips, E. and Rosenberg, Y. (1994). *Equidistant letter sequences in the Book of Genesis*. *Statist. Sci.* 9 No. 3 429-438.

2. הנה כמה מן העבודות שנעשו בספר התורה (גם מחוץ לספר בראשית): באתר הנ"ל, במדור "פרסומים מדעיים", http://www.torahcode.co.il/pub_index_heb.htm, מאמרים 5, 8, 9, וכן במאמרים http://www.torahcode.co.il/kriah1_heb.htm, http://www.torahcode.co.il/weis2_heb.htm.

3. זאת ניתן לראות בצפנים המחברים בין החומשים (למשל, בשני הקישורים האחרונים).

4. ד. ויצטום (התשע"ה): סיגנאל של מערכת צפני ELS: דעיכה מול הגברה.

5. עבודת MBBK:

McKay, B. D., Bar-Natan, D., Bar-Hillel, M. and Kalai, G. (1999). *Solving the Bible Code puzzle*. *Statist. Sci.* 14 No. 2 150-173.

6. ד. ויצטום (התשס"א): על בחירת התאריכים למדגמים של WRR. ראה באתר הנ"ל בקישור:

http://www.torahcode.co.il/date_heb.htm

7. "מסמך 2" בקישור http://www.torahcode.co.il/pdf_files/oppose/docum2h.pdf

ראה שם שאלה מס' 4.