

לפי מסורת חז"ל ארבעה שערים לתורה:
פשט, רמז, דרש וסוד (פרד"ס).

כאן נתמקד באחת מדרכי הרמז:
דילוג שווה של אותיות

דוגמא לביטוי בדילוג שווה של אותיות:

הכנס השישי למדע ותודעה
הרמוניה בין מדע, תודעה ויהדות, האומנם?

בתוך הכותרת שניתנה לכנס הנוכחי התגלה **עב"ם** (**עצם בלתי מזוהה**), ולא סתם עב"ם, אלא **העב"ם** בה"א הידיעה. **העב"ם** מופיע כאן בדילוג שווה של 5 אותיות לאחור ($d = -5$).

נא לשים לב:

1. איננו סופרים את הרווחים בין המלים.
2. איננו מבדילים בין אותיות סופיות לרגילות.
3. אנו מתעניינים בדילוג שווה קדימה וגם בדילוג אחורה.

ביטויים בדילוג שווה צפויים להופיע **באקראי** בכל טקסט:
ספר, מכתב, עיתון, או טקסט אקראי שנוצר על ידי מכונת הגרלה.

המחקר המדעי של הצופן בתורה

[שם מקוצר של "המחקר המדעי של שיטת הרמז בדילוג השווה בתורה".]

המחקר המדעי של הצופן בתורה החל בהשראת עבודתו החלוצית של הרב חיים מיכאל דב וייסמנדל זצ"ל. לפני למעלה משבעים שנה (וללא עזרת מכונות חישוב), הוא הדגים מערכות מסוימות של מלים וביטויים בדילוגים שווים בתורה - תבניות (patterns) - הנראות כמתוכננות.

המחקר המדעי של הצופן בתורה בודק:

האם אפשר להבחין בכלים מדעיים - אובייקטיביים, כי בספר התורה מופיעים ביטויים בדילוג שווה שלא באקראי?

השימוש במודל האקראי

המחקר המדעי של הצופן בתורה עושה שימוש בהשוואה **למודל האקראי** (דהיינו, למצב הנוצר באקראי על ידי מכונת הגרלה). ההשוואה למודל האקראי היא שימושית ומקובלת לגבי טקסטים או מאגרים גדולים של נתונים אלפביתיים או מספריים.

כידוע, ישנם תחומי מחקר שונים שבהם שואפים לגלות **תבניות סמויות** (hidden patterns) למיניהן:

1. המחקר לפענוח הקוד הגנטי של שרשראות הנוקליאוטידים.
2. המחקרים לגילוי תשדורות בין-כוכביות.
3. המחקר ביחידות הפענוח בכל חילות המודיעין.

היסוד המתודולוגי בתחומים הללו הוא חיפוש תבניות שיש בהן משום סטייה משמעותית מן המודל האקראי.

המודל האקראי משמש כבסיס למאמצים האדירים המושקעים בתחומים אלה, והמסקנות המתקבלות בדרך זו משמשות יסוד להחלטות חשובות.

[עוד בעניין זה - בספר **צופן בראשית**]

העבודה של ויצטום, ריפס ורוזנברג בספר בראשית

מכאן ואילך נתמקד בעבודה של דורון ויצטום, אליהו ריפס ויואב רוזנברג, בספר בראשית. עבודה זו פורסמה בסטטיסטיקאל סאיינס:

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

מינוח

מופע מינימלי: ביטוי עשוי להופיע כמה פעמים בדילוג שווה בטקסט, כאשר **גודל** הדילוג משתנה ממופע למופע. למופע בעל **הדילוג הקצר ביותר בקטע נתון** בטקסט נקרא: **המופע המינימלי** של הביטוי בקטע.

צפני ELS: למופעים של ביטויים, שהם **מינימליים** **בקטעים גדולים בטקסט**, נקרא להלן "**צפני ELS**" -

ELS = Equidistant Letter Sequences.

בעבודה זו התמקדו החוקרים **בצפני ELS**

מפגשים של צפני ELS המייצגים ביטויים קשורים מושגית על פני טבלאות דו-ממדיות

דוגמא למפגשים על פני טבלה דו-ממדית של צפני ELS המייצגים ביטויים קשורים מושגית:
לפנינו קטע מספר בראשית, כתוב כסדרו, אות אחר אות ללא רווחים ביניהן, בצורת טבלה
ובה 49 טורים:

ישכחהווייהימקצשנתימימימופרעהחלמוהנהעמדעלהיארונהמנ
היארעלתשבפרותיפותרמראהובריאתבשרותתירעינהבאחווהנהשבע
פרותאחרותעלותאחריהנמנהיאררעותמראהודקויתבשרותעמדנהא
צלהפרותעלשפתהיארותאכלנההפרותרעותהמראהודקתהבשראתשב
עהפרותיפתהמראהוהבריאיתוייקצפרעהויישנויחלמשניתוהנהש
בעשבלימעלותתבנהאחדבריאאותוטבותוהנהשבעשבלימדקויתושדו
פתקדימצמחותאחריהנותבלענההשכלימהדקויתאתשבעהשכלימהבר
יאותוהמלאותווייקצפרעהוהנהחללומוייהיבבקרויתפערמרוחווישל
חוויקראאתכלחרטמימצרמימואתכלחכמיהויספרפרעהלהמאתחלמוו
אינפותראותמלפרעהוידברשרהמשקמימאתפרעהלאמראתחטאיאנימ
זכירהוימפרעהקצפעלעבדיוויתנאתיבמשמרביתשרהטבחמימאתיו
אתשרהאפימונחלמהחללומבלילהאחדאניוהואאישכפתרונחלמוחל
צפן חבוי, בתורתי

בטבלה זו קיים מפגש של זוג צפני ELS קשורים מושגית:

צפן חבוי - בתורתי

ידוע כי:

מפגשים של צפני ELS קשורים מושגית על פני טבלאות דו-ממדיות עשויים להתרחש **באקראי** גם בטקסטים אחרים. **שכיחות המפגשים תתאים למודל האקראי**.

השערת האפס

מפגשים של צפני ELS קשורים מושגית על פני טבלאות דו-ממדיות מתרחשים **גם בספר בראשית כצפוי** לפי המודל האקראי.

השערת המחקר

מפגשים של צפני ELS קשורים מושגית על פני טבלאות דו-ממדיות מתרחשים **בספר בראשית יותר מן הצפוי** לפי המודל האקראי.

[הרחבה בעניין זה - בספר **צופן בראשית** ובאתר **צופן בראשית**]

בדיקת ההשערה

ארבעה שלבי הבדיקה:

שלב א:

מגדירים את "מידת הקרבה המכויילת" בין שתי מלים. זה מספר בין 0 ל-1. ככל שמופעי המלים כצפני ELS מינימליים בקטעים גדולים יותר, וככל שהם קרובים יותר זה לזה ופחות מפוזרים על פני הטבלאות, אזי "מידת הקרבה המכויילת" קטנה וקרובה יותר ל-0.

שלב ב:

מגדירים שיטת סיכום של ערכי "מידת הקרבה" עבור קבוצה של זוגות מלים. הסיכום נעשה באמצעות סטטיסטי (פונקציה) המודד את "מידת הקרבה הכוללת". הסטטיסטי מקבל ערך בין 0 ל-1. אם בקבוצה יש הרבה זוגות שלהם "מידת קרבה" קטנה - הסטטיסטי יקבל ערך קרוב יותר ל-0. [למעשה, הוגדרו שני סטטיסטים שיסומנו P1 ו-P2].

שלב ג:

בוחרים רשימת זוגות מלים קשורים מושגית לבדיקה ומחשבים את המידות שהוגדרו בשלבים א-ב.

שלב ד:

בודקים במבחן מובהקות האם הערך של הסטטיסטי עבור רשימת זוגות המלים קטן ממה שצפוי לקרות באקראי, ומה הסיכוי לקבל ערך כל כך קטן.

[הרחבה בנושא זה - במאמר הנ"ל, בספר צופן בראשית ובאתר צופן בראשית]

רשימת זוגות המילים

המשימה בשלב ג: לבחור **רשימת זוגות מילים** קשורים מושגית לבדיקה.

שאלה: האם לפי השערת המחקר צפויה **כל** רשימה כזו להצליח בספר בראשית?

- תשובה: לא.** גם אם קיימת מערכת צפני ELS מתוכננת בספר בראשית -
- יתכן שחלק חשוב מן המילים ברשימה אינו כלול באוצר המילים שבמערכת הצפנים.
 - יתכן שרוב המילים אכן כלולות באוצר המילים, רק הקשר המושגי במערכת הצפנים אינו אותו קשר מושגי שנבחר על ידינו.

[עוד בעניין זה - בספר **צופן בראשית**]

הרעיון לרשימת זוגות המילים

סוג זוגות המילים: החוקרים בחרו בזוגות ביטויים מן הסוג:
שם אדם (או כינוי) - - - **תאריך** הלידה/הפטירה (יום וחודש)

הרציונל:

- א. בספר בראשית כתובה כותרת "**זה ספר תולדות אדם**", ולפי מסורת חז"ל הראה הקב"ה לאדם הראשון "**דור דור ודורשיו, דור דור וחכמיו, דור דור וסופריו**" - אולי זה רמוז בצפני ELS.
- ב. סוג הקשר המושגי הוא טכני, והוא **אחיד** עבור כל זוגות הרשימה.
- ג. היתה אינדיקציה מממצא לגבי אישיות מסוימת, שיתכן שנושא מסוג זה מוצפן.

סוג האישים: החוקרים בחרו לבדוק רשימה של גדולי חכמי ישראל.

הרציונל: נראה סביר יותר שבתורה מוצפנים שמות של חכמי התורה, מאשר שמות נהגי קטרים בשבדיה או שמות דייגים במלזיה. יתר על כן, המסורת הנ"ל מדברת במפורש על דורשי התורה, חכמיה וסופריה.

[עוד בעניין זה - בספר **צופן בראשית**]

הרשימה הראשונה של זוגות מילים

הכנת רשימת זוגות המלים **חייבת להעשות מראש, באופן א-פריורי:**

א. קבוצת החכמים נקבעה לפי **קריטריון אובייקטיבי**. (כל החכמים שערכם באנציקלופדיה לגדולי ישראל אורכו משלשה טורים ומעלה, ואשר צויין בערכם תאריך לידה/פטירה).

ב. רשימת שמות החכמים וכינוייהם נקבעה **מראש, לפני הניסוי**, ע"י **מומחה חיצוני** בתחום הביבליוגרפיה, **פרופסור שלמה זלמן הבלין**, שהיה בעת הכנת הרשימה ראש המחלקה לביבליוגרפיה ולספרנות באוניברסיטת בר אילן. הכנת הרשימה נעשתה לפי כללים קבועים.

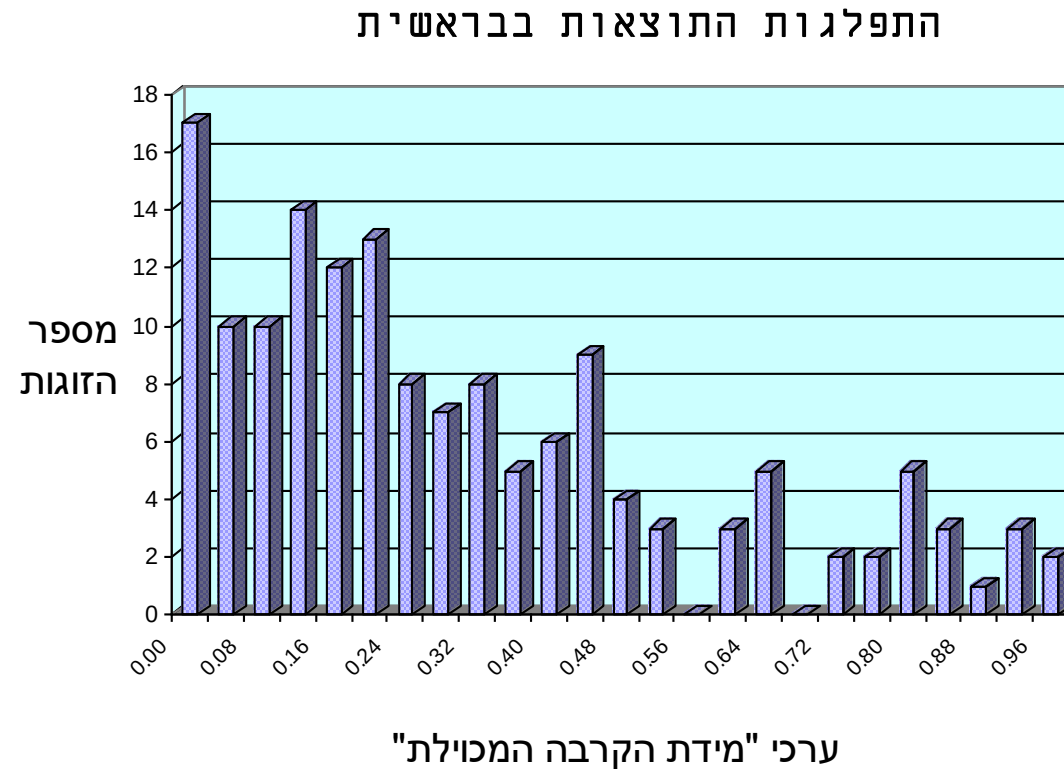
ג. כללי האיות וצורת כתיבת התאריכים נקבעו **מראש, לפני הניסוי**, ע"י **מומחה חיצוני**, הבלשן **יעקב אורבך ז"ל**.

ד. המומחים פעלו ללא שום ידיעה איזה שם או כינוי "יצליח" בניסוי, ואיזה "יכשל". הם אפילו לא ידעו פרטים חיוניים לגבי התיאור המדויק של התופעה הנחקרת, וכלל לא התמצאו בשיטת המדידה.

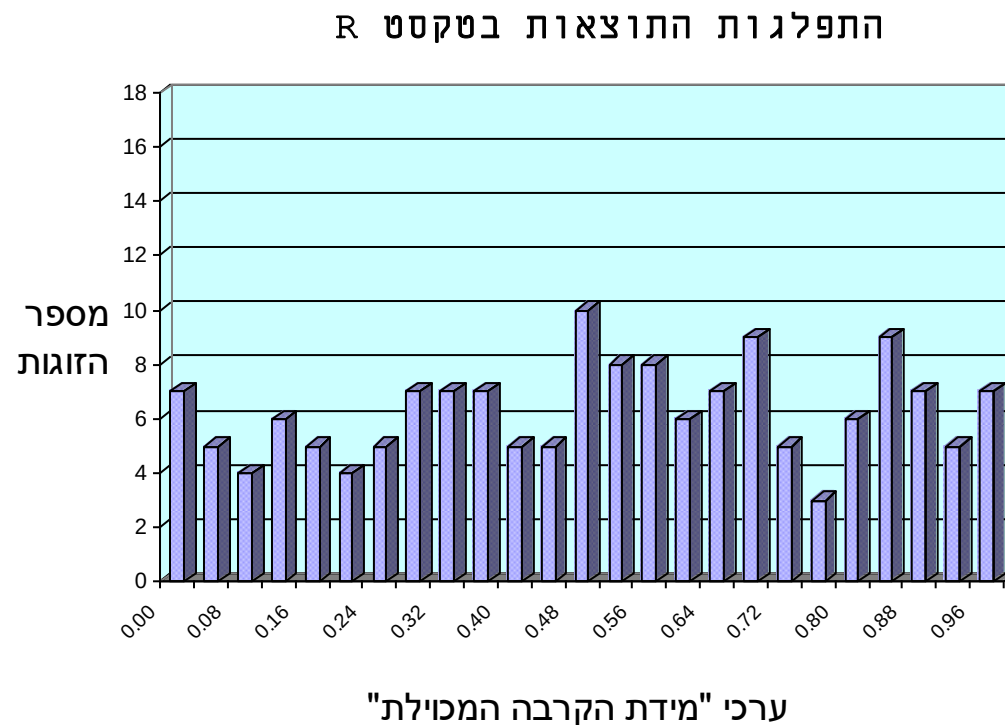
רשימת זוגות המלים היתה מורכבת מזוגות (שם, תאריך). למשל:
(הרמב"ם, י"ד ניסן) - י"ד ניסן היה יום הולדת הרמב"ם.

הרשימה הראשונה של זוגות מילים: תוצאות

יישום שלב א. הנה התפלגות 152 ערכי "מידת הקרבה המכוילת" בקטע $(0, 1]$:



לשם השוואה נציג את התוצאות עבור אותה רשימה בטקסט R שנוצר מספר בראשית על ידי ערבוב אקראי של האותיות:



ההבדל בין ההיסטוגרמות ברור. **הצטברות התוצאות בקרבת ה-0 עבור ספר בראשית**
- יוצאת דופן ביותר.

יישום שלב ב. שני הסטטיסטים $P1$ ו- $P2$, שהוגדרו מראש בשלב ב, נועדו למדוד את ההסתברות הגולמית (לא מנורמלת) של הצטברות אפשרית כזאת. שניהם קיבלו ערך של **אחד למיליארד** (בערך).

הרשימה הראשונה של זוגות מילים: תוצאות

כדי למצוא את ההסתברות המדויקת (מנורמלת) היה צורך להפעיל את **שלב ד** - מבחן המובהקות. מבחן המובהקות המתבקש הוא להשוות את ערכי הסטטיסטיים $P1$ ו- $P2$ לאלה המתקבלים באוסף גדול של טקסטים "דומים".

שלב ד לא בוצע: לפי אמצעי החישוב שעמדו לרשות החוקרים משך העבודה לכל טקסט היה שלשה שבועות לפחות. (1000 טקסטים = 58 שנים.)

הרשימה השניה של זוגות מילים

מבקר חיצוני, פרסי דיאקוניס, טען שצריך להכין רשימה **חדשה** של זוגות מילים ולהריץ עליה ניסוי **בדיוק באותו אופן** בו נעשה הניסוי הראשון.

הוכנה רשימה חדשה:

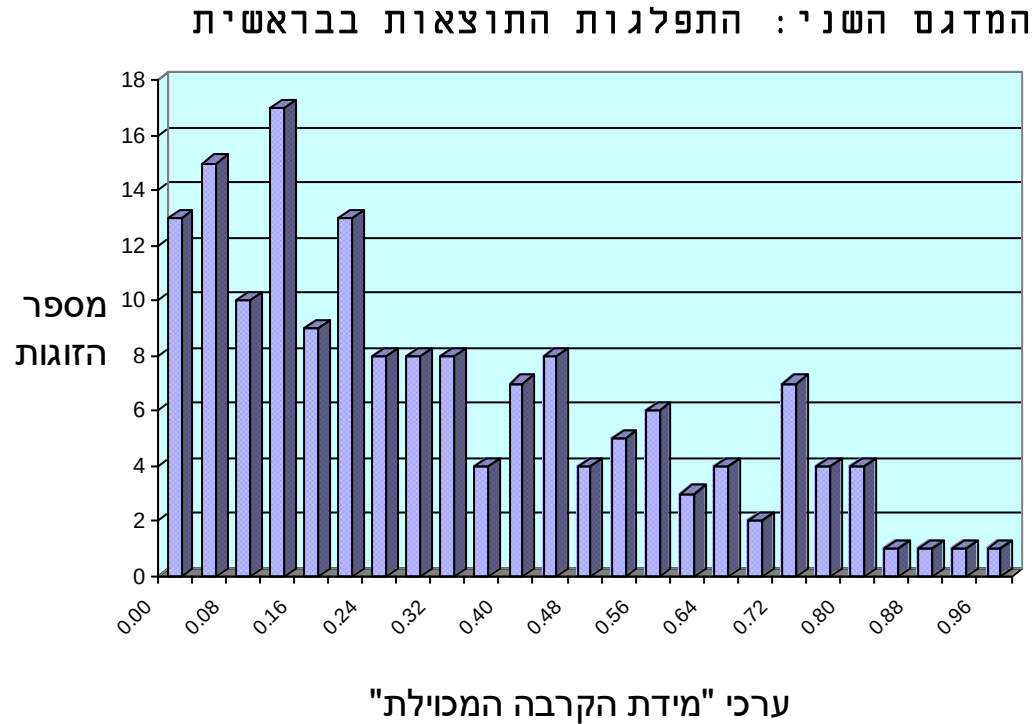
א. קבוצת החכמים החדשה כוללת את החכמים שערכם באנציקלופדיה לגדולי ישראל אורכו מטור וחצי עד שלשה טורים, ואשר מצוין בערכם תאריך לידה/פטירה.

ב. רשימת שמות החכמים וכינוייהם נקבעה **מראש, לפני הניסוי**, ע"י **פרופסור שלמה זלמן הבלין**, **לפי הכללים שנקבעו כבר**.

ג. כללי האיות וצורת כתיבת התאריכים - **כמו ברשימה הראשונה**.

ד. גם הפעם, המומחה פעל ללא שום ידיעה איזה שם או כינוי "יצליח" בניסוי, ואיזה "יכשל", לא ידע פרטים חיוניים לגבי התיאור המדויק של התופעה הנחקרת, וכלל לא התמצא בשיטת המדידה.

הרשימה השניה של זוגות מילים: תוצאות



הניסוי בוצע בדיוק באותו אופן
כמו הניסוי הראשון.

יישום שלב א. הנה התפלגות
163 ערכי "מידת הקרבה
המכוילת" בקטע $(0, 1]$:

יישום שלב ב. הפעם קיבל הסטטיסטי P_2 , ערך של אחד למיליארד (בערך).
הסטטיסטי השני P_1 קיבל ערך גדול ממנו, בערך פי 10.

גם כאן, שלב ד לא בוצע, בהעדר אמצעי החישוב מספיקים.

מבחן הפרמוטציות

א. פרסי דיאקוניס הציע לבצע את **שלב ד**, **מבחן המובהקות**, באמצעות **מבחן פרמוטציות**.

ב. כדי ליישם **מבחן פרמוטציות** היה צורך להבחין ברשימה:

- בין כינויים בעלי אופי כללי **המשותפים לכמה אישים** מסוג "רבי משה" או "רבי דוד" (כולם נמצאים בעמודה מיוחדת בטבלת השמות והכינויים שהכין המומחה). נסמן רשימת כינויים אלה ב- **L'2**,

- לבין שאר השמות והכינויים, שהם ספציפיים, כמו "הרמב"ם" (נסמן רשימת כינויים אלה ב- **L2**).

$$\text{List2} = \text{הרשימה הכוללת} = \text{L'2} + \text{L2}$$

ג. פרטי **מבחן הפרמוטציות** סוכמו בכתב בין ישראל אומן לפרסי דיאקוניס. המבחן נועד למדוד את המובהקות של ארבעה סטטיסטים: $P1$ ו- $P2$ לגבי הרשימה כולה, ו- $P1$ ו- $P2$ לגבי **L2**.

ד. פרטי הניסוי המתוכנן **נכתבו מראש**, ונשלחו לפרסי דיאקוניס ועוד ארבעה סטטיסטיקאים ידועי שם וכולם **אישרו את פרטי הניסוי** וקבעו את הרף להצלחה.

הצלחה במבחן הפרמוטציות

שלב ד בוצע במבחן של 1,000,000 פרמוטציות והתוצאה החזקה ביותר היתה עבור הסטטיסטי P2 המיושם לגבי L2: 4 למיליון. [מבקרי המחקר מדדו אחר כך, במבחן של 100,000,000 פרמוטציות, מובהקות של פחות מ-1 למיליון].

[הרחבה בנושא זה - במאמר הנ"ל, בספר צופן בראשית ובאתר צופן בראשית]

הקרב על הצופן

מבקרי המחקר טענו מראש, בטרם בדקו משהו – כי הכל רמאות. במאמר "הפרכה", שפרסמו חמש שנים אחרי כן בסטיסטיקאל סאיינס, הם טוענים כי הוכיחו את המסקנה שנקבעה מראש... "ההפרכה" מבוססת על טענת קונספירציה. הם טוענים כי הכל רמאות וכי

המומחים הבלין ואורבך שיתפו פעולה עם החוקרים ויצטום וריפס כדי לרמות את כל העולם.

- הם טוענים: הוכחנו שאפשר לרמות, על ידי "תפירה" מוצהרת של רשימת שמות וכינויים עבור אישי הרשימה השניה, "כמו" L2, שמצליחה לא פחות ממנה, אבל בספר "מלחמה ושלום".
- אומנם הם מודים, שהשיגו זאת הודות לחריגה מן הכללים הביבליוגרפיים והלשוניים ששימשו ליצירת "הרשימה הראשונה" והיוו אילוצים על יצירת L2.
 - אבל הם טוענים, שגם המשתתפים בקונספירציה חרגו מן האילוצים הללו "באותה מידה" ביצירת L2.

בקצרה, הם טוענים: עשינו "אותו דבר" בספר "מלחמה ושלום".

הקרב על הצופן - הפרכת "ההפרכה"

טענת המבקרים כי עשו "אותו דבר" - מבוססת על אריתמטיקה מוזרה ושקרית של טענות ביבליוגרפיות ולשוניות שהם המציאו לצורך העניין. שום מומחה בתחום הביבליוגרפי והלשוני אינו חתום עליה. (המבקרים עצמם בורים בתחום זה - גם לפי דבריהם.)

הפרכנו באופן פרטני את הטענה, כי המבקרים עשו "אותו דבר". הראינו, כי כל מה שהם עשו אינו אלא בחירה שרירותית של כינויים ללא אילוצים. - ברור שבדרך זו אפשר לייצר "הצלחה" בכל ספר, מכתב או עיתון...

אבל, למרות שהפרכנו את טענתם המרכזית, נוצרה בעיה הסברתית: הוויכוח הוא בתחום הביבליוגרפי והלשוני ולא בתחום הסטטיסטי. קהל היעד של המחקר, שהוא בתחום המתמטי-סטטיסטי, בדרך כלל רחוק מנושאים אלה, ובדרך כלל אף אינו דובר עברית.

לכך נועד מאמרנו "על מדע ועל פרודיה: הפרכה גמורה של הטענה המרכזית של המבקרים". מאמר זה מפריך את טענת המבקרים כי עשו "אותו דבר", וזאת מן הבחינה הלוגית והעובדתית, מבלי להיכנס לדיון ביבליוגרפי או לשוני.

הקרב על הצופן - הפרכת "ההפרכה"

כבר בשלב מוקדם של הוויכוח, הצענו **מבחן מדעי, פשוט ונכון** לבדיקת טענת המבקרים: **הצענו שמומחה חיצוני אובייקטיבי יפעל באופן מוסכם על שני הצדדים ויכין את L2 מחדש על פי האילוצים.** **L2 החדשה** תיבדק בספר בראשית, ואם היא תצליח, יוכח שלא חופש תימרון ולא שבירת הכללים הם שיצרו את ההצלחה.

המבקרים סרבו לעמוד באתגר. חזרנו והצענו זאת בתחילת תש"ס, כאשר אנו מלווים הצעה זו בהימור של **מיליון דולר** על כך שהרשימה שיכין המומחה החיצוני לפי האילוצים תצליח בספר **בראשית יותר** מאשר ב"**מלחמה ושלום**". [ההצעה פורסמה כותרת במאמר במוסף שבת (וסוכות) של ידיעות אחרונות]. תגובת **מיה בר-הלל** הובאה באותו מאמר: "למה שמישהו יעשה את הניסוי **הטיפשי** הזה?" (ההדגשה שלי).

הסיבה לתגובה מוזרה זו, התבררה רק אחר כך. כעבור זמן קצר פורסמה עבודת המבקרים בסטטיסטיקל סאינס. שם הם דיווחו כי ד"ר **שמחה עמנואל**, מומחה בהיסטוריה רבנית באוניברסיטת תל-אביב, נשכר על ידם והכין לפי הזמנתם כמה רשימות של שמות וכינויים. **העבודה נעשתה בהדרכתם וללא ידיעת החוקרים.** אחת מן הרשימות שהכין נועדה על ידם **"לחקות" את L2.** לטענת המבקרים, הרשימה החדשה שהכין **עמנואל נכשלת במבחן הפרמוטציות**, ומפריכה בכך את טענת החוקרים שקיים צופן בספר בראשית. המבקרים כתבו: "[זה] **אולי הניסוי החשוב ביותר שעשינו**".

הקרב על הצופן - הפרכת "ההפרכה"

- אבל, במאמרנו "הוכחה סטטיסטית לקיום צופן בספר בראשית", המבוסס -
- על בדיקת הרשימות שפירסמו המבקרים (באיחור של שנה),
 - על מידע שמסר ד"ר עמנואל במהלך שיחות שהיו לנו עמו,
 - ועל מה שכתבו המבקרים במקומות שונים בשם המומחה שלהם,
- אנו מוכיחים:

- א. כי המבקרים ביימו "ניסוי בלתי תלוי" בשיטות הנראות כתרמית גמורה. מטרתם היתה להציג תמונה לפיה הניסוי שלהם מוליך לכשלון, ולהעלים את העובדה החשובה -
- ב. כי מן הנתונים של ד"ר עמנואל נובעת מסקנה מוכרחת: שהצלחת החוקרים לא נבעה מ"תפירה" של שמות וכינויים (כטענת המבקרים), אלא מקיומו של צופן בספר בראשית.

עכשיו ברור לנו למה הם סרבו לניסוי המדעי הפשוט שהצענו:

אם ניקח את השמות והכינויים שהכין ד"ר עמנואל, המומחה שהם העסיקו באופן חד-צדדי ובהדרכתם, נקבל עבור L2 החדשה הצלחה בספר בראשית וכישלון ב"מלחמה ושלום".

[הרחבה בנושא זה, וכן הפרכת טענות משניות של המבקרים - באתר צופן בראשית במדור "ההתנגדות למחקר".]

הקרב על הצופן - התפתחויות חדשות

מבחן טקסטים

כחמש וחצי שנים לאחר ביצוע **מבחן הפרמוטציות** בעבודתנו המקורית, סבר אליהו ריפס שהתקדמותם המהירה של אמצעי החישוב מאפשרת לשוב **לרעיון הפשוט ביותר של מדידת המובהקות**: מדידה באמצעות **השוואה לטקסטים רבים "דומים"**.

במכתב לדוד קשדן הציע אליהו ריפס את העקרונות הבאים לגבי מדידת מובהקות של רשימה בספר בראשית:

- להשתמש ב-1000 (או יותר) טקסטים T_j , כל T_j נוצר ע"י **ערבוב אקראי של מלים** בתוך פסוקים (הדבר נעשה בדיוק כפי שנעשה במאמרנו לגבי טקסט U).
- להעריך את מובהקות הסטטיסטים P_i בספר בראשית, באמצעות השוואתם לערכי P'_i המתקבלים בטקסטים T_j .

מבחן מסוג זה יכולה להלן "מבחן טקסטים".

הקרב על הצופן - התפתחויות חדשות

מבחן טקסטים

השוואה בין מבחן הפרמוטציות למבחן הטקסטים:

כל אחד משני המבחנים מודד דבר **שונה**:

- **מבחן הפרמוטציות** מודד עד כמה "מוצלחים" המפגשים של שמות האישים עם תאריכי הלידה ו/או הפטירה שלהם **יותר** מן המפגשים של אותם שמות אישים עם תאריכים אחרים (שהם תאריכי לידה/פטירה של אישים אחרים מאותה רשימת נתונים).
- **מבחן הטקסטים** מודד מה הסיכוי לקבל במקרה P_i כה קטנים. דהיינו: **מה הסיכוי שהצטברו במקרה כל כך הרבה מפגשים "מוצלחים"**.

במלים אחרות: **מבחן הטקסטים** מודד את **מובהקות הכמות המצטברת** (אקומולציה) של מפגשים "מוצלחים", ואילו **מבחן הפרמוטציות** מודד את המובהקות של "ספציפיות הצופן".

לכן אנו מצפים **אפריורי** להבדל בתוצאות שני המבחנים **עבור אותה רשימת נתונים**, שהרי הם מודדים **דברים שונים**.

הקרב על הצופן - התפתחויות חדשות

מבחן טקסטים

מהו המבחן המתאים למצבו העכשווי של מחקר הצופן בתורה?
המחקר המדעי של הצפן בתורה, עד כמה שבא לידי ביטוי במדידת רשימות גדולות של נתונים, עסק עד כה **בהוכחת קיומו** של צופן מסוים. **הוא לא עסק באיפיון מדוייק או בפיענוח של הצפנים**. הדבר **הוצהר במפורש** בהקדמה למאמרנו.

לכן, לעבודות העוסקות במדידת רשימות גדולות של נתונים מתאים דווקא **מבחן הטקסטים**, ולא **מבחן פרמוטציות** המודד את המובהקות של "ספציפיות הצופן".

הקרב על הצופן - התפתחויות חדשות

מבחן טקסטים

1. מדידת רשימת הנתונים של בארי סיימון:

מבחן הטקסטים	מבחן הפרמוטציות
מובהקות חזקה ($p < 0.001$)	אין מובהקות

2. מדידת "רשימה F" של הוועדה האוניברסיטאית בראשות ישראל אומן:

מבחן הטקסטים	מבחן הפרמוטציות
מובהקות חזקה ($p < 0.001$)	אין מובהקות

3. מדידת רשימת הנתונים של גנז - ענבל - בומבר:

מבחן הטקסטים	מבחן הפרמוטציות
מובהקות חזקה יותר	מובהקות חזקה

בשנת התשע"ד התחלתי להשתמש במבחן הטקסטים, כדי להעריך את המובהקות של רשימות נתונים גדולות. הדבר נעשה על ידי השוואה לאוסף של 50,000 טקסטים "דומים" (= עבודה של חודשים אחדים של מחשב ביתי חדיש לכל רשימת נתונים). הנה סיכום שלוש עבודות חשובות (מדגמי "הקהילות"):

[כל הפרטים במאמר: מדידת מובהקות במחקר הצפנים בתורה: תוצאות מדידות במבחן הטקסטים - באתר צופן בראשית.]

הקרב על הצופן - התפתחויות חדשות

סיגנאל של מערכת צפני ELS: דעיכה מול הגברה

מחקר חדש

בסייעתא דשמיא התברר, כי צירוף של רעיון פשוט ורמת מהירות המחשבים בימינו, מאפשר לתת תשובה ברורה לשאלה מרכזית:

האם ניתן להבחין בין **הצלחה אמיתית** של מערכת צפני ELS לבין **הצלחה מלאכותית** הנוצרת כתוצאה מ"**תפירה**" של נתונים?

הגדרה - סיגנאל:

יהא T טקסט, D קטע ממנו, $LIST$ רשימה של זוגות ביטויים. נגדיר **סיגנאל** S_i באופן הבא:

$$S_i(LIST, D) \equiv 1 / P_i(LIST, D)$$

כאשר P_1 ו- P_2 הם הסטטיסטים שהוגדרו בעבודת החוקרים והוזכרו בראשית ההרצאה. כזכור, P_1 וגם P_2 מקבלים **ערך הקטן** ככל שהמפגשים בין צפני ה-ELS יותר "מוצלחים". לכן, הסיגנאל יקבל **ערך הגדל** ככל שהמפגשים בין צפני ה-ELS יותר "מוצלחים". (הגדרת הסיגנאלים S_i אינה מחדשת דבר מהותי לעומת ההגדרה המקורית של P_i , אבל עשויה לסייע לתפישה אינטואיטיבית של הנדון כאן.)

הקרב על הצופן - התפתחויות חדשות

סיגנאל של מערכת צפני ELS: דעיכה מול הגברה

"תפירה" של נתונים:

"תפירה" של נתונים מאפשרת לקבל סיגנאל גבוה באופן מלאכותי בקטע D מטקסט T , למרות שבקטע D הופעת צפני ELS **אקראית לחלוטין**. הדבר נעשה באמצעות **התאמה של הנתונים** למערכת הצפנים. [למשל, על ידי בדיקת זוגות רבים מאד, בחירת "המצליחים" והסתרת "הנכשלים".]

רעיון הניסוי החדש:

נניח שהסיגנאל של רשימה $LIST$ בקטע D נתקבל **באופן מלאכותי**, על ידי "תפירה" של נתוני $LIST$. **שאלה:** מה צפוי לקרות לסיגנאל של רשימה $LIST$ אם נמדוד אותו בקטע D' בטקסט T , שהוא **ארוך מ- D** והמכיל אותו?

תשובה: במעבר לקטע D' יש שינויים רבים, הנגרמים על ידי צפני ELS **חדשים המופיעים באקראי**, ואשר הנתונים **לא הותאמו אליהם**. כך נוצר "**רעש**" אקראי **המקטין את הסיגנאל**.

בקצרה: במדידה החדשה בקטע המורחב D' , אנו מצפים **לדעיכת הסיגנאל**:

$$S_i(LIST, D') < S_i(LIST, D)$$

הקרב על הצופן - התפתחויות חדשות

סיגנאל של מערכת צפני ELS: דעיכה מול הגברה

נגדיר את מדד 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 .

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 **במדד ההגברה**
המרבית Q :

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

דוגמא:

BM2, רשימה הנתונים שנמדדה בעבודת המבקרים בספר
"מלחמה ושלום", הוכנה באופן מוצהר על ידי "תפירת" נתונים.
תחום המדידה המקורי, D , היה הקטע $[1, 78064]$ מתחילת
הספר הנ"ל. תחום המדידה החדש, D' , הכפול באורכו, הינו
הקטע $[1, 156128]$ מתחילת "מלחמה ושלום".

מתברר, כי **בתחום המוגדל** הסיגנאל אכן **דועך באופן בולט**:

$$Q_1 = 1.4577E-4, \quad Q_2 = 9.8629E-4,$$

ולכן **מדד ההגברה המרבית בתחום המוגדל** הוא:

$$Q = \max\{Q_1, Q_2\} = 9.8629E-4.$$

הקרב על הצופן - התפתחויות חדשות

סיגנאל של מערכת צפני ELS: דעיכה מול הגברה

שאלה: מה יקרה אם קיימת הצפנה אמיתית של $LIST$ ב- T ?

תשובה: ישנן שתי אפשרויות:

1. אם תחום ההצפנה של $LIST$ ב- T מוכל בקטע D , אזי הרחבת הקטע מוסיפה רק "רעש" אקראי, ולכן הסיגנאל ייחלש וידעך.
2. אם תחום ההצפנה של $LIST$ ב- T גדול מן הקטע D , אזי הרחבת הקטע לא תחליש את הסיגנאל, וייתכן שאף תגביר אותו.

ספר בראשית כקטע מספר התורה

לפי נסיון שהצטבר מאז פורסמה עבודת החוקרים נראה הדבר, כי התופעה לה טענו החוקרים אינה מוגבלת לספר בראשית, אלא מצוייה בתורה כולה. אנו רואים את כל התורה כחטיבה אחת לעניין התופעה שלפנינו, כאשר נושאים שונים מוצפנים באזורים שונים בספר. בראיה זו, ספר בראשית בו נמדדה רשימה $L2$, אינו אלא הקטע $[1, 78064]$ מספר התורה.

הסיגנאל של $L2$

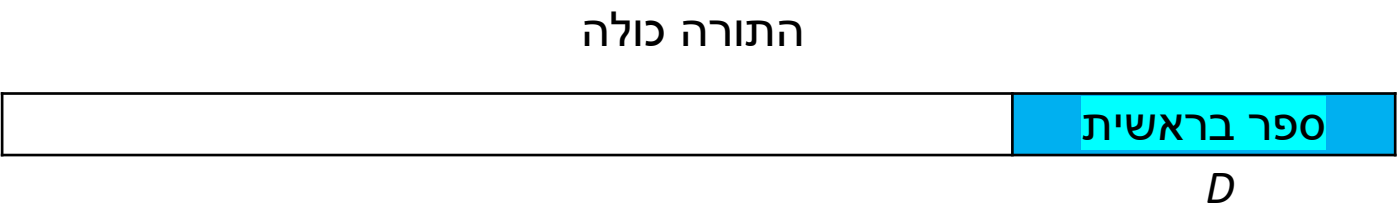
נבדוק את $L2$ בקטע גדול מתוך ספר התורה, המכיל את ספר בראשית. לפי טענת המבקרים כי $L2$ אינה אלא רשימת נתונים "תפורה", אנו מצפים אך ורק לדעיכת הסיגנאל המקורי.

סיגנאל של מערכת צפני ELS: דעיכה מול הגברה

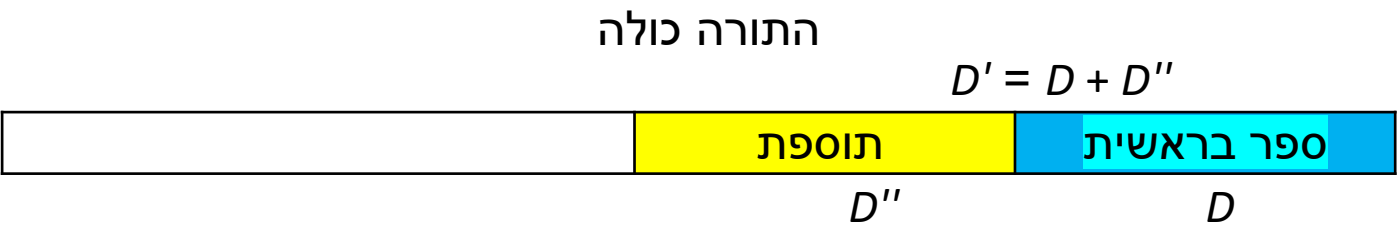
מדידה 1: בדיקת L2 בקטע גדול מספר התורה.

הקטע המקורי, D , היה ספר **בראשית**. אנו צריכים להרחיב את D בקטע "גדול דיו" כדי לצפות לדעיכת הסיגנאל. כ"כלל אצבע" בחרנו ב- D' **הכפול באורכו** מ- D . כלומר, הוספנו קטע D'' באורך ספר בראשית.

ספר בראשית כקטע מן התורה:



הקטע המורחב D' כקטע מן התורה:



א. תוצאת המדידה:

$Q = 2,307$

מדידת L2 בקטע המורחב D' מעלה כי **מדד ההגברה המרבית** עבור הקטע המורחב הוא:

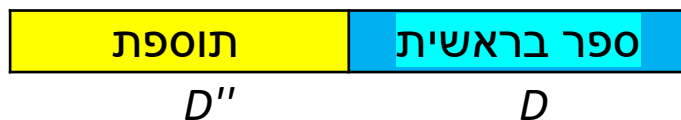
סיגנאל של מערכת צפני ELS: דעיכה מול הגברה

ב. מובהקות התוצאה:

1. צריך להעריך מהי מובהקות התוצאה, דהיינו: "מה הסיכוי שיתקבל **במקרה** Q **כה גדול?**"
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 , הוא ספר בראשית בדיוק,
- ומתוספת – הקטע D'' מן התורה.

$$D' = D + D''$$



$$T_j = D + D''_j$$



- כל טקסט T_j אף הוא בנוי משני חלקים:
- מחלקו הראשון, D , הוא ספר בראשית בדיוק,
 - ומתוספת – הקטע D''_j - שהוא **ערבוב אקראי** של הקטע D'' מן התורה (ערבוב אקראי של מלים בתוך פסוקים. הדבר נעשה **בדיוק** כפי שנעשה במאמרנו לגבי טקסט U).

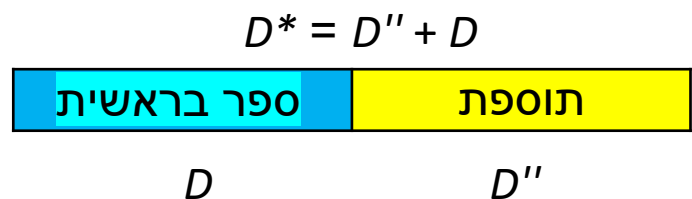
סיגנאל של מערכת צפני ELS: דעיכה מול הגברה

2. כאשר מודדים את L2 בטקסט T_j , 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-2,307. הסיכוי לקבל כזו תוצאה במקרה הוא $p = n/N$.

3. המובהקות שנתקבלה בהרצת 50,000 טקסטים T_j : $p < 0.001$

מדידה 2: בדיקות ביקורת.

נעשתה סדרת בדיקות ביקורת:



א. היפוך סדר הקטעים:

יצרנו טקסט נוסף, D^* , על ידי הפיכת הסדר של הקטעים D ו- D'' :

- החלק הראשון הוא הקטע D'' מן התורה.

- החלק הנוסף הוא ספר בראשית.

התוצאה בקטע D^* מראה על דעיכה: $Q = 8.937E-4$

סיגנאל של מערכת צפני ELS: דעיכה מול הגברה

ב. דעיכת הנפילים:

במבחן הפרמוטציות בעבודת החוקרים, היו 3 פרמוטציות מתוך 1,000,000 אשר "נצחו" את הרשימה המקורית: הסיגנאל עבורן היה חזק יותר מן הסיגנאל המקורי. מעניין לבדוק מה יקרה לסיגנאלים אלה כאשר נרחיב את תחום המדידה מספר בראשית לקטע המורחב D' .

1. הפרמוטציה שהיתה ראשונה מתוך מיליון פרמוטציות (PER1).

$$Q = 9.3927E-7$$

מתברר שהסיגנאל נחלש מאד:

תוצאה זו מפתיעה. לא זו בלבד שהביטויים ב-PER1 הם השמות, הכינויים והתאריכים הנמצאים ב L2, אלא אף שני אישים "שהצליחו" בעבודה המקורית במפגשים עם תאריכיהם, נמצאים עם תאריכיהם, ללא כל שינוי, גם ב-PER1.

2. הפרמוטציה שהיתה שניה מתוך מיליון פרמוטציות (PER2).

$$Q = 1.7945E-5$$

גם כאן מתברר שהסיגנאל נחלש מאד:

מה שנאמר לעיל לגבי PER1 רלוונטי גם כאן. במקרה זה מדובר על ארבעה אישים "שהצליחו" בעבודה המקורית במפגשים עם תאריכיהם, הנמצאים עם תאריכיהם, ללא כל שינוי, גם ב-PER2.

[כל הפרטים במאמר: סיגנאל של מערכת צפני ELS: דעיכה מול הגברה - באתר צופן בראשית.]

סיגנאל של מערכת צפני ELS: דעיכה מול הגברה

3. הפרמוטציה שהיתה **שלישית** מתוך מיליון פרמוטציות (PER3).

$$Q = 1.95E-3$$

מתברר שהסיגנאל **נחלש**:

מה שנאמר לעיל לגבי PER1 ו-PER2 רלוונטי גם כאן. במקרה זה מדובר על **שלשה** אישים "שהצליחו" בעבודה המקורית במפגשים עם תאריכיהם, **הנמצאים עם תאריכיהם**, ללא כל שינוי, גם ב-PER3.

ג. בדיקת רשימת המבקרים ב"מלחמה ושלום":

כבר ראינו לעיל, כי הסיגנאל עבור הרשימה "התפורה" **BM2**, **דער** כאשר עברנו לתחום המוגדל:

$$Q = 9.8629E-4.$$

סיגנאל של מערכת צפני ELS: דעיכה מול הגברה

מדידה 3: בדיקת L2M בקטע גדול מספר התורה.

במאמר העלו המבקרים טענות אחדות בנוגע לתאריכים ולצורות התאריך שננקטו ברשימות הנתונים של החוקרים. מטענות אלה רלוונטית רק טענה אחת **ויחידה** לגבי **L2**: כי **עבור 4 אישים יש להוסיף או לשנות תאריך**. נגדיר את הרשימה **L2M** כתת הרשימה **L2** עם השינויים הנ"ל בתאריכים (כל שאר התאריכים והפרטים, בפרט השמות והכינויים - נשארים כפי שהם, ללא שינוי).

א. המדידה: נחזור על **מדידה 1** כאשר אנו מודדים את **L2M** במקום **L2**.
התוצאה: הסיגנאל **התחזק מאד**

$$Q = 2.114E+5.$$

ב. מובהקות התוצאה:
המובהקות שנתקבלה בהרצת 50,000 טקסטים T_j :

$$p = 0.00026.$$

הקרב על הצופן - התפתחויות חדשות

סיגנאל של מערכת צפני ELS: דעיכה מול הגברה

מסקנות

א. תוצאת הניסוי בתורה מורה, כי הסיגנאל מתחזק בקטע המורחב D' . עובדה זו **מפריכה את טענת המבקרים** כי הסיגנאל שנתקבל בקטע D המקורי היה מלאכותי ונוצר ממניפולציה בנתוני L2.

ב. תוצאה א של מדידה 2 מראה, כי התגברות הסיגנאל אינה תולדה של סכום ההצפנות בקטעים D בפני עצמו ו- D'' בפני עצמו. סדר החיבור הנכון בין הקטעים הוא **קריטי** להתגברות הסיגנאל.

ג. נראה, שתחום ההצפנה של L2 קרוב יותר לתחום המורחב, D' , מאשר לתחום המקורי, D (ספר בראשית). לכן, יש עניין רב לערוך ניסויים נוספים בנושא L2 בתחום המורחב. [יש להדגיש, כי אין אנו טוענים כי בתחום D' מוצפנת L2 באופן אופטימאלי. דבר זה כלל לא נבדק.]

[כל הפרטים במאמר: סיגנאל של מערכת צפני ELS: דעיכה מול הגברה - באתר **צופן בראשית**.]

נשוב לטבלה שראינו בתחילה:

לפנינו קטע מספר בראשית, כתוב כסדרו, אות אחר אות ללא רווחים ביניהן, בצורת טבלה ובה 49 טורים:

ישכחהווייהימקצשנתימימימופרעהחלמוהנהעמדעלהיארונהמנ
היארעלתשבועפרותיפותמראהובריאתבשרותתיעינהבאחונהנהשבע
פרותאחרותעלותאחריהנמנהיארעיותמראהודקותבשרותעמדנהא
צלהפרותעלשפתהיארותאכלנההפרותתרעותהמראהודקתהבשראתשב
עהפרותיפתהמראהוהברייאתוייקצפרעהויישנויחלמשניתוהנהש
בעשבלימעלותבתבנהאחדברייאותוטבותוהנהשבעשבלימדקותושדו
פתקדימצמחותאחריהנותלענההשבלימהדקותאתשבועהשבלימהבר
יאותוהמלאותוייקצפרעהוהנהחלומויהיבבקרותפעמרוחווישל
חוויקראאתכלחרטמימצימואתכלחכמיהויספרפרעהלהמאתחלמו
אינפותראותמלפרעהוידברשרהמשקמאתפרעהלאמראתחטאינימ
זכירהיומפרעהקצפעלעבדיווייתנאתיבמשמרבתשרהטבחמאתיו
אתשרהאפימונחלמהחלומבלילהאחדאניוהואאישכפתרונחלמוחל
כל חכמי, הסכוי, יראו בברור, צפן חבוי, בתורתי

בטבלה זו ראינו מפגש של זוג צפני ELS קשורים מושגית:

צפן חבוי - בתורתי

עתה אנו מוסיפים לראות:

כל חכמי, הסכוי, יראו בברור, צפן חבוי, בתורתי.

חכמי יחיד בטקסט, כל הצפנים הם המינימליים בכל ספר בראשית.