

הטיה מגדרית בוויקיפדיה העברית : הערכה כמותית

מרב מרקוביץ וגילה פריבור

1. מבוא

ויקיפדיה (Wikipedia), האנציקלופדיה המקוונת אשר נוסדה בתחילת המאה הנוכחית, הפכה תוך שנים ספורות למאגר מידע שיתופי וחופשי עצום, המשמש כמקור מידע ראשון שאליו פונות אוכלוסיות מגוונות. ערכי ויקיפדיה קיימים בלמעלה מ-200 שפות, ביניהן עברית, והאתר ניצב בעקביות ברשימת עשרת האתרים הנצפים ביותר באינטרנט בעולם ובישראל. ויקיפדיה בשפה העברית הוקמה בשנת 2005, ומאז צמחה בקצב מהיר ביותר. בשנת 2013 מנתה למעלה מ-140,000 ערכים (גבע, 2013), ונכון לספטמבר 2023 קיימים כ-341,662 ערכים בוויקיפדיה העברית ("ויקיפדיה העברית", 2023). קצב הגידול של ויקיפדיה העברית, כמו גם הנתונים על שיעור הצפייה בדפי הערכים השונים שלה, מלמדים על חשיבותה כמקור מידע מקוון (אלון ובר אילן, 2012; לין, 2015; Wadewitz, 2013; Top Websites Ranking, 2023).

החל משנת 2011 החלו נחקרות טענות על הטיה מגדרית בוויקיפדיה, הטיה שיש בה כדי לפגוע הן באיכות המידע, הן באמינותו והן ביומרה של ויקיפדיה לספק מידע חסר פניות (Neutral Point Of View) (Lir, 2019). כפי שיפורט להלן, היבטים שונים של פער מגדרי בוויקיפדיה בשפות שונות נבדקו ונתחו במחקרים קודמים רבים. עם זאת, בפלטפורמה של ויקיפדיה בשפה העברית טרם בוצעו הערכה וניתוח כמותי של הטיות מגדריות בתוכן. מטרתה של עבודה זו היא בחינה כמותית של השאלה: האם תוכני אתר ויקיפדיה בשפה העברית (להלן: ויקיפדיה העברית) מוטים מבחינה מגדרית? בליבה של עבודה זו הצגת מתודה לבחינה כמותית של הטענה כי תוכני ויקיפדיה מוטים מבחינה מגדרית.

2. מצב המחקר

ויקיפדיה נוסדה בתחילת שנות האלפיים בארצות הברית ומבוססת על פלטפורמת ה-Wiki, ששירתה את קהילת הקוד הפתוח. אתרי ויקיפדיה הם הפרויקט הבולט של "קרן ויקימדיה" העולמית (Wikimedia Foundation), ארגון אמריקני ללא מטרת רווח שמטרתו המוצהרת היא הנגשת והפצת ידע לכול (Wikimedia Foundation, n.d.). לשם כך נוצר גוף מידע עצום בהיקפו בשאיפה מוצהרת להוות אנציקלופדיה המתעדת את תחומי הדעת האנושיים. בתחומים מסוימים ויקיפדיה מתוארת כמאגר המידע הגדול ביותר בתולדות האנושות, ויש המתארים אותה כ"אנציקלופדיה אחרונה", שיתרונותיה מאיימים ליתר את כלל המיזמים האנציקלופדיים שקדמו לה (לין, 2015).

מעבר לנתוני השימוש והפופולריות של ויקיפדיה, מחקרים הראו כי מבחינת אמינות המידע היא אינה נופלת ממקורות מידע אנציקלופדיים מסורתיים, כגון "אנציקלופדיה בריטניקה" (Giles, 2005). זאת למרות שיטת העריכה ה"דמוקרטית" של ויקיפדיה, המאפשרת לכל משתמש לערוך ולהוסיף תוכן לרשומותיה.

2.1 הטיה מגדרית בוויקיפדיה – מצב המחקר

מראשית צמיחתה של ויקיפדיה כמאגר מידע חשוב ובולט החלו עולות טענות על הטיות שונות במידע המופיע בדפי ויקיפדיה השונים, הטיות הפוגעות הן באיכות המידע, הן באמינותו והן בטענה

של ויקיפדיה לספק מידע חסר פניות. בין שאר החלו מצטברות טענות כי הפלטפורמה מוטה (Biased) מבחינה מגדרית (Graells-Garrido et al., 2015).

אחת הטענות היא שהכותבים בוויקיפדיה (Editors) הם רובם גברים, ולכל הפחות שגברים נהנים מייצוג יתר בקרב עורכי ויקיפדיה. טענה עובדתית זו בוססה במספר מחקרים, הן בזמנים שונים הן בוויקיפדיות בשפות שונות: בוויקיפדיה בשפה השוודית נמצא כי קיימת הטיה מגדרית בשיעור ההשתתפות של נשים בעריכת תוכני ויקיפדיה (Helgeson, 2015), בוויקיפדיה הפרסית נחקרו תמריצים שיש לעורכים מקבוצות שונות לתרום ולערוך תוכן בוויקיפדיה (Asadi et al., 2013), בוויקיפדיה בשפה היוונית נמצאו הבדלים מגדריים הן בשיעור השימוש מצד הצופים בתכנים והן בשיעור התרומה של עורכי תכנים לוויקיפדיה (Protonotarios et al., 2016), בוויקיפדיה בשפה הגרמנית נחקרו ההשלכות האפשריות של שיעור ההשתתפות הנמוך של נשים ביצירת ערכי ויקיפדיה (Sichler & Prommer, 2014) ובוויקיפדיה הספרדית נמצא שבכל התחומים מספר הגברים עולה על הנשים בכל מרחבי השם של ויקיפדיה (מרחבי השם מחלקים את דפי אתר הוויקי לקבוצות, כך שניתן לבצע פעולות ולהחיל הגדרות על דפים המשתייכים לקבוצה מסוימת, ולא על אחרות) ביחס שתמיד שווה ל-1:3 או גדול מ-1:3 (Minguillón et al., 2021). אמנם מחקרים שונים העלו נתונים מספריים שונים, אך המשותף לכולם הוא הממצא ששיעור הנשים בקרב עורכי ויקיפדיה אינו עולה על 20%, שיעור הנופל בהרבה משיעורן של נשים באוכלוסייה (Konieczny & Klein, 2018). קרן ויקימדיה מפעילה פרויקט שמטרתו עיבוי הערכים והמטא-דאטה על רשומות ביוגרפיות של נשים (https://www.wikidata.org/wiki/Wikidata:WikiProject_Women). מטרת הפרוייקט כפי שמנוסחת בראש עמוד אתר הפרוייקט היא: "to get every item about a woman described properly on Wikidata."1. כמו כן טענת משנה היא שבקרב קהילות העורכים שבהן יש רוב של גברים התקבלו והשתרשו נורמות חברתיות המאפיינות סביבה "גברית": תחרות, היררכיה, ביקורת אגרסיבית וכד'. נורמות אלה כשלעצמן מקשות או מונעות השתלבות של עורכות נשים בקרב עורכי ויקיפדיה, ובכך נשמר שיעורן הנמוך של הנשים בקרב העורכים (Lir, 2019).

טענה נוספת היא שהטיה מגדרית מתבטאת בתוך תוכני הרשומות (entries) של ויקיפדיה, וטענה זו עומדת במוקד עבודה זו. מטרת מחקר זה היא ניסיון לכמת את קיומה ושיעורה של הטיה כפי שהיא משתקפת בתוצרים של עורכי הויקיפדיה השונים, כלומר – המידע שנגלה לעיני המשתמשים והצופים בדפי ויקיפדיה.

נושא ההטיה המגדרית בוויקיפדיה העברית נזכר במקורות אקדמיים שונים כעובדה שאין עליה עוררין זה מספר שנים (גבע, 2013; Lir, 2019) ושיעורם הגבוה של גברים בקרב עורכי ויקיפדיה הוא בגדר עובדה מבוססת מבחינה מחקרית (Konieczny, 2018), אך למיטב ידיעתנו טרם נערך בירור כמותי סדור של קיום ההטיה המגדרית ועוצמתה. כמו כן, גם אם זהות המשתתפים בוויקיפדיה אינה מאוזנת מבחינה מגדרית, האם ניתן להסיק מכך שהתכנים יהיו לא מאוזנים? התיאוריה המקובלת גורסת שהתשובה לשאלה זו חיובית, שכן אנשים כותבים על תחומים ועניינים הקרובים לליבם ולסביבתם. בהתאם לכך, מספר חוקרים עמדו על הקשר בין זהותם המגדרית של

¹ https://www.wikidata.org/wiki/Wikidata:WikiProject_Women [ההדגשה לא במקור]

כותבי הערכים בויקיפדיה ובין פער מגדרי בתוכני הערכים עצמם (Hinnosaar, 2019; Jemielniak, 2016), ואף עלתה טענה שהשיעור הגבוה של גברים בקרב העורכים מביא לייצוג יתר של נושאים ותכנים שמעניינים עורכים גברים (Falenska & Çetinoğlu, 2021).

פער מגדרי בתוכן יכול ללבוש צורות מגוונות. מחקרים אחדים בדקו הטיה מגדרית בויקיפדיה האנגלית בעזרת ניתוח תוכן של ערכים ביוגרפיים, שכללו היבטים של מטא-דאטה, אוצר מילים, מבנה רשת הקשרים ומבנים לשוניים, ומצאו הבדלים משמעותיים בניסוח הערכים בויקיפדיה בשל הבדלים מגדריים (Graells-Garrido et al., 2015; Wagner et al., 2016). כמו כן, נמצא כי ערכים העוסקים בנשים הם קצרים יותר, נתונים יותר לוויכוח (בין עורכים) ואף למחיקה.

כיום קרן ויקימדיה מפעילה את פרויקט WikiProject Women, שמטרתו, כפי שהיא מנוסחת בראש עמוד אתר הפרויקט: "WikiProject Women is an effort to get every item about a woman described properly on Wikidata."² דומה שניסוח זה של מטרת הפרויקט, על העמימות שבהגדרת המילה "כיאות" (properly), ממחיש עד כמה קשה למצוא הגדרה ממצה של ייצוג מאוזן מבחינה מגדרית. בפועל הפרויקט עוסק בעיבוי הערכים והמטא-דאטה על רשומות ביוגרפיות של נשים, והוא אכן סייע רבות בצמצום הפער בין המינים – מספר הביוגרפיות של נשים בויקיפדיה האנגלית גדל באופן משמעותי ונשים רבות נוספו כעורכות (Kandek, 2023).

2.2 מדדים להערכת הטיה מגדרית

ההגדרה המדויקת של המונח "הטיה מגדרית" משפיעה על המדד שייבחר כדי להעריך את קיומה ועוצמתה של ההטיה. בהתאם לכך, פותחו כלים מחקריים שונים להערכת איזון מגדרי (Hube & Fetahu, 2018; Konieczny & Klein, 2018).

עקרונית ניתן לטעון שכיוון ששיעור הנשים באוכלוסייה הוא מחצית, כ-50%, כל ייצוג במאגר מידע שאינו ייצוג שוויוני הוא ייצוג לא מאוזן. בויקיפדיה העברית נכון ל-2016 שיעור הביוגרפיות (ערכי ויקיפדיה שהנושא שלהם הוא אישיות) של נשים הוא כ-19%, ואילו 81% מהביוגרפיות הן ביוגרפיות של גברים. לגישת קרן ויקימדיה, העובדה ששיעור הביוגרפיות העוסקות בנשים הוא 19% מכלל הביוגרפיות היא כשלעצמה אינדיקציה ל"הטיה מגדרית בערכי ויקיפדיה" ("הפער המגדרי בויקיפדיה", 2023).

עם זאת, לצורך מחקרנו פער זה כשלעצמו אינו מצביע על הטיה מגדרית, משום שויקיפדיה אינה מתיימרת לשקף את המאזן הדמוגרפי באוכלוסייה אלא את ההיבטים האנציקלופדיים של מציאות זו, ובתחומים מסוימים אין מחלוקת שהיחס המגדרי אינו מקביל ליחס הדמוגרפי. כך, לדוגמה, אם ויקיפדיה מתיימרת להציג את האישים הבולטים בתחום המדיניות והפוליטיקה, לא סביר לצפות לשיעור של 50% ערכים המוקדשים לנשים מבין ערכי ויקיפדיה של אישים שהם ראשי מדינות (מלכים, נשיאים, ראשי ממשלה וכד'), משום שאין מחלוקת שבמציאות אין יחס שווה-מגדרית בין ראשי המדינות.

לפיכך תנאי חיוני לבחינה כמותית של הטיה במאגרי מידע הוא קו בסיס (Baseline) אשר ביחס אליו יימדד הייצוג במאגר המידע (Pitoura et al., 2018). ובענייננו, הגדרת המונחים "איזון מגדרי" ו"הטיה מגדרית" צריכה להתחשב הן במטרותיה של קרן ויקימדיה – הנגשת והפצת ידע לכול, והן

² https://www.wikidata.org/wiki/Wikidata:WikiProject_Women [ההדגשה לא במקור]

במהותה של ויקיפדיה – מיזם אנציקלופדי שתכליתו להוות **שיקוף** או **ייצוג** של תחומי הדעת האנושיים. מטרת אלה מניחות את קיומו של ידע אובייקטיבי לגבי המציאות, שמן הראוי להנגישו ולהפיצו בציבור. כלומר, קו הבסיס הרצוי, אשר ביחס אליו תיבחן קיומה של הטיה מגדרית הוא שיעור הייצוג המגדרי **בעולם המציאות**, שויקיפדיה שואפת לשקף. מטרת המחקר ושאלות המחקר המוצע נוסחו בהתאם לעיקרון השיקוף.

3. מטרת המחקר

מטרת המחקר היא לבדוק אם ובאיזו מידה קיימת הטיה מגדרית בויקיפדיה העברית. הטיה זו יכולה להתבטא באחת משתי דרכים:

ביטוי אחד הינו סטייה מובהקת סטטיסטית – כלומר ייצוג-יתר או תת-ייצוג לעומת קו הבסיס (Baseline). קו הבסיס נקבע על ידי מקורות חיצוניים לויקיפדיה אשר נהנים מאמינות ואובייקטיביות גבוהות.

לפיכך שאלת המחקר הראשונה היא: האם קיים הבדל בשיעור הייצוג של נשים (לעומת גברים) מבחינת מספר הרשומות שלהן בויקיפדיה, בהשוואה לשיעור הייצוג של נשים (לעומת גברים) באותה קטגוריה, כפי שהוא קיים במציאות?

ביטוי שני הינו הבדל בין ערכים המתייחסים לגברים ובין ערכים המתייחסים לנשים במדדים כמותיים הנוגעים לתוכן הרשומות.

לפיכך שאלת המחקר השנייה היא: האם ישנם הבדלים בתוכן הרשומות המתייחסות לביוגרפיות של נשים לעומת אלה של גברים?

לאור ריבוי התחומים שבהם עוסקת ויקיפדיה ועל מנת לבסס השוואה תקפה נבחרו קטגוריות מובחנות של רשומות ויקיפדיה ועבורן נבחנו התשובות לשאלות אלה. כך לדוגמה, אחת הקטגוריות שנבחרה היא קטגוריית עורכי הדין. עריכת דין כשלעצמה אינה קריטריון המשמש להכללת קורותיו של אדם כערך בויקיפדיה (דוגמה מנוגדת היא דרגת קצונה: כללי ויקיפדיה קובעים כי החזקה בדרגה קצונה גבוהה היא כשלעצמה סיבה לכלול ביוגרפיה של נושא הדרגה בויקיפדיה). כיוון שכך, אפשר להתייחס לקטגוריית עורכי הדין כמדגם אקראי של ערכים הנוגעים לאישים בויקיפדיה ולבדוק אם קיימים הבדלים מגדריים בקטגוריה זו.

4. השערות והנחות המחקר

הטיה מגדרית מוגדרת במחקר הנוכחי כמצב שבו היחס בין גברים ונשים, כפי שהוא מתבטא בויקיפדיה, שונה באופן משמעותי מהיחס באותו תחום **כפי שהוא במציאות**.

השערת המחקר באשר לשאלת המחקר הראשונה: בקטגוריה נתונה קיים הבדל בשיעור הייצוג של נשים (לעומת גברים) מבחינת מספר הרשומות שלהן בויקיפדיה בהשוואה לשיעור הייצוג שלהן באותה קטגוריה כפי שהוא קיים במציאות, כך שנשים מיוצגות ייצוג חסר בויקיפדיה.

בהתייחס לשאלת המחקר השנייה, השערת המחקר היא: קיים הבדל בין גברים ונשים, באופן שבקטגוריה נתונה לרשומות של נשים יש נפח רשומה קטן יותר משל גברים בכל המדדים הכמותיים (ראו פירוט בסעיף 5.6).

5. שיטת המחקר

5.1 שיטת המחקר

שיטת המחקר היא כמותית והיא מבוססת על איסוף נתונים כמותיים מתוך ערכי ויקיפדיה וניתוח תוכן של הרשומות. גם ניתוח התוכן של הרשומות הוא ניתוח תוכן כמותי: תוכן הרשומות עצמן לא

נבחן. לא נבחנו בחירת המילים, השפה והיבטים לקסיקליים או לינגוויסטיים אחרים של תוכן הרשומות, אלא היבטים כמותיים של תוכן הרשומות. מטבעו ניתוח כזה חשוף פחות להטיות החוקר או להטיות שמקורן בניתוח לקסיקלי או לינגוויסטי שגוי.

5.2. מערך המחקר

על מנת לקבוע אם הייצוג המגדרי במאגר המידע של ויקיפדיה³ משקף נאמנה את המציאות שאותה הוא מתיימר לייצג, נדרשה מלאכת איסוף נתונים משני סוגים:

1. חולצו נתוני מאגר המידע של ויקיפדיה העברית עצמה.
 2. חולצו נתוני בסיס ממקורות חיצוניים – נתוני יחס מגדרי בתחומים שונים במציאות, לשם השוואה לנתונים של ויקיפדיה.
- נתוני ויקיפדיה עצמם שימשו אותנו בשני אופנים: ראשית, כדי לחשב את היחס בין המגדרים (Gender Balance) ולהשוות ליחס שחושב לפי נתוני הבסיס, על מנת לבדוק אם היחס בויקיפדיה משקף את היחס הקיים בנתוני המציאות. שנית, שימוש בנתונים כמותיים שונים כדי להעריך את "נפח" או "עומק" הרשומה. כך, לדוגמה, בדקנו כמה מילים מכילה הרשומה וערכנו השוואה מגדרית.

5.3. חילוץ נתונים מתוך ויקיפדיה

בשלב ראשון נדגמו כלל הרשומות בויקיפדיה המתייחסות לאישים (להבדיל מערכים ויקיפדיים אחרים).

לשם כך נכתב סקריפט בשפת התכנות python המזהה ערכי ויקיפדיה שהסוג שלהם הוא "אדם" (instance of: human). עבור כל אחת מהרשומות שזוהו, נדגמו מספר נתונים:

1. שם הרשומה / האישי (זיהוי חד-חד-ערכי).
2. המגדר של האישי.
3. מספר המילים בגוף הרשומה (לא כולל הערות שוליים).
4. מספר הקישורים הפנימיים – מספר הקישורים לערכי ויקיפדיה אחרים המצויים בתוך הרשומה.
5. מספר הקישורים החיצוניים – קישורים לאתרים שאינם ערכי ויקיפדיה מתוך הרשומה.
6. מספר השפות הנוספות שבהן קיימת הרשומה.
7. מספר הקטגוריות של ויקיפדיה שאליהן משויכת הרשומה.
8. מספר ערכי הויקיפדיה המפנים אל הרשומה.
9. תאריך יצירת הרשומה.
10. מספר העריכות שעברה הרשומה מיום יצירתה.

חלק מהנתונים נשאבו בעזרת ממשק השאלות על מאגר הנתונים של ויקיפדיה (Wikidata), וחלק

בעזרת סקריפט ייעודי לשם שאיבת נתונים מדפי הויקיפדיה של הרשומות הרלוונטיות.⁴

³ החל מפרק זה, בכל מקום שבו מוזכרת ויקיפדיה, הכוונה לויקיפדיה בשפה העברית.

⁴ בשל גודלו העצום של המאגר לא היתה אפשרות טכנית לעשות שימוש בשאלות sparql, שבהן נעשה שימוש בממשק השאלות על Wikidata.

הנתונים נשאבו הן מתוך הטקסט של הרשומה עצמה, והן מתוך נתוני הויקיפדיה של הרשומה: זיהוי המגדר של האישיות ברשומה ושנת הלידה והפטירה נשאבו מנתוני הויקיפדיה של הרשומות, בעוד ששאר הנתונים חולצו בעזרת ניתוח דף הויקיפדיה של הרשומות עצמן.

בטבלה 1 מפורטים הקריטריונים שלפיהם חולצו הקטגוריות ששימשו במחקר מתוך מאגר הרשומות של ויקיפדיה, וכן מספר הרשומות שחולצו בכל קטגוריה.

טבלה 1 – קריטריונים לחילוץ רשומות ויקיפדיה ויחס מגדרי בקטגוריות נבחרות

קטגוריה	קריטריון לחילוץ רשומות	מספר רשומות גברים	מספר רשומות נשים	סה"כ רשומות
עורכי דין	נתוני ויקיפדיה: מקצוע = עורכי דין	465	119	584
קציני צבא	נתוני ויקיפדיה: דרגה צבאית = אלוף, דרגה צבאית = תת-אלוף, דרגה צבאית = אלוף משנה, דרגה צבאית = סגן אלוף, דרגה צבאית = רב סרן, דרגה צבאית = סרן	1,413	27	1,440
דירקטורים	קטגוריית ויקיפדיה: "דירקטורים ישראלים" + קטגוריית ויקיפדיה: "דירקטורים ישראליות"	200	108	308
ראשי ערים ומועצות	רשומות בקטגוריות ובקטגוריות המשנה של קטגוריות ויקיפדיה: "ראשי ערים בישראל" + "ראשי מועצות מקומיות בישראל"	398	18	416
רואי חשבון	קטגוריית ויקיפדיה: "רואי חשבון ישראלים" + קטגוריית ויקיפדיה: "רואות חשבון ישראליות"	41	6	47

5.4. חילוץ נתוני בסיס להשוואה

נתוני הבסיס להשוואה עם ויקיפדיה נשאבו ממקורות מגוונים, מקורות מוסמכים וניטרליים ככל הניתן. המקורות העיקריים שבהם נעשה שימוש הם נתוני מחקרים ייעודיים שבוצעו על ידי מרכז המחקר והמידע של הכנסת, אשר חלקם הגדול נשאבו מנתוני הלשכה המרכזית לסטטיסטיקה.

בסופו של דבר התקבלו נתונים על ייצוג גברים לעומת נשים בקטגוריות האלה:

1. בעלי דרגות קצונה בכירה בצבא (כאן התקבלו רק הנתונים על היחס בין המגדרים, באחוזים, ולא נתונים נומינליים) – בשל שיקולי גודל המדגם (הן בויקיפדיה והן בנתוני האמת) נבחר היחס המגדרי החל מדרגת סרן ועד דרגת אלוף (אבגר, 2019א).
2. עורכי דין (שחק, 2017).
3. רואי חשבון (שחק, 2017).
4. ראשי מועצות מקומיות וראשי ערים (אבגר, 2019ב).
5. חברי דירקטוריון (מזרחי, 2013).

בטבלה 2 מובאים נתוני הבסיס בקטגוריות שנבחרו כפי שנשאבו מתוך המקורות השונים.

קטגוריה	מספר גברים	מספר נשים	סה"כ
עורכי דין	33,167	25,525	58,692
קציני צבא	אין נתונים מספריים		
דירקטורים	280	234	514
ראשי ערים ומועצות	398	18	416
רואי חשבון	12,896	6,222	19,118

5.5. שאלת המחקר הראשונה: ההשוואה בין היחס המגדרי בויקיפדיה ובנתונים

החיצוניים

שאלת המחקר הראשונה היא האם הייצוג המגדרי הקיים בויקיפדיה מהווה שיקוף נאמן של הייצוג המגדרי במציאות. לשם כך בחנו את הייצוג המגדרי בקטגוריות ספציפיות. זהו שלב הכרחי – אין די בבדיקת יחס הגברים לעומת הנשים במכלול ערכי האישים בויקיפדיה, משום שבתחומים מסוימים היחס המגדרי אינו זהה ליחס הדמוגרפי בכלל האוכלוסייה.

על מנת להתגבר על מכשלת הקריטריונים של ויקיפדיה לכתיבת רשומה על אדם זה או אחר, בחרנו במכוון חלק מהקטגוריות כך שהמכנה המשותף שלהן **אינו תנאי ליצירת רשומה בויקיפדיה**. למשל, קטגוריות עורכי דין, רואי חשבון והדירקטורים הינן קטגוריות לתארים שאינם מצדיקים כשלעצמם כתיבת רשומה בויקיפדיה (הדברים שונים מעט בקטגוריות של קציני הצבא וראשי הרשויות). אולם, אם ימצא שהחלוקה המגדרית של ערכי ויקיפדיה המוקדשים לעורכי דין, לרואי חשבון ולדירקטורים דומה לחלוקה במציאות, סביר להניח שויקיפדיה היא אכן מדגם המייצג היטב את המציאות בתחומים שונים.

מכיון שהמשתנה הבלתי תלוי – המגדר – הוא משתנה שמי, המבחן הסטטיסטי ששימש אותנו היה מבחן χ^2 (כי-בריבוע), כדי לבדוק אם השכיחות של נשים לעומת גברים באוכלוסיית המדגם (ויקיפדיה) סוטה באופן משמעותי מהשכיחות הנצפית באוכלוסיית הבסיס (על פי הנתונים שהתקבלו ממקורות חיצוניים, ראו טבלה 2).

5.6. שאלת המחקר השנייה: השוואת נפח הערכים בחתך מגדרי

שאלת המחקר השנייה היא: האם ישנם הבדלים בנפח הערכים (בהתאם לרשימת המדדים שלהלן) של ערכי הויקיפדיה המתייחסים לביוגרפיות של נשים, לעומת אלה המתייחסים לגברים? לשם בחינת שאלה זו השתמשנו בנתונים כמותיים של ויקיפדיה, כדי להשוות את נפח הרשומות בויקיפדיה המתייחסות לאישים. הנחנו שעצם קיומה של רשומה לגבר או אישה, אינו מהווה ייצוג מספק של אותה אישיות. רשומות ויקיפדיה נבדלות זו מזו בנפחן ובתוכן. עם זאת, סברנו שיש מספר מדדים כמותיים שיש בהם כדי להעיד על הנפח של הרשומה ועל האיכות שלה, וביקשנו להשוות את המדדים הללו בין הרשומות הנוגעות לשני המגדרים.

המדדים שבדקנו על מנת לבחון את נפח ותוכן הרשומות הם המדדים המפורטים להלן:

1. מספר המילים (Number Of Words) – ההנחה היא שיש קשר בין מספר המילים ברשומה לבין איכותה. מספר מילים נמוך יכול להעיד הן על רשומה שאיננה נהנית מתשומת לב רבה של עורכים והן על איכות נמוכה של הרשומה. מספר מילים נמוך אף קשור בהעדרם של קישורים פנימיים וחיצוניים לרשומה (כפי שיפורט להלן).

2. מספר הקישורים הפנימיים (Internal Links) – מספר הקישורים לערכי ויקיפדיה נוספים מעיד על נוכחות הרשומה ברשת המידע של ויקיפדיה וכן על השקעת זמן ומאמץ של עורך כדי לקשר בין הרשומה ובין ערכי ויקיפדיה אחרים.
 3. מספר הקישורים החיצוניים (External Links) – מספר הקישורים למקומות מחוץ לויקיפדיה מעיד הן על נוכחות של האישיות המיוצגת במקומות נוספים במרחב המידע והן על השקעה של עורך הרשומה בקישור המידע המופיע בה למקורות חיצוניים. לרוב קישורים חיצוניים ברשומה הם למראי מקום, כך שמקומם ברשומה הוא בהערות השוליים או בהפניות למקורות נוספים, מחוץ לגוף התוכן (Content) של הרשומה. כיוון שכך, יש בקישורים חיצוניים רבים כדי להעיד על מהימנות התוכן של הרשומה.
 4. מספר קטגוריות הויקיפדיה שהרשומה משויכת אליהן (Wikipedia Categories) – שיוך רשומה לקטגוריות רבות יכול להעיד הן על חשיבותה של האישיות שהיא נושא הרשומה והן על תשומת הלב שניתנה לתוכן הרשומה. ככל שרשומה שייכת לקטגוריות רבות יותר, מתרבים הסיכויים שייחשפו אליה קוראים רבים יותר, שמתעניינים לפחות באחת מקטגוריות ויקיפדיה אלה.
 5. מספר השפות הנוספות שבהן קיימת הרשומה (Languages) – קיומה של רשומה בשפות נוספות יכול להעיד כי נמצאו מי שסברו שהאישיות מעניינת גם דוברי שפות נוספות, ויש בכך אינדיקציה לחשיבות הרשומה.
 6. מספר ההפניות לרשומה מערכים אחרים (References) – נתון זה מעיד הן על נוכחות של הרשומה ברשת הערכים של ויקיפדיה והן על חשיפה רבה יותר למשתמשי ויקיפדיה (קוראים ועורכים). ככל שרשומה מקושרת ליותר ערכים אחרים מתרבים הסיכויים שקוראי ויקיפדיה ייחשפו אליה באמצעות גישה דרך רשומות ויקיפדיה אחרות.
 7. מספר עריכות של הרשומה ביחס למשך קיומה (Edits) – ערך זה חושב כמנה של מספר העריכות ומשך קיום הרשומה בשנים. עריכות רבות של רשומה יכולות להעיד על עניין של עורכים ברשומה ובתוכנה. מנגד, עריכות רבות של רשומה יכולות להעיד כי הרשומה ותוכנה שנויים במחלוקת.
- ביחס למדדים 1–6, המפורטים לעיל, הנחנו שככל שהמדד גבוה יותר, כך נפח ואיכות הרשומה גבוהים יותר. יוצא דופן הוא מדד 7 – מספר עריכות של הרשומה – שאינו נתון מדד ברור באשר לאיכותה. לעיתים ריבוי עריכות מעיד דווקא על רשומה שחשיבותה פחותה, שעברה עריכות רבות לשם ניקוי מתוכן או אפילו ניסיונות למחוק אותה כליל. לבד מכך, ידוע לנו שערכים שיש לגביהם מחלוקות הם ערכים שנערכים עריכות ועריכות-שכנגד פעמים רבות, כך שעצם העריכות הרבות של רשומה מסוימת אינו מעיד בהכרח על נפח הרשומה, ולעיתים אף להפך, יכול להעיד על העדר מהימנות של תוכן הרשומה, ועל קיצוצים ומחיקות (Wagner et al., 2016).
- לשם ביצוע ההשוואה בין המדדים הטקסטואליים של רשומות העוסקות בגברים ובנשים נעשה שימוש במבחן הסטטיסטי t לקבוצות בלתי תלויות. במקרים שבהם נצפו מדדים סטטיסטיים של צידוד (Skewness) וגבנוניות (Kurtosis), המעידים על כך שההתפלגות של המדדים הללו איננה התפלגות נורמלית, נעשה שימוש במבחן Mann-Whitney, שהערך הסטטיסטי שלו מסומן באות

Z. עם זאת, כפי שיפורט להלן, במרבית המקרים כשנצפו תוצאות מובהקות הן נצפו בשני המבחנים, וכך היה גם כשלא נצפו תוצאות מובהקות.

ממצאים

5.7. כלל ערכי האישים

בדגימת כל הערכים בויקיפדיה שהם מסוג "אישיות" (instance of: human) נמצאו כ-91 אלף רשומות של ערכי אישים בויקיפדיה העברית.

לאחר מכן בוצע תהליך של טיוב הרשומות – הסרת רשומות כפולות, רשומות שאין בהן תוכן מלבד הפניה לרשומות אחרות, או רשומות שנתוני הויקיפדיה שלהן היו שגויים. בתהליך זה הוסרו כ-6,000 רשומות.

לאחר תהליך זה נותרה במדגם רשימת האישים שמנתה 85,374 אישים. מתוכם כ-68,000 גברים ו-17,000 נשים. כלומר, כ-20% מתוך הביוגרפיות הקיימות בויקיפדיה העברית הן ביוגרפיות של נשים ו-80% הן ביוגרפיות של גברים.

בשל גודל המדגם העצום, לא נערך מבחן t להשוואה בין המדדים הטקסטואליים של המגדרים השונים בכלל המדגם.

5.8. עורכי דין

במהלך המחקר נדגמו 584 רשומות של עורכי דין ישראלים באמצעות שאילתא ייעודית על מאגר ויקינתונים. 119 (20%) מתוך הרשומות הללו היו של נשים ו-465 (80%) של גברים.

5.8.1. סטייה מרמת הייצוג

לצורך בחינת שאלת המחקר הראשונה, בדקנו האם היחס בין נשים וגברים ברשומות ויקיפדיה סוטה באופן מובהק מהיחס הקיים בעולם האמיתי. לשם כך נועצנו בנתוני לשכת עורכי הדין, שלפיהם עד שנת 2009 הוסמכו בישראל 58,692 עורכי דין. 25,525 מהם הם נשים (45%) ו-33,167 גברים (55%) (שחק, 2017).

על בסיס נתונים אלה בוצע מבחן χ^2 כדי לבחון את הקשר בין המגדר ובין השיעור באוכלוסיית הבסיס ובאוכלוסיית המדגם – ויקיפדיה. תוצאת המבחן מוצגת להלן:

$$\chi^2(1, N=584) = 125.85, p < 0.001$$

נמצא שלמשתנה המגדר יש קשר לשיעור הנוכחות ברשומות ויקיפדיה. נמצא שמשנתה המגדר קשור באופן מובהק למידת הנוכחות ברשומות ויקיפדיה – רשומות של גברים יוצגו יותר בויקיפדיה בהתחשב בשיעור הייצוג שלהם בקטגוריית עורכי דין, ורשומות של נשים יוצגו פחות בויקיפדיה בהתחשב בשיעור הייצוג שלהן בקטגוריית עורכי הדין.

5.8.2. בחינת עומק של תוכן הערכים

לצורך בחינת שאלת המחקר השנייה בוצעה השוואה של תוכן רשומות ויקיפדיה. לבדיקת ההשערה כי קיים הבדל בממוצעי המדדים הטקסטואליים של הרשומות בין גברים ובין נשים, נערך מבחן t למדגמים בלתי תלויים. כמו כן, נערך מבחן Mann-Whitney במקרים שבהם התגלו אינדיקציות למשתנים שאינם מתפלגים התפלגויות נורמליות. התוצאות מוצגות בטבלה 3.

טבלה 3 – השוואת מדדים כמותיים ברשומות בחתך מגדר – קטגוריית עורכי דין

פרמטר	מגדר	M	SD	ערך t	Mann-Whitney
	גברים	519.92	527.19		
				t(155)= 1.003, p > 0.05	

Mann-Whitney	ערך t	SD	M	מגדר	פרמטר
Z = -0.008, p > 0.05		687.43	587.74	נשים	מספר מילים
Z = -0.689, p > 0.05	t(582) = 0.675, p > 0.05	17.97	20.36	גברים	מספר קישורים
		21.53	21.66	נשים	חיצוניים
Z = -0.882, p > 0.05	t(582) = 0.441, p > 0.05	80.58	109.06	גברים	מספר קישורים פנימיים
		83.67	105.39	נשים	
* Z = -2.326, p < 0.05	t(582) = 0.911, p > 0.05	7.15	3.18	גברים	מספר שפות נוספות
		6.04	2.53	נשים	
Z = -1.913, p > 0.05	t(582) = 0.577, p > 0.05	78.16	52.99	גברים	מספר הפניות
		81.00	48.32	נשים	
* Z = -2.261, p < 0.05	* t(190) = 2.138, p < 0.05	6.64	11.23	גברים	מספר קטגוריות
		6.31	9.82	נשים	
** Z = -4.910, p < 0.01	** t(155) = 3.234, p < 0.01	18.64	13.38	גברים	עריכות (כפונקציה של הזמן)
		24.35	21.12	נשים	

מקרא:

- *p<0.05 - **p<0.01

ניתוח הנתונים בטבלה 3

לבדיקת ההשערה כי קיים הבדל בין גברים ונשים בממוצע מספר השפות של הרשומה, נערך מבחן Mann-Whitney למדגמים שאינם מתפלגים נורמלית. נמצא כי קיים הבדל מובהק בין הממוצעים ($Z = -2.326, p < 0.05$), כך שהממוצע של מספר השפות של רשומות גברים ($M = 3.18, SD = 7.15$) גבוה מהממוצע של מספר השפות של רשומות נשים ($M = 2.53, SD = 6.04$).

לבדיקת ההשערה כי קיים הבדל בין גברים ונשים בממוצע מספר הקטגוריות של הרשומה, נערך מבחן t למדגמים בלתי תלויים. נמצא כי קיים הבדל מובהק בין הממוצעים ($t(190) = 2.138, p < 0.05$) כך שממוצע מספר הקטגוריות של רשומות גברים ($M = 11.23, SD = 6.64$) גבוה מממוצע מספר הקטגוריות של רשומות נשים ($M = 9.82, SD = 6.31$).

לבדיקת ההשערה כי קיים הבדל בין גברים ונשים בממוצע מספר העריכות של הרשומה, נערך מבחן t למדגמים בלתי תלויים. נמצא כי קיים הבדל מובהק בין הממוצעים ($t(155) = 3.234, p < 0.01$) כך שממוצע מספר העריכות של רשומות גברים ($M = 13.38, SD = 18.64$) נמוך מממוצע מספר העריכות של רשומות נשים ($M = 21.12, SD = 24.35$).

ביתר המדדים הטקסטואליים – מספר מילים, מספר קישורים, מספר קישורים חיצוניים, מספר הפניות – לא נמצאו הבדלים מובהקים בין גברים ונשים.

על מנת לבסס את התוצאות שקיבלנו, בדקנו אם המדדים הכמותיים של הרשומות שבדקנו מצויים במתאם זה עם זה, כלומר, אם המדדים הטקסטואליים השונים שנבדקו עבור הרשומות בודקים את אותו עולם תוכן, וממילא מהימנים יותר במתן מענה לשאלת המחקר. התוצאות מוצגות בטבלה

.4

פרמטר	מספר מילים	קישורים פנימיים	קישורים חיצוניים	שפות	קטגוריות	הפניות	עריכות
מספר מילים	1	***0.712	***0.717	***0.385	***0.431	***0.568	***0.453
קישורים פנימיים	1	***0.736	***0.736	***0.544	***0.742	***0.904	***0.494
קישורים חיצוניים			1	***0.473	***0.488	***0.652	***0.515
שפות				1	***0.504	***0.614	***0.333
קטגוריות					1	***0.736	***0.270
הפניות						1	***0.448

מקרא:

$***p < 0.001$

ניתוח הנתונים בטבלה 4

מטבלה 4 עולה כי כלל המדדים הטקסטואליים מדגימים מתאם מובהק וחיובי זה עם זה. בכלל המדדים המתאם בינוני עד חזק מאוד, למעט המדד של מספר העריכות, שנמצא במתאם חלש יחסית עם כל אחד מהמדדים האחרים (בין 0.23 ל-0.52).

5.9. בעלי דרגות צבאיות

במהלך המחקר נדגמו רשומות של 1,440 אישים שהם בעלי דרגת קצונה בצה"ל, החל מדרגת סרן ועד דרגת אלוף.⁵ הרשומות נדגמו באמצעות מספר שאילתות על מסד הנתונים של ויקינתונים. מתוך רשומות אלה, 1,413 (98%) היו של גברים ו-27 (2%) של נשים.

5.9.1. סטייה מרמת הייצוג

מטעמי שמירה על סודיות, ביחס לאנשי צבא לא ניתן היה לקבל נתונים נומינליים לגבי היחס המגדרי בקרב בעלי ובעלות דרגת קצונה בצה"ל, אלא נתוני אחוזים בלבד (אבגר, 2019). מטעם זה לא בוצע מבחן χ^2 על מנת לבדוק אם יש הטיה ביחס המגדרי של מספר הרשומות לעומת נתוני הבסיס באשר לאוכלוסיית הקצינים. כיוון שכך, בקטגוריה אנשי הצבא לא נבדקה שאלת המחקר הראשונה.

5.9.2. בחינת עומק של תוכן הרשומות

לצורך בחינת שאלת המחקר השנייה בוצעה השוואה של תוכן רשומות הויקיפדיה.

⁵ בעלי דרגות רב-אלוף לא נדגמו משום שאין ביניהם רשומות של נשים כלל, ועל כן בהכנסת רשומות אלה היה משום עיוות של נתוני המדגם.

לבדיקת ההשערה כי קיים הבדל בין גברים ונשים בממוצעי המדדים הטקסטואליים של הרשומות, נערך מבחן t למדגמים בלתי תלויים. כמו כן נערך מבחן Mann-Whitney במקרים שבהם התגלו אינדיקציות למשתנים שאינם מתפלגים התפלגויות נורמליות. תוצאות המבחן מוצגות בטבלה 5.

טבלה 5 – השוואת מדדים טקסטואליים ברשומות בחתך מגדר – קטגוריית אנשי צבא

פרמטר	מגדר	M	SD	ערך t	Mann-Whitney
מספר מילים	גברים	417.66	509.91	t(1438) = 1.479, p > 0.05	Z=-1.792, p>0.05
	נשים	272.33	158.51		
קישורים חיצוניים	גברים	10.82	14.03	t(1438) = 0.452, p > 0.05	Z=-0.117, p>0.05
	נשים	9.59	8.55		
קישורים פנימיים	גברים	111.26	64.81	** t(1438) = 2.267, p < 0.01	*** Z=-3.651, p<0.001
	נשים	78.26	55.66		
שפות נוספות	גברים	0.79	4.42	t(1438) = 0.057, p > 0.05	Z=-1.040, p>0.05
	נשים	0.74	1.61		
הפניות	גברים	49.15	56.64	* t(1438) = 2.032, p < 0.05	** Z=-3.227, p<0.01
	נשים	26.89	39.38		
קטגוריות	גברים	8.51	4.55	T(1438) = 1.415, p > 0.05	* Z=-2.039, p<0.05
	נשים	7.26	3.69		
עריכות (כפונקציה של הזמן)	גברים	13.05	16.36	t(1438) = -0.13, p > 0.05	Z=-1.681, p>0.05
	נשים	13.46	7.71		

- * $p < 0.05$ - ** $p < 0.01$ - *** $p < 0.001$

ניתוח הנתונים בטבלה 5

לבדיקת ההשערה כי קיים הבדל בממוצע מספר הקישורים הפנימיים של הרשומה בין גברים ונשים, נערך מבחן t למדגמים בלתי תלויים. נמצא כי קיים הבדל מובהק בין הממוצעים ($t(1438) = 2.267, p < 0.01$), כך שממוצע הקישורים הפנימיים של רשומות גברים ($M = 111.26, SD = 64.81$) גבוה מממוצע הקישורים הפנימיים של רשומות נשים ($M = 78.26, SD = 55.66$).

לבדיקת ההשערה כי קיים הבדל בין גברים ונשים בממוצע מספר ההפניות אל הרשומה, נערך מבחן t למדגמים בלתי תלויים. נמצא כי קיים הבדל מובהק בין הממוצעים ($t(1438) = 2.032, p < 0.05$), כך שממוצע ההפניות אל רשומות גברים ($M = 49.15, SD = 56.64$) גבוה מממוצע ההפניות אל רשומות נשים ($M = 26.89, SD = 39.38$).

לבדיקת ההשערה כי קיים הבדל בין גברים ונשים בממוצע מספר הקטגוריות של הרשומה, נערך מבחן Mann-Whitney למדגמים שאינם מתפלגים נורמלית. נמצא כי קיים הבדל מובהק בין הממוצעים ($Z = -2.039, p < 0.05$), כך שממוצע מספר הקטגוריות של רשומות גברים ($M = 8.51, SD = 4.55$) גבוה מממוצע מספר הקטגוריות של רשומות נשים ($M = 7.26, SD = 3.69$).

ביתר המדדים הטקסטואליים לא נמצאו הבדלים מובהקים בין גברים ונשים.

על מנת לבסס את התוצאות שקיבלנו, בדקנו אם המדדים הכמותיים של הרשומות שבדקנו מצויים במתאם זה עם זה, כלומר, אם המדדים הטקסטואליים השונים שנבדקו עבור הרשומות בודקים את אותו עולם תוכן, וממילא מהימנים יותר במתן מענה לשאלת המחקר. התוצאות מוצגות בטבלה 6.

טבלה 6 – מתאם בין מדדים טקסטואליים של רשומות – קטגוריית אנשי צבא

פרמטר	מספר מילים	קישורים פנימיים	קישורים חיצוניים	שפות	קטגוריות	הפניות	עריכות
מספר מילים	1	***0.569	***0.632	***0.610	***0.521	***0.451	***0.657
קישורים פנימיים		1	***0.646	***0.478	***0.725	***0.868	***0.380
קישורים חיצוניים			1	***0.703	***0.634	***0.642	***0.647
שפות				1	***0.635	***0.577	***0.608
קטגוריות					1	***0.738	***0.417
הפניות						1	***0.368

$$***p < 0.001$$

מטבלה 6 עולה כי כלל המדדים הטקסטואליים מקיימים מתאם מובהק וחיובי זה עם זה. בכלל המדדים המתאם בינוני עד חזק, למעט המדד של מספר העריכות, שנמצא במתאם חלש יחסית ($R < 0.5$) עם חלק מהמדדים הטקסטואליים האחרים.

5.10. דירקטורים

במהלך המחקר נדגמו רשומות של 308 חברי וחברות דירקטוריון בישראל באמצעות חילוץ שמות מרשומות השייכות לקטגוריית ויקיפדיה "דירקטורים ישראלים" ו"דירקטוריות ישראליות". מתוך רשומות אלה 200 (65%) היו של גברים ו-108 (35%) של נשים.

5.10.1 סטייה מרמת הייצוג

על מנת לבחון את שאלת המחקר הראשונה, השווינו את השיעורים הללו לשיעורים בנתוני הבסיס: 514 חברי דירקטוריון, מתוכם 234 (46%) נשים ו-280 (54%) גברים (מזרחי, 2013). על בסיס נתונים אלה בוצע מבחן X^2 כדי לבחון את הקשר בין המגדר ובין השיעור באוכלוסיית הבסיס ובאוכלוסיית המדגם – ויקיפדיה. תוצאת המבחן מוצגת להלן:

$$X^2(1, N=308) = 8.6738, p\text{-value} < 0.05$$

נמצא שלמשתנה המגדר יש קשר לשיעור הנוכחות ברשומות הויקיפדיה. נמצא שמשנתנה המגדר קשור באופן מובהק למידת הנוכחות ברשומות ויקיפדיה, באופן שרשומות של גברים יוצגו יותר בוויקיפדיה בהתחשב בשיעור הייצוג שלהם בקטגוריית הדירקטורים, ורשומות של נשים יוצגו פחות בוויקיפדיה בהתחשב בשיעור הייצוג שלהן בקטגוריית הדירקטורים.

5.10.2 בחינת עומק של תוכן הרשומות

לצורך בחינת שאלת המחקר השנייה בוצעה השוואה של תוכן רשומות הויקיפדיה. לבדיקת ההשערה כי קיים הבדל בממוצעי המדדים הטקסטואליים של הרשומות בין גברים ובין נשים, נערך מבחן t למדגמים בלתי תלויים. כמו כן, נערך מבחן Mann-Whitney במקרים שבהם התגלו אינדיקציות למשתנים שאינם מתפלגים התפלגויות נורמליות. תוצאות המבחן מוצגות בטבלה 7.

טבלה 7 – השוואת מדדים טקסטואליים ברשומות בחתך מגדר – קטגוריית דירקטורים

פרמטר	מגדר	M	SD	ערך t	Mann-Whitney
מספר מילים	גברים	567.93	643.67	* $t(299) = 2.562, p < 0.05$	$Z = -1.620, p > 0.05$
	נשים	430.53	294.67		
קישורים חיצוניים	גברים	17.79	18.17	$t(306) = 0.906, p > 0.05$	$Z = -0.386, p > 0.05$
	נשים	16.03	12.04		

פרמטר	מגדר	M	SD	ערך t	Mann-Whitney
קישורים פנימיים	גברים	104.77	88.41	*** t(305) = 3.582, p < 0.001	** Z = -2.977, p < 0.01
	נשים	76.49	50.12		
שפות נוספות	גברים	2.13	5.76	t(306) = 1.333, p > 0.05	* Z = -2.230, p < 0.05
	נשים	1.31	3.86		
הפניות	גברים	45.49	80.00	** t(305) = 3.199, p < 0.01	*** Z = -3.550, p < 0.001
	נשים	23.12	42.68		
קטגוריות	גברים	10.23	5.46	t(306) = 1.085, p > 0.05	Z = -1.055, p > 0.05
	נשים	9.56	4.56		
עריכות (כפונקציה של הזמן)	גברים	16.69	17.74	** t(157) = 3.247, p < 0.01	*** Z = -4.047, p < 0.001
	נשים	26.10	27.12		

מקרא:

- *p<0.05 - **p<0.01 - ***p<0.001

ניתוח הנתונים בטבלה 7

לבדיקת ההשערה כי קיים הבדל בין גברים ונשים במוצע הקישורים הפנימיים של הרשומה, נערך מבחן t למדגמים בלתי תלויים. נמצא כי קיים הבדל מובהק בין הממוצעים ($t(305) = 3.582, p < 0.001$) – מספר הקישורים הפנימיים של רשומות גברים ($M = 104.77, SD = 88.41$) גבוה ממספר הקישורים הפנימיים של רשומות נשים ($M = 76.49, SD = 50.12$).

לבדיקת ההשערה כי קיים הבדל בין גברים ונשים בין מספר השפות של רשומות, נערך מבחן Mann-Whitney למדגמים שאינם מתפלגים התפלגות נורמלית. נמצא כי קיים הבדל מובהק בין הממוצעים ($Z = -2.230, p < 0.05$) – מספר השפות של רשומות גברים ($M = 2.13, SD = 5.76$) גבוה ממספר השפות של רשומות נשים ($M = 1.31, SD = 3.86$).

לבדיקת ההשערה כי קיים הבדל במוצע ההפניות לרשומה בין גברים ונשים, נערך מבחן t למדגמים בלתי תלויים. נמצא כי קיים הבדל מובהק בין הממוצעים ($t(305) = 3.199, p < 0.01$) – מספר ההפניות לרשומות גברים ($M = 45.49, SD = 80.00$) גבוה ממספר ההפניות לרשומות נשים ($M = 23.12, SD = 42.68$).

לבדיקת ההשערה כי קיים הבדל בין גברים ונשים בממוצע העריכות של הרשומה, נערך מבחן t למדגמים בלתי תלויים. נמצא כי קיים הבדל מובהק בין הממוצעים ($t(157) = 3.247, p < 0.01$) – מספר העריכות של רשומות גברים ($M = 16.69, SD = 17.74$) נמוך ממספר העריכות של רשומות נשים ($M = 26.10, SD = 27.12$).

ביתר המדדים (מספר מילים, מספר קישורים חיצוניים, מספר קטגוריות) לא נמצאו הבדלים מובהקים בין גברים ונשים.

על מנת לבסס את התוצאות שקיבלנו, בדקנו אם המדדים הכמותיים של הרשומות שבדקנו מצויים במתאם זה עם זה, כלומר, אם המדדים הטקסטואליים השונים שנבדקו עבור הרשומות בודקים את אותו עולם תוכן, וממילא מהימנים יותר במתן מענה לשאלת המחקר. התוצאות מוצגות בטבלה 8.

טבלה 8 – מתאם בין מדדים טקסטואליים של רשומות – קטגוריית דירקטורים

עריכות	הפניות	קטגוריות	שפות	קישורים חיצוניים	קישורים פנימיים	מספר מילים	פרמטר
***0.333	***0.590	***0.522	***0.644	***0.578	***0.657	1	מספר מילים
***0.305	***0.909	***0.670	***0.699	***0.641	1		קישורים פנימיים
***0.418	***0.591	***0.515	***0.592	1			קישורים חיצוניים
***0.248	***0.755	***0.610	1				שפות
***0.322	***0.687	1					קטגוריות
***0.288	1						הפניות

מקרא:

*** $p < 0.001$

מטבלה 8 עולה כי כלל המדדים מקיימים מתאם מובהק זה עם זה. בכלל המדדים המתאם בינוני עד חזק מאוד, למעט המדד של מספר העריכות, שנמצא במתאם חלשה יחסית ($R < 0.5$) עם כל אחד מהמדדים הטקסטואליים האחרים.

5.11. ראשי ערים ומועצות

במהלך המחקר נדגמו רשומות של 416 ראשי ערים ומועצות בישראל באמצעות חילוץ שמות הרשומות השייכות לקטגוריית ויקיפדיה "ראשי ערים בישראל" וקטגוריית המשנה שלה. מתוך רשומות אלה 398 (96%) היו של גברים ו-18 (4%) של נשים.

5.11.1. סטייה מרמת הייצוג

לצורך בחינת שאלת המחקר הראשונה, בדקנו אם היחס בין נשים וגברים ברשומות ויקיפדיה סוטה באופן מובהק מהיחס הקיים, לפי נתוני הבסיס.

לפי נתוני הבסיס, בשנת 2018 היו בישראל 250 ראשי ערים ומועצות, מתוכם 14 (6%) נשים ו-236 גברים (94%) (אבגר, 2019).

בוצע מבחן χ^2 כדי לבחון את הקשר בין המגדר ובין השיעור באוכלוסיית הבסיס ובאוכלוסיית המדגם – ויקיפדיה.

$$\chi^2(1, N=416) = 0.5533, p\text{-value} > 0.05$$

ממבחן זה עולה כי לא נמצא הבדל מובהק בין הייצוג המגדרי בוויקיפדיה ובין הייצוג המגדרי באוכלוסיית הבסיס, כלומר לא נמצא הבדל מובהק בין מידת הייצוג של נשים וגברים ברשומות ויקיפדיה לעומת מידת הייצוג שלהם בנתוני הבסיס.

5.11.2. בחינת עומק של תוכן הרשומות

לצורך בחינת שאלת המחקר השנייה השווינו את הפרמטרים התוכניים השונים של הרשומות בעזרת מבחן t לקבוצות בלתי תלויות. התוצאות מוצגות בטבלה 9.

טבלה 9 – השוואת מדדים טקסטואליים ברשומות בחתך מגדר – קטגוריית ראשי מועצות

מגדר	M	SD	ערך t	Mann-Whitney	פרמטר
גברים	428.91	467.33	t(414) = 0.095, p > 0.05	Z=-1.436, p > 0.05	מספר המילים
נשים	418.22	449.60			
גברים	93.00	71.39	t(414) = 0.486, p > 0.05	Z=-0.803, p > 0.05	מספר קישורים פנימיים
נשים	90.33	71.18			
גברים	13.33	13.32	t(414) = 1.343, p > 0.05	Z=-1.273, p > 0.05	מספר קישורים חיצוניים
נשים	17.67	14.90			
גברים	2.37	5.44	t(17) = 1.306, p > 0.05	Z=-1.099, p > 0.05	

Mann-Whitney	ערך t	SD	M	מגדר	פרמטר
		10.10	5.50	נשים	מספר שפות נוספות
Z=-0.246, p > 0.05	t(414) = 0.443, p > 0.05	5.72	8.88	גברים	מספר קטגוריות
		4.57	8.28	נשים	
Z=-0.488, p > 0.05	t(414) = 0.405, p > 0.05	70.61	45.37	גברים	מספר הפניות לרשומה
		49.00	38.56	נשים	
*Z=-2.532, p<0.05	t(414) = 1.307, p > 0.05	16.39	13.64	גברים	מספר עריכות (כפונקציה של הזמן)
		14.37	19.83	נשים	

מקרא:

- p < 0.05 .

ניתוח הנתונים בטבלה 9

לבדיקת ההשערה כי קיים הבדל בין גברים ונשים בעריכות של רשומות, נערך מבחן Mann-Whitney למדגמים שאינם מתפלגים התפלגות נורמלית. נמצא כי קיים הבדל מובהק בין הממוצעים (Z = -2.532, p < 0.05) – מספר העריכות של רשומות גברים (M = 13.64, SD = 16.39) נמוך ממספר העריכות של רשומות נשים (M = 19.83, SD = 14.37). בשאר המדדים הטקסטואליים של הרשומות – מספר המילים, הקישורים הפנימיים, הקישורים החיצוניים, ההפניות, השפות והקטגוריות – לא נמצאו הבדלים מובהקים בין נשים וגברים. על מנת לבסס את התוצאות שקיבלנו, בדקנו אם המדדים הכמותיים של הרשומות שבדקנו מצויים במתאם זה עם זה. התוצאות מוצגות בטבלה 10.

פרמטר	מספר מילים	קישורים פנימיים	קישורים חיצוניים	שפות	קטגוריות	הפניות	עריכות
מספר מילים	1	***0.705	***0.679	***0.557	***0.459	***0.610	***0.461
קישורים פנימיים		1	***0.741	***0.627	***0.775	***0.903	***0.410
קישורים חיצוניים			1	***0.688	***0.580	***0.656	***0.541
שפות				1	***0.583	***0.676	***0.400
קטגוריות					1	***0.749	***0.232
הפניות						1	***0.405

מקרא:

*** $p < 0.001$

מטבלה 10 עולה כי כלל המדדים מקיימים מתאם מובהק וחיובי זה עם זה. בכלל המדדים המתאם בינוני עד חזק מאוד, למעט המדד של מספר העריכות, שנמצא במתאם חלש יחסית ($R < 0.5$) עם כל אחד מהמדדים הטקטסואליים האחרים.

5.12. רואי חשבון

במהלך המחקר, נדגמו רשומות של 47 רואי חשבון באמצעות חילוף שמות הרשומות השייכות לקטגוריית ויקיפדיה "רואי חשבון בישראל", או "רואות חשבון בישראל". 41 (87%) מתוך הרשומות הללו היו של גברים ו-6 (13%) של נשים.

5.12.1. סטייה מרמת הייצוג

לצורך בחינת שאלת המחקר הראשונה, בדקנו אם היחס בין נשים וגברים ברשומות ויקיפדיה סוטה באופן מובהק מהיחס הקיים לפי נתוני העולם האמיתי.

לפי נתוני הבסיס שבידינו, בשנים 1960–2009 היו בישראל 19,118 בעלי רישיון לעסוק בראיית חשבון. מתוכם 12,896 (67%) גברים ו-6,222 (33%) נשים.

על בסיס נתונים אלה בוצע מבחן χ^2 כדי לבחון את הקשר בין המגדר ובין השיעור באוכלוסיית הבסיס ובאוכלוסיית המדגם – ויקיפדיה. תוצאת המבחן מוצגת להלן:

$$\chi^2(1, N=47) = 8.3616, p\text{-value} < 0.05$$

נמצא שלמשתנה המגדר יש קשר לשיעור הנוכחות ברשומות הויקיפדיה. נמצא שמשתנה המגדר קשור באופן מובהק למידת הנוכחות ברשומות ויקיפדיה – רשומות של גברים יוצגו יותר בוויקיפדיה, בהתחשב בשיעור הייצוג שלהם בקטגוריית רואי החשבון, ורשומות של נשים יוצגו פחות בוויקיפדיה, בהתחשב בשיעור הייצוג שלהן בקטגוריית רואי החשבון.

5.12.2. בחינת עומק של תוכן הרשומות

לשם בחינת שאלת המחקר השנייה בוצע מבחן t להשוואת ממוצעי המדדים התוכניים של הרשומות. תוצאות המבחן מוצגות בטבלה 11.

טבלה 11 – השוואת מדדים טקסטואליים ברשומות בחתך מגדר – קטגוריית רואי חשבון

פרמטר	מגדר	M	SD	ערך t	Mann-Whitney
מספר מילים	גברים	459.71	455.21	t(45) = 0.398, p > 0.05	Z=-1.020, p > 0.05
	נשים	381.67	388.81		
קישורים פנימיים	גברים	79.22	58.03	t(45) = 0.776, p > 0.05	Z=-0.367, p > 0.05
	נשים	60.50	21.60		
קישורים חיצוניים	גברים	14.29	15.11	t(45) = 0.916, p > 0.05	Z=-1.006, p > 0.05
	נשים	20.67	21.26		
שפות נוספות	גברים	2.29	7.91	t(45) = 0.6, p > 0.05	Z=-0.312, p > 0.05
	נשים	0.33	0.82		
הפניות	גברים	27.76	49.29	t(45) = 0.791, p > 0.05	Z=-0.128, p > 0.05
	נשים	11.67	6.41		
קטגוריות	גברים	7.17	4.78	t(45) = 0.249, p > 0.05	Z=-0.787, p > 0.05
	נשים	7.67	2.07		
עריכות (כפונקציה של הזמן)	גברים	18.42	22.35	t(45) = 0.323, p > 0.05	Z=-0.988, p > 0.05
	נשים	15.41	7.52		

מהנתונים בטבלה 11 עולה שלא נמצא הבדל מובהק בין גברים לנשים בכלל המדדים הטקסטואליים שנבדקו.

על מנת לבסס את התוצאות שקיבלנו, בדקנו אם המדדים הכמותיים של הרשומות שבדקנו מצויים במתאם זה עם זה, כלומר האם המדדים הטקסטואליים השונים שנבדקו עבור הרשומות בודקים את אותו עולם תוכן, וממילא מהימנים יותר במתן מענה לשאלת המחקר. התוצאות מוצגות בטבלה 12.

טבלה 12 – מתאם בין מדדים טקסטואליים של רשומות – קטגוריית רואי חשבון

פרמטר	מספר מילים	קישורים פנימיים	קישורים חיצוניים	שפות	קטגוריות	הפניות	עריכות
מספר מילים	1	***0.691	***0.809	***0.541	***0.647	***0.732	***0.552
קישורים פנימיים		1	***0.680	***0.470	***0.740	***0.913	***0.645
קישורים חיצוניים			1	***0.574	***0.624	***0.700	***0.585
שפות				1	***0.494	***0.555	***0.227
קטגוריות					1	***0.767	***0.423
הפניות						1	***0.656

מקרא:

$***p < 0.001$

מטבלה 12 עולה כי כלל המדדים מקיימים קורלציה מובהקת וחיובית זה עם זה. בכלל המדדים המתאם בינוני עד חזק מאוד, למעט המתאם בין מספר השפות ומספר הקטגוריות, ולמעט המדדים של מספר העריכות ומספר השפות, שנמצאים במתאם חלש יחסית עם כל אחד מהמדדים הטקסטואליים האחרים (R קרוב ל-0.5 או נמוך מכך).

6. דיון

במהלך המחקר נבדקו חמש קטגוריות של בעלי ביוגרפיות, כלומר ערכי ויקיפדיה המתייחסים לאישים: עורכי דין, קציני צבא, דירקטורים, ראשי מועצות ורואי חשבון.

בהתייחס לשאלת המחקר הראשונה, בכל אחת מהקטגוריות (למעט קציני הצבא) נבדק ההבדל בין שיעור הייצוג המגדרי ברשומות ויקיפדיה לעומת שיעור הייצוג המגדרי בנתוני בסיס שהתקבלו ממקורות חיצוניים, על מנת לבדוק אם היחס המגדרי (Gender Balance) בויקיפדיה משקף את היחס המגדרי הקיים במציאות.

מהנתונים עולה שמתוך ארבע הקטגוריות שבהן נבדקה שאלה זו, בשלוש קטגוריות – עורכי דין, דירקטורים ורואי חשבון – נמצא הבדל מובהק בין היחס המגדרי במציאות ליחס המגדרי כפי שהוא נמצא בויקיפדיה, ובכל המקרים מדובר היה בהבדל המבטא תת-ייצוג לנשים לעומת ייצוג-יתר לגברים. על משמעותו של ממצא זה נרחיב להלן.

על מנת לענות לשאלת המחקר השנייה, בכל אחת מהקטגוריות נבדקו שבעה מדדים טקסטואליים המתאייחים אל תוכן הרשומה הביוגרפית: מספר המילים ברשומה, מספר הקישורים הפנימיים ברשומה, מספר הקישורים החיצוניים ברשומה, מספר הקטגוריות שאליהן משתייכת הרשומה, מספר השפות שבהן קיימת הרשומה, מספר ההפניות לרשומה ומספר העריכות שנערכה הרשומה ביחס לזמן שהיא קיימת. עבור כל אחד ממדדים אלה הושוּו הממוצעים של הרשומות הנוגעות לגברים ולנשים.

מהנתונים שהתקבלו עולה מסקנה מעורבת.

באחת הקטגוריות, קטגוריית רואי החשבון, לא נמצאו כל ראיות להבדלים מגדריים במדדים הטקסטואליים של הרשומות. בקטגוריה נוספת, קטגוריית ראשי המועצות, התגלה פרמטר אחד בלבד מתוך שבעה שבו נצפו הבדלים מובהקים. עם זאת, בשלוש הקטגוריות האחרות שנבחנו, התגלו 2–5 מדדים שבהם נמצא הבדל מובהק במדדים בין המגדרים: קציני צבא, עורכי דין, דירקטורים.

ניתוח הממצאים דלעיל מעלה מסקנה מיידיית אחת: אין ראיות להטיה מגדרית מתמשכת, עקבית ויציבה בויקיפדיה, אלא ניתן לדבר על "כיסים" של הטיה במדדים הרלוונטיים, וכולם לרעת נשים. מצד אחד, בהתייחס לשאלת המחקר הראשונה התקבל ממצא ברור המרמז על הטיה. בכלל הקטגוריות שנבדקו למעט אחת, בחינה של הייצוג המגדרי לעומת נתוני הבסיס העלתה כי נשים זוכות לייצוג חסר מבחינת מספר הרשומות בויקיפדיה לעומת השיעור שלהן באוכלוסיית הבסיס. יוצאת מכלל זה קטגוריית ראשי הערים, שבה לא נמצאו ראיות להטיה. עם זאת, יש לזכור שנתוני הבסיס בקטגוריה זו – רשימת ראשי המועצות המכהנים נכון לשנת 2019 (אבגר, 2019) – היו דלים יחסית, ולא הכילו מדגם גדול מספיק של נתוני בסיס להשוואה, ועל רקע עובדה זו יש לבחון את התוצאות. השלכותיו האפשריות של ממצא זה יידונו בפירוט בפרק הבא.

מצד שני, בהתייחס לשאלת המחקר השנייה, בחינת העומק של תוכן הרשומות, הממצאים אמביוולנטיים ומטיבם נתונים לפרשנות. ככלל, המדדים הטקסטואליים שנבחנו הציגו סטיית תקן (SD) גדולה מאוד, כפי שניתן לראות בטבלאות המפורטות מעלה (טבלה 3, טבלה 5, טבלה 7, טבלה 9, טבלה 11). נתון זה, המצביע על שונות גבוהה מאוד בין נתוני הרשומות השונים, הקשה על חילוץ מסקנה מובהקת מבחינה סטטיסטית. כמו כן, בחלק מהקטגוריות היה מספר הרשומות של נשים קטן מאוד – בקטגוריית רואי החשבון 6 בלבד ובקטגוריית קציני הצבא 27 רשומות נשים מתוך 1,440 רשומות – כך שקשה היה לבסס ממצא סטטיסטי מהימן על השוואה בין מדדי הקבוצות.

עם זאת, מעיון בממצאים ניתן לזהות מגמות מובחנות לעניין קיומה של הטיה.

בקטגוריית הדירקטורים נמצאו 5 מדדים טקסטואליים (מתוך 7 מדדים שנבחנו) שבהם נמצא שוני מובהק סטטיסטי בין תוכני הערכים בחיתוך על פי מגדר. בקטגוריית עורכי הדין נמצאו 3 מדדים

טקסטואליים שמגלים הטיה מגדרית; בקטגוריית אנשי הצבא נמצאו 3 מדדים טקסטואליים שמגלים הטיה מגדרית; בקטגוריית ראשי המועצות נמצא מדד טקסטואלי אחד שמגלה הטיה מגדרית, והוא מספר העריכות של הרשומה. המדדים שבהם נמצאה הטיה במרבית הפעמים היו מספר העריכות של הרשומה (נמצאה הטיה ב-3 מתוך 5 קבוצות שנבדקו), מספר הקישורים הפנימיים ברשומה (2), מספר ההפניות אל הרשומה (2), מספר הקטגוריות (2) ומספר השפות (2). במדד מספר הקישורים החיצוניים של הרשומה ומדד מספר המילים ברשומה לא התגלו ראיות מובהקות להטיה מגדרית באף אחת מהקטגוריות שנבחנו. מנתונים אלה ניתן לגזור מספר מסקנות רוחב.

6.1. הטיה מגדרית במספר העריכות

ב-3 מתוך 5 הקטגוריות שנבדקו, מספר העריכות של רשומות של נשים גבוה ממספר העריכות של רשומות של גברים באופן מובהק מבחינה סטטיסטית. בהתאם להשערות המחקר שניסחנו, ניתן לומר שערכי נשים שנויים יותר במחלוקת ומקבלים תשומת לב רבה יותר מעורכים בהשוואה לערכים השייכים לגברים. מסקנה זו מתחזקת מבחינה של מדדי הרשומות האחרים. בכלל הקטגוריות שנבחנו, גם במקומות שבהן רשומות נשים נערכו באופן מובהק יותר מרשומות של גברים, התגלו מדדים אחרים המעידים על נפח הרשומה, אשר היו נמוכים באופן מובהק אצל הנשים בהשוואה לאלה של הגברים (קישורים פנימיים, שפות, הפניות, קטגוריות). הניגוד בין ערכי מדד העריכות ובין ערכי יתר המדדים שבהם נמצאו הבדלים מובהקים התבטא גם בערכי המתאם בין מדד העריכות ובין יתר המדדים הטקסטואליים, שהיו נמוכים ביחס למתאמים בין המדדים הטקסטואליים האחרים ובין עצמם. הדבר מעיד שהעריכות הנוספות כשלעצמן אינן מגדילות את נפח הרשומה. כיוון שכך, יש מקום להניח שחלקן הגדול הוא עריכות של שינוי או הסרת תוכן מהרשומה.

6.2. מבחנים שונים – הטיות שונות

מסקנה נוספת היא שזיהוי הטיה מגדרית בנוגע ליחס המגדרי שאינו תואם את המציאות אינו קשור בהטיה מגדרית כפי שבאה לידי ביטוי בתוכן הרשומות. לדוגמה, בקטגוריית רואי החשבון, שבה נמצאה הטיה מובהקת מבחינת הפער בין היחס המגדרי במדגם של ויקיפדיה ובין היחס המגדרי בנתוני האמת, לא נמצאה כל הטיה מגדרית במדדים הטקסטואליים שנבדקו. גם בקטגוריית עורכי דין, העובדה שנמצאה הטיה מובהקת בנושא היחס המגדרי לעומת נתוני הבסיס, אין משמעה מציאת הטיה מגדרית בנתונים הטקסטואליים של הרשומה עצמה. בפועל בקטגוריית עורכי הדין נמצאו 3 מדדים (מתוך 7 שנבדקו) שבהם נמצאה הטיה מגדרית מובהקת. לעניין זה יצוין שהממצאים בקטגוריית עורכי הדין נהנים ממהימנות גבוהה יחסית, כיוון שבקטגוריה זו גם אוכלוסיית המדגם וגם אוכלוסיית הבסיס היו גדולות, בשני המגדרים, ואפשרו חילוץ מסקנות תקפות מבחינה סטטיסטית. המסקנה הנובעת מכך היא שיהא אשר יהא מקורן של הטיות אלה, הוא לא בהכרח זהה. העובדה שנשים, מסיבה זו או אחרת (והדברים יידונו בפרק הבא) אינן מיוצגות לפי שיעורן באוכלוסיית הבסיס גם ברשומות ויקיפדיה, אין משמעה בהכרח שהערכים שלהן סובלים הזנחה או עריכת יתר ביחס לרשומות של גברים.

6.3. מתאם גבוה בין המדדים הטקסטואליים

כפי שעולה מהטבלאות בפרק התוצאות, בכלל הקטגוריות שנבדקו נמצאו מתאמים מובהקים בין שבעת המדדים הטקסטואליים שנבחנו ובין עצמם. ככלל, המתאמים היו בינוניים עד גבוהים, ובחלקם אף הגיעו לשיעור של 90%.

בהתאם לכך, ניתן לראות שבקטגוריות שבהן נמצאו הטיית מובהקות, הן בדרך כלל נמצאו בכמה מדדים יחד – ראו קטגוריית הדירקטורים, קטגוריית עורכי הדין. מנגד, נמצאו קטגוריות שבהן כלל המדדים הראו על העדר הטייה מובהקת – קטגוריית רואי חשבון וקטגוריית ראשי מועצה.

מתאמים אלה חשובים, כיוון שיש בהם כדי להעיד שכלל המדדים הטקסטואליים שנבדקו בודקים עולם תוכן אחד, עולם תוכן שכונה "נפח הרשומה". באופן תיאורטי, אין הכרח שרשומות ארוכות (כלומר, שמספר המילים בהן גבוה) יהיו גם רשומות שמתורגמות למספר רב של שפות. בדומה לכך, אין סיבה להניח שרשומות שמקושרות לרשומות מועטות, יהיו גם רשומות שמשויכות לקטגוריות ויקיפדיה מועטות. אולם בפועל, כפי שמלמדים ערכי המתאמים בכלל הקטגוריות שנבחנו, קיים קשר סטטיסטי חזק בין כלל המדדים הטקסטואליים שנבחנו (למעט מדד מספר העריכות, אשר כפי שפורט לעיל, ערכים גבוהים שלו נושאים משמעות אמביוולנטית לעניין נפח הרשומה). יש בכך כדי להראות כי מדידה אפקטיבית של "נפח" או "איכות" של רשומה צריכה להיגזר משקלול של מספר היבטים כמותיים של הרשומה.

6.4. אפיון ההבדלים בין הקטגוריות

כפי שפורט לעיל, בחלק מהקטגוריות שנבדקו נמצאו ראיות רבות יותר להטייה מגדרית ובקטגוריות אחרות נמצאו ראיות מעטות יחסית או שלא נמצאו ראיות כלל. האם ניתן להסביר הבדלים אלה בהסתמך על הקטגוריות עצמן?

הקטגוריות שבהן נמצאו ההבדלים המגדריים המשמעותיים ביותר מבחינת נפח הן קטגוריות עורכי הדין, קציני הצבא והדירקטורים, ואילו בקטגוריות של רואי החשבון וראשי המועצות לא נמצאו ראיות רבות להטייה.

הסבר אפשרי אחד להבדלים אלה הוא גודל המדגמים השונה – בקטגוריית רואי החשבון הדבר בולט במיוחד.

הסבר אפשרי נוסף נעוץ במהותן של הקטגוריות שנבחרו. כזכור, הקטגוריות נבחרו במתכוון כקטגוריות שאין בהן כשלעצמן כדי לחייב יצירה של רשומה בויקיפדיה על אדם המחזיק בהן. אולם, יש רגליים לסברה שעורכי דין וקציני צבא אינם נמצאים בויקיפדיה בהכרח בגלל הכשרתם ככאלה, אלא בגלל פועלם בתחומים אחרים. לדוגמה, רבים מחברי הכנסת בישראל מחזיקים בדרגת קצונה צבאית (Barak & Tsur, 2012) ורבים מהמחזיקים בתואר עורך דין עושים זאת במסגרת עיסוק בניהול (Zer-Gutman, 2012). ייתכן שדפוס פעילות רב-תחומי או בין-תחומי, שמאפיין חברי חלק מהקטגוריות שנבחרו ואינו מאפיין חברי קטגוריות אחרות, ישפיע גם על היבטים של הטייה מגדרית. ראוי לציין כי גם במחקר של וגנר ואחרים (Wagner et al., 2016) הרשומות הביוגרפיות בויקיפדיה מסווגות לביוגרפיות של "כוכבי העל" (Superstars) ולביוגרפיות של "כוכבים מקומיים" (Local Heroes), זאת לפי מספר השפות שבהן מופיעה הרשומה. כך נמצאו יותר ראיות להטייה מגדרית בקרב ה"כוכבים המקומיים" לעומת אצל "כוכבי העל" בעלי הפרופיל הבולט יחסית. עם זאת, במחקר הנוכחי אם נסווג את קטגוריות עורכי הדין, אנשי הצבא והדירקטורים כ"כוכבי על", שיש להם פעילות במספר תחומים הבאים לידי ביטוי ברשומות

ויקיפדיה, ואת קטגוריות רואי החשבון וראשי המועצות כ"כוכבים מקומיים", שפעילותם מתרכזת בשדה אחד שמתועד ברשומות ויקיפדיה, הממצאים העולים מהמחקר כאן עומדים בניגוד לממצאים של וגנר ואחרים (Wagner et al., 2016). בעבודה זו נמצאו ראיות רבות יותר להטיה מגדרית בביוגרפיות של "כוכבי העל" לעומת אצל ה"כוכבים המקומיים".

7. סיכום ומסקנות

מהי משמעותם של ממצאינו בדבר הטיה מגדרית? אפשר לתאר קשת של פרשנויות הנמצאות על מעין ציר אידיאולוגי, שמשפיע על פרשנות הנתונים שהתקבלו.

בקוטב האחד תהא פרשנות הרואה בממצאים שהתקבלו עדות לכך שמסיבות שונות נשים זכות לייצוג לא הוגן, הן מבחינת עצם הנוכחות שלהן במרחב הויקיפדי והן מבחינת אופן הצגת פועלן ברשומות של ויקיפדיה. לפי פרשנות זו, התוכן הויקיפדי עושה עוול לנוכחותן המציאותית של נשים במרחב, משום שהוא מציג תמונה המוטה באופן עקבי לרעתן ונכשל בהבאת מידע "עיוור מגדרית" בנושאים שבהם ויקיפדיה מתיימרת לעסוק (Halfaker, 2017). לצורכי נוחות בלבד נכנה פרשנות זו פרשנות ליברלית.

לפי הגישה הליברלית, יש לברך על פרויקטים ייעודיים שונים להגדלת שיעורן של נשים במרחב הויקיפדי ולשיפור הערכים של נשים, מבחינת נפח, תוכן ועומק (גבע, 2013; Lir, 2019).

בקוטב האחר ניצבת פרשנות שאפשר לכנותה לשם הנוחות פרשנות שמרנית. לפי פרשנות זו, הבדלי ייצוג וכן הבדלי נוסחים ונפח ערכים בין המגדרים מעידים כי אכן במציאות שויקיפדיה מתיימרת לייצג – המציאות האנציקלופדית, של בולטות מבחינת הישגים, תרומה לחברה, תפקיד היסטורי או ערכים דומים שמקור אנציקלופדי מתיימר לתעד – נשים תופסות מקום משני. לפי פרשנות זו, לדוגמה, גם אם מבחינה מספרית נשים מצויות בשיעורים נכבדים בין עורכי הדין, הרי שלדרגות הגבוהות של השפעה, הישגים או מעמד מקצועי בפרופסיה הן אינן מגיעות. לפי גישה זו, מציאות זו תשתקף גם בערכי ויקיפדיה. לפיכך פרויקטים שמטרתם שיפור היחס המגדרי, הן בכמות הערכים בוויקיפדיה והן בתוכנם, הם ניסיון מלאכותי ליצור תדמית של ערך אנציקלופדי לרשומות נשים במקום שאינו קיים במציאות, ומכאן גם השיעורים הגבוהים של עריכות של רשומות המתייחסות לנשים.

תומך הגישה השמרנית יכול שלא לייחס כל ערך לממצאים שהתקבלו במחקר זה ביחס לשאלת המחקר הראשונה, שכן לדידו, גם אם שיעורן של נשים בקרב עורכי הדין בוויקיפדיה, לדוגמה, נופל בהרבה משיעורן באוכלוסייה, אין הדבר מצביע על הטיה. ייתכן, יאמר תומך הגישה השמרנית, שמצויות עורכות דין רבות שלא עשו או השיגו דבר שמצדיק את הכנסתן לוויקיפדיה, ואילו עורכי הדין שמגיעים בשיעור גבוה יותר להישגים, כגון הצלחה ופרסום במקצועם או כהונה כשופטים או כחברי כנסת, הינם בעיקר גברים.

תומכי הגישה השמרנית יכולים ליטול את הנתונים שעלו בעבודה זו כתמיכה לגישתם: לא נמצאו בעבודה זו עדויות להטיה שיטתית, עקבית או יציבה בתוכן הרשומות ויקיפדיה, והעדרן של נשים שנבחרו על סמך קריטריון שאינו קריטריון ויקיפדי לקבלת רשומה (לדוגמה, עורכות דין או דירקטוריות) אינו מהווה ראיה לכך שנשים שראויות לכך, לא קיבלו את רשומתן במאגר המידע.

מנגד, גם תומכי הגישה הליברלית יכולים להסתייע בממצאים שעלו כאן כאסמכתא לגישתם – שכן קשה להצדיק את שלל המדדים שבהם נמצאה הטיה (ובכולם, לרעת הערכים של נשים) ולפטור אותם כטעות סטטיסטית.

שתי הגישות הללו הן גישות קוטביות, ובין הגישה השמרנית לגישה הליברלית יש מרחב של פרשנויות ביניים לנתונים על הטיה. נדרשים מחקר נוסף ונתונים נוספים על מנת לבסס או להפריך כל אחת משתי גישות אלה. לדוגמה, מחקר מפורט על השיקולים של עורכי ויקיפדיה בהכנסת תוכן לאתר, עריכתו או מחיקתו; חקירת המיזמים השונים של יצירת ערכים במטרה לשנות את האיזון המגדרי; מחקר על נטייתם של עורכים שונים לעסוק בעריכה בהתאם למגדר של האישייות שהיא נושא הרשומה; ועוד.

עם זאת, דומה שאפשר לחלץ, בזהירות הראויה, מסקנה אחת והשערה אחת שיש ביניהן קשר. מסקנה אחת, כמעט מובנת מאליה – ישנם הבדלים מובהקים בייצוג של נשים וגברים בוויקיפדיה, והבדלים אלה כמעט כולם נוטים לרעת הרשומות המתעדות נשים. מקורם של הבדלים אלה, כאמור, שנוי במחלוקת וחורג ממסגרתה של עבודה זו. אך הממצא נותר בעינו – אדם מן השורה המחפש מידע ראשוני ולוחץ לפי תומו על היפר-קישור לאתר ויקיפדיה העברית ימצא שם מידע אשר באופן סטטיסטי מוטה מבחינה מגדרית. סביר יותר שיתרשם שבתחומים מסוימים אין פעילות של נשים, או שיתרשם מדלות התוכן המאפיינת את תוכן של רשומות המתארות ביוגרפיות של נשים. ממצא זה לבדו מחייב ביקורתיות וזהירות בשימוש בוויקיפדיה כמקור מידע. ממצאות זה אפשר לשער כי לאור המשקל הגדול שיש לרשומות ויקיפדיה בקרב מחפשי המידע הראשוני במרחב המקוון, ויקיפדיה לא רק מהווה ייצוג לא מאוזן מבחינה מגדרית, אלא אף תורמת לחוסר איזון זה במעין מעגל שוטה, בדומה לדינמיקה המתוארת, בהקשר אחר, אצל ליר (Lir, 2019).

במה דברים אמורים? לשם הבהרת הדברים נשתמש בדוגמאות: ועדת פרס כלשהי מעוניינת לחלק פרס בתחום מדעי; תחקירנית רדיו מעוניינת להעלות לשידור מומחה בתחום מקצועי; מנהל בכיר מקבל קורות חיים של מועמד לעבודה המציג את עצמו כמומחה בתחומו.

אם כל אחד ממקבלי ההחלטות האלה נעזר בוויקיפדיה כדי לחפש מידע ביוגרפי, קיים סיכוי לא מבוטל שהוא יקבל מידע מוטה – ואם מדובר במועמדת, תהא זו הטיה לרעה. הוא ימצא פחות רשומות של נשים העוסקות בתחום ביחס לשיעור הנשים שקיים במציאות בתחום זה (כפי שמצאנו בתשובה לשאלת המחקר הראשונה) וגם הרשומות שאליהן יגיע יהיו מקושרות פחות לרשומות אחרות, חסרות שיוכים לקטגוריות ויקיפדיה רלוונטיות וכתובות בפחות שפות (כפי שמצאנו בתשובה לשאלת המחקר השנייה). ממילא, גדול יותר הסיכוי שהפרס יינתן למועמד ולא למועמדת, הוא יוזמן כאורח לראיונות ופאנלים מקצועיים או יקבל את המשרה הבכירה והנחשקת. וכל אחד מאלה, בתורו, יביא לשדרוג ערך הויקיפדיה שלו בהתאם. בצורה זו הפער המגדרי בוויקיפדיה הופך למעין נבואה המגשימה את עצמה בהיזון חוזר. דינמיקה זו יכולה להסביר את הממצא המחקרי שלפיו הטיה בערכי ויקיפדיה פוחתת ככל שהאישייות מפורסמת, נודעת, או בולטת יותר (Konieczny & Klein, 2018; Wagner et al., 2016).

לפי השערה זו, דווקא מעמדה הנוכחי של ויקיפדיה כמקור מידע הופך אותה לצומת נוסף ברשת של מקורות מידע המזינים זה את זה ומשפיעים זה על זה, וכך מי שמחזיק במשאבים מספיקים כדי

[7%D7%A7%D7%A4%D7%A4%D7%A4&oldid=37116077.](https://www.wikipedia.org/w/index.php?title=%D7%A7%D7%A4%D7%A4%D7%A4&oldid=37116077)

ויקיפדיה העברית. (2023, ספטמבר 21). ויקיפדיה.

<https://he.wikipedia.org/w/index.php?title=%D7%A7%D7%A4%D7%A4%D7%A4&oldid=37144235>

לין, ד' (2015). קריאת התגייסות למהפכת הידע שמחוללת 'ויקיפדיה'. בתוך ז' מן (עורך). עלי זית וחרב, כרך ט"ו: סלע המחלוקת (11–22). הוצאת כרמל, העמותה לחקר כוח המגן ע"ש ישראל גלילי.

מזרחי, ש' (2013). נתונים על נשים בדירקטוריונים של חברות ממשלתיות וציבוריות. הכנסת. מרכז

המחקר והמידע של הכנסת. https://fs.knesset.gov.il/globaldocs/MMM/e2ef6d8d-f1f7-e411-80c8-00155d01107c/2_e2ef6d8d-f1f7-e411-80c8-00155d01107c_11_9948.pdf

שחק, מי (2017). נתונים על עורכי דין ורואי חשבון בפילוח לפי מגדר וקבוצות אוכלוסייה. הכנסת. מרכז המחקר והמידע של הכנסת.

https://fs.knesset.gov.il/globaldocs/MMM/7cef995d-0c03-e711-80cc-00155d0206a2/2_7cef995d-0c03-e711-80cc-00155d0206a2_11_8452.pdf

Asadi, S., Ghafghazi, S., & Jamali, H. R. (2013). Motivating and discouraging factors for Wikipedians: The case study of Persian Wikipedia. *Library Review*, 62(4/5), 237-252. <https://doi.org/10.1108/LR-10-2012-0114>

Barak, O., & Tsur, E. (2012). The military careers and second careers of Israel's military elite. *The Middle East Journal*, 66(3), 473-492. <https://doi.org/10.1353/mej.2012.0083>

Falenska, A., & Çetinoğlu, Ö. (2021). Assessing gender bias in Wikipedia: Inequalities in article titles. In M. Costa-jussa, H. Gonen, C. Hardmeier, & K. Webster (Eds.). *Proceedings of the 3rd Workshop on Gender Bias in Natural Language Processing* (pp. 75-85). <https://doi.org/10.18653/v1/2021.gebnlp-1.9>

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, 438(15), 900-901. <https://doi.org/10.1038/438900a>

Graells-Garrido, E., Lalmas, M., & Menczer, F. (2015). First women, second sex: Gender bias in Wikipedia. In *Proceedings of the 26th ACM Conference on Hypertext & social media* (pp. 165-174). Association for Computing Machinery. <https://doi.org/10.1145/2700171.2791036>

Halfaker, A. (2017). Interpolating quality dynamics in Wikipedia and demonstrating the Keilana effect. In *Proceedings of the 13th International Symposium on Open*

- Collaboration* (pp. 1-9). Association for Computing Machinery.
<https://doi.org/10.1145/3125433.3125475>
- Helgeson, B. (2015). The Swedish Wikipedia gender gap. [Master's Thesis, KTH, School of Computer Science and Communication, Media Technology and Interaction Design].
- Hinnosaar, M. (2019). Gender inequality in new media: Evidence from Wikipedia. *Journal of Economic Behavior & Organization*, 163, 262-276.
<https://doi.org/10.1016/j.jebo.2019.04.020>
- Hube, C., & Fetahu, B. (2018). Detecting biased statements in Wikipedia. In *Companion Proceedings of the The Web Conference 2018* (pp. 1779-1786). International World Wide Web Conferences Steering Committee.
<https://doi.org/10.1145/3184558.3191640>
- Jemielniak, D. (2016). Breaking the glass ceiling on Wikipedia. *Feminist Review*, 113(1), 103-108.
- Kandek, B., (2023, March 1) Closing the gender gap: Women in Red's efforts to add more women to Wikipedia [Web Blog message]. Retrieved October 3, 2023 from <https://wikimediafoundation.org/news/2023/03/01/closing-the-gender-gap-women-in-reds-efforts-to-add-more-women-to-wikipedia/>
- Konieczny, P., & Klein, M. (2018). Gender gap through time and space: A journey through Wikipedia biographies via the Wikidata Human Gender Indicator. *New Media & Society*, 20(12), 4608-4633. <https://doi.org/10.1177/1461444818779080>
- Lir, S. A. (2019). Strangers in a seemingly open-to-all website: The gender bias in Wikipedia. *Equality, Diversity and Inclusion: An International Journal*, 40(7), 801-818. <https://doi.org/10.1108/EDI-10-2018-0198>
- Minguillón, J., Meneses, J., Aibar, E., Ferran-Ferrer, N., & Fàbregues, S. (2021). Exploring the gender gap in the Spanish Wikipedia: Differences in engagement and editing practices. *PloS ONE*, 16(2), e0246702.
<https://doi.org/10.1371/journal.pone.0246702>
- Pitoura, E., Tsaparas, P., Flouris, G., Fundulaki, I., Papadakos, P., Abiteboul, S., & Weikum, G. (2018). On measuring bias in online information. *ACM SIGMOD Record*, 46(4), 16-21. <https://doi.org/10.1145/3186549.3186553>
- Protonotarios, I., Sarimpei, V., & Otterbacher, J. (2016). Similar gaps, different origins? Women readers and editors at Greek Wikipedia. *Proceedings of the*

- International AAAI Conference on Web and Social Media*, 10(2), 80-87.
<https://doi.org/10.1609/icwsm.v10i2.14827>
- Top Websites Ranking*. (2023). Similarweb. retrieved October 3, 2023 from:
<https://www.similarweb.com/top-websites/>
- Sichler, A., & Prommer, E. (2014). Gender differences within the German-language Wikipedia. *ESSACHESS – Journal for Communication Studies*, 7(2 (14)), 77-93.
- Wadewitz, A. (2013). Wikipedia is pushing the boundaries of scholarly practice but the gender gap must be addressed. [Blog post]. *Impact of Social Sciences Blog*.
<http://blogs.lse.ac.uk/impactofsocialsciences>
- Wagner, C., Graells-Garrido, E., Garcia, D., & Menczer, F. (2016). Women through the glass ceiling: Gender asymmetries in Wikipedia. *EPJ Data Science*, 5(1), 5.
<https://doi.org/10.1140/epjds/s13688-016-0066-4>
- Wikimedia Foundation. (n.d). *About*. Retrieved October 3, 2023 from:
<https://wikimediafoundation.org/about/>
- Zer-Gutman, L. (2012). Effects of the acceleration in the number of lawyers in Israel. *International Journal of the Legal Profession*, 19(2-3), 247-263.
<https://doi.org/10.1080/09695958.2013.782308>