

פוקו והרשת הסמנטית

שי אופיר

בית הספר למדעי המתמטיקה ובית הספר למדעי המחשב, אוניברסיטת תל-אביב

אפיסטמה בהתהוות

בספרו המילים והדברים¹ מציג מישל פוקו "סדרי דברים" שונים, המאפיינים תקופות היסטוריות שונות. במונח "סדר דברים" מתכוון פוקו לסדירות הקיימות בין יסודות בשדה השיח, כגון היחסים בין האובייקטים בשיח ובין המסמן למסומן בעבור האובייקט הספציפי, נוהלי השיח (כגון הרשאות – למי מותר להגיד מה ומתי), ואופן השחזור וההפצה של אמירות שונות. נראה שכל מאפייני השיח הללו משתנים בצורה משמעותית במדיום החדש – האינטרנט. השיח הופך לאנונימי במהותו ומאפשר שימוש בזהויות שאולות. הוא שוויוני יותר ונגיש לכל, ומתבסס על הטקסט הכתוב ועל כלי שיח ממוחשבים ומוֹכָּנִים (structured), דבר המחייב כללי שיח שונים (כך למשל, אי-אפשר להתפרץ ולקטוע משתתפים אחרים). פוקו מבחין בסדרים שונים המאפיינים את תקופת הרנסנס, את התקופה הקלאסית ואת התקופה המודרנית, ויוצרים אפיסטמות שונות שתהום מושגית פעורה ביניהן. "מהפכת המידע" מבשרת שינויים רבים בתחום הנגישות למידע (נגיש לכל), הקניין (קוד פתוח), יצירות האמנות (השעתוק הטכני) והיווצרותו המואצת של הכפר הגלובלי. נוכח זאת עולה השאלה: האם אנו ניצבים בפתחה של אפיסטמה חדשה, כפי שהגדיר אותה פוקו?

האפיסטמה של הרנסנס מציבה במרכז את מושג הדמיון (resemblance). היחסים בין האובייקטים סובבים כולם סביב מושג הדמיון: דמיון מרחבי, דמיון איכותי, דמיון הנוצר על ידי האדם בתהליך של אנלוגיות ואמולציות (חיקוי, הדמיה). האפיסטמה של התקופה הקלאסית (מדון קישוט עד סוף המאה ה-18) עוסקת בהגדרת ההיררכיה בין זהויות ובאפיון ההבדלים (differences) ביניהן. בתקופה זו מתנהל תהליך של מתמטיזציה ומכניזציה של הטבע, וכל אובייקט ניתן למיפוי בסדר הדברים הכללי. יחסי הדמיון בין האובייקטים הופכים ליחסי ייצוג מורכבים הרבה יותר, מסוגים שונים ובעלי ממד של הסתברות. האפיסטמה השלישית שפוקו מנתח, המודרנית, עוברת מסדר יציב ומוגדר היטב של מעין טבלה של זהויות והבדלים לסדר דינמי, המתייחס לפונקציונליות של האובייקטים ולתפקידם במכלול. היחסים מוגדרים על ידי הכוחות הפועלים במרחב ולכן הם זמניים ויוצרים תמונת עולם

¹ Michel Foucault, 1973. *The Order of Things: An Archaeology of the Human Sciences*, New York:

.Vintage Books

דינמית ומשתנה תדיר. הייצוג אינו משמש עוד את האדם רק להתבוננות בעולם, אלא הופך למושא התבוננות בפני עצמו. השפה, שנחשבה לאמצעי ייצוג שקוף למשתמש, הופכת לבעיה ולמושא מרכזי של המחשבה (פוקו סבור כי קאנט, שחקר את תנאי האפשרות של הייצוג, הוא הדמות שתרמה יותר מכל למעבר אל המודרניות). מרחב הידע, שהיה הומוגני ויישם אותם כללי סדר בכל התחומים, מתפצל לשלושה מרחבים: הפורמלי (מתמטיקה, פיזיקה), האמפירי (ביולוגיה, כלכלה, פילולוגיה) והפילוסופי. האדם עצמו הופך למושא חקירה וכך נוצרים מדעי האדם (למשל פסיכולוגיה, סוציולוגיה וביקורת הספרות). כל מדעי האדם מתרכזים למעשה בניסיון להראות כיצד ייצוגים בתת-מודע משפיעים על פני השטח. לדידו של פוקו, הלינגוויסטיקה היא המדע התשתיתי של העתיד. בהיותה מסגרת פורמלית של מסמנים, היא תדחק לשוליים את העיסוק בייצוגים שבתת-מודע ותתרכז בכיוון הפורמליסטי הטהור. היא תשמש לפיכך מצע שיקשר את מדעי האדם לתבנית אחת. עם זאת, היא תרחיק את האדם מתפקידו המקורי כאמצעי הייצוג של המציאות, תהליך שיביא בסופו של דבר לצמצום תפקידו של האדם.

חזונו של פוקו לגבי הלינגוויסטיקה הוא החוט המקשר בין האפיסטמה המודרנית לזו החדשה. אפיסטמה זו אינה מוגדרת עדיין, ומוקדם מדי לאפיינה ולתת לה שמות ותארים. אפיסטמות מטבען מתאפיינות בראשית דרכן בעמימות. השאלה היא אם יש בכלל אפשרות להגדירה בשלמות מתוך עצמה, או רק מתוך האפיסטמות העוקבות.

סדר דברים חדש

בשנים האחרונות החלו ניסיונות להגדיר עולמות ידע על ידי מערכות פורמליות הקרויות אונטולוגיות – במשמעות לינגוויסטית של קטלוג וסידור כל ה"ישים" הקיימים בשפה, ולא במשמעות מטאפיזית. השאיפה היא להניח בסיס לניהול הכמות העצומה של הידע ההולך ונאגר. מאגר הידע הגדול ביותר כיום הוא כמובן הרשת, והניסיון המרכזי להגדרת אונטולוגיה כזו הוא פרויקט הקרוי Web Ontology (או "הרשת הסמנטית"), המתבצע בידי חלוץ ה-WWW טים ברנרס לי (Berners-Lee). הרעיון העומד בבסיסה של הרשת הסמנטית הוא הצמדת מאפיינים סמנטיים למידע, כך שמידע לא יהיה עוד טקסט גולמי בלבד, אלא יצביע גם על סיווג ועל קטלוגו. מהפכה זו תאפשר ניהול ידע מתקדם ויעיל לאין שיעור מזה המתבצע כיום. כך למשל, חיפוש במאגרי מידע יוכל לשלוף רק את המידע הרלוונטי, ולא להציף אותנו באלפי פיסות מידע המחייבות מיון ידני. המידע הסמנטי יאפשר שיתוף פעולה טוב יותר בין אנשים למחשבים, אך גם יצמצם את חלקו של האדם בפרשנות הטקסט ואולי אף את מרחב הפרשנות. הרשת הסמנטית תוכל לשמש מצע ליישומים אינטליגנטיים רבים שעד כה היו נחלת האדם בלבד. לדוגמה: תרגום אוטומטי משפה לשפה (מאחר שמשמעות הטקסט ברמת הסינטקטית והסמנטית תהיה חלק בלתי נפרד ממנו), תמצות אוטומטי של טקסטים וכריית מידע על פי קריטריונים מורכבים במיוחד.

ממד נוסף במהפכת הרשת הסמנטית הוא אחדות הייצוג של כל תחומי הידע. פיצול הידע שפוקו הצביע עליו באפיסטמה המודרנית ילך וייעלם תחת כנפי הרשת, שתאפשר מידע נגיש בכל התחומים ובכל עומק של פירוט. כיום, מנועי החיפוש ברשת ממפים רק את דפי האתרים, שהם ברוב המקרים מעטפת שטחית לתוכן ותו לא. חזון הרשת הסמנטית שואף להגיע לכל פיסת מידע. מנועי החיפוש הקיימים אינם חודרים לספריות, לכתבי-עת ולספרים. נתאר לעצמנו שכל כתב-עת מדעי וכל ספר שנכתב אי פעם יהיה נגיש לכל, וחלקי כל המשפטים בהם ימוינו על פי תפקידם התחבירי. זהו אינו חזון מופרך מבחינה טכנולוגית. למעשה, זוהי המטרה של פרויקט ארוך טווח של חברת IBM, הקרוי Web Fountain.

כדי לממש ייצוג אחיד לכל תחומי הידע, הגדירה הרשת הסמנטית שלוש אבני יסוד:

א. URIs (Uniform Resource Identifiers): המייצגים את עצמם ואינם תלויים בסוג המידע;

ב. XML (Extensible Markup Language): התשתית הסינטקטית לשפה האחידה; ג. RDF (Resource Description Framework): המגדירה את אוצר המילים ואת היחסים של שפת התיאור, את האונטולוגיה, ולכן היא החלק המעניין והחשוב של התשתית כולה.

סכמת ה-RDF מסדרת את האובייקטים על פי שלושה מאפיינים כלליים:² הקבוצה, התכונות של האובייקט ויחסי הדמיון והשונות עם אובייקטים אחרים. הקבוצה מכילה אובייקטים בעלי תכונות משותפות (למשל יצורי אנוש). התכונות מוגבלות על ידי הגדרת תחום או טווח (למשל, התכונה "היות בן של" יכולה להיות מוגבלת לתחום היונקים בלבד). יחסים עשויים להתקיים בין אובייקטים, בין קבוצות של אובייקטים ובין תכונות, תוך שימוש במבנים הבאים:

קבוצה א יכולה להיות *זוהי* לקבוצה ב אם כל האלמנטים בקבוצה זהים.

תכונה א יכולה להיות *זוהי* לתכונה ב אם הן משייכות אובייקטים שונים לאותה הקבוצה.

עצם א יכול להיות *זוהי* לעצם ב אם מדובר באותו העצם (תחת שמות שונים).

עצם א יכול להיות *שונה* מעצם ב אם לפחות אחת מתכונותיהם שונה.

עצמים א, ב ו-ג יכולים להיות כולם *שונים* זה מזה אם לפחות תכונה אחת אינה משותפת לכולם.

בין התכונות עשויים להתקיים יחסי היפוך, טרנזיטיביות וסימטריה:

היפוך: אם תכונה 1 היא היפוכה של 2, ואם עצם א מתייחס לעצם ב על ידי 1, אזי עצם ב מתייחס לעצם א על ידי 2.

טרנזיטיביות: אם תכונה 1 מוגדרת כטרנזיטיבית, ואם הזוגות א-ב ו-ב-ג מקיימים 1, אזי גם א-ג מקיים 1.

סימטריות: אם תכונה 1 מוגדרת כסימטרית, ואם הזוג א-ב מקיים 1, אזי גם הזוג ב-א מקיים 1.

² לתיאור מפורט של הגדרות היחסים, ראו

<http://www.w3.org/TR/4002/REC-owl-features-20040210/#domain>

ניתן לראות שיחס הדומות/שונות הוא היחס המרכזי באונטולוגיה המוצעת. מושג ההבדל הוא מושג מרכזי בסטרוקטורליזם, מושג שהוגדר בידי פרדינן דה־סוסיר ופותח בידי ז'אק דרידה ואחרים. פוקו ניסה להיבדל ממסורת זו, אך רבים רואים קשר הדוק בינו לבינה. ההבדל (difference) שלאחר מכן הפך בידי דרידה ל"הבדל תחת השהייה" (difference) — הבדל שנוצר בין המסמן למסומן בציר הזמן של הפרשנות (difference + deferred) — רואה בשונות שבין המילים את מקור היווצרות המשמעות. אי־אפשר להתעלם מההקבלה המפתיעה לתורת האינפורמציה שניסח קלוד שאנון בשנות הארבעים, המגדירה תקשורת כאנטרופיה שלילית, כ"רעשים" הנוצרים במערכת או כרמת אי־האחידות במערכת. האם מקורות אלו שאבו זה מזה? בעקבות שאנון, הגדיר גרגורי בייטסון בשנות השבעים את המידע כ"הבדל שעושה את ההבדל" (information is a difference that makes a difference). כל המקורות הללו מחזקים לדעתי את הקשר בין האונטולוגיה המתגבשת בעולם המידע לבין המסורות הפוסטמודרניות שפוקו היה שותף פעיל בהתגבשותן.

התשתית הסמנטית של מדעי האדם

כפי שצינתי קודם, פוקו סבור שהלינגוויסטיקה היא התשתית שעתידה לאחד את מדעי האדם. אפשר למצוא חיזוק להערכה זו בעבודות לשוניות שנכתבו בעשורים האחרונים בפסיכולוגיה ובסוציולוגיה.

העבודות נועדו לבנות אֶגְרוֹנִים (thesaurus) לתחומים אלו, המשמשים כלים הכרחיים לאחזור מידע יעיל. המידע הרב המצטבר מחייב אותנו למצוא שיטות יעילות לחיפוש ממוחשב. שיטות החיפוש המתקדמות נעזרות באגרונים ומאפשרות לאתר מידע על ידי שימוש במילים נרדפות (synonyms) ואף ביחסים השונים מיחס הזהות הפשוט המוגדר בין המילים הנרדפות. באופן מפתיע, במחקרים שונים בתחומים מגוונים של מדעי האדם התגלתה תשתית סמנטית דומה מאוד. דמיון לינגוויסטי זה מחזק לדעתי את הערכתו של פוקו.

חוקרי מדעי החברה ג'וזף קסגרנדה וקנת' האל³ אספו כ־800 הגדרות בשפת הפימה (Pima) של שבט הפפגו (Papago) וסיננו בסופו של דבר 13 יחסים בסיסיים. התברר שאלו הן בדיוק הקטגוריות שמצאו ו' ראסל וג'ר' ג'נקינס⁴ במחקר על טקסטים חברתיים בשפה האנגלית (נוסף על ארבעה יחסים שהופיעו בשכיחות נמוכה). היחסים מתוארים בלוח שלהלן

3 J. Casagrande and K. Hale, 1967. "Semantic Relations in Papago Folk Definitions," in *Studies in Southwestern Ethnolinguistics*, eds. Dell Hymes and William E. Bittle. Mouton: The Hague, pp. 165–196.

4 W. Russel and R. Jenkins, 1954. "The Complete Minnesota Norms for Responses for 100 Words on the Kent-Rasanoff Word Association Test," *Studies in the Role of Languages in Behaviour*, Technical Report, no. 11. Minneapolis: Minnesota University Press.

ורשומים על פי שכיחות הופעותיהם בקורפוס שנותח (הארבעה הנוספים מופיעים בדירוג 14-17).

יחסי הגדרה על פי קסגרנדה והאל, ראסל וג'נקינס

	יחס	הגדרת האובייקט
1	אטריבוטיבי	מוגדר על ידי היותו בעל תכונה
2	קונטינגנטי	מוגדר על ידי היותו בעל תכונה הכרחית
3	פונקציונלי	מוגדר על ידי היותו בעל שימוש מסוים
4	מרחבי	מוגדר על ידי היחס לאובייקט אחר במרחב
5	אופרטיבי	מוגדר על ידי אופן ההפעלה
6	השוואתי	מוגדר על ידי השוואה לאובייקט אחר
7	המחשתי	מוגדר על ידי המחשה של אובייקט אחר
8	השתייכותי	מוגדר על ידי השתייכות לקבוצה
9	סינונימי	מוגדר על ידי שימוש במילה נרדפת
10	הופכי	מוגדר על ידי היפוך ליחס אחר
11	מוצאי	מוגדר על ידי המקור שממנו הוא עשוי/נגזר
12	מדרגי	מוגדר על ידי השתייכות למדרג של ערכים
13	מעגלי	מוגדר על ידי שימוש בהטיה אחרת של התואר
14	חלקי	מוגדר על ידי היותו חלק ממערכת שלמה
15	קואורדינטיבי	מוגדר על ידי תיאום ידוע עם מושג אחר
16	צלילי	מוגדר על ידי צליל דומה של מושג אחר
17	סדרתי	מוגדר על ידי עקיבה מקובלת אחר מושג אחר

אציין עוד שהעבודות שהביאו לבניית האגרונים נעשו תוך שימוש בסריקה ידנית ומאמצת של קורפוסים גדולים בתחומים הרלוונטיים. בשנים האחרונות הולכות ומתפתחות שיטות

לניתוח אוטומטי של טקסטים בטכנולוגיות של NLP (Natural Language Processing) ו-Statistical NLP. אלו הן טכניקות לזיהוי משמעויות של מילים בטקסט נתון על ידי הגדרת חוקי שפה למערכת התוכנה ועל ידי ניתוח סטטיסטי של מופעי מילים לאחר הסריקה. שיטות אלו ירחיבו מאוד את יכולות חיפוש המידע ואחזורו, ויעבירו אינטליגנציה נוספת מהאדם אל מערכות התוכנה.

הרשת כמושג מרכזי באפיסטמה המתהווה

האונטולוגיה היא אפוא סדר הדברים של הרשת הסמנטית, רשת העשויה להיות תשתית גלובלית ומקיפה לידע האנושי. מושג הרשת, שהשתרש רק בעשורים האחרונים, יהיה מושג מרכזי בכל אפיסטמה חדשה שתיווצר, ועל כן הוא מחזק את הקביעה שאפיסטמה כזו הולכת ומתגבשת. הרשת מייצגת מבנה הוליסטי, שבו המשמעות נגזרת בעיקרה מארגון האלמנטים ומהיחסים ביניהם, ולא מהאלמנטים הבדידים. תפיסה זו מתאימה לייצוג מבוזר של מידע, כאשר חלקי מידע מצביעים על חלקים אחרים והתמונה השלמה אינה מרוכזת בצומת בודד. המייצגים של אונטולוגיית הרשת מוגדרים בפורמט של URI, שהוא למעשה פורמט אוניברסלי של קישור (link) לכתובת http ברשת.

בשנים האחרונות הולכת ומתגבשת תפיסה של מערכות הוליסטיות לא רק בעבור ייצוג מידע, אלא גם כתפיסה של הטבע. תורת המורכבות (complexity) עוסקת בסימולציות של חוקי הפיזיקה, חוקים חברתיים וביולוגיים במערכות "אורגניות", המנסות ליצור חיים מלאכותיים וחברות מלאכותיות. ההנחה היא שהתופעות המורכבות התפתחו bottom-to-top, מהיחידות האלמנטריות אל אלו המורכבות, על ידי מספר מצומצם של חוקים פשוטים. המורכבות של הטבע ושל תופעותיו היא בשיתוף, במערכת כולה וביחסי הגומלין שבה. תפיסה זו, כפי שראינו, משקפת גם את הארגון העיתי של המידע והיא תשמש נדבך חשוב באפיסטמה החדשה.

לתפיסה זו יש גם השלכות אתיות ניכרות. חוקי המוסר אינם עוד חוקים אבסולוטיים שבאו "מלמעלה", מסמכות עליונה כלשהי, אלא הם צומחים "מלמטה", בתהליך אבולוציוני, תוך התאמה לצורכי המערכת כולה. אין הדבר אומר ש"הכל הולך", ושאינן בסיס חיצוני לחוקי המוסר, כפי שטוענת הגישה הפוסטמודרנית. יש בהחלט בסיס אוניברסלי למוסר, אך הוא אינו נגזר מהצו הקטגורי או מכל ציווי דתי אחר, אלא מהצרכים הפיזיקליים של הפריטים במערכת. יש בסיס לציווי "לא תרצח", אך הציווי אינו בא מלוחות הברית. זוהי כמובן מסקנה מרחיקת לכת, העשויה לעורר ויכוחים רבים. אין היא בהכרח המסקנה היחידה הנובעת מתורת המורכבות לגבי האתיקה, ואני סמוך ובטוח שיהיו ניסיונות לגזור ממנה מסקנות שונות לחלוטין.