

הערות מתודולוגיות על כינון מאגר העברית המדוברת בישראל (מעמ"ד)

רגינה א' ורום

מחקר בין-תחומי הוא אחד מאותם דברים שקל יותר לומר מאשר לעשות. האתגר הגדול בעריכת פרויקטים מן הסוג הזה נשען על גרעין משותף של ידע מתודולוגי, תיאורטי ומהותי, ההופכת את התוצר לנגיש לאחרים. במילים אחרות, מחקר בין-תחומי דורש עושר אינטלקטואלי ומיומנויות תקשורת מעולות. מטרת מאמר זה היא להציג כמה מן השיקולים המתודולוגיים המכריעים שעיצבו את תוכנית המחקר ואת שלבי איסוף המידע של מאגר העברית המדוברת בישראל (מעמ"ד) של מתכנניו. טענתי היא כי הבחירות המתודולוגיות שנעשו בתכנון מעמ"ד מדגימות את העוצמה הפוטנציאלית שבפרוייקטים מחקריים בין-תחומיים.

כאשר עמיתי בנימין הרי (אוניברסיטת אמורי) ושלמה זרעאל (אוניברסיטת תל-אביב) ביקשו ממני לראשונה לייעץ ייעוץ מתודולוגי לפרוייקט, חשתי מוחמאת ועם זאת, ספקנית. אחרי הכול, מה לי ולסוציולוגיוסטיקה? אף שהרקע המקצועי שלי הוא בין-תחומי במהותו (אני סוציולוגית עם נטייה בולטת להיסטוריה משווה), הרי שורשי הכשרתי המתודולוגית נטועים עמוק במדעי החברה, במיוחד נתונים כמותיים וניתוחם. אך כאשר מתכנני מעמ"ד הציגו בפני את הפרוייקט ותיארו את האתגרים שהם מתמודדים איתם, הבנתי כי המוקשים הפוטנציאליים שיעמדו בדרכם בתכנון מעמ"ד, הם למעשה מוקשים אופייניים לפרוייקטים בין-תחומיים: מינוח (עגה דיסציפלינרית), מתודולוגיה (כללי ראיות) והתמקדות תימתית.

המוקש הראשון הוא בעיית תקשורת קלאסית, שלפתרונה נדרשו פגישות

ספורות בלבד. בראש ובראשונה היה עלינו להגדיר מונחים כגון "מאגר", "דיאלקט", "מישלב", "מידגם", "יחידת ניתוח", "משתנים (בבלתי-תלויים)". לאחר שהגדרנו את היקף הפרוייקט, התחלנו לדון בעריכת הדגימה ובנושאים לניתוח. מכאן שמאמר זה בנוי כדלהלן:

- היקף הפרוייקט
- תכנון המחקר
- מוקדי הניתוח

אמנם דגש המאמר הוא על נושאים מתודולוגיים, אך הרשו לי לנסח תחילה את המסגרת האקדמית לפרוייקט זה מנקודת המבט הסוציולוגית.

1. היקף הפרוייקט

הבלשן השוויצרי פרדיננד דה-הוסר (1857–1913) מוכר בדרך כלל כמייסד התחום. סוציולוגים מוסיפים לשאוב מתובנותיו הבסיסיות בנוגע להבדלים שבין *langue* (לשון) לבין *parole* (דיבור). אנו ממשיגים את הראשון כמונחים של שפה כמערכת פורמלית, תוך התמקדות בהיבטים המבניים שלה, ואילו את האחרון אנו רואים כדוגמאות של שימוש אישי. בלשון ההדיוטות עוסקת ההבחנה בהבדל בין שפה כתובה לשפה מדוברת.

ומכאן, אל לב לכו של פרוייקט מיוחד זה. לפני כחצי יובל שנים הבחין פישמן (1974) בשני חסרונות בבלשנות העברית מבוססת המאגר: (1) הוא נוטה להתמקד במודל של שפה יותר מאשר לתעד את השימוש בשפה; (2) מחקרי העברית הקיימים התמקדו בשימוש ההיסטורי שלה יותר מאשר בשימוש המודרני. אמנם הספר שאתם מחזיקים כעת בידכם מוכיח כי המחקר הבלשני מבוסס המאגר השתנה מאז כתב פישמן את מאמרו (גם Crowdy, 1993; Berman, 1997; McCarthy, 1999), אך דה-בוגראנד (De Beaugrande, 1999) חזר וטען לאחרונה כי הנטייה להעדיף 'לשון אידיאלית' על פני 'לשון אמיתית' עדיין שלטת בקרב החוקרים. יש לציין כי מתכנני מעמ"ד הם הראשונים שנענים לקריאתו של פישמן לכוונן מאגר של העברית המדוברת. לדברי פישמן, השאלות העיקריות הן: "אילו דגמים של 'עברית טובה' מזהה הציבור בעיתונות, ברדיו, בספרות? מהי 'המפה הקוגניטיבית' הפופולרית מול סוגי העברית המדוברים כיום בישראל (סוגי איכות? סוגים אזוריים? סוגים מעמדיים?)" (עמ' 11).

ראגין (Ragin, 1994) מצייין כי רוב הפרוייקטים המחקריים חותרים למטרה

הערות מתודולוגיות

ספציפית אחת. חלק מייחודו של מעמ"ד טמון בעובדה כי הוא מסוגל למעשה להתמודד עם שלוש מטרות מחקר בריזמנית:

- מעמ"ד יאפשר לנו לשאוב מידע תיאורי, לדוגמה על הבדלים אתניים או אזוריים בשימוש בעברית המדוברת שאליהם רומז פישמן בציטוט לעיל. אך חשוב עוד יותר,
- למעמ"ד הפוטנציאל לפרש תופעות משמעותיות מבחינת הקשרי השיח, וכן
- נתוני מעמ"ד יאפשרו לנו לבחון יחסים סיבתיים בסוגיות של ריבוד חברתי.

שלוש מטרות אלה עולות בקנה אחד עם מה שראגין מכנה "חקר הגיון", "פירוש תופעות תרבות משמעותיות", ו"זיהוי דפוסים משותפים ויחסי גומלין" (Ragin, 1994: 32 ואילך). יתר על כן, המטרות ממחישות כיצד מצליח מעמ"ד לגשר בין מטרות ושיטות ספציפיות לכל אחד מהתחומים, שבאופן מסורתי לרוע המזל קוממו את חוקרי מדעי החברה נגד חוקרי מדעי הרוח (ראה גם Griffin, 1992; 1995). לסיכום, מעמ"ד משקף ניסיון ייחודי לאיסוף חומר עם פוטנציאל מחקרי עבור מומחים מתחומים שונים מלבד בלשנות, החל מאנתרופולוגיה וסוציולוגיה וכלה בחינוך ועיתונות. כעת אדון במאפיינים המתודולוגיים עליהם מושתתת הזיקה הבין-תחומית של מעמ"ד.

2. תכנון המחקר

פרוייקט זה, שהוא כחוד החנית מבחינה מתודולוגית, מבוסס על התפתחויות אחרונות בתחום המחקר (ראה Ragin, 1994; 1998) כאשר הוא משלב את עקרונות היסוד של שיטות כמותיות ואיכותניות לאיסוף חומר. בחלקים א' וב' להלן אביא הגדרות בסיסיות ודוגמאות ממשיות הקשורות לתכנון מעמ"ד. הדיון במערך המחקר מוביל לעקרונות הדגימה והמדידה, הקודמים מבחינה לוגית לשלב איסוף החומר עצמו. ארחיב בנושאים אלה בחלקים ג' וד' להלן.¹

א. שיטות ככלים/אמצעים

מטרת מעמ"ד היא לאסוף נתונים של העברית הישראלית המדוברת מכל תחומי החיים ולהציג אותם בצורה שיטתית. מעמ"ד נבדל ממאגרים אחרים בהיבטים

1 לפרטים נוספים על מונחים טכניים ראה Babbie, 1995; Baker, 1994; Neuman, 1997.

כפופות לכללים. רוב הזמן אנו משתמשים בשיטות מוכרות לאיסוף נתונים, שאלות וסולמות מן המוכן בסקרים, נוסחאות מתימטיות לבדיקת ההשפעות של תופעה אחת על רעותה. חשוב עוד יותר, מעטים הם החוקרים המתכננים פרוייקטים המגשרים על הפער הקלאסי בין גישות דדוקטיביות ואינדוקטיביות. בדרך כלל אף איננו משתמשים בשיטות דגימה כמותיות ואיכותניות בר-זמנית. במקרה זה נאלצו מתכנני מעמ"ד להמציא כלים מסוימים כדי לממש את השקפתם לגבי מאגר העברית המדוברת.

מתכנני מעמ"ד יבחנו השערות הנובעות מן ההנחה כי הקשר האירוע משפיע על השימוש בשפה המדוברת. מאמצים מן הסוג הזה לבדיקת התיאוריה נחשבים בעינינו כדדוקטיביים, אך תהליך איסוף החומר עצמו כולל רישום נתונים איכותניים שייצר מידע רב-שכבתי עשיר על העברית הנהוגה במסגרות שונות. קידוד קטעי דיבור, אלה, הנבחרים באקראי ומהווים את המאגר העיקרי ($n=950$ תאים), כסיטואציות, ולא כמילים או כצירופי מילים ערטילאיים, יאפשר לחוקרים אחרים בהמשך לגבש תיאוריות. למשל, כאלה שיתבססו על דפוסי שיחה או על דפוסי לשון. מעניין שאנו קושרים בדרך כלל סוג זה של מחקר בחשיבה אינדוקטיבית. האמת היא שהחלטתם לכלול קטעי דיבור נוספים, שייבחרו באמצעות דגימה מכוונת במאגר משנה ($n=50$ תאים), תספק למערך הנתונים יסוד אינדוקטיבי נוסף. ולבסוף, תוכניתם לקודד את כל "תאי" קטעי הדיבור על-פי הקשרי שיח ולא על-פי קריטריונים בלשניים טהורים נובעת בחלקה מכורח הנסיבות, שכן לא קיים כל מערך נתונים שיוכל לשמש להשוואה, וקטגוריות המשמשות במאגרי לשון אחרים יוכלו לסייע כאן במידה מוגבלת בלבד.

האופן החדשני שבו משלבת התוכנית המחקרית של מתכנני מעמ"ד מחשבה דדוקטיבית במחשבה אינדוקטיבית מקביל לבחירה בצירוף שתי שיטות דגימה, שרוב החוקרים היו רואים כסותרות יותר מאשר כמשלימות זו את זו. לאור שאפתנותן של פרוייקט בסדר גודל כזה (אלף קטעי דיבור מוקלטים, בערך 30 דקות אורכו של כל אחד מהם, מה שמסתכם בחמישה מיליון מילים), יכלו מתכנני מעמ"ד לתת לנוחות המתודולוגית לקבוע את היקף המאגר. לדוגמה, קל יותר היה לקודד אלף קטעי דיבור מוקלטים קיימים (לדוגמה — שידורי חדשות, דרשות, הרצאות), או לבקש מאלף איש לקרוא את אותו טקסט ולקודד את הנתונים על-פי הבדלים בדיאלקט או בהנגנה. תחת זאת, מתכנני מעמ"ד נוקטים סדרה יוצאת דופן של אמצעים משלימים. הם משלבים שיטת דגימה אחת, הקשורה בדרך כלל למחקר כמותי (מידגם אקראי), עם שיטת דגימה

אחדים: ראשית, אמנם מאגרי לשון קיימים בשפות רבות, אך איש לא ניסה עדיין לכוון מאגר של העברית המדוברת. למעשה, לא קיים אף מאגר של שפה שניתן להשוותה לעברית (ראה יזרעאל בכרך זה). אף על פי כן, השכילו מתכנני מעמ"ד לעצב היבטים של מאגר זה על סמך מאגרים קיימים של שפות מדוברות (Crowdy, 1993). שנית, מאגרי לשון מתוכננים בדרך כלל על-ידי חוקרי דקדוק ואטימולוגיה ועבורם (Biber et al., 1994). משום כך, מילים ומבעים היו למוקד העניין וסביבם נערכו נתונים אחרים. בניגוד לכך, מתכנני מעמ"ד מעוניינים בעיקר בהקשר החברתי שבו משמשת העברית. כתוצאה מכך, נותר מעמ"ד יוצא דופן בכל הקשור לממדי המאגר, לשילוב שיטות הדגימה הנקוטות (דהיינו, האופן שבו דוגמאות או "תאים" — קטעי דיבור מוקלט — נבחרים כדי להיכלל במאגר) ולמשתנים הדמוגרפיים של הקשרי השיח המשמשים כדי לאפשר גישה לנתונים מצד חוקרי מדעי החברה.

בנקודה זו עלינו לציין כי בחירותיהם המתודולוגיות של מתכנני מעמ"ד נקבעו על-ידי שיקולים תיאורטיים וענייניים. ברור כי השאיפה היא להשיג ייצוג מקיף של העברית המדוברת על-ידי הקלטת דיבור מפי דוברים שונים במצבים שונים (מונולוג מול דיאלוג, שיחת טלפון מול שיחה פנים אל פנים, ריאיון בטלוויזיה וכו'). אך הנחה סוציולוגית מרכזית עומדת בבסיס ההקלטות האלה: המחברים מניחים כי הקשר האירוע קובע לגבי השימוש בלשון. שלושת הצירים התיאורטיים החשובים — או "משתני ההקשר" — שעל פיהם תוכנן מעמ"ד, עוסקים ביחסים בין הדוברים, במבנה השיח ובנושא השיח (ראה בסופו של דבר, מעמ"ד יהיה נגיש ופתוח לחוקרים מתחומים שונים, כולל אנשי מדעי החברה המעוניינים להסביר יחסים סיבתיים, למשל, בין הרקע החברתי של הדוברים (מצב סוציו-אקונומי, עדה, גיל, מין) לבין השימוש שהם עושים בביטויים ספציפיים. מחקרים אחרים יוכלו לעסוק בתוצאות האינטראקציה בין רקע אישי, הקשר חברתי ושימוש לשוני.

ב. החוקרים כמפתחי אמצעים

כל שיטות איסוף החומר, החל מתצפית בלתי-משתתפת וכלה בראיונות מובנים,

2 המחברים מגדירים את המשתנים הדיכוטומיים הללו במקום אחר בספר זה. בקצרה, משתנה 'היחסים הבין-אישיים' מבדיל שיחות בין קרובים/חברים משיחות בין אחרים, 'מבנה בשיח' עוסק בהימצאותם/היעדרם של הבדלי סמכות בין דוברים, 'נושא השיח' מבדיל שיחות העוסקות בעניינים אישיים משיחות אחרות.

אחרת, המשמשת במחקר איכותני (מידגם מכוון). אי לכך, איסוף החומר ושלבי הקידוד יחייבו עבודה רבה ואינטנסיבית. עם זאת, לכישושלם, אמור מעמ"ד להוכיח את עצמו כמשאב רב-ערך לחוקרים מחייבי הגישות הכמותיות והאיכותניות גם יחד.

ג. המידגם

ההחלטה על עריכת מידגם חייבה לקבוע ממי, מה, מתי והיכן לאסוף נתונים. ההחלטה חייבה גם להתוות שיטות דגימה חלופיות ולגבש הסכמה בדבר יחידות הניתוח. בעוד שסוגי הדגימה מכריעים בשלב איסוף החומר, יחידות הניתוח משרתות הן את איסוף החומר והן את הקידוד.

(1) סוגי דגימה

בעת הדיון במטרת המאגר ובהיקפו, התברר שמתכנני מעמ"ד אינם מבקשים לספק "מפה קוגניטיבית" ממצה של עולם העברית המדוברת. הם התחבטו בשאלת בניית המידגם, דהיינו, אילו דוגמאות של עברית מדוברת לאסוף, וכמה (Biber, 1990).

מחקרים שנערכו לאחרונה על מאגרי לשון אחרים התבססו על מידגמים שונים מאוד, שנאספו בדרכים שונות. לדוגמה, במאמר שבדק את השימוש במחמאות במצרים ובארצות הברית, המחברים ניסחו שאלונים בעקבות מחקר טרומי, שכלל ארבעים ראיונות עומק. החוקרים גייסו את המשתתפים בדגימה מבין הסטודנטים שלהם, ובכך ערכו מה שמכונה בפנינו מידגם נוחות. הן להיבט האיכותני (ראיון) והן להיבט הכמותי (סקר) של הפרוייקט. לבסוף זיהו החוקרים שישים מחמאות שונות על בסיס $n=240$ שאלונים מכל אחת משתי המדינות המשתתפות (Nelson, El Bakary and Al Batal, 1993). מחקר אחר עשה שימוש חדשני ב- $n=338$ שאלונים ובמספר זהה של מאמרים. החוקר ערך מה שנקרא "מידגם אקראי" של מתבגרים בעיר מסוימת, שאמורים היו לייצג את כלל האוכלוסייה. המידע שסיפקו שימש לניתוח רגרסיה³ של דפוסי לשון מבוססי מעמד באיסלנד (Thorlindsson, 1987).

סוציולוגים מגדירים את העולם כ"אוכלוסייה" שאליה מבקשים להכליל את הממצאים והמסקנות. זהו עניין תיאורטי ולא דווקא אמפירי. לדוגמה, במחקרים

3 רגרסיה ליניארית ורגרסיה מרובה הן שיטות סטטיסטיות מקובלות לניתוח נתונים כמותיים. הן מאפשרות לייצג משתנה אחד (או את ערכו הצפוי של אותו משתנה) כפונקציה ליניארית של משתנה אחר או של מספר משתנים אחרים.

שפירטתי לעיל, כוונת החוקרים יכולה להיות לומר משהו על כלל הסטודנטים או על כלל בני-העשרה במדינות הללו, או להכליל לכלל האוכלוסייה במדינות אלה. זאת ועוד: המונח 'אוכלוסייה' עשוי להתייחס לא רק לאנשים, אלא לסך כל מוקדי העניין שמעסיקים את החוקר. במקרים שצוטטו לעיל, נאסף החומר מאנשים, אך "האוכלוסייה" עשויה לכלול גם יחידות ניתוח אחרות, כגון מילים, משפטים, שיחות — או קטעי דיבור מוקלט. עוד אשוב לסוגיית יחידות הניתוח.

לעתים, החוקרים במדעי החברה יודעים מהי האוכלוסייה שלהם. במקרה זה, מתכנני מעמ"ד מבחינים כי במונחים אנושיים, האוכלוסייה שלגביה הם מבקשים להסיק מסקנות כלליות מונה כחמישה וחצי מיליון התושבים בישראל (לחלופין, הם יכלו לבחור רק באזרחים הישראלים [שעשויים להשתייך לעדות שונות] או רק ביהודים בישראל [שעשויים להיות מארצות מוצא שונות] כאוכלוסייה האנושית שעליה יבוסס הניתוח). אך במושגים תיאורטיים, "האוכלוסייה" שלהם מורכבת מן העברית המדוברת בישראל בתחילת המאה ה-21. שימו לב כי מספר האנשים המשתמשים בשפה אינו רלוונטי כאן וכי, למעשה, מה שחשוב הוא מיגוון חלופות הדיבור שבשימוש.

האתגר הפך להיות בניית מידגם גדול ומגוון דיו כדי לתאר נכונה את רוב (אם לא את כל) החלופות של העברית המדוברת, מתוך מודעות לעובדה שאף מידגם או מאגר אינו מושלם. כלומר, כדי להיצמד למוקד תימטי ולהבטיח כי הנתונים יוכלו להניב דפוסי קידוד חסכניים, נאלצו מתכנני מעמ"ד להחליט על אילו צורות דיבור (דיאלקטים, סיטואציות) לוותר. לאחר דיון בקטגוריות שיהוו את גרעין המאגר, החליטו כי המידגם יחולק לשני תת-מאגרים או תת-מידגמים. כל אחד מהם ממחיש סוג דגימה שונה.

עד כה, החוקרים נהגו לראות כל אחת משיטות דגימה אלה כמשקפת דרך שונה לעריכת מחקר. גישת הדגימה שבבסיס המאגר העיקרי משקפת כנראה קו חשיבה דדוקטיבי, האופייני למחקר הכמותי, ואילו המידע שעשוי להתקבל מהמאגר המשלים נראה כיישום של לוגיקה אינדוקטיבית, המאפיינת על פי רוב פרוייקטים של מחקר איכותני. עם זאת, כפי שמדגים הדיון שלהלן (והמאמר המתפרסם בקובץ זה), התכנון המוקפד של מתכנני מעמ"ד מצליח לשזור את השניים יחד. כתוצאה מכך, החלק הכמותי לכאורה של המאגר מושפע מאוד משיקולים איכותניים באשר למשקלו של החומר הנאסף; ואילו תת-המאגר האיכותני לכאורה עומד בבירור בקריטריונים המעניינים חוקרים כמותיים: ראשית, המאגר עוצב למזער את החששות ביחס למספר מקרים קטן, חששות

שעלולים לפגוע באפשרויות ההכללה והתוקף של הנתונים. שנית, הוא מביא בחשבון שלא כל הדיבור מתרחש/מטביע חותם בנסיבות של קבוצה קטנה או של תקשורת בינאישית. למעשה, כלי התקשורת וגופי הממשל ממלאים תפקיד מכריע בקביעת השימוש בשפה.

א. המאגר הראשי: מידגם אקראי-שיטתי

המאגר הראשי במעמ"ד מהווה 95% מכלל המאגר: 950 "תאים" לשוניים, ובכל תא בערך 5,000 מילות טקסט. חוקרים במדעי החברה מבחינים בין צורות שונות של מידגמים אקראיים, אף שכולם מבקשים ליצור תמונה 'מייצגת' של האוכלוסייה.⁴ מידגם שיטתי-מקרי מסתמך על פרודורות שנקבעות מראש, לפיהן נבחר למידגם כל אדם (או מקרה) ה. במקרה שלפנינו, צוות הפרוייקט מתכנן לדגום דוברי עברית על סמך מקום מגוריהם. קשיים אפשריים בהשגת מידגם מייצג עלולים להיות במהלך הניסיונות להגיע אל הדיירים או להשיג את הסכמתם והשתתפותם. לכן חשבו החוקרים כיצד להבטיח את עושר הדיאלקטים המדוברים. לשם כך, הם מתכננים לוודא את מספר התאים הלשוניים המוקלטים לפי קטגוריות דמוגרפיות שאמורות להשפיע על הדיבור (Crowdy, 1993). הקטגוריות הדמוגרפיות העיקריות:

- (1) עדה או דת/מקום לידה/ארץ מוצא המשפחה = שתחולק לחמש תת-קטגוריות מנועות אהרדי;
- (2) גיל = תחולק לשלוש תת-קטגוריות מנועות אהרדי.
- (3) חינוך = תחולק לשלוש קטגוריות מנועות אהרדי.⁵

4 ככל הנראה, איסוף החומר יכול גם יסוד של דגימת אשכולות. במקום לבחור את הדוגמאות מתוך כמה שיותר מקומות, ייתכן שמגבלות מימון זמן יצמצמו את איסוף החומר לאזורים עירוניים וכפריים נבחרים בלבד. למידע נוסף אודות יתרונות וחסרונות מידגם האשכולות, ראה לדוגמה Babbie, 1995; Kmenta, 2000: 157.

5 משתכרר כי ניתן לרכז את המידגם גם בדרכים אחרות/נוספות (לדוגמה על-פי הכנסה/מעמד סוציו-אקונומי, או על-פי מין). החליטו המחברים שלא להתמקד בשימוש על-פי שיוכם המיני של דוברי העברית. זאת ועוד, עבור חוקרים המתעניינים בנושאים אלה, האופי האקראי של המידגם צריך להבטיח חלוקה שווה פחות או יותר בין גברים לנשים. אם נצרך לכך את היקפו המלא של המאגר, הרי שהדבר יצמצם את החשש מפני סטייה פוטנציאלית בדגימה. כמו כן, מתוך מודעות לקשר שבין השכלה להכנסה החליטו המחברים כי להבדלים ברמת ההשכלה תהיה השפעה מכרעת יותר לגבי ההבדלים בשימוש בשפה, בהשוואה להשפעת גובה ההכנסה בלבד. זוהי דוגמה מובהקת לשיקולים שחוקרים נדרשים להפעיל כאשר הם מתכננים מחקרים מן הסוג הזה. ככל שהמחקר קטן יותר, כך גדלה הדאגה לשיקול מדיק יותר של המשתנים, וכך תוצאות הסטייה המיידגמית עלולות להיות הרסניות יותר.

בסך הכול, תבנית זו מספקת 45 צירופים שונים (3x3x5). אך העניין מסתבך כאשר מביאים בחשבון כי מעמ"ד שואף לתעד שונות בדיבור המוקלט. בתיאור הפרוייקט (המופיע בקובץ זה) משתמשים במונח "קטגוריות הקשרי שיח" לציון הבחנות אנליטיות מכריעות, כגון, בין הקשר פיזי/מרחבי (מונולוג מול דיאלוג, או שיחה טלפונית לעומת שיחה פנים אל פנים), הקשר תימטי (נושא השיחה) והקשר חברתי (היחסים בין הדוברים). המונח "הקשרי שיח" מקביל למונח "משתנים עצמאיים" המשמש את חוקרי מדעי החברה. מנקודת מבט אמפירית, האתגר העיקרי יהיה להשיג קטעי דיבור רבים לכמה שיותר קטגוריות כתבנית זו (Biber, 1990). אם לא תהיה ברירה, ייאלצו המחברים (והמשתמשים העתידיים) לאחד כמה מן הקטגוריות הדמוגרפיות או כמה מקטגוריות ההקשר. החלטתם של מתכנני מעמ"ד להשתמש בדגימה אקראית שיטתית מצטיירת כטובה ביותר בכאנו להשוות את מעמ"ד לאופן שבו כוננו מאגרי לשון אחרים. יש לציין כי החלטה זאת לא בהכרח מיטיבה עם התוכנית המחקרית שלהם עצמם. הדבר מעיד על ראייתם ארוכת הטווח, שכן כך תועצם תועלתו של המאגר לחוקרי מדעי החברה, המבקשים לבחון מארג יחסי סיבתיות ו/או לעשות זאת באמצעות ניתוח כמותי. אחרי הכול, דגימה אקראית שיטתית תהיה מחויבת המציאות כאשר מצפים ליחסים שונים בין "משתנים תלויים" ובין "משתנים עצמאיים" שונים בתת-קבוצות שיש להן אותם מאפיינים שיוכיים או דמוגרפיים.

ב. המאגר המשלים: מידגם מכון

לעומתו, המאגר המשלים יורכב מחמישים "תאים" נוספים (גם הם בני כ-5,000 מילה כל אחד). למטרה זו קטעי הדיבור לא ייבחרו על בסיס קריטריונים דמוגרפיים, שמטרתם לספק "תמונה מייצגת" של העברית המדוברת. תחת זאת, ייבחרו הדוגמאות על פי קריטריונים תיאורטיים מפורשים (Denzin and Lincoln, 1994). מתוך ידיעה שלסוגי דיבור מסוימים יש השפעה אדירה על כלל דפוסי הדיבור, למרות העובדה שמעט אנשים יחסית מפיצים דיבור כזה באופן פעיל, יידגמו שלוש קטגוריות רחבות של דיבור ציבורי: אמצעי תקשורת ההמונים (טלוויזיה, רדיו), הכנסת ומערכת המשפט (לרציונל, ראה O'Bar, 1982; Crowdy, 1993).

(2) יחידות ניתוח

הבנת חשיבותה של בחירת יחידות ניתוח הולמות היא המפתח להצלחת כל פרויקט מחקר. בעגה הסוציולוגית המקובלת, השימוש ביחידות ניתוח לא

הערות מתודולוגיות

ידועים, שממדיהם או מאפייניהם יהיו מעניינים מבחינה בין-תחומית, היה חשוב למתכנני מעמ"ד דווקא על שום שאיפתם להפוך את מערך הנתונים שלהם לזמין לכול.

בניית המשתנים כוללת למעשה שני שלבים שאנו מכנים "המשגה" ו"אופרציונליזציה" (או "הגדרה תצפיתית"). השניים כרוכים זה בזה (לדוגמה Babbie, 1995; Neuman, 1997), אך הראשון מציב אתגר תיאורטי (לדוגמה, כיצד נגדיר "עדה"? לפי דת? לשון-אם? מוצא? אולי צירוף של כולם?). והשני, לעומתו, מציב אתגר אמפירי של דיוק במדידה. במקרה זה, הסתמכו מתכנני מעמ"ד על הגדרת העדה המשמשת בדרך כלל בהקשר הישראלי, המבוססת על צירוף גורמי דת, מקום לידה וארץ מוצא. המשתנה שהתקבל בנוי מחמש קטגוריות אמפיריות מנועות אהדרי:

- יהודים ילידי הארץ שאבותיהם מאסיה/אפריקה
- יהודים ילידי הארץ שאבותיהם ממקומות אחרים
- יהודים ילידי חו"ל שעלו לארץ לפני 1965
- יהודים ילידי חו"ל שעלו לארץ אחרי 1965
- לא-יהודים (כגון מוסלמים, נוצרים, דרוזים)

קטגוריות אלה משמעותיות בהקשר הישראלי: הן מבחינות בין אשכנזים לספרדים, הן מבחינות בין גלי עלייה שהושפעו משינוי מכריע במדיניות (1965) והן מזהות דוברים ממיצוטים אתניים. ברור שמאגר של שפה מדוברת אחרת יתבסס על סמנים (markers) אתניים אחרים. אך ייתכן גם שאילו כונן מאגר של עברית מדוברת בישראל לפני תום המלחמה הקרה, היה מספק דפוסי דיבור שונים לחלוטין מאלה שיספק מאגר המכונן בשנים אלה. נקודת תורפה אפשרית אחת בדרך שבה העדתיות מיתרגמת אצל מתכנני מעמ"ד נעוצה בהעדר הבחנה בין תושבים ותיקים לעולים חדשים, בעיקר לאור ממדיה הנרחבים של עליית יהודי ברית המועצות לשעבר, שללא ספק הטביעה את חותמה על העברית המדוברת בעשר השנים האחרונות. אם כך, המשתנה של מתכנני מעמ"ד לא יצליח לתאר הבחנה אתנית מכריעה זו.

גם המשתנים תלויי ההקשר שמתכנני מעמ"ד משתמשים בהם בדגימת קטעי הדיבור מערבים החלטות תיאורטיות ואמפיריות. לדוגמה, הם מגדירים את "מבנה השיח" במונחים של היחסים בין המשתתפים שהוקלטו בכל קטע דיבור. לכל מושג עשויים להיות "אינדיקטורים" או משתנים מרובים. אך מכל הדרכים השונות המשמשות לאפיון יחסים בין-אישיים, בחרו מתכנני מעמ"ד להדגיש

נכונות עשוי להוביל למה שמכונה בסוציולוגיה "כשל אקולוגי". הכוונה היא שיש להימנע מהקשת היקשים מרמת ניתוח אחת לשנייה. אנו טועים טעות אקולוגית כאשר אנו עושים שימוש בנתונים קיבוציים כדי להקיש על מניעהם האישיים של אנשים. הרשו לי לנקוב בדוגמה מפורסמת, המבוססת על המחקר הקלאסי של דורקהיים: אם חוקר מגלה ששיעורי ההתאבדות (=נתונים ברמה הקיבוצית) גבוהים יותר במדינות פרוטסטנטיות מאשר במדינות קתוליות, ומקיש מכך שפרוטסטנטים מגלים יותר סובלות כלפי התאבדות (=נתונים ברמה האישית), הרי שהוא טועה טעות אקולוגית.

בדומה לכך, בהתחשב בעובדה שקטעי הדיבור במעמ"ד מהווים את יחידת הניתוח העיקרית, על החוקרים להקפיד שלא להקיש היקשים על השימוש במילים או במשפטים מסוימים, ואף לא על האנשים שדיבורם הוקלט. זה עשוי להתגלות כאחד האתגרים הגדולים ביותר הניצבים בפני חוקרים המבקשים להשתמש בנתוני מעמ"ד. אחרי הכול, כאשר 5,000 מילים מרכיבות קטע/תא/דוגמת דיבור יחיד/ה, קשה לנתח מספר גדול של דוגמאות כאלה זו לצד זו. לשם המחשה, טקסט המאמר שאתם קוראים כעת מכיל בערך 5,000 מילה. מכאן שאני מתארת לעצמי שחוקרים המעוניינים ביחידות ניתוח קטנות יותר — למשל, השימוש בשמות גוף מסומנים מינית לעומת אלה שאינם מסומנים מינית בנאומי הכנסת, בתוכניות טלוויזיה, או בשיחות טלפון — עשויים לבחור בקידוד מחדש של חלקים מן החומר המקורי. עם זאת, תוכנית הקידוד הראשונית של מתכנני מעמ"ד צריכה לאפשר בחירה נוספת בתאים ספציפיים שתתבסס, לדוגמה, על קריטריונים דמוגרפיים או תלויי הקשר.

ד. מדידות

כדי לפתור סוגיות מדידה, שקלנו את תוכן האירועים שיידגמו: באיזו מידה עשוי השימוש בעברית המדוברת להשתנות ממצב למצב ובתחומי האירוע עצמו? הדבר חייב אותנו לחשוב במונחים של משתנים, המכונים גם "תופעות שיש בהן עניין", "קטגוריות אנליטיות", "אינדיקטורים" או "מבארים ומבארים". שוב, הקוראים עשויים להעדיף מונחים אלה או אחרים בהתאם לתחום עיסוקם. היות שמתכנני מעמ"ד היו שקועים בספרות כלשנות המאגר ולא היו ביניהם חילוקי דעות בדבר הממדים האנליטיים החיוניים של הפרוייקט, אופן מדידת המשתנים לא היווה קושי מיוחד בתכנון מעמ"ד.⁶ אף על פי כן, השימוש באינדיקטורים

6 לדין מעמיק יותר באופן שבו עשויות סוגיות מדידה להשפיע על מהימנותם ועל תקפותם של הנתונים, ראה טקסטים על שיטות במדעי החברה, כגון Babbie, 1995; Neuman, 1997.

אין לי ספק כי ישלימו את איסוף החומר, יקטלגו את קטעי הדיבור השונים לפי המשתנים הדמוגרפיים והמשתנים תלויי ההקשר שהגדירו ויעמידו את מערך הנתונים לרשות חוקרי העולם כולו.

עם זאת, נראה שקידוד לשוני מושלם של כל קטע לא ייעשה בעתיד הקרוב. חלקית, מפני שקשה לקבוע מראש מסגרת זמן לשלב זה של כינון המאגר, שכן על אף התקדמויות אחרונות בתחום התוכנה, עדיין לא ברור באיזו מידה ניתן להשתמש בטכנולוגיה ממוחשבת לצורך פרוייקט זה. אם הקידוד הלשוני יישאר בגדר עבודה ידנית, כולו או אפילו רק חלק ממנו, הרי שזמן רב מאוד עשוי להידרש כדי להשלים את מעמ"ד. חשוב עוד יותר, המחברים טרם פירטו את הקריטריונים, או אפילו את המשתנים, שלפיהם יקודד תוכנו של כל קטע דיבור. שוב, מעמ"ד יוכל לשמש לבחינת נושאים לשוניים שהורים רבים הקשורים בדקדוק, בתחביר, בהגייה או במישלבים. סביר להניח כי חוקרים שיתעניינו בעתיד בנושאים אלה, יוכלו לשלוף תת-מידגמים ממעמ"ד ולקודד את קטעי הדיבור המסוימים להפקת מידע לנושאי מחקרם. אמנם הדבר עלול לפגום בלכידות הכללית של מערך הנתונים, אך הוא גם יקדם את השימוש בנתוני מעמ"ד ויגביר את השפעתו הבין-תחומית. בקצרה, אני צופה כי שלב איסוף החומר עצמו יביא לתוצאה הרצויה, דהיינו, השגת המטרה האמפירית, שהיא מתן תיאור מדויק של השונות בעברית המדוברת בישראל. ימים יגידו אם קידוד הנתונים ומאמצי הניתוח ישיגו את המטרות המוצגות במאמרים בקובץ זה.

השאלה השנייה היא: האם קיים קו מושגי ואנליטי ברור המנחה את התוכנית המחקרית של צוות מעמ"ד ומהן המטרות התיאורטיות והאמפיריות שלו? יש לציין כי המתכננים הצליחו להבדיל בין מטרתם האמפירית לכונן מאגר של העברית המדוברת בישראל לבין התוכנית המחקרית שלהם עצמם. כפי שתואר לעיל, המטרה העיקרית מאחורי שלב איסוף החומר נותרת תיאורית: הצגת תמונה של השונות בשפה המדוברת. הם תיכננו את מעמ"ד באופן שיאפשר להם לחקור נתח מסוים של החומר, אך יתאים מאוד גם לחוקרים שעניינם שונה במידה ניכרת משלהם ושעשויים לנחץ את הנתונים תוך שימוש במיגון שיטות או "כלים", החל מניתוח תוכן וכלה ברגרסיה מרובה.⁷

בניגוד לכך, פרוייקט המחקר שלהם עצמם, שלטובתו ישתמשו בנתוני מעמ"ד, מתברר כבעל מניע הרכה יותר תיאורטי. בהתמקדם באופן שבו ההקשר החברתי והמדמים השונים בהייררכיה של כוח מעצבים את השימוש בשפה, מבקשים מתכנני מעמ"ד לענות על שאלה סוציולוגית בסיסית. הלוגיקה

ממד ספציפי אחד: הם מבחינים בין שני סוגי שיח לפי קיומם או היעדרם של הבדלי כוח בין הדוברים. ההנחה התיאורטית שבבסיס אופרציונליזציה זו היא כי השפה המדוברת שונה במצבים שבהם שני הדוברים מתייחסים זה לזה בצורה הייררכית (לדוגמה, הם עשויים להשתמש בשפה רשמית יותר, מילות חלל מעטות יותר, סייגים — hedges — המאפיינים דיבור של בני מין מסוים). כמו לגבי המשתנים הדמוגרפיים המשמשים לריבוד המידגם, כך גם בבחירת המשתנים תלויי ההקשר ממחישים מתכנני מעמ"ד את מחויבותם ליצירת מעמ"ד כמסד נתונים שישרת את חוקרי מדעי החברה השונים: "עדתיות" ו"סמכות", שניהם מושגים מרכזיים בסוציולוגיה.

3. מוקדי הניתוח

בתחילת מאמר זה הזכרתי מכשלה פוטנציאלית אחת. היא מוכרת לכל החוקרים והיא נוטה לפגוע במיוחד בפרוייקטים בין-תחומיים: העדר מוקד מושגי ואנליטי, כאשר מנסים להיות כל דבר עבור כולם. למאמצים לאיסוף החומר אורכות לפחות שתי סכנות: או שהאיסוף לא יושלם לעולם; או שיתברר כי הנתונים מכסים תחום נרחב של נושאים משמעותיים מעבר למה שתוכנן. הדבר יחייב את החוקרים לצאת מתחומי התמחותם, ואם לא יעזרו באופן מהותי באנשי מקצוע אחרים, יסתכנו בשימוש באינדיקטורים שאינם תקפים או שיש בהם פגמים אחרים. כאשר פרוייקטים אדירי ממדים כאלה מגיעים לידי השלמה, גוף הנתונים שמתקבל עלול להתגלות כמסורבל מדי לשימוש. על אף שאנו מצפים מכולם להיות מתוחכמים טכנולוגית ומתודולוגית, הידידותיות למשתמש — או היעדרה — יכריעו לגבי המידה והאופן שבהם ינוצלו הנתונים. אי-לכך, לשם השפעה מרבית על תחומי הידע השונים וביניהם, כדאי לקחת בחשבון שמערך נתונים המתיימר להשיג פחות, עשוי בסופו של דבר להשיג הרכה יותר (ראה Biber, 1993).

מעמ"ד הוא פרוייקט שאפתני מאוד: הכוונה היא לקודד סכום כולל של חמישה מיליון מילים באלף קטעי דיבור. שתי שאלות מתעוררות כאן. ראשית, האם ישלימו את פרוייקט מעמ"ד? האתגר הראשון הניצב בפני צוות החוקרים הוא איסוף החומר עצמו. עד היום שוררת אי-ודאות באשר למסגרות הזמן והמקום המתאימות לאיסוף החומר. פרוייקטים מחקריים הדורשים כמות כזו של עבודת שדה ולאחריה קידוד סוציולינגוויסטי מפורט, זקוקים לתמיכה כלכלית רצינית וארוכת טווח. אם יצליחו המחברים להשיג את המימון הדרוש,

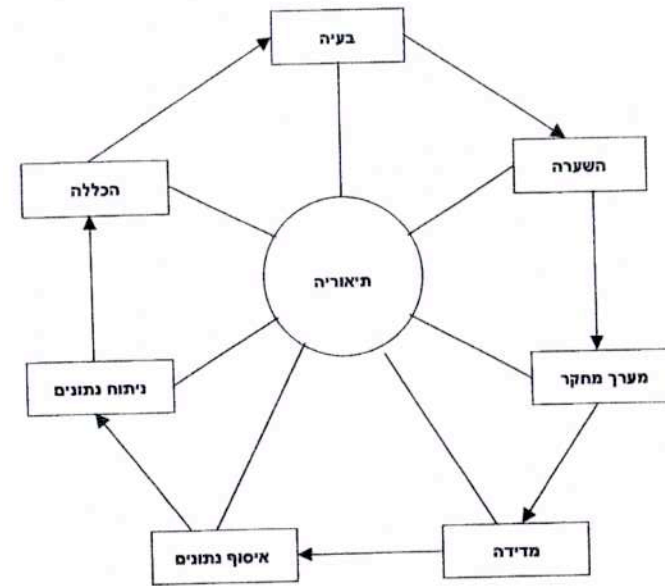
הערות מתודולוגיות

בשל תחום ההתמחות שלי,⁸ היחשפותי למחקר שמקור השראתו במסורת זו הצטמצמה לחוקרי מדעי החברה הבוחנים כיצד גוע, עדה, מעמד חברתי או מין משפיעים על השימוש בשפה ועל משמעותה (ראה גם Thorne et al., 1983; Thorlindsson; 1987; Lareau; 1989; Lakoff; 1990; Calhoun; 1994; Zaretsky; 1994; Laureau and Horvat; 1999). אם להסתמך על דוגמה למחקר אודות הקשר בין מין לשפה, ג'ונסון (Johnson, 1994) מוכיח כי להקשר, ובמיוחד להבדל במעמד בין חברי קבוצה כלשהי, יש השפעה גדולה יותר על דפוסי שיחה מאשר למין עצמו. בניגוד לכך, התנהגויות בלתי-מילוליות מושפעות מאוד ממין של בני השיח, ללא קשר למעמדם הפורמלי בקבוצה (ראה גם Kollock et al., 1985; Carli; 1990). מחקרים נוספים מוכיחים כי הבדלי מין בשימוש בשפה אינם מצטמצמים למבוגרים בלבד, אלא באים לידי ביטוי גם בקרב ילדים (Maltz and Borker, 1982).

לסיכום, מה יוכל מעמ"ד להציע לסוציולוגים? מהנאמר לעיל ניתן להבין כי התחום שלנו עוסק בקו שאלות שונה. "הדמיון הסוציולוגי" (Mills, 1959) רואה את היחיד הן כתוצר של סביבתו והן כמייצר מציאות חברתית. מטבעו עוסק המחקר שלנו בהשוואה פחות או יותר מפורשת בין קבוצות אוכלוסין שונות, כמו למשל קבוצות הנבדלות זו מזו במאפייני עדה, רובד חברתי או מין (Ragin, 1994). מבחינת היקף ומורכבות, מאגר העברית המדוברת בישראל (מעמ"ד) יהיה אידיאלי לסוציולוגים המתעניינים בשפה כצורה של הון תרבותי, "מטבע" שערכו תלוי, לדוגמה, ברקע האישי של הדובר ובנסיבות או בהקשר שבהם הוא משמש.

⁸ לא אוכל לספק תיאור מדויק של ההשפעה שהייתה לדה-סוסיר על תחומים בתיאוריה הספרותית או הביקורתית. אך סוציולוגים תרבותיים ואנתרופולוגים מסתמכים גם הם על דה-סוסיר ועל בני זמנו, במיוחד על אמיל דורקהיים – אחד ממייסדי הסוציולוגיה. לדוגמה, וונדי גריזוולד (Griswold, 1992; 1993; 1994; 2000), הידועה בעיקר בזכות עבודתה בתחום הסוציולוגיה של הספרות ותרבות ה"פופ", מתמקדת במפורש בהבדלים היסטוריים ובין-תרבותיים בשפה הכתובה. כך גם מרי דאגלס (Douglas, 1966; 1983; 1984; 1994), הידועה בעיקר בזכות תרומתה לחקר המנהגים התרבותיים והדת, כשטיפלה במבנה חברתי של המעמדות ושל הקבוצות האתניות. לדיון נוסף בקשרים שבין דורקהיים לדה-סוסיר – כולל מקבילות אפשריות הנוגעות לדיכטומיה של דורקהיים בין "חברה" ל"יחיד" וההבחנה של דה-סוסיר בין היסודות המבניים ליסודות האידיויסינקרטיים של השפה – ראה לדוגמה Gane, 1983; Dinneen; 1990; Wagner; 1990; Reckwitz; 1997. לדיון כללי נוסף בתחום זה, ראה, לדוגמה, Thorne et al., 1983; Heritage; 1984; Lakoff; 1990; Ochs; 1992.

של המחקר מעוגנת בתיאוריה ומשלבת את מיטב היסודות של החשיבה הדדוקטיבית והאינדוקטיבית כאחד. התרשים דלהלן ממחיש כיצד ספרים העוסקים בשיטות מחקר עשויים לתאר סוג זה של חשיבה: הצגת תהליך המחקר בצורת גלגל ולא דווקא כרצף ליניארי הנע מהנחת היסוד, דרך איסוף החומר לעבר ניתוח הנתונים (לדוגמה, Neuman, 1997).



4. סיכום

בעוד התופעות המעניינות סוציולוגים שונות למדי מאלה המעניינות סוציולינגוויסטים, הרי אנו משתמשים במושגים ובתיאוריות דומים, המתפתחים כלי הרף. לדוגמה, סוציולוגים דגולים אחדים הסתמכו על עבודתו המוקדמת של דה-סוסיר, אולי הבולטים שבהם בזיל ברנשטיין (Bernstein, 1971) ופייר בורדיה (Bourdieu, 1982; 1991). מתכנני מעמ"ד שילבו את גישתו התיאורטית של דה-סוסיר עם תיאוריות סוציולוגיות מקובלות בהתמקד מפורשות

- בהבדלי מעמדות חברתיים בשימוש בשפה
- ובשפה כצורה של "הון תרבותי" המשמש את חברי העלית כסמך מעמדי.

- Douglas, Mary. 1996. *Purity and Danger: An Analysis of Concepts of Pollution and Taboo*. New York: Praeger.
- . 1994. The Stranger in the Bible. *European Journal of Sociology/Archives Européennes de Sociologie* 35: 283–298.
- (ed.). 1984. *Food in the Social Order: Studies of Food and Festivities in Three American Communities*. New York: Russell Sage.
- and Steven Tipton (eds.). 1983. *Religion and America: Spiritual Life in a Secular Age*. Boston: Beacon Press.
- Dinneen, Francis. 1990. Ferdinand de Saussure. *Georgetown Journal of Languages and Linguistics* 1: 31–53.
- Fishman, Joshua. 1974. Introduction: The Sociology of Language in Israel. *International Journal of the Sociology of Language* 1: 9–13.
- Gane, Mike. 1983. Durkheim: The Sacred Language. *Economy and Society* 12: 1–47.
- Griffin, Larry. 1995. How is Sociology Informed by History. *Social Forces* 73: 1245–1254.
- . 1992. Temporality, Events, and Explanation in Historical Sociology: An Introduction. *Sociological Methods and Research* 20: 403–427.
- Griswold, Wendy. 2000. *Bearing Witness: Readers, Writers, and the Novel in Nigeria*. Princeton, NJ: Princeton University Press.
- . 1994. *Cultures and Societies in a Changing World*. Pine Forge Press.
- . 1993. Recent Moves in the Sociology of Literature. *Annual Review of Sociology* 19: 455–467.
- . 1992. The Sociology of Culture: Four Good Arguments (and One Bad One). *Acta Sociologica* 35: 323–328.
- Johnson, Cathryn. 1994. Gender, Legitimate Authority, and Leader-Subordinate Conversations. *American Sociological Review* 59: 112–135.
- Kmenta, Jan. 2000. 2nd ed. *Elements of Econometrics*. Ann Arbor: University of Michigan Press.
- Kollock, Peter, Philip Blumstein and Pepper Schwartz. 1985. Sex and Power in Interaction: Conversational Privileges and Duties. *American Sociological Review* 50: 34–46.
- Lareau, Annette. 1989. *Home Advantage*. London: Falmer Press.

- Atkins, Sue, Jeremy Clear and Nicholas Ostler. 1992. Corpus Design Criteria. *Literary and Linguistic Computing* 7: 1–16.
- Babbie, Earl. 1995. 7th ed. *The Practice of Social Research*. Belmont: Wadsworth.
- Baker, Therese. 1994. 2nd ed. *Doing Social Research*. New York: McGraw-Hill, Inc.
- Berman, Ruth. 1997. Modern Hebrew. In: R. Hetzron (ed.). *The Semitic Languages*. London: Routledge. 312–333
- Bernstein, Basil. 1971. *Class, Codes and Control*. London: Routledge and Kegan Paul.
- Biber, Douglas. 1990. Methodological Issues Regarding Corpus-Based Analyses of Linguistic Variation. *Literary and Linguistic Computing* 5: 257–269.
- . 1993. Representativeness in Corpus Design. *Literary and Linguistic Computing* 8: 243–257.
- , Susan Conrad and Randi Reppen. 1994. Corpus-Based Approaches to Issues in Applied Linguistics. *Applied Linguistics* 15: 169–189.
- Bourdieu, Pierre. 1982. *Ce que Parler Veut Dire: L'économie des échanges Linguistiques*. Paris: Fayard.
- . 1991. *Language and Symbolic Power*. J.B. Thompson (ed.). Cambridge, MA: Harvard University Press.
- Calhoun, Craig. 1994. *Social Theory and the Politics of Identity*. Oxford: Blackwell.
- Carli, Linda. 1990. Gender, Language and Influence. *Journal of Personality and Social Psychology* 56: 565–576.
- Crowdy, Steve. 1993. Spoken Corpus Design. *Literary and Linguistic Computing* 8: 259–265.
- De Beaugrande, Robert. 1999. Linguistics, Sociolinguistics, and Corpus Linguistics: Ideal Language versus Real Language. *Journal of Sociolinguistics* 3: 128–139.
- Denzin, Norman and Yvonna Lincoln (eds.). 1994. *Handbook of Qualitative Research*. Thousand Oaks: Sage Publications.

Remarks on the Problem of Social Order. *Zeitschrift für Soziologie* 19: 13–25.

Zaretsky, Eli. 1994. Identity Theory, Identity Politics: Psychoanalysis, Marxism, Post-Structuralism. In: Craig Calhoun (ed.). *Social Theory and the Politics of Identity*. Oxford: Blackwell. Chapter 7: 198–215.

- and Erin Horvat. 1999. Moments of Social Inclusion and Exclusion: Race, Class, and Cultural Capital in Family-School Relationships. *Sociology of Education* 72: 37–53.
- Maltz, Daniel and Ruth Borker. 1982. A Cultural Approach to Male-Female Miscommunication. In: J.J. Gumperz (ed.). *Language and Social Identity*. Cambridge: Cambridge University Press. 196–216.
- McCarthy, Michael. 1999. *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press.
- Mills, C. Wright. 1959. *The Sociological Imagination*. New York: Oxford University Press.
- Nelson, Gayle, Waguida El Bakary and Mahmoud Al Batal. 1993. Egyptian and American Compliments: A Cross-Cultural Study. *International Journal of Intercultural Relations* 17: 293–313.
- Neuman, W. Lawrence. 1997. 3rd ed. *Social Research Methods: Qualitative and Quantitative Approaches*. Boston: Allyn and Bacon.
- O'Barr, William. 1982. *Linguistic Evidence: Language, Power, and Strategy in the Courtroom*. New York: Academic Press.
- Ochs, Elinor. 1992. Indexing Gender. In: A. Duranti and C. Goodwin (eds.). *Rethinking Context: Language as an Interactive Phenomenon*. Cambridge: Cambridge University Press. 335–358.
- Ragin, Charles. 1994. *Constructing Social Research: The Unity and Diversity of Method*. Thousand Oaks: Pine Forge Press.
- . 1998. The Logic of Qualitative Comparative Analysis. *International Review of Social History* 43: suppl. 6: 105–124.
- Reckwitz, Andreas. 1997. Cultural Theory, Systems Theory and the Social-Theoretical Pattern of an Inside-Outside Distinction. *Zeitschrift für Soziologie* 26: 317–336.
- Thorlindsson, Thorolfur. 1987. Bernstein's Sociolinguistics: An Empirical Test in Iceland. *Social Forces* 65: 695–718.
- Thorne, Barrie, Cheri Kramarac and Nancy Hendley (eds.). 1983. *Language, Gender, and Society*. Rowley: Newbury House.
- Wagner, Gerhard. 1990. Emile Durkheim and Ferdinand de Saussure. Some