

משרד המדע והטכנולוגיה

בלשנות חיונית עברית

קובץ מאמרים מימי עיון שנערכו בשנים 1988, 1989, 1990

ע"י משרד המדע והטכנולוגיה

בעריכת: עוזי אורנן
גדעון אריאלי
עידית דורון

1992

הצרוף השמני

עידית דורון
האוניברסיטה העברית בירושלים

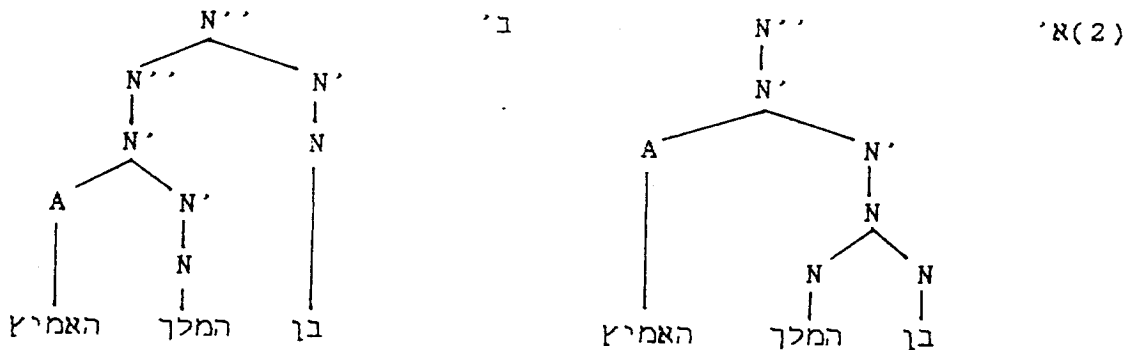
כדי לגשת לעיבוד שפה ממוחשב ברמה של צרופי מילים, הן למטרות ניתוח והן למטרות ייצור, יש להניח קיומו של תחביר מפורש לשפה. מאמר זה הוא נדבך במאמץ, המצוי עדיין בחיתוליו, לנסח תחביר מפורש לשפה העברית. מאמר זה מציע חוקי תחביר עבור מבנים מסוימים בצרוף השמני.

1. לואי התאור

למרות שהדברייכ כבר הובהרו פעמים רבות, מן הראוי לעמוד שוב על החשיבות של חוקי התחביר, במקרה זה תחביר הצרוף השמני, לעיבוד אוטומאטי של שפה. לשם המחשה נטפל בדוגמא מתחום הדו-משמעות. בהנתן הצרוף השמני (1):

(1) בן המלך האמיץ

חשוב שהתוכנה לעיבוד ממוחשב תוכל להציע שני מבנים תחביריים לשם ניתוח צרוף זה. שני המבנים השונים יוכלו לשמש בסיס לדו-משמעות הסמנטית שכל דובר של השפה מזהה אותה כאשר הוא מפרש את הצרוף השמני שב- (1). על פי פרוש אחד, מדובר בנסיך אמיץ, ועל פי הפרוש האחר, מדובר בבנו של מלך אמיץ, כאשר הבן אינו בהכרח אמיץ. הדו-משמעות הזאת היא תולדה של הדו-מבניות המוצגת ב-(2):

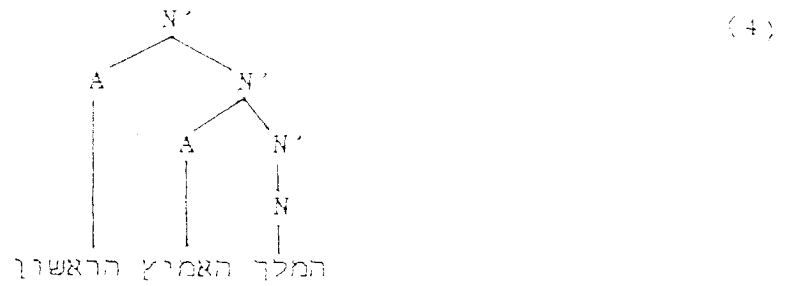


מערכת חוקים שהיא נאותה על פי קנה המידה הבלשני, תתן תשובה הולמת לבעיות של ניתוח אוטומאטי, בכך שתספק את הניתוחים שיסבירו את הדו-משמעות. תת-מערכת נאותה תכיל את החוקים הבאים:

N	N	<---	N	'א(3)
	N	<---	N'	'ב
A	N'	<---	N'	'ג
(N'')	N'	<---	N''	'ד

כלל ('א3') מייצר את הצרוף בן-המלך ב-(2) א', כצמידות, שהקטגוריה שלה N, ומבדילה בינו לבין מבנה סמיכות, שהקטגוריה שלו N', המיוצר ע"י כלל ('ד3') ומופיע ב-(2) ב'. ההבחנה בין סמיכות לצמידות היא של רוזן בספרו "עברית טובה".

כלל (א3) מיוצר לואי תאור באופן רקורסיבי, כפי שנתון למשל בדוגמה הבאה:

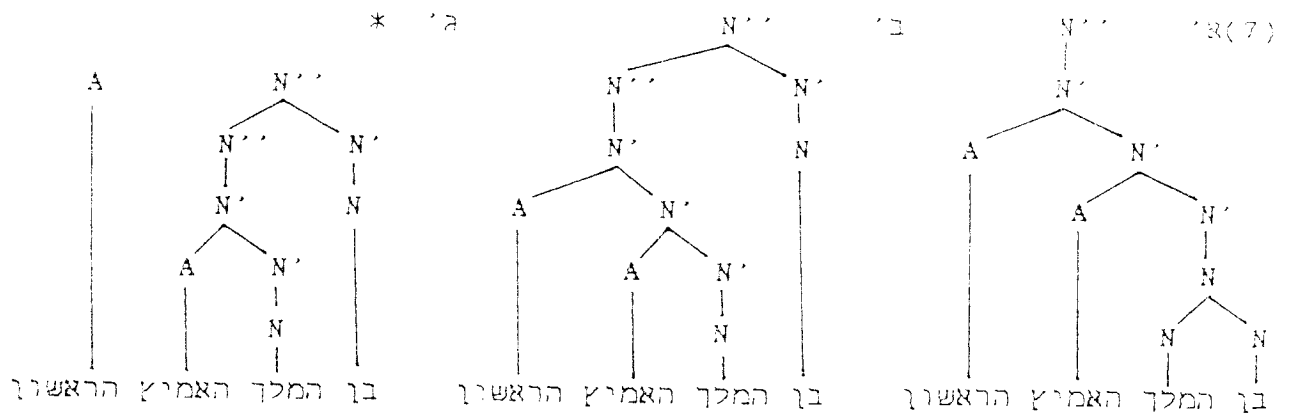


יש לשים לב לכך שלמרות שהיא מאפשרת דו-מבניות מוצדקת עבור דוגמה (1), מערכת הכללים ב-(3) מתוחמת היטב ומגבילה דו-משמעות מבנית למינימום הדרוש. לדוגמה, לואי התאור A מיוצר כצומת אחת ל-N' בלבד, ולא ל-N''. ההגבלה הזאת היא בעלת חשיבות במקומות בהם מספר המשמעויות של צרוף שמני קטן למעשה ממספר המשמעויות האפשריות לוגיות. דוגמה (5), למשל, היא בעלת שתי משמעויות (א5) ו-(ב5), אך אין לה משמעות (ג5), למרות שגם זו אפשרות לוגית:

(5) בן המלך האמיץ הראשון

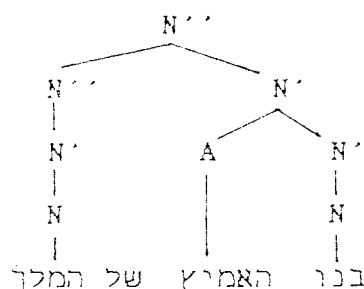
- א5) בנו האמיץ הראשון של המלך
- ב5) בנו של המלך האמיץ הראשון
- ג5) בנו הראשון של המלך האמיץ

הסיבה לכך היא שמבנה (ג7), שמתאים למשמעות (א5) אינו מיוצר ע"י המערכת הנתונה, כיוון שב-(ג7) יש צורך לצרף את A ל-N'', דבר שאין המערכת מאפשרת. לביטוי (5) יש רק המבנים (א7) ו-(ב7), ועל כן רק משמעויות (א5) ו-(ב5):



המערכת הפשוטה שב-(3) היא על כן נאותה מבחינה בלשנית, בכך שאוסף המבנים שהיא מאפשרת, עבור דוגמה (5) למשל, הוא עשיר מספיק אך ביוחד עם זאת גם מתוחם.

כושר היוצור של המערכת ב-(3) כולל גם ביטויים שבהם צרופי הקניין אינם סומכים, אלא מסומנים בסימן היחסה של. אני מניחה, כפי שהוצע ע"י בורר בספר "Parametric Syntax" כי של, בניגוד למילות יחס, אינו מוסיף נדבך בעץ הגזירה, אלא משמש סימן יחסה שאינו משנה את הקטגוריה שאותה הוא מסמן:



בצרוף השמני שב- (8) נעשה שימוש בכלל (ד3'), המאפשר לצרף את לואי האמיץ לשם בנו, שהוא גרעין הצרוף, עוד לפני שמצרפים את המשלים המלך. מה שמיוחד את הברירה הזאת הוא השימוש בסימן היחסה ש במקום בצורת הסמיכות. ההבדל הזה אינו מופיע כחלק מכללי הגזירה, אלא כחלק ממשוואות סימנים נלוות. גם ההתאמה ביידוע בין השם לבין לואי התאור אינה נעשית בגזירה כי אם במשוואות נלוות.

לסיכום, הדיון לעיל מדגים כיצד פועלת ההנמקה הבלשנית לנאותות של כללי גזירה. במקרה זה, נומקה נאותות גזירתו של לואי-התאור כצומת אחות ל-N', אך לא ל-N''.

2. לואי הכמות

מערכת הכללים הפשוטה שב- (3) מבחינה בין סמיכות לצמידות, ומייצרת לואי-תאור וצרופי קניין במקומות הנכונים, אך חסרה עדיין כללים לייצור כמתים. לפני שניגש לניסוח הכללים לייצור כמתים, ברצוני להציע סוג של הכמתים המופיעים בצרופים שמניים לשתי קבוצות על פי תכונותיהם.

נפתח בדוגמאות אחדות של צרופים שמניים מכומתים. הדוגמאות האלה לקוחות מתוך הדיסטרציה של אורנג', אולם כאן נסווג אותן לשתי קבוצות, (9'א') ו-(9'ב'):

(9'א')

שלושה אבות/ שלושת האבות
יותר בעיות
רבע שעה
הרבה חלב
מעט תשומת לב
כמה אנשים
חלק מהם
ק"ג עגבניות
מטר בד

(9'ב')

מרבית הלילה
רוב האנשים
מחצית המשחק
עיקר דבריך
שאר העבודה
מבחר הקצינים
מיטב הנוער
יתר הקהל
כל הילדים
כל ילד
אותם האנשים

הצרופים השמניים המכומתים לעיל טווגו על פי הכונות הכמת. נסמן את הכמתים שבקבוצה (A9) ודומיהם בסימן Q, Quantifier, ואת אלה שבקבוצה (B9) ודומיהם בסימן D, Determiner, ונציה את הנימוקים המכתובים את חלוקת הכמתים לשתי קטגוריות.

ההבחנה הראשונה מבוססת על כך שכמתי Q יכולים לרצטרף ל-A גם באמצעות מילת היחס מן, ואין הדבר נכון עבור כמתי D:
(10) A' שלושה מתלמידיו של ק"ג מהעבניות

ב' * רוב מתלמידיו
* שאר מהעבודה

ההבחנה השניה מבוססת על כך שכמתי Q קובעים את ההתאמה במין ובמספר של הצרוף השמני כולו שאותו הם מכמתים (בתנאי כמובן שיש להם סימני התאמה ואין הם תארי-פעל כגון הרבה, מעט). כמתי D, לעומת זאת, לעולם לא יקבעו את סימני ההתאמה של הצרוף השמני כולו:

(11) A' ק"ג עגבניות עולה/עולות 2 ש"ח
מטר מלמלה יעלה/אתעלה החל ממטר 40 ש"ח
שלוש מאתנו יושבות/איושבים על הספסל

ב' מרבית הבחורות אוהבות/אוהבת להתחבק עם בני-זוגן/*בן-זוגר
כל הבחורות אוהבות/אוהבים להתחבק עם בני-זוגן/*זוגם

ההבחנה השלישית מבוססת על כך שכמתי Q יכולים להיות מלווים או מכומתים בעצמם, ואילו כמתי D לעולם אינם מלווים או מכומתים. בחוקי הגזירה שבהם נגזר את Q, יופיעו על כן כמתי Q כחלק מצרוף כמות Q, ואילו כמתי D יופיעו שלא כחלק מצרוף, כפי שמודגם ב-(12):

(12) A' [הרבה מאוד] חלב
[אלפים רבים] של אנשים
[פי שניים יותר] בעיות
[מעט מידי] תשומת לב

ב' * [הכול הרב] של הילדים
* [פי שניים מבתר] הקצינים

ישנן דוגמאות הסותרות לכאורה הכללה זו:
(13) הרוב המכריע של הכיתה
המתצית הראשונה של הסרט
החצי המשכיל של העם

בדוגמאות אלה נראה כביכול כמת D כגון רוב או מתצית כשהוא מלווה בתאור. אולם אין הדבר כך, כיוון שבכל מקרה עלינו לקטלג את הכמתים רוב, חצי ומתצית גם בקטגוריה Q. כמו כמתי Q אחרים, הם יכולים לקבוע את סימני ההתאמה של הצרוף כולו:

(14) הרוב המכריע של הכיתה תמך/אתמכה במורה.
המתצית הראשונה של הסרט נחרטה/אנחרט בזכרונו.

דוגמה נוספת הסותרת לכאורה את חוסר היכולת לכמת את D מופיעה ב-(15):

(15) כמעט כל הילדים

אולם בדוגמא זו מצורף כמעט לצרוף השמני כולו, ולא ל-D, כפי שניתן לראות מן הניגוד ב-(16):

(16) א' * פגשתי את כמעט כל הילדים
ב' פגשתי כמעט את כל הילדים

הדבר מלמד כי מבנה הצרוף ב-(15) הוא (17א') ולא (17ב'):

(17) א' כמעט [כל הילדים]
ב' * [כמעט כל] הילדים

לעומת זאת, עבור Q אפשרי מבנה גם כמו ב-(18):

(18) [כמעט אלף] ילדים

ההבחנה הרביעית מבוטסת על כך שכמתי Q יכולים להופיע במקום צרוף שמני, ואילו כמתי D אינם יכולים:

(19) א' יש לנו רוב/ מחצית/ שלושה/ יותר/ רבע/ חצי/ הרבה/ כמה/ מעט/ פי
שניים/ חלק/ מספיק/ המון/ כפול/ הכי הרבה

ב' * יש לנו מרבית/ עיקר/ שאר/ אותם/ מבוחר/ כל/ אף/ שום

הבחנה חמישית בין Q ל-D מבוטסת על כך שבכל צרוף שמני מופיע לכל היותר Q יחיד, ולעומת זאת יכולים להופיע מספר כמתי D - וכולם יקדימו את Q:

(20) א' כל שאר מאות המפגינים
שאר כל מאות המפגינים

ב' מרבית אותם עשרת התלמידים
אותם מרבית עשרת התלמידים

כפי שמראות הדוגמאות ב-(20), ייתכנו סדרים הדדיים שונים של כמתי D, אולם (21) מראה כי עליהם להקדים כולם את Q:

(21) א' * מאות כל שאר המפגינים
* שאר מאות כל המפגינים

ב' * אותם עשרת מרבית התלמידים
* עשרת אותם מרבית התלמידים

הבחנה ששית בין D ל-Q היא כי D הוא בהכרח נסמך לשאר הצרוף השמני, ואילו Q אינו בהכרח נסמך, אפילו כשאין דבר המצטרף ל-Q בתוך QP. כך למשל שמות המספר הבלתי מיוחדים בדרך כלל אינם נסמכים, כפי שרואים ב-(22א'), אם כי הם ייתכנו גם בסמיכות, ראה (22ב'):

(22) א' ארבעה ילדים
ב' ארבעת אלפים

וכן ביטויי כמות אחרים גם כן אינם בהכרח נסמכים: (אם כי סמיכות תיתכן עבור ביטויים אלה כאשר הם מקוטלגים כ-N)

(23) א' מנה חומוס
ב' מנת הרואין

