

עיבוד שפות טבעיות – תרגיל בית 5

חלק א' – מימוש CYK

בחלק זה מימשת את האלגוריתם כפי שראינו בהרצאה.

יצרתי את המחלקה Node, שמתארת צומת בעץ הגזירה של ה-CFG. כל Node מכיל:

- symbol – ה-POS של הצומת
- probability – ההסתברות המקסימלית של תת העץ
- child1 – הבן הראשון של הצומת
- child2 – הבן השני של הצומת

יצרתי מחלקת Parser שתהיה אחראית על הפרסור באמצעות CYK. המחלקה מכילה טבלה תלת מימדית

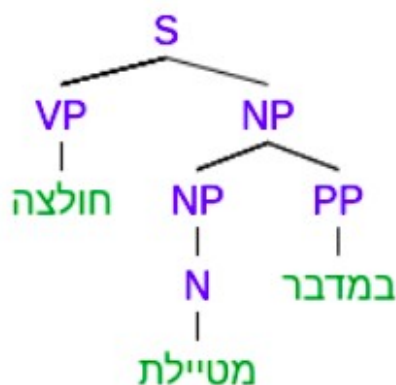
בגודל $|G| \times (n+1) \times N$, כל שזוהי בעצם טבלה דו מימדית, ובכל תא יש Node המתאים ל-POS, כאשר אנחנו שומרים את ההסתברות הגבוהה ביותר, ולפיה בונים את שאר העץ.

לאחר שהאלגוריתם סיים עם בניית הטבלה, ניתן לקבל את עץ הגזירה ע"י סריקת העץ מן האיבר S בטבלה באופן רקורסיבי.

חלק ב' – משפטי Garden Path

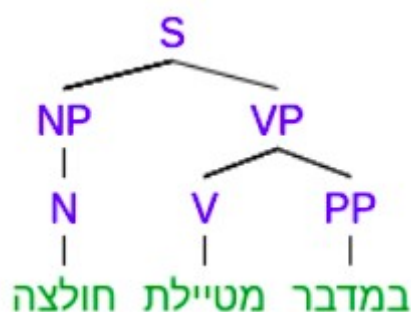
1. הנה עץ הגזירה שבו המשפט מובן כ-A traveler was rescued in the desert:

[[VP] S [[V חולצה] NP] NP] N מטיילת] PP במדבר]



הנה עץ הגזירה שבו המשפט מובן כ-A shirt is traveling in the desert:

S] NP] N חולצה] VP] V מטיילת] PP במדבר]



ניתן לראות שהעצים שונים במבנה שלהם, בעיקר כי באחד מהם מטיילת מובן כפועל, ובאחר כשם עצם (וגם חולצה בהתאם).

2. חשבתי על המשפטים הבאים:

1. I convinced her children are noisy

ניתן לראות את האפקט בכך שאפשר להבין את her children בשני אופנים:

- ניתן לייחס את her ל-children, ואז נבין שמדובר בילדים שלה. ואז אין הגיון תחבירי במשפט, כי אנחנו מצפים למשהו שיהיה מיוחס ל-are.
- ניתן לייחס את her ל-convinced, כלומר, הדובר שכנע אותה (ולא את הילדים שלה, כמו במשפט הקודם). במצב כזה יש הגיון תחבירי, וכוונת המשפט היא שהדובר שכנע אותה שילדים הם מעצבנים.

2. The old man the boat

- פה ניתן לפרש את המילה old בשני מובנים:
- The old, כלומר, הכוונה היא הכללה לאנשים זקנים. במקרה זה, המשפט אומר שאנשים זקנים מתחזקים (הפועל man) סירות.
 - The old man – הכוונה לאיש זקן מסוים. במקרה זה אין הגיון תחבירי למשפט, שכן אחרי old man אנחנו מצפים לפועל.
- משפטי Garden Path יכולים לאתגר מודל של שפה כי שנוצר לנו מצב של Ambiguity. אמנם בחלק מהמשפטים אין הגיון תחבירי אם נבין אותם בצורה אחרת (כמו המשפטים הנ"ל), אבל בחלקם יש, ואפשר להבין את המשפט בשתי דרכים שונות ונכונות תחבירית (למשל, חולצה מטיילת במדבר), כך שלמודל השפה לא ברורה כוונת המשורר.

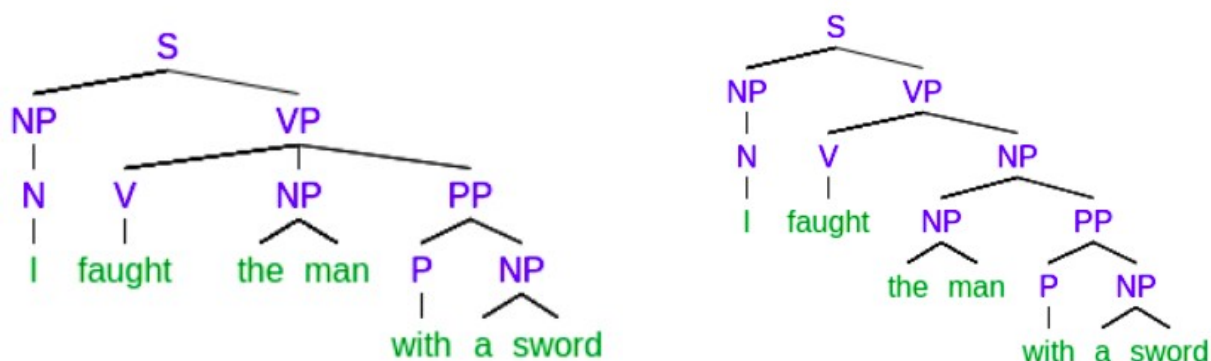
3. לקחתי שני משפטים דו משמעיים.

דו משמעות סינטקנטי:

I faught the man with a sword.

- לא ברור אם sword מתייחס לאיש, כלומר, שהאיש החזיק חרב (1), או שהדובר נלחם באיש באמצעות חרב (2). זהו דו משמעות סינטקנטי מכיוון שלא ברור השיוך של המילה sword.

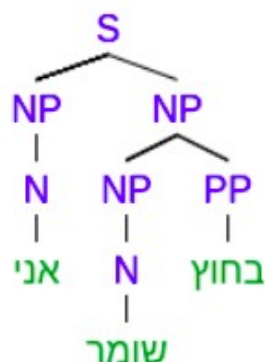
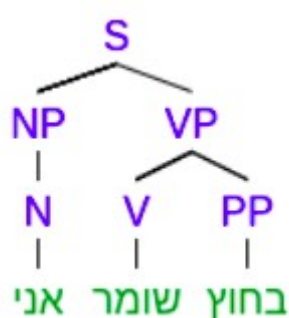
ניתוח של sword עבור (1) מצד ימין, ושל (2) מצד שמאל:



דו משמעות לקסיקלי:**אני שומר בחוץ.**

המילה שומר יכולה להיות מפורשת כשם עצם (guard), או כפועל (guarding). מדובר בדו משמעות לקסיקלי, כי לא ברור מהו ה-POS של "שומר".

ניתוח של שומר כפועל (ימין) וכשם עצם (שמאל):



דוגמה נוספת:

Flying planes can be dangerous.

יש כאן דו משמעות לקסיקלי מכיוון שלא ברור אם Flying הוא שם תואר או פועל.

1. אם הוא שם תואר, כוונת המשפט היא שמטוסים מעופפים הם מסוכנים – כלומר,

המטוסים עצמם.

2. אם הוא פועל, כוונת המשפט היא שהטסת מטוסים היא דבר מסוכן – פעולת

ההטסה עצמה.

עץ הגזירה של 1 משמאל, ושל 2 מימין:

