

אקדמיה אופטימלית

יובל פינטר
התוכנית הבינתחומית ע"ש עדי לאוטמן
מאי 2009

תחום המחקר

● המטרה: לבדוק האם ניתן לחזות הצלחה של חידוש בעברית מודרנית על בסיס תיאוריה פונולוגית (אופטימליות)

● תיחום:

- מילים מ-50 השנים האחרונות
- בעיקר חידושי אקדמיה
- מילים שידועות לציבור הרחב
- מוכר שימוש בצורה הלועזית

● נסיון למצוא מילים עם יותר משני מועמדים

תיאוריית האופטימליות

◎ קיים סט אילוצים אוניברסלי

◎ שפות העולם נבדלות ע"י דירוג האילוצים

◎ קלט נכנס למערכת החישובית, מיוצרים מועמדים שעוברים בדירוג האילוצים כמו במסננות; האחרון שנשאר הוא הפלט

◎ סוגי אילוצים


• מוסמנות (Markedness) – היררכיות כלליות, נוגעות לקושי לבטא ולתפוס צלילים ורצפי צלילים

• נאמנות (Faithfulness) – הצורך ליצור בשפה ניגוד בין משמעויות

אופטימליות – דוגמה

- אנגלית – סיומת גוף שלישי בפועל בהווה
- נתון 1: עיצור אחרון קולי – סיומת z
- נתון 2: עיצור אחרון בלתי-קולי – סיומת s

input:	/brag+z/	(M) קוליות של שני עיצורים רצופים זהה	(F) שמירה על קוליות ה"שורש"	(F) שמירה על קוליות המוספית
	bragz			
	brags	*!		*

input:	/breik+z/	(M) קוליות של שני עיצורים רצופים זהה	(F) שמירה על קוליות ה"שורש"	(F) שמירה על קוליות המוספית
	breikz	*!		
	breiks			*
	breigz		*!	

מסגרת עבודה

- Comparative Tableaux (Tesar 2002)
- אלגוריתם שמיועד במקור להסביר יכולת רכישה:
 - רושמים זוגות מנצח-מפסיד (W-L pairs) בראש שורות הטבלה
 - רושמים את האילוצים בראש הטורים
 - רושמים ברובריקה W / L לפי המילה מהזוג אותה "בוחר" האילוץ (או e אם יש תיקו)
 - אילוץ שלא בוחר אף L מועבר לראש רשימת האילוצים ומוצא מהטבלה
 - זוגות המילים בהן האילוץ בחר W מוצאים מהטבלה
 - וחוזר חלילה עד קבלת דירוג במרקם אחיד

קביעת המנצחים

⊙ סקר (מייל)

- כ-60 נשאלים בגילאי 20-70
- 12 זוגות ושלוש מילים:
 - האם מוכרת לכם המילה במשמעות זו?
 - האם אתם משתמשים במילה בדיבור רגיל?
- נוסח בלשון רבים בלי שבכלל שמתילב

⊙ המילים שנבחרו ($L < L, L$):

בייטים < בתים, בייטים; עמלה < קומישן, קומיסיון; קרן < קורן, קורנר;
מונית < טקסי; פנצ'ר < תקר; כָּבֵל < כָּבֵל; אסאמאס < מסרון;
גיבוי < בקאפ; להיט < היט; בדיון < פיקציה, פיקשן; יורו < אירו;
הנגאובר < חמרמורת; דיסק < סי-די, תקליטור; תייר < טוריסט.

אילוצים מוצעים

● מודל בעל ארבע קטגוריות:

● **M: מסומנות פונולוגית**

● IAMB, TROCHEE – מערכת הטעם

● *COMPLEX – אין צרור עיצורים בראש הברה או בזנבה

● *CODA – אין זנב הברה

● PRWD=FT – מילה היא בת שתי הברות בדיוק

● **F₁: נאמנות לצורת המקור (מותאמת אקוסטית)**

● MAX-C – אין השמטת עיצורים.

● DEP-C – אין החדרת עיצורים.

● IDENT-C₁ – כל עיצור בפלט תואם את עיצור הקלט המתאים לו.

אילוצים מוצעים – המשך

⦿ F_2 : נאמנות לשורש המוצע

- IDENT- C_2 – כל עיצור בפלט תואם את עיצור הקלט השורשי המתאים לו
- SEM-FIELD – קיימות מילים בשפה בעלות צליל דומה ובאותו שדה סמנטי

⦿ דמיון למבנה העברית (המדוברת)

- OUT-PARADIGM – המילה היא במשקל הקיים בעברית, ובעל יכולת יצרנית
- OUT-SUFFIX – למילה יש מוספית הקיימת בעברית ומתאימה לקטגוריה הלקסיקלית
- OUTPUT-CLASH – לא קיימת מילה זהה בשפה
- OUTPUT-CLASH+CATEGORY – לא קיימת בשפה מילה זהה בעלת אותה קטגוריה לקסיקלית (שם / פועל / ...)

⦿ סה"כ 19 זוגות מילים, 14 אילוצים

הטבלו...

Input-L1	Winner	Loser	IAMB	* CX	* CODA	PRWD = FT	MAX-C1	DEP-C1	IDENT-C1	IDENT-C2	SEMFIELD	* OUT-CLASH+ CAT	OUT-SUFFIX	OUT-MISHKAL
/bajt + PL/	baj.tim	ba.tim	L	e	L	e	W	e	W	W	e	W	e	L
/bajt + PL/	baj.tim	bajts	L	W	W	W	e	L	L	e	e	e	W	e
/kor.ner/	ke.ren	kor.ner	e	e	W	e	e	e	L	e	W	L	e	W
/tak.si/	mo.nit	tak.si	W	e	e	e	L	L	L	W	W	e	W	W
/pan.čer/	pan.čer	te.ker	e	e	L	e	W	W	W	L	L	e	e	L
/ka.bel/	ka.bel	ke.vel	e	e	e	e	e	e	W	L	L	e	e	L
/e.se.mes/	e.se.mes	mis.ron	e	e	W	L	W	W	W	L	L	e	L	L
/be.kap/	gi.buy	be.kap	W	e	e	e	L	L	L	e	W	L	e	W
/hit/	la.hit	hit	e	e	e	W	e	L	e	W	W	e	e	W
/fik.šen/	bid.yon	fik.šen	W	e	e	e	L	L	L	W	W	e	W	W
/fik.tsya/	bid.yon	fik.tsya	W	W	L	e	L	L	L	W	W	e	e	W
/yu.ro/	yu.ro	e.ro	e	e	e	e	W	e	W	L	W	e	e	e
/disk/, /si.di/	disk	si.di	W	L	L	L	e	e	e	e	W	L	e	e
/disk/	disk	tak.li.tor	L	L	e	e	W	W	W	L	e	L	e	e
/tu.rist/	ta.yar	tu.rist	e	e	W	e	e	L	L	e	e	e	e	W

...ושברו

Input-L1	Winner	Loser	IAMB	* CX	* CODA	PRWD = FT	MAX-C1	DEP-C1	IDENT-C1	IDENT-C2	SEM-FIELD	* OUT- CLASH+ CAT	OUT- SUFFIX	OUT- MISHKAL
/bajt + PL/	baj.tim	ba.tim	L	e	L	e	W	e	W	W	e	W	e	L
/bajt + PL/	baj.tim	bajts	L	W	W	W	e	L	L	e	e	e	W	e
/kor.ner/	ke.ren	kor.ner	e	e	W	e	e	e	L	e	W	L	e	W
/tak.si/	mo.nit	tak.si	W	e	e	e	L	L	L	W	W	e	W	W
/pan.čer/	pan.čer	te.ker	e	e	L	e	W	W	W	L	L	e	e	L
/ka.bel/	ka.bel	ke.vel	e	e	e	e	e	e	W	L	L	e	e	L
/e.se.mes/	e.se.mes	mis.ron	e	e	W	L	W	W	W	L	L	e	L	L
/be.kap/	gi.buy	be.kap	W	e	e	e	L	L	L	e	W	L	e	W
/hit/	la.hit	hit	e	e	e	W	e	L	e	W	W	e	e	W
/fik.šen/	bid.yon	fik.šen	W	e	e	e	L	L	L	W	W	e	W	W
/fik.tsya/	bid.yon	fik.tsya	W	W	L	e	L	L	L	W	W	e	e	W
/yu.ro/	yu.ro	e.ro	e	e	e	e	W	e	W	L	W	e	e	e
/disk/, /si.di/	disk	si.di	W	L	L	L	e	e	e	e	W	L	e	e
/disk/	disk	tak.li.tor	L	L	e	e	W	W	W	L	e	L	e	e
/tu.rist/	ta.yar	tu.rist	e	e	W	e	e	L	L	e	e	e	e	W

סטטיסטיקה?

נסיונות להסיר מהטבלה מקרים גבוליים לא צלחו

- במקרה הכי טוב ללא שני זוגות מועמדים, האלגוריתם נתקע אחרי שלוש איטרציות

נוור סיי נוור

⊙ על סמך המידע שאספתי והנסיונות שניסיתי, אין לתיאוריית האופטימליות דרך לחזות הצלחה או כשלון של חידוש אקדמיה.

⊙ איך בכל זאת ניתן להמשיך? (וניתן!)

- לנסות עם עוד אילוצים קיימים
- לייצר "אילוצים סוציולוגיים"
- ליצור אילוצים מחוברים (קוניונקציות)
- לאנדקס אילוצים (ריבוד)
- לאמץ מודל סטוכאסטי

קווסטטיינז?