

שפה טבעית מנקודת מבט חישובית (67127)

פרופ' ארי רפופרט

מועד א', 13-07-2015

משך הבחינה: שעה וחצי.

יש לענות על כל השאלות.

שאלה 1 (60 נקודות)

מה הדברים העיקריים שלמדנו בקורס על דרך פעולת המוח, השוני בינה לבין דרך פעולת המחשב, וכיצד ממומשת שפה במוח?
יש לכתוב לא יותר מארבע פסקאות כשבכל פסקה לכל היותר 100 מילים.

שאלה 2 (40 נקודות)

נעסוק באלגוריתם לתרגום מאנגלית לצרפתית, `TRANSLATE_WORD_TO_WORD`. בהנתן סדרת הטוקנים:
`a big eagle flies in the sky`.
נרצה שהאלגוריתם `TRANSLATE_WORD_TO_WORD` יחזיר את סדרת הטוקנים:
`un grand aigle vole dans le ciel`.
באופן כללי, אלגוריתם זה תמיד יחזיר סדרה באורך זהה לאורך סדרת הקלט.

i. (10 נקודות) תנו דוגמה למשימה נוספת שנלמדה בקורס, אשר אלגוריתם שפותר אותה מחזיר גם הוא סדרה באורך זהה לאורך סדרת הקלט שלו.

לאלגוריתם שלושה פרמטרים: בנוסף לסדרת הקלט (נקרא לה בשם `TOKENS`, והיא סדרת מחרוזות), האלגוריתם יקבל טבלה בשם `DICTIONARY` וטבלה בשם `TRANSITIONS`. שתי הטבלאות ממפות זוגות של מחרוזות למספרים. לצורך השאלה נניח שהטבלאות כבר נתונות ולא צריך ליצור אותן.

הטבלה `DICTIONARY` מתקבלת מספירת המופעים בקורפוס מתורגם מאנגלית לצרפתית, שבהם מלה מסוימת באנגלית מתורגמת למלה מסוימת בצרפתית. דוגמאות לשורות בטבלה (מלה באנגלית, מלה בצרפתית ומספר):

["big", "grand"]	86
["big", "grande"]	74
["sky", "ciel"]	60
["flies", "vole"]	9
["flies", "mouches"]	12
["eagle", "aigle"]	16

הטבלה TRANSITIONS מתקבלת מספירת המופעים באותו קורפוס, שבהם מלה מסוימת בטקסט שתורגם לצרפתית באה אחרי מלה אחרת באותו טקסט. דוגמאות לשורות בטבלה (זוג מלים בצרפתית ומספר):

["un", "grand"]	26
["aigle", "vole"]	15
["aigle", "mouches"]	1
["le", "ciel"]	22
["le", "monde"]	114

האלגוריתם **TRANSLATE_WORD_TO_WORD** ישתמש בפונקציית עזר בשם **TRANSLATE_SINGLE_WORD**, אשר מקבלת ארבעה פרמטרים:

- **TOKEN**, המלה הנוכחית באנגלית שיש לתרגם;
 - **LAST_TRANSLATED_TOKEN**, המלה בצרפתית שהיא התרגום למלה הקודמת בסדרה;
 - **DICTIONARY**, כמו ב-**TRANSLATE_WORD_TO_WORD**;
 - **TRANSITIONS**, כמו ב-**TRANSLATE_WORD_TO_WORD**.
- פונקציית העזר **TRANSLATE_SINGLE_WORD** מחשבת מהו התרגום הסביר ביותר ל-**TOKEN** לצרפתית, באמצעות חישוב ציון לכל תרגום אפשרי ומציאת התרגום בעל הציון המירבי. הציון מחושב כמכפלת הספירה מהטבלה **DICTIONARY** והספירה מהטבלה **TRANSITIONS**.

ii. (15 נקודות) נניח שהשורות הבאות הן כל השורות בטבלה **DICTIONARY** שמתחילות ב-"flies":

["flies", "vole"]	9
["flies", "mouches"]	12

ונניח שהשורות הבאות הן כל השורות בטבלה **TRANSITIONS** שמתחילות ב-"aigle":

["aigle", "vole"]	15
["aigle", "mouches"]	1

מה תחזיר הפונקציה **TRANSLATE_SINGLE_WORD** אם ערך הפרמטר **TOKEN** הוא "flies" וערך הפרמטר **LAST_TRANSLATED_TOKEN** הוא "aigle"? פרטו את החישוב שמתבצע.

iii. (15 נקודות) בפסודו-קוד הבא לאלגוריתם **TRANSLATE_WORD_TO_WORD** נפלו שלוש טעויות. מהן?

TRANSLATE_WORD_TO_WORD(TOKENS, DICTIONARY, TRANSITIONS):

TRANSLATED_TOKENS $\leftarrow$ []

LAST_TRANSLATED_TOKEN $\leftarrow$ "<s>"

for each TOKEN in TOKENS:

 LAST_TRANSLATED_TOKEN $\leftarrow$ TRANSLATED_TOKEN

 TRANSLATED_TOKEN $\leftarrow$ **TRANSLATE_SINGLE_WORD**(TOKEN)

add TRANSLATED_TOKEN to TRANSLATED_TOKENS

return TRANSLATED_TOKENS

iv. בונים (10 נקודות): כתבו פסודו-קוד לפונקציית העזר **TRANSLATE_SINGLE_WORD**.