
האוניברסיטה העברית בירושלים

סילבוס

נושאים בבלשנות חישובית ובלשנות קורפוס - 36635

תאריך עדכון אחרון 05-10-2015

נקודות זכות באוניברסיטה העברית: 4

תואר: בוגר

היחידה האקדמית שאחראית על הקורס: בלשנות

השנה הראשונה בתואר בה ניתן ללמוד את הקורס: 0

סמסטר: שנתי

שפת ההוראה: עברית

קמפוס: הר הצופים

מורה אחראי על הקורס (רכז): ד"ר עינת רובינשטיין

דוא"ל של המורה האחראי על הקורס: aynat.rubinstein@mail.huji.ac.il

שעות קבלה של רכז הקורס: בתיאום מראש

מורי הקורס:

תאור כללי של הקורס:

זמינותם של קורפוסים דיגיטליים רבים ומגוונים ופיתוחם של אלגוריתמים מתוחכמים לעיבוד אוטומטי של קורפוסים אלה הגדילו באופן ניכר את מגוון השאלות הבלשניות שנחקרות בהסתמך על נתונים מקורפוסים. בקורס זה נעסוק בשאלות הנוגעות לשינוי לשוני, גרמטיקליזציה ומגע בין שפות, תוך התמקדות במקרה הפרטי של התהוות העברית החדשה. הקורס מהווה מבוא תאורטי ומעשי למחקר מבוסס-קורפוס ולמודלים חישוביים של שינוי לשוני, ודן ביישומם לחקר העברית.

מטרות הקורס:

מטרת הקורס ללמד שיטות מחקר בבלשנות קורפוס ולהדגים את תרומתן לסיווג שפות ולמודלים חישוביים של שינוי לשוני. הסטודנטים יונחו שלב אחר שלב ביצירתו של קורפוס דיגיטלי חדש של טקסטים בעברית. במהלך השיעור יעשה שימוש בקורפוס זה כדי לחקור מאפיינים דקדוקיים שונים של העברית בעת התגבשותה בתחילת המאה ה-20.

תוצרי למידה

בסיומו של קורס זה, סטודנטים יהיו מסוגלים:

- לערוך חיפושים במאגרים טקסטואליים אינטרנטיים קיימים באנגלית ובעברית
- להשתמש בחוקי ה-TEI כדי לייצג את תוכנו של מסמך דיגיטלי
- לבנות קורפוסים דיגיטליים חדשים הכוללים מידע לשוני על הטקסט
- לתאר את היבטיו השימושיים של מחקר מבוסס-קורפוס בבלשנות וכן את מגבלותיו
- לנסח השערות ולבחון שאלות תאורטיות באמצעות מחקר קורפוס
- לנתח נתונים לשוניים ברמות מורכבות ושימוש שונות
- לנסח טיעונים בלשוניים ברורים
- לדווח על מחקר בלשוני מבוסס-קורפוס תוך התייחסות לתכנון המחקר, לממצאי המחקר ולמסקנותיו

דרישות נוכחות (%) :

100

שיטת ההוראה בקורס: השיעורים משלבים הוראה פרונטלית עם פעילות אינטראקטיבית והתנסות מעשית. בין היתר יידרשו הסטודנטים להשתתף באופן פעיל בהצעת "שאלות ותשובות" על חומר הקריאה. יתכן סיור לספרייה הלאומית במהלך השנה.

רשימת נושאים / תכנית הלימודים בקורס:

- i. קולוקציות, קונקורדנציות, שכיחויות רשימת, (רגולריים ביטויים) חיפוש: קורפוס בבלשנות שיטות
- ii. תחביר-במורפולוגיה יישומים: קולוקציה עוצמת
- iii. לשונית בטיפולוגיה יישומים: אשכולות מציאת
- iv. וטקסט תמונה סנכרון, נתונים במסדי שאילתות, המונים מיקור: קורפוס של דיגיטציה
- v. סמנטי תיג, תחבירי-מורפולוגי תיג, טוקניזציה: לשוני במידע העשרה

חומר חובה לקריאה:

- Adler, Menahem (Meni). 2007. *Hebrew morphological disambiguation: An unsupervised stochastic word-based approach*. PhD thesis, Ben-Gurion University of the Negev.
- Deo, Ashwini. 2015. *Diachronic semantics*. *Annual Review of Linguistics* 1: 179-197.
- Fadida, Hanna, Alon Itai, and Shuly Wintner. 2014. *A Hebrew verb-complement dictionary*. *Language Resources and Evaluation* 48: 249-278.
- Longobardi, Giuseppe, Cristina Guardiano, Giuseppina Silvestri, Alessio Boattini, and Andrea Ceolin. 2013. *Toward a syntactic phylogeny of modern Indo-European languages*. *Journal of Historical Linguistics* 3: 122-152.
- McMahon, April. 2010. *Computational models and language contact*. In Hickey, R. (ed.), *The Handbook of Language Contact*, 128-147. Wiley-Blackwell.
- Piotrowski, Michael. 2012. *Natural Language Processing for historical texts*. Morgan & Claypool.
- Reshef, Yael. 2015. *Revival of Hebrew: Grammatical Structure and Lexicon*. In Khan, G. (ed.), *Encyclopedia of Hebrew language and linguistics*. Brill Online. <http://referenceworks.brillonline.com/entries/encyclopedia-of-hebrew-language-and-linguistics/revival-of-hebrew-grammatical-structure-and-lexicon-EHLL_COM_00000702>
- Reshef, Yael. 2015. *Revival of Hebrew: Sociolinguistic Dimension*. In Khan, G. (ed.), *Encyclopedia of Hebrew language and linguistics*. Brill Online. <http://referenceworks.brillonline.com/entries/encyclopedia-of-hebrew-language-and-linguistics/revival-of-hebrew-sociolinguistic-dimension-EHLL_COM_00000703>
- Yang, Charles D.. 2000. *Internal and external forces in language change*. *Language Variation and Change* 12: 231-250.
- Zeldes, Amir. 2013. *Is Modern Hebrew Standard Average European? The View from European*. *Linguistic Typology* 17: 439-470.

חומר לקריאה נוספת:

הערכת הקורס - הרכב הציון הסופי :

- מבחן מסכם בכתב/בחינה בעל פה 50 %
הרצאה 0 %
השתתפות 0 %
הגשת עבודה 0 %
הגשת תרגילים 40 %
הגשת דו"חות 10 %
פרויקט מחקר 0 %
בחנים 0 %
אחר 0 %

מידע נוסף / הערות: