



# *The Hebrew University of Jerusalem*

## *Syllabus*

### *Topics in computational and corpus linguistics - 36635*

*Last update 05-10-2015*

*HU Credits: 4*

*Degree/Cycle: 1st degree (Bachelor)*

*Responsible Department: linguistics*

*Academic year: 0*

*Semester: Yearly*

*Teaching Languages: Hebrew*

*Campus: Mt. Scopus*

*Course/Module Coordinator: Dr. Aynat Rubinstein*

*Coordinator Email: [aynat.rubinstein@mail.huji.ac.il](mailto:aynat.rubinstein@mail.huji.ac.il)*

*Coordinator Office Hours: By appointment*

*Teaching Staff:*

---

Dr. Aynat Rubinstein

Course/Module description:

The existence of a wide variety of readily available digital corpora of natural language and the development of sophisticated algorithms to automatically process these corpora have greatly expanded the range of linguistic questions that are currently being addressed using corpus data. This course focuses on questions relating to language change, grammaticalization, and language contact, taking the emergence of Modern Hebrew as a test case. It provides an hands-on introduction to corpus methods and computational models of language change, discussing their applicability to Hebrew.

Course/Module aims:

This course aims to familiarize students with corpus methods in linguistics and to demonstrate their usefulness in studying language classification and change from a computational perspective. Students will be guided step by step in the creation and annotation of a new digital corpus of Hebrew. This corpus will be used in class to study various emergent grammatical properties of Hebrew in the beginning of the 20th century.

Learning outcomes - On successful completion of this module, students should be able to:

- Search existing web-based corpora in English and Hebrew
- Employ the TEI standard to encode the content of a digital document
- Create new digital corpora with linguistic annotation
- Describe the usefulness and limitations of corpus methods in linguistics
- Formulate hypotheses and test theoretical questions using corpus methods
- Analyze linguistic data at various levels of structure and use
- Construct clear linguistic arguments
- Report on the design, findings, and conclusions of corpus-based linguistic research

Attendance requirements(%):

100

Teaching arrangement and method of instruction: Class periods will consist of lectures as well as more interactive and practical ("hands-on") modules. Students are required to participate actively in frequent Q&A sessions about the assigned readings. A visit to the National Library might take place during the year.

---

### Course/Module Content:

- i. Methods in corpus linguistics: search (regular expressions), frequency lists, concordances, collocations
- ii. Collocation strength: applications in morphosyntax
- iii. Clustering: applications in language typology
- iv. Corpus digitization: crowdsourcing, database queries, text-image alignment
- v. Linguistic enrichment: tokenization, morphosyntactic markup, semantic annotation
- vi. Computational models of language change

### Required Reading:

- 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](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](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.

### Additional Reading Material:

### Course/Module evaluation:

End of year written/oral examination 50 %  
Presentation 0 %

---

*Participation in Tutorials 0 %*  
*Project work 0 %*  
*Assignments 40 %*  
*Reports 10 %*  
*Research project 0 %*  
*Quizzes 0 %*  
*Other 0 %*

*Additional information:*