ISCL Hauptseminar (Winter semester 2021, Meurers)

NLP for Computer-Assisted Language Learning

Abstract:

This seminar explores the relevance and uses of NLP in the context of language learning, combining conceptual discussion with hands-on programming tasks.

We distinguish two broad uses of NLP related to language learning (cf. Meurers 2020): On the one hand, NLP can be used to analyze learner language, i.e., words, sentences, or texts produced by language learners. This includes the development of NLP techniques for the analysis of learner language by intelligent tutoring systems in Intelligent Computer-Assisted Language Learning, automated scoring in language testing, as well as the analysis and annotation of learner corpora for researech and learning analytics. On the other hand, NLP for the analysis of native language can also play an important role in the language learning context. Applications in this second domain support the search for and the enhanced presentation of native language reading material for language learners as well as the generation of exercises and tests based on authentic materials.

Complementing the conceptual side, where we delineate the uses of NLP, motivate the relevance, and discuss current research strands, we will develop web-based applications in hands-on sessions. While for these tasks we envisage heterogeneous student teams combining different backgrounds, some programming experience is required for successful participation.

Detmar Meurers (2020). Natural Language Processing and Language Learning. The Concise Encyclopedia of Applied Linguistics, edited by Carol A. Chapelle. Wiley. 817-831. http://purl.org/dm/papers/Meurers-19.pdf

Instructor: Prof. Dr. Detmar Meurers
- *Office:* 1.28 in the SfS
- *Office hours:* Tuesdays 10:00–12:00 (please arrange slot by email beforehand)

Tutor: Valentin Pickard

- *Email:* valentin.pickard@student.uni-tuebingen.de

Course meets: 4 SWS
- Tuesdays and Thursdays 8ct–10 (i.e., 8:15-9:45)
- In lecture room 0.01 in the SfS, Wilhelmstr. 19. If for reasons beyond your control, like visa or health issues, you cannot join in person, you can paticipate in https://zoom.us/j/96719107835.

Credit Points:
- Core Computatonal Linguistics Hauptseminar with 6 or, with term paper, 9 CP.

Syllabus: http://purl.org/dm/21/ws/hs (http://purl.org/dm/21/ws/hs/syllabus.pdf)

Moodle page: https://moodle.zdv.uni-tuebingen.de/course/view.php?id=2150

Please enroll in this course by logging into this moodle course with your ordinary ZDV university login.

Nature of course and our expectations: This is an overview-oriented Hauptseminar, in which we jointly introduce and explore the topic, perspectives and approaches. You are expected to

1. regularly and actively participate in class, with everyone's camera turned on to allow for meaningful two-way interaction in class, read the papers assigned by any of the presenters and post a meaningful question on Moodle to the "Discussion Forum" on each reading *at the latest on the day before it is discussed* in class.

2. explore and present a topic:
   - select one of the sub-topics by the Nov 5
   - thoroughly research the topic, starting from the references provided
   - prepare the presentation with slides, send them to me by email and discuss them with me in a half hour slot during my office hour *at least a week before the presentation*
   - start a new Moodle thread on the "Discussion Forum" specifying what every course participant should read to prepare for your presentation *a week before your presentation*
   - present and discuss the topic in class

3. if you pursue the 9 CP option, work out a project term paper
   - *by February 4* select a topic and submit a one-page abstract (and a revised version by February 11, 2022)
   - *by April 1, 2022, i.e., by the beginning of the next semester* email the term paper in pdf format to the instructor.
     - Note for Computational Linguistics students: The term paper must be produced in LaTeX using the ACL conference format or the Computational Linguistics journal format; BibTeX must be used for the bibliography.

Academic conduct and misconduct: Research is driven by discussion and free exchange of ideas, motivations, and perspectives. So you are encouraged to work in groups, discuss, and exchange ideas. At the same time, the foundation of the free exchange of ideas is that everyone is open about where they obtained which information. Concretely, this means you are expected to always make explicit when you've worked on something as a team – and keep in mind that being part of a team always means sharing the work.

For text you write, you always have to provide explicit references for any ideas or passages you reuse from somewhere else. Note that this includes text "found" on the web, where you should cite the url of the web site in case no more official publication is available.

Sessions

1. until 2.12: Overview (Detmar)
2. 7.12. Hands-On Discussion
3. 9.12. Exercise Generation (Tanja Heck, Leander Girrbach)

   - to be read: (Perez & Cuadros 2017)
   - (Heck 2021), Language Muse (Burstein et al. 2012, 2014; Burstein & Sabatini 2016; Madnani et al. 2016; Burstein et al. 2017)

4. 14.12. Vocabulary learning (Hoa Do, Detmar Meurers)

   - (Santhi Ponnusamy & Meurers 2021)
   - (Utiyama et al. 2005)
   - (Dela Rosa & Eskenazi 2011; Brown et al. 2005; Pino & Eskenazi 2009; Heilman et al. 2008)
   - measuring (Yoon et al. 2012)
   - (Knowles et al. 2016),"macaronic" texts (Renduchintala 2020)
   - (Michelony 2011)

5. 16.12. NLP in established Language Learning Apps (Ann-Katharina Dick, Kristina Koch)

   - Duolingo, Rosetta Stone, Memrise, Babbel, . . .

6. 21.12. Input Enhancement and Enrichment (Anika Ott, Alexander Schwab)

   - foundational research: (Schmidt 1995; Sharwood Smith 1993; Han et al. 2008)
   - computational linguistic approaches: (Meurers et al. 2010; Chinkina & Meurers 2017; Ziegler et al. 2017), (De Hertog et al. 2017)

7. 23.12. Hands-On Discussion
8. 11.1. Systems teaching pronunciation, prosody (Declan O'Donnell, Valentin Pickard)

   - overview: (Agarwal & Chakraborty 2019; Cucchiarini & Strik 2017)
   - (Chen & Li 2016; Thomson 2011; Neri et al. 2002; Witt 2012)
   - suprasegmental properties: (Bozorgian & Shamsi 2020)
   - high variability training: (Thomson 2018)

9. 13.1. Spoke language systems (e.g., dialog systems) (Qin Gu) + Hands-On

   - literature pointers: cf. above

10. 18.1. Readability/Linguistic Complexity Analysis (Mourhaf Kazzaz, Hao Lu, Leixin Zhang)
11. 20.1. Systems teaching argumentation, coherent writing (Jan-Felix Klumpp, Maren Reiber)
12. 25.1. Question generation for education (Mario Arcione, Benjamin Starzec)

References

Agarwal, C. & P. Chakraborty (2019). A review of tools and techniques for computer aided pronunciation training (CAPT) in English. *Education and Information Technologies* 24(6), 3731–3743.

Bozorgian, H. & E. Shamsi (2020). Computer-assisted pronunciation training on Iranian EFL learners' use of suprasegmental features: A case study. *Computer-Assisted Language Learning Electronic Journal* 21(1), 93–113.

Brown, J. C., G. A. Frishkoff & M. Eskenazi (2005). Automatic question generation for vocabulary assessment. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*. Ann Arbor, MI: Association for Computational Linguistics, pp. 819–826.

Burstein, J., N. Madnani, J. Sabatini, D. McCaffrey, K. Biggers & K. Dreier (2017). Generating language activities in real-time for English learners using language muse. In *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale*. pp. 213–215.

Burstein, J. & J. Sabatini (2016). The Language Muse Activity Palette. In S. A. Crossley & D. S. McNamara (eds.), *Adaptive educational technologies for literacy instruction*, Routledge, pp. 275–280.

Burstein, J., J. Shore, J. Sabatini, B. Moulder, S. Holtzman & T. Pedersen (2012). The language musesm system: Linguistically focused instructional authoring. *ETS Research Report Series* 2012(2), i–36.

Burstein, J., J. Shore, J. Sabatini, B. Moulder, J. Lentini, K. Biggers & S. Holtzman (2014). From Teacher Professional Development to the Classroom: How NLP Technology Can Enhance Teachers' Linguistic Awareness to Support Curriculum Development for English Language Learners. *Journal of Educational Computing Research* 51(1), 119–144.

Chen, N. F. & H. Li (2016). Computer-assisted pronunciation training: From pronunciation scoring towards spoken language learning. In *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*. IEEE, pp. 1–7.

Chinkina, M. & D. Meurers (2017). Question Generation for Language Learning: From ensuring texts are read to supporting learning. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications (BEA)*. Copenhagen, Denmark, pp. 334–344. URL http://aclweb.org/anthology/W17-5038.pdf.

Cucchiarini, C. & H. Strik (2017). Automatic Speech Recognition for second language pronunciation training. In *The Routledge handbook of contemporary English pronunciation*, Routledge, pp. 556–569.

De Hertog, D., F. Cornillie & P. Desmet (2017). Context-aware Automatic Input Enhancement for Language Learners. In J. Colpaert, A. Aerts, R. Kern & M. Kaiser (eds.), *CALL in CONTEXT*. Colpaert, Jozef, Universiteit Antwerpen, pp. 209–214. URL http://call2017.language.berkeley.edu/wp-content/uploads/2017/07/CALL2017_proceedings.pdf.

Dela Rosa, K. & M. Eskenazi (2011). Effect of Word Complexity on L2 Vocabulary Learning. In *Proceedings of the Sixth Workshop on Innovative Use of NLP for Building Educational Applications*. Portland, Oregon: Association for Computational Linguistics, pp. 76–80. URL http://aclweb.org/anthology/W11-1409.

Han, Z., E. S. Park & C. Combs (2008). Textual Enhancement of Input: Issues and Possibilities. *Applied Linguistics* 29(4), 597–618. URL http://applij.oxfordjournals.org/content/29/4/597.abstract.

Heck, T. (2021). Combining Automatic Generation of Form-based Grammar Exercises from Authentic Texts with Language Aware Text Search. Master's thesis, International Studies in Computational Linguistics, University of Tübingen. URL http://purl.org/dm/papers/Heck-21.pdf.

Heilman, M., L. Zhao, J. Pino & M. Eskenazi (2008). Retrieval of Reading Materials for Vocabulary and Reading Practice. In *Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications (BEA-3) at ACL'08*. Columbus, Ohio, pp. 80–88.

Knowles, R., A. Renduchintala, P. Koehn & J. Eisner (2016). Analyzing learner understanding of novel L2 vocabulary. In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*. pp. 126–135.

Madnani, N., J. Burstein, J. Sabatini, K. Biggers & S. Andreyev (2016). Language Muse: Automated linguistic activity generation for English language learners. *Proceedings of ACL-2016 System Demonstrations* pp. 79–84.

Meurers, D., R. Ziai, L. Amaral, A. Boyd, A. Dimitrov, V. Metcalf & N. Ott (2010). Enhancing Authentic Web Pages for Language Learners. In *Proceedings of the 5th Workshop on Innovative Use of NLP for Building Educational Applications (BEA)*. Los Angeles: ACL, pp. 10–18. URL http://aclweb.org/anthology/W10-1002.pdf.

Michelony, A. (2011). Predicting clicks in a vocabulary learning system. In *Proceedings of the ACL 2011 Student Session*. pp. 99–104.

Neri, A., C. Cucchiarini, H. Strik & L. Boves (2002). The pedagogy-technology interface in computer assisted pronunciation training. *Computer assisted language learning* 15(5), 441–467.

Perez, N. & M. Cuadros (2017). Multilingual CALL Framework for Automatic Language Exercise Generation from Free Text. In *Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. Valencia, Spain: Association for Computational Linguistics, pp. 49–52. URL https://aclanthology.org/E17-3013.

Pino, J. & M. Eskenazi (2009). An application of latent semantic analysis to word sense discrimination for words with related and unrelated meanings. In *Proceedings of the Fourth Workshop on Innovative Use of NLP for Building Educational Applications*. pp. 43–46.

Renduchintala, A. (2020). A Language Learning Framework based on Macaronic Texts. Ph.D. thesis, Johns Hopkins University. URL http://jhir.library.jhu.edu/handle/1774.2/63709.

Santhi Ponnusamy, H. & D. Meurers (2021). Employing distributional semantics to organize task-focused vocabulary learning. In *Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications*. Online: Association for Computational Linguistics, pp. 26–36. URL https://www.aclweb.org/anthology/2021.bea-1.3.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (ed.), *Attention and awareness in foreign language learning*, Honolulu, HI: University of Hawaii, pp. 1–63.

Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition* 15, 165–179. URL https://doi.org/10.1017/s0272263100011943.

Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *Calico Journal* 28(3), 744–765.

Thomson, R. I. (2018). High variability [pronunciation] training (HVPT): A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation* 4(2), 208–231.

Utiyama, M., M. Tanimura & H. Isahara (2005). Organizing English Reading Materials for Vocabulary Learning. In *Proceedings of the ACL Interactive Poster and Demonstration Sessions*. pp. 117–120.

Witt, S. M. (2012). Automatic error detection in pronunciation training: Where we are and where we need to go. In *International Symposium on automatic detection on errors in pronunciation*

training.

Yoon, S.-Y., S. Bhat & K. Zechner (2012). Vocabulary profile as a measure of vocabulary sophistication. In *Proceedings of the seventh workshop on building educational applications using NLP*. pp. 180–189.

Ziegler, N., D. Meurers, P. Rebuschat, S. Ruiz, J. L. Moreno Vega, M. Chinkina, W. Li & S. Grey (2017). Interdisciplinary research at the intersection of CALL, NLP, and SLA: Methodological implications from an input enhancement project. *Language Learning* 67, 209–231.