A roadmap connecting NLP research and language learning

Detmar Meurers
LEAD graduate school, Universität Tübingen
Universitetet i Tromsø

Third Workshop on NLP for Computer-Assisted Language Learning
Fifth Swedish Language Technology Conference
Uppsala, Sweden – 13. Novembe 2014

Connecting NLP and Language Learning

Detmar Meurers

Introduction

NLP analysis of learner language
Intelligent Tutoring Systems
Learner Corpora
Writer's aids and testing
Under the hood: Interpreting learner data

Native language analysis
Searching materials
Under the hood: Readability ranking the web

Conclusion/Outlook

References

EBERHARD KARLS
UNIVERSITÄT TÜBINGEN

LEAD Graduate School

The talk is based on:

- Detmar Meurers (2012). Natural Language Processing and Language Learning. *Encyclopedia of Applied Linguistics*, edited by Carol A. Chapelle. Blackwell. 4193–4205.

- Detmar Meurers (2015). Learner Corpora and Natural Language Processing. *The Cambridge Handbook of Learner Corpus Research*, edited by Sylviane Granger, Gaëtanelle Gilquin and Fanny Meunier. Cambridge University Press.

- Luiz Amaral & Detmar Meurers (2011). On Using Intelligent Computer-Assisted Language Learning in Real-Life Foreign Language Teaching and Learning. *ReCALL* 23(1), 4–24.

- Markus Dickinson, Chris Brew & Detmar Meurers (2013). *Language and Computers, Ch. 3: Language Tutoring Systems*. Blackwell.

Connecting NLP and Language Learning

Detmar Meurers

Introduction

NLP analysis of learner language
Intelligent Tutoring Systems
Learner Corpora
Writer's aids and testing
Under the hood: Interpreting learner data

Native language analysis
Searching materials
Under the hood: Readability ranking the web

Conclusion/Outlook

References

EBERHARD KARLS
UNIVERSITÄT TÜBINGEN

LEAD Graduate School

Introduction

- Computers are widely used in foreign language teaching to help learners experience a foreign language & culture.
  - multimedia, web-based TV, chat with native speakers, . . .
  - includes Computer-Assisted Language Learning tools, but such CALL tools do not analyze language

- Despite decades of research, NLP plays no role in real-life foreign language teaching and learning.

- This is striking given that NLP can address real-life needs for **interactive feedback** and **adaptive materials** in
  - schools with increasingly heterogeneous students
  - individualized and distance education
  - life-long learning

Connecting NLP and Language Learning

Detmar Meurers

Introduction

NLP analysis of learner language
Intelligent Tutoring Systems
Learner Corpora
Writer's aids and testing
Under the hood: Interpreting learner data

Native language analysis
Searching materials
Under the hood: Readability ranking the web

Conclusion/Outlook

References

EBERHARD KARLS
UNIVERSITÄT TÜBINGEN

LEAD Graduate School

The need for a roadmap

- NLP supports the analysis of
  - **learner language** to provide feedback, assess capabilities
  - **native language** to search, adapt and enhance materials

- Needed: a roadmap connecting NLP with
  - Second Language Acquisition (SLA) research
    - tasks, instructional interventions, relevance of input/output and meaning, focus-on-form, developmental sequences
  - Foreign Language Teaching and Learning (FLTL)
    - address teacher needs, while keeping them in charge
  - Cognitive Psychology
    - attention, memory, learning, motivation, lab studies
  - Empirical Educational Science
    - intervention studies, real-life evaluation, multi-level modeling

Connecting NLP and Language Learning

Detmar Meurers

Introduction

NLP analysis of learner language
Intelligent Tutoring Systems
Learner Corpora
Writer's aids and testing
Under the hood: Interpreting learner data

Native language analysis
Searching materials
Under the hood: Readability ranking the web

Conclusion/Outlook

References

EBERHARD KARLS
UNIVERSITÄT TÜBINGEN

LEAD Graduate School

Why analyze learner language?

Intelligent Tutoring Systems

- online analysis of learner language aimed at supporting language acquisition
- provide immediate, individualized scaffolding feedback:
  - meta-linguistic feedback in form-focused activities
  - incidental Focus-on-Form in meaning-based activities
  - feedback on meaning (rare in ITS, but needed in TBL)
- determine progression through pedagogical material

Connecting NLP and Language Learning

Detmar Meurers

Introduction

NLP analysis of learner language

Intelligent Tutoring Systems
Learner Corpora
Writer's aids and testing
Under the hood: Interpreting learner data

Native language analysis

Searching materials
Under the hood: Readability ranking the web

Conclusion/Outlook

References

EBERHARD KARLS
UNIVERSITÄT TÜBINGEN

LEAD
Graduate School

Why analyze learner language?

Learner Corpora

- offline analysis of learner language
- gold-standard training & testing data for NLP development
- support effective search in annotated data
  - to provide insights into typical student needs in FLT
  - to provide empirical evidence for SLA research
    - developmental sequences
    - linguistic correlates of CEFR proficiency levels
    - L1 transfer ($\rightarrow$ NLI shared task)
    - ...

Connecting NLP and Language Learning

Detmar Meurers

Introduction

NLP analysis of learner language

Intelligent Tutoring Systems
Learner Corpora
Writer's aids and testing
Under the hood: Interpreting learner data

Native language analysis

Searching materials
Under the hood: Readability ranking the web

Conclusion/Outlook

References

EBERHARD KARLS
UNIVERSITÄT TÜBINGEN

LEAD
Graduate School

Why analyze learner language?

Writer's aid tools

- feedback aimed at *producing text*
- identify and correct errors in orthography, grammar, usage
- could support tailoring of material to specific audiences (readability analysis, manual or automatic simplification)

Language Testing

- support assessment of *learner competence*
  - automate (some) grading
  - support more efficient grading by grouping learner answers
- draw valid inferences about a learner's state of knowledge
  - also central (but little discussed) for ITS

Connecting NLP and Language Learning

Detmar Meurers

Introduction

NLP analysis of learner language

Intelligent Tutoring Systems
Learner Corpora
Writer's aids and testing
Under the hood: Interpreting learner data

Native language analysis

Searching materials
Under the hood: Readability ranking the web

Conclusion/Outlook

References

EBERHARD KARLS
UNIVERSITÄT TÜBINGEN

LEAD
Graduate School

Under the hood: Interpreting learner data

- How would you analyze the following sentences from the Hiroshima English Learners' Corpus (Miura 1998)?

  (1) *I didn't know*
  (2) *I don't know his lives.*
  (3) *I know where he lives.*

  They are taken from a translation task, for the Japanese of

  (4) *I don't know where he lives.*

- $\Rightarrow$ To reliably interpret learner language in ITS and learner corpus research, we should more seriously consider
  - the particular **task**, and
  - the **learner** characteristics.

Connecting NLP and Language Learning

Detmar Meurers

Introduction

NLP analysis of learner language

Intelligent Tutoring Systems
Learner Corpora
Writer's aids and testing
Under the hood: Interpreting learner data

Native language analysis

Searching materials
Under the hood: Readability ranking the web

Conclusion/Outlook

References

EBERHARD KARLS
UNIVERSITÄT TÜBINGEN

LEAD
Graduate School

Under the hood: Interpreting learner data
Interpreting learner data requires task and learner modeling

- The best approach to grammatical error correction only reaches 39.7% precision, 30.1% recall (Ng et al. 2014)

- Inter-annotator agreement for error annotation of learner corpora is only starting to be reported (Rosen et al. 2014).
  - agreement good on some error tags (agreement, subcat.)
  - target hypotheses required for reliable error diagnosis

- Effective analysis of learner language requires modeling the expected **well-formed and ill-formed variation**.

- We need to constrain the space of variation using
  - well-designed tasks (Amaral & Meurers 2011; Quixal 2012)
  - learner modeling (Michaud et al. 2001; Amaral & Meurers 2008)

EBERHARD KARLS
UNIVERSITÄT TÜBINGEN

LEAD
Graduate School

Why analyze native language for learners?

Searching for appropriate materials for learners

- learner or teacher specifies a keyword as in web search
  - to obtain authentic materials
  - on a current topic of individual interest
  - fully contextualized language use
- making search engines language-aware allows us to target
  - materials appropriate in readability
  - including language forms to be learned, in representative contexts

Enhanced presentation of materials, adapted to learner

- visual input enhancement supporting noticing
- generation of exercises
- generation of annotations (e.g., vocabulary)

EBERHARD KARLS
UNIVERSITÄT TÜBINGEN

LEAD
Graduate School

Under the hood: Readability-ranking the web
(Vajjala & Meurers 2013)

- Are state-of-the-art readability models actually useful for classifying texts as found on the web?
  - Can we re-rank search results based on reading levels?

- Implementation details:
  - feature set inspired by SLA measures
  - WEKA linear regression, since we want output on a scale
  - trained model on 5-level WeeBit corpus

- We applied the readability model to search results obtained through BING search API.
  - took 50 search queries from a public query log
  - computed reading levels for Top-100 results

EBERHARD KARLS
UNIVERSITÄT TÜBINGEN

LEAD
Graduate School

Results: Reading levels of top search results
Vajjala & Meurers (2013)

Result Rank:	1	2	3	4	5	6	7	8	9	10	Avg. Top100
Query:											
copyright copy law	1.8	4.6	1.4	2.7	4.6	6.2	2.7	**1.1**	3.9	5.6	4.6
halley comet	**1.7**	4.5	4.5	4.2	2.4	4.1	4.9	3.6	4.2	3.6	4.0
europe union politics	3.6	4.9	6.3	4.0	2.2	4.5	**1.5**	1.6	4.9	6.3	4.3
shakespeare	2.4	2.9	4.2	4.7	4.7	3.9	**1.5**	2.1	2.6	4.0	3.6
euclidean geometry	3.9	4.7	4.7	4.3	4.5	4.6	4.0	4.1	3.5	**2.6**	3.2
. . .											

- Results:
  - avg. reading level of search results high (5 = GCSE)
  - full range of reading levels among most relevant results returned by search engine

- Readability-based re-ranking of results is potentially useful.
  → language-aware search engines (Ott & Meurers 2010)
  · authentic web materials for Nordic languages (Nilsson & Borin 2002)

EBERHARD KARLS
UNIVERSITÄT TÜBINGEN

LEAD
Graduate School

Conclusion/Outlook

- In language learning, there are real-life and research needs that NLP can help address.

- Development and evaluation requires collaboration with SLA, FLTL, Empirical Educational Science.
  - LEAD graduate school is a step in this direction
    - ReadingDemands: Is school book language taking the grade level and school type into account?
    - CogCorr: Linking linguistic and cognitive measures of text complexity
  - NLP benefits, but also feeds back into SLA & education
  - iCALL facilitates experimentally testing SLA hypotheses
    - WERTi/VIEW (Meurers et al. 2010)
    - eCALL (Presson et al. 2013)

Connecting NLP and Language Learning

Detmar Meurers

Introduction

NLP analysis of learner language

Intelligent Tutoring Systems

Learner Corpora

Writer's aids and testing

Under the hood: Interpreting learner data

Native language analysis

Searching materials

Under the hood: Readability ranking the web

Conclusion/Outlook

References

EBERHARD KARLS
UNIVERSITÄT TÜBINGEN

LEAD
Graduate School

References

Amaral, L. & D. Meurers (2008). From Recording Linguistic Competence to Supporting Inferences about Language Acquisition in Context: Extending the Conceptualization of Student Models for Intelligent Computer-Assisted Language Learning. *Computer-Assisted Language Learning* 21(4), 323–338. URL http://purl.org/dm/papers/amaral-meurers-call08.html.

Amaral, L. & D. Meurers (2011). On Using Intelligent Computer-Assisted Language Learning in Real-Life Foreign Language Teaching and Learning. *ReCALL* 23(1), 4–24. URL http://dx.doi.org/10.1017/S0958344010000261.

Dickinson, M., C. Brew & D. Meurers (2013). *Language and Computers*. Wiley-Blackwell.

Meurers, D. (2012). Natural Language Processing and Language Learning. In C. A. Chapelle (ed.), *Encyclopedia of Applied Linguistics*, Oxford: Wiley, pp. 4193–4205. URL http://purl.org/dm/papers/meurers-12.html.

Meurers, D. (2015). Learner Corpora and Natural Language Processing. In S. Granger, G. Gilquin & F. Meunier (eds.), *The Cambridge Handbook of Learner Corpus Research*, Cambridge University Press.

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-5) at NAACL-HLT 2010*. Los Angeles, pp. 10–18. URL http://aclweb.org/anthology/W10-1002.pdf.

Michaud, L. N., K. F. McCoy & L. A. Stark (2001). Modeling the Acquisition of English: An Intelligent CALL Approach. In *Proceedings of The 8th International*

Connecting NLP and Language Learning

Detmar Meurers

Introduction

NLP analysis of learner language

Intelligent Tutoring Systems

Learner Corpora

Writer's aids and testing

Under the hood: Interpreting learner data

Native language analysis

Searching materials

Under the hood: Readability ranking the web

Conclusion/Outlook

References

EBERHARD KARLS
UNIVERSITÄT TÜBINGEN

LEAD
Graduate School

Conference on User Modeling. Sonthofen, Germany, pp. 14–25. URL http://www.eecis.udel.edu/research/icicle/pubs/MiMcSt01.ps.

Miura, S. (1998). Hiroshima English Learners' Corpus: English learner No. 2 (English I & English II). Department of English Language Education, Hiroshima University. http://purl.org/icall/helc.

Ng, H. T., S. M. Wu, T. Briscoe, C. Hadiwinoto, R. H. Susanto & C. Bryant (2014). The CoNLL-2014 Shared Task on Grammatical Error Correction. In *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*. Baltimore, Maryland: Association for Computational Linguistics, pp. 1–14.

Nilsson, K. & L. Borin (2002). Living off the land: The Web as a source of practice texts for learners of less prevalent languages. In *Proceedings of LREC 2002, Third International Conference on Language Resources and Evaluation, Las Palmas: ELRA*. pp. 411–418.

Ott, N. & D. Meurers (2010). Information Retrieval for Education: Making Search Engines Language Aware. *Themes in Science and Technology Education. Special issue on computer-aided language analysis, teaching and learning: Approaches, perspectives and applications* 3(1–2), 9–30. URL http://purl.org/dm/papers/ott-meurers-10.html.

Presson, N., C. Davy & B. MacWhinney (2013). Experimentalized CALL for adult second language learners. In J. W. Schwieter (ed.), *Innovative Research and Practices in Second Language Acquisition and Bilingualism*, John Benjamins, pp. 139—-164. URL http://talkbank.org/SLA/pubs/eCALL.pdf.

Quixal, M. (2012). Language Learning Tasks and Automatic Analysis of Learner Language. Connecting FLTL and NLP in the design of ICALL materials supporting effective use in real-life instruction. Ph.D. thesis, Universitat Pompeu Fabra, Barcelona and Eberhard-Karls-Universität Tübingen.

Connecting NLP and Language Learning

Detmar Meurers

Introduction

NLP analysis of learner language

Intelligent Tutoring Systems

Learner Corpora

Writer's aids and testing

Under the hood: Interpreting learner data

Native language analysis

Searching materials

Under the hood: Readability ranking the web

Conclusion/Outlook

References

EBERHARD KARLS
UNIVERSITÄT TÜBINGEN

LEAD
Graduate School

Rosen, A., J. Hana, B. Štindlová & A. Feldman (2014). Evaluating and automating the annotation of a learner corpus. *Language Resources and Evaluation* 48(1), 65–92.

Vajjala, S. & D. Meurers (2013). On The Applicability of Readability Models to Web Texts. In *Proceedings of the Second Workshop on Predicting and Improving Text Readability for Target Reader Populations*. pp. 59–68.

Connecting NLP and Language Learning

Detmar Meurers

Introduction

NLP analysis of learner language

Intelligent Tutoring Systems

Learner Corpora

Writer's aids and testing

Under the hood: Interpreting learner data

Native language analysis

Searching materials

Under the hood: Readability ranking the web

Conclusion/Outlook

References

EBERHARD KARLS
UNIVERSITÄT TÜBINGEN

LEAD
Graduate School