

Computational Approaches to Functional Elements

Abstract:

Functional elements such as determiners and prepositions have received only little attention in computational linguistics, yet they contribute significantly to the syntactic, semantic, and pragmatic structure of a sentence in context. As such, functional elements offer an exciting perspective into the interplay of different modules of linguistic analysis – and the question how the relevant properties can reliably be identified in the data. At the same time, there are concrete applications for accurate models of functional elements, such as natural language generation or intelligent computer-assisted language learning, where functional elements have been identified as the single most important source of errors.

The course will begin with a discussion of the linguistic aspects and their computational modeling and then turn into a project-driven seminar, in which student groups participate in a challenge to design and implement approaches for predicting preposition and determiner usage.

Sessions

1. Wednesday, 4. November:

- Syllabus and course motivation
- Ⓟ Establishing Majority Baseline. Subtask 1:
<http://purl.org/dm/09/ws/func/ex1.pdf>

2. Monday, 9. November:

- H&P Articles

3. Wednesday, 11. November:

- Ⓟ Subtask 1 (cont)
- H&P Articles (cont)

4. Monday, 16. November:

- H&P Prepositions

5. Wednesday, 18. November:

- Ⓟ Majority Baseline for German. Subtask 2:
<http://purl.org/dm/09/ws/func/ex2.pdf>
- H&P Prepositions (cont)

6. Monday, 23. November:

- Subtask 2 and the insight it provides on the empirical landscape of German

7. Wednesday, 25. November:
 - Intro to Language Modeling [Kilian]
 - Ⓟ Subtask 3: SRI Toolkit and Using Language Models
<http://purl.org/dm/09/ws/func/ex3.pdf>
8. Monday, 30. November:
 - Language Modeling in **Turner & Charniak (2007)** [Kilian] (cont)
9. Wednesday, 2. December:
 - Ⓟ Subtask 3 (cont)
 - **De Felice (2008)** [Anas, Katya]
10. Monday, 7. December:
 - **De Felice (2008)** [Anas, Katya] (cont)
11. Wednesday, 9. December:
 - Ⓟ Subtask 3 (cont)
 - Ⓟ Subtask 4: Using Language Models for German.
<http://purl.org/dm/09/ws/func/ex4.pdf>
12. Monday, 14. December:
 - **De Felice (2008)** [Anas, Katya] (cont)
13. Wednesday, 16. December:
 - Ⓟ Wrapping up Subtask 4
 - **De Felice (2008)** [Anas, Katya] (cont)
14. Monday, 11. January:
 - TiMBL (as a tool used by **Lee & Seneff 2006**; **Lee & Knutsson 2008**) [Aleks]
<http://purl.org/dm/09/ws/func/ex5.pdf>
15. Wednesday, 13. January:
 - Ⓟ TiMBL activity
 - **Lee & Seneff (2006)**; **Lee & Knutsson (2008)** [Ekaterina]
16. Monday, 18. January:
 - **Lee & Seneff (2006)**; **Lee & Knutsson (2008)** [Ekaterina] (cont)
17. Wednesday, 20. January:
 - Ⓟ Subtask 5: adding linguistic insight to T1, T2

- [Chodorow et al. \(2007\)](#); [Tetreault & Chodorow \(2008a,b\)](#) [Nomi]
18. Monday, 25. January: [Gamon et al. \(2008\)](#) [Janina]
 19. Wednesday, 27. January: Subtask 5: T3
 20. Monday, 1. February: T3
 21. Wednesday, 3. February: T4
 22. Monday, 8. February: T4 (cont)
 23. Wednesday, 10. February: T4 (cont)
 24. Monday, 15. February: T5
 25. Wednesday, 17. February: T5 (cont)

Instructors:

- Detmar Meurers
 - *Office:* Room 1.28, Blochbau (Wilhelmstr. 19)
 - *Email:* dm@sfs.uni-tuebingen.de
 - *Office hours:* Mondays 11:30–12:30
- Holger Wunsch
 - *Office:* Room 1.26, Blochbau (Wilhelmstr. 19)
 - *Email:* Wunsch@sfs.uni-tuebingen.de
 - *Office hours:* Tuesdays, 14:00–15:00

Course meets: in Seminarraum 1.13, Blochbau (Wilhelmstr. 19); towards the later part mostly in the Computer Lab 2.28

- Mondays, 14ct-16
- Wednesdays, 10ct-12
- Wednesdays, 14ct-16

Credits and Campus Prüfungsnummer:

- 6 SWS, register for both of the following to obtain 15 Credit Points: 2100 Hauptseminare I + 2400 Hauptseminare II

Syllabus (this file):

- [html-Version](http://purl.org/dm/09/ws/func) (<http://purl.org/dm/09/ws/func>)
- [pdf-Version](http://purl.org/dm/09/ws/func/syllabus.pdf) (<http://purl.org/dm/09/ws/func/syllabus.pdf>)

Moodle page: <http://moodle01.zdv.uni-tuebingen.de/course/view.php?id=295>

Nature of course and my expectations: This is a project-driven research-oriented seminar, where we jointly explore a current research topic and develop our own approaches to it. Each participant is expected to

1. regularly and actively participate in class, read the papers assigned by any of the presenters and post a question on Moodle to the “Reading Discussion Forum” on each reading *at the latest on the day before it is discussed* in class. (30% of grade)

Note: Following the rules of the Neuphilologische Fakultät, missing more than two meetings unexcused, automatically results in failing the class.

2. explore and present a topic (30% of grade):
 - select one of the sub-topics, at the latest by our meeting on November 11
 - thoroughly research the topic, taking our literature pointers *as a starting point*
 - prepare the presentation with slides and discuss the presentation with Holger or Detmar during the office hours *a week before the presentation*
 - start a new Moodle thread on the “Reading Discussion Forum” specifying what every course participant should read to prepare for your presentation *a week before your presentation*
 - present the topic in class
3. successfully complete the practical projects assigned in class and write up the final project in a term paper (40% of grade):
 - form a two person team and work as part of the team throughout the seminar
 - after the seminar, each team writes up a term paper describing the main project, which needs to be handed in *before the beginning of the next semester*.

Each term paper must include an explicit specification of a) who did what in the project and b) what on the term paper, and c) which grade each group member would give him/herself for this effort. The guideline is that the *effort spent on the project and term paper* should be equally divided.

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.

Class etiquette: Please do not read or work on materials for other classes in our seminar. Come to class on time and do not pack up early. When our seminar meets in the computer lab, only use the computers when you are asked to do a specific activity – do not read email or browse the web. All portable electronic devices such as cell phones should be switched off for the entire length of the flight, oops, class. If for some reason, you must leave early or you have an important call coming in, or you have to miss class for an important reason, please let Detmar or Holger know *before* class.

Topics

- The empirical landscape in English:
 - Articles: pp. 354–399 of [Huddleston & Pullum \(2002, Chapter 5\)](#)
 - Prepositions: pp. 598–661 of [Huddleston & Pullum \(2002, Chapter 7\)](#)
- Computational models of Article and Preposition Use in English
 - [De Felice \(2008\)](#), [De Felice & Pulman \(2007\)](#)
 - [Turner & Charniak \(2007\)](#)
 - [Chodorow et al. \(2007\)](#); [Tetreault & Chodorow \(2008a,b\)](#)
 - [Gamon et al. \(2008\)](#)
 - [Lee & Senefc \(2006\)](#); [Lee & Knutsson \(2008\)](#) (= [Lee 2009](#))
- The empirical landscape in German

Progression of Project Tasks

- T1:
 - Given: Native English with all articles (or prepositions) replaced by a dummy
 - Predict which article (or preposition) is correct for each dummy position
 - * majority baseline
 - * language models
 - * adding linguistic modeling
- T2: As with T1, but using native German texts.
- T3: As with T1 and 2, but instead of replacing target with dummy, the target is removed so that the location where an article (or preposition) needs to be inserted has to be determined as well.
- T4: As with T1 and 2 (i.e., using artificially introduced dummy locations), but instead of native language use learner language. corpora
- T5: As with T3, but use learner corpora to detect naturally occurring missing or incorrect determiners (or prepositions).

References

- Chodorow, M., J. Tetreault & N.-R. Han (2007). Detection of Grammatical Errors Involving Prepositions. In *Proceedings of the 4th ACL-SIGSEM Workshop on Prepositions*. Prague, Czech Republic: Association for Computational Linguistics, pp. 25–30. URL <http://www.aclweb.org/anthology-new/W/W07/W07-1604>.
- De Felice, R. (2008). Automatic Error Detection in Non-native English. Ph.D. thesis, St Catherine’s College, University of Oxford.

- De Felice, R. & S. Pulman (2007). Automatically Acquiring Models of Preposition Use. In *Proceedings of the 4th ACL-SIGSEM Workshop on Prepositions*. Prague, Czech Republic: Association for Computational Linguistics, pp. 45–50. URL <http://www.aclweb.org/anthology-new/W/W07/W07-1607>.
- Gamon, M., J. Gao, C. Brockett, A. Klementiev, W. Dolan, D. Belenko & L. Vanderwende (2008). Using Contextual Speller Techniques and Language Modeling for ESL Error Correction. In *Proceedings of IJCNLP*. Hyderabad, India. URL <http://www.mt-archive.info/IJCNLP-2008-Gamon.pdf>.
- Huddleston, R. & G. K. Pullum (2002). *The Cambridge Grammar of the English Language*. Cambridge University Press.
- Lee, J. & O. Knutsson (2008). The Role of PP Attachment in Preposition Generation. In A. Gelbukh (ed.), *Proceedings of CICLing 2008*. URL <http://www.springerlink.com/index/5524470312j28317.pdf>.
- Lee, J. & S. Seneff (2006). Automatic Grammar Correction for Second-Language Learners. In *INTER_SPEECH 2006 – ICSLP*. URL <http://groups.csail.mit.edu/sls/publications/2006/IS061299.pdf>.
- Lee, J. S. Y. (2009). Automatic Correction of Grammatical Errors in Non-native English Text. Ph.D. thesis, Massachusetts Institute of Technology. Department of Electrical Engineering and Computer Science.
- Tetreault, J. & M. Chodorow (2008a). Native Judgments of Non-Native Usage: Experiments in Preposition Error Detection. In *Proceedings of COLING-08*. Manchester. URL <http://www.ets.org/Media/Research/pdf/h4.pdf>.
- Tetreault, J. & M. Chodorow (2008b). The Ups and Downs of Preposition Error Detection in ESL Writing. In *Proceedings of COLING-08*. Manchester. URL <http://www.ets.org/Media/Research/pdf/r3.pdf>.
- Turner, J. & E. Charniak (2007). Language Modeling for Determiner Selection. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers*. Rochester, New York: Association for Computational Linguistics, pp. 177–180. URL <http://www.aclweb.org/anthology-new/N/N07/N07-2045>.