

# Introduction to Computational Linguistics

## Session 8: Computer-Assisted Language Learning

Stephen Bodnar

Universität Tübingen

December 20, 2023

- Introduction
- NLP for Analyzing Language for Learners
- NLP for Analyzing Learner Language

# Introduction

- NLP tools typically developed and optimized for language of native speakers
- allowing for variability (e.g. spelling correction), but target always well-formed language by native speaker
- NLP for language learning requiring different and creative approaches

2 applications of NLP in language learning and instruction

- ① analyzing (native) language productions for learners
- ② analyzing learner productions

# NLP for Analyzing Language for Learners

- learners need input in target language
  - reading material, listening texts, authentic material, ...
  - exercises for practicing
- tasks with NLP:
  - searching relevant and appropriate texts and examples
  - enhancing texts

# Why Analyze Language for Learners with NLP?

- off-the-shelf NLP tools work for native language
- text books: static content
- NLP for dynamically annotating new texts
- allowing learners to practice on material
  - they are interested in (e.g. news)
  - that is at their current level of proficiency (readability, vocabulary, constructions)
  - that contains the constructions learners want to practice (e.g. comparative forms) → exercise generation

# NLP for LL - Example

- FLAIR system Chinkina and Meurers [2016]
- linguistically aware search engine
- re-rank search engine results based on linguistic constructions

<http://sifnos.sfs.uni-tuebingen.de/FLAIR/>

# FLAIR - Linguistically Aware Search Engine

The screenshot displays the FLAIR web interface. The top navigation bar is orange with the FLAIR logo. The main content area shows search results for the query 'cars'. On the left, there are filters for 'Text Characteristics' (Length, Levels) and 'Constructions' (Sentences, Parts of Speech, Verbs). The search results are listed in a table with columns for rank, title, and description. The first result is 'The Cars - Wikipedia'. The second result is 'Cars (song) - Wikipedia'. The third result is 'Car Games - Play Car Games on CrazyGames'. The fourth result is 'Cars & Trucks for sale | eBay'. The fifth result is 'Used Cars for Sale Online Near Me | Cars.com'. The sixth result is 'Cars For Sale - Used Cars For Sale - Used Cars ...'. The seventh result is 'Car Games - Y8.COM'. On the right, there is a sidebar with a 'Cars (song) - Wikipedia' entry and a 'Construction' section showing 'Verbs > Verb forms > Contr...' with a count of 6 and weight of (3). Below this, there are buttons for 'ALL CONSTRUCTIONS', 'Export document', 'COPY TO CLIPBOARD', and 'EXPORT AS WORD'.

7 Results (3 Filtered)

VISUALIZE 38

SHARE SEARCH SETUP

Text Characteristics:

Length:

☐ Prefer shorter texts.

Levels:

☒ A1-A2 0.57

☒ B1-B2 0.70

☐ C1-C2 0.72

Constructions:

▼ Sentences

▼ Parts of Speech

▼ Verbs

▼ Verb Forms

☒ Contracted auxiliary 0.57

☒ Full auxiliary 0.70

☒ Auxiliaries 0.57

1 The Cars - Wikipedia  
[https://en.wikipedia.org/wiki/The\\_Cars](https://en.wikipedia.org/wiki/The_Cars)  
The Cars were an American rock band formed in Boston in 1976. Emerging from the new wave scene in the late 1970s, the line-up consisted of Elliot Easton (lead guitar), Greg Hawkes (rhythm guitar), Benjamin Orr (bass guitar), and David Robinson (drums). The band split into two groups, and the Cars were at the forefront of ...

2 Cars (song) - Wikipedia  
[https://en.wikipedia.org/wiki/Cars\\_\(song\)](https://en.wikipedia.org/wiki/Cars_(song))  
"Cars" is a song by English musician Gary Numan. His debut single, it was released as the lead single from his debut studio album, *The Pleasure Principle* (1979), on 21 August 1979. It reached the top of the charts in several countries, and today is considered a new wave staple. The song was the first release credited solely to Gary Numan after he dropped the band name Tubeway Army, under ...

3 Car Games - Play Car Games on CrazyGames  
<https://www.crazygames.com/cr>  
Simulator games - drive cars and other vehicles in more realistic settings, sometimes completing tasks. There's even unique car games that combine elements like idle and merge gameplay into the core mechanics of the game. Road Crash is a car game that does this, putting you in races where you must crash against your opponents to get ahead.

4 Cars & Trucks for sale | eBay  
<https://www.ebay.com/b/Cars-Trucks/3007>  
But why put yourself through that stress when you can shop cars for sale from the comfort of your own home? eBay has an extensive selection of new, used and certified pre-owned cars and trucks, from Audi and Subaru to Mercedes-Benz. You can also easily sell your used car on eBay, helping raise the cash you need for your next car or truck.

5 Used Cars for Sale Online Near Me | Cars.com  
<https://www.cars.com/shopping>  
Browse cars for sale, shop the best deals near you, find current loan rates and read FAQs about financing and warranties at Cars.com.

6 Cars For Sale - Used Cars For Sale - Used Cars ...  
<https://www.carsforsale.com>  
About Carsforsale.com ®. For over 20 years, Carsforsale.com ® has simply been one of the fastest and easiest ways to buy or sell new and used cars online. With millions of vehicles listed from thousands of State-Verified Auto Dealers nationwide, finding your perfect vehicle is a snap.

7 Car Games - Y8.COM  
<https://www.y8.com/tag/Car>

Cars (song) - Wikipedia

B1-B2 125 Sentence(s) 1936 Word(s)

"Cars" is a song by English musician Gary Numan. His debut single, it was released as the lead single from his debut studio album, *The Pleasure Principle* (1979), on 21 August 1979. It reached the top of the charts in several countries, and today is considered a new wave staple. The song was the first release credited solely to Gary Numan after he dropped the band name Tubeway Army, under which name he had released four singles and two LPs, including the number one UK hit "Are Friends Electric?"; and its parent album, *Replicas*. Musically, the new song was somewhat lighter and more pop-oriented than its predecessors, Numan later conceding that he had chart success in mind: "This was the first time I had written a song with the intention of 'maybe it could be a hit single'; I was writing this before 'Are Friends Electric?' happened." He has since described "Cars" as "a pretty average song" in the UK charts. It reached number 1 in 1979, and in 1980 hit number 1 in Canada two weeks running on the RPM

Construction

Count Weight

Verbs > Verb forms > Contr... 6 (3)

ALL CONSTRUCTIONS

Export document

☒ Include text formatting

COPY TO CLIPBOARD

EXPORT AS WORD



- Introduction
- NLP for Analyzing Language for Learners
- => NLP for Analyzing Learner Language

# Native-Speaker Language vs. Learner Language

- Díaz Negrillo et al. [2010]: analyzing learner language
- learner language:
  - divergences between linguistic levels
  - categories for native language not applicable
  - difficult to determine *target hypotheses*
  - often more than one error

# A Note about Terminology

- The term **distribution** has a special meaning in this context and refers to grammatical “slots” in the structure of a sentence (See Díaz Negrillo et al. [2010])
- For a given distributional slot, only certain word classes are allowed
- E.g., \_\_\_\_\_walked to the store.
- E.g., Those shoes ? Aren't they too \_\_\_\_\_?

# Learner Language - Examples of Errors

- “they are very kind and friendship”
- distribution mismatch  
(noun occurring where adjectives normally occur)

Examples (from Díaz Negrillo et al. [2010])

## Learner Language - Examples of Errors 2

- “to be choiced for a job”
- distribution-stem and distribution-morphology mismatch  
(a noun appears with verbal inflection, and in a slot reserved for verbs)

Examples (from Díaz Negrillo et al. [2010])

# Learner Language - Examples of Errors 3

Examples (from Díaz Negrillo et al. [2010])

- “television, radio are very subjectives”
- stem-morphology mismatch  
(word occurring in adjective slot, but is inflected either as plural noun or present tense verb)

Examples (from Díaz Negrillo et al. [2010])

# Learner Language - Examples of Errors 4

- “I didn’t went to the playground yesterday.”
- distribution morphology mismatch  
(In its present slot, the verb “go” should not appear in the past tense)

# Language Instruction & Practical Challenges

- How does one best learn a language?
- Research in Second Language Acquisition and Foreign Language Teaching (FLT) has shown:
  - individualized, immediate feedback is very important for learner production.
- Problem: limited opportunity for individual, immediate feedback
- In the classroom, the teacher is generally the only source of reliable and accurate feedback available to students.
  - no time to focus on individual students
  - difficult to take heterogeneity of students into account
- Outside of class, how can students be supported in a fair way (not relying on parents)?



# Intelligent Tutoring Systems

- Intelligent Tutoring Systems (ITS) can help address this issue
    - automated, immediate feedback while learner is working on the task
    - many students can use the system simultaneously at any time
  - Goals:
    - close gap between ITS research, FLT insights, and real-life classroom
    - address real formal education needs using current NLP technology
- We are developing and testing FeedBook
- an interactive online workbook for secondary school English learners

- Starting point: *Camden Town Gymnasium 3* Workbook
  - approved for 7th grade English classes in German secondary schools
  - existing workbook, already integrated into real-life formal education
- FeedBook system provides a web-based implementation of the traditional print workbook enabling
  - students to complete activities online
  - teachers to give formative and summative feedback
- Purpose of the system:
  - to provide individualized, immediate scaffolding feedback to learners
  - to guide them towards solutions for a number of different activities.
- Automatic feedback on
  - 7th grade grammar topics (form-oriented)
  - reading/listening exercises (meaning-oriented)

Approach for detecting errors and generating feedback:

- analyze target answers with rich linguistic analysis
- use linguistic analyses and search for target forms
- apply rules to introduce errors for target forms
- save edited forms with diagnoses
- use flexible matching to align learner answer with target answer(s)

→ Start out from the expected target forms and systematically transform them into well-formed and ill-formed variations of the target [Rudzewitz et al., 2018]

# Generating Variants: Examples

*What are you doing in front of the TV ?* (target answer)



*What is you doing in front of the TV* (wrong agreement)

*What are you do in front of the TV ?* (base form)

...

## Generating Variants: Examples cont.

*What are you doing in front of the TV ?*

(target answer)



*What did you do in front of the TV ?*

(simple past)

*What have you done in front of the TV ?*

(past perfect)

...



*What did you dos in front of the TV ?*

(simple past, extra s)

*What has you been doing in front of the TV ?*

(present perfect progressive, wrong agreement)

...

# NLP Tasks in Analyzing Learner Language

task	tool
segmentation	ClearNLP [Choi and Palmer, 2012]
part-of-speech tagging	ClearNLP
dependency parsing	ClearNLP
lemmatization	Morpha [Minnen et al., 2001]
morphological analysis	Sfst [Schmid, 2005]
synonym annotator	WordNet [Leacock and Chodorow, 1998]
chunking	OpenNLP [Baldrige, 2005]
keyWordFromDepExtractor	CoMiC [Meurers et al., 2011a]
word embeddings	ELMO [Peters et al., 2018]

Table: Adapted from Rudzewitz et al. [2018]

# FeedBook - Looking Forward

- Extending FeedBook to support individualized instruction
- Same learning goals for everyone, but different learning paths there
- Idea of Scaffolding
  - Accomodate weaker students with easier, but more exercises
  - Challenge stronger students with less
- Requires explicit
  - learner models
  - activity models
  - domain models
  - pedagogy models
- First test in schools completed last spring; Second test completed fall 2022; A third test planned for February 2024.
- more information: Quixal et al. [2021]

# References

Slides based on Björn Rudzewitz's slides from 2021/22, which are in turn based on NLP for language learning Meurers [2012] and Dickinson et al. [2013, chapter 3]

Further reading about the FeedBook in Rudzewitz [2021], Ziai et al. [2018], Rudzewitz et al. [2020]

See Reference list below for more information.



# Reference List I

- Jason Baldridge. The OpenNLP Project. URL: *<http://opennlp.apache.org/index.html>*, (accessed 25 August 2015), 2005.
- Maria Chinkina and Detmar Meurers. Linguistically-aware information retrieval: Providing input enrichment for second language learners. In *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications (BEA)*, pages 188–198, San Diego, CA, 2016. ACL. URL <http://aclweb.org/anthology/W16-0521.pdf>.
- Jinho D Choi and Martha Palmer. Fast and robust part-of-speech tagging using dynamic model selection. In *Proceedings of the 50th Annual Meeting of the ACL*, pages 363–367, 2012.

## Reference List II

- Ana Díaz Negrillo, Detmar Meurers, Salvador Valera, and Holger Wunsch. Towards interlanguage POS annotation for effective learner corpora in SLA and FLT. *Language Forum*, 36(1–2):139–154, 2010. ISSN 0253-9071. URL <http://purl.org/dm/papers/diaz-negrillo-et-al-09.html>.
- Markus Dickinson, Chris Brew, and Detmar Meurers. *Language and Computers*. Wiley-Blackwell, 2013.
- Claudia Leacock and Martin Chodorow. Combining local context and wordnet sense similarity for word sense identification. In C. Fellbaum, editor, *WordNet, An Electronic Lexical Database*, pages 265–283. MIT Press, 1998.
- Detmar Meurers. Natural language processing and language learning. In Carol A. Chapelle, editor, *Encyclopedia of Applied Linguistics*, pages 4193–4205. Wiley, Oxford, 2012. URL <http://purl.org/dm/papers/meurers-12.html>.

## Reference List III

- Detmar Meurers, Ramon Ziai, Niels Ott, and Stacey Bailey. Integrating parallel analysis modules to evaluate the meaning of answers to reading comprehension questions. *IJCELL. Special Issue on Automatic Free-text Evaluation*, 21(4):355–369, 2011a. URL <http://purl.org/dm/papers/Meurers.Ziai.ea-11.pdf>.
- Detmar Meurers, Ramon Ziai, Niels Ott, and Janina Kopp. Evaluating answers to reading comprehension questions in context: Results for German and the role of information structure. In *Proceedings of the TextInfer 2011 Workshop on Textual Entailment*, pages 1–9, Edinburgh, 2011b. URL <http://aclweb.org/anthology/W11-2401.pdf>.
- Guido Minnen, John Carroll, and Darren Pearce. Applied morphological processing of English. *Natural Language Engineering*, 7(3):207–233, 2001.

## Reference List IV

- Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*, 2018.
- Martí Quixal, Björn Rudzewitz, Elizabeth Bear, and Detmar Meurers. Automatic annotation of curricular language targets to enrich activity models and support both pedagogy and adaptive systems. In *Proceedings of the 10th Workshop on NLP for Computer Assisted Language Learning*, pages 15–27, 2021. URL <https://aclanthology.org/2021.nlp4call-1.2.pdf>.

## Reference List V

Björn Rudzewitz, Ramon Ziai, Kordula De Kuthy, Verena Möller, Florian Nuxoll, and Detmar Meurers. Generating feedback for English foreign language exercises. In *Proceedings of the 13th Workshop on Innovative Use of NLP for Building Educational Applications (BEA)*, pages 127–136. ACL, 2018. URL <http://aclweb.org/anthology/W18-0513.pdf>.

Björn Rudzewitz, Ramon Ziai, Florian Nuxoll, Kordula De Kuthy, and Detmar Meurers. Enhancing a web-based language tutoring system with learning analytics. In Luc Paquette and Cristobal Romero, editors, *Joint Proceedings of the Workshops of the 12th International Conference on Educational Data Mining co-located with the 12th International Conference on Educational Data Mining (EDM 2019)*, volume 2592, pages 1–7, Montréal, Canada, 2020. CEUR Workshop Proceedings. URL <http://ceur-ws.org/Vol-2592/paper1.pdf>.

# Reference List VI

- Björn Rudzewitz. *Learning Analytics in Intelligent Computer-Assisted Language Learning*. PhD thesis, Eberhard-Karls Universität Tübingen, 2021. URL <http://hdl.handle.net/10900/117358>.
- Helmut Schmid. A programming language for finite state transducers. In *Proceedings of the 5th International Workshop on Finite State Methods in Natural Language Processing*, pages 308–309, 2005.
- Ramon Ziai, Björn Rudzewitz, Kordula De Kuthy, Florian Nuxoll, and Detmar Meurers. Feedback strategies for form and meaning in a real-life language tutoring system. In *Proceedings of the 7th Workshop on Natural Language Processing for Computer-Assisted Language Learning (NLP4CALL)*, pages 91–98. ACL, 2018. URL <http://aclweb.org/anthology/W18-7110>.