GECKO Software project, work update 1

Eduardo Calò, Léo Jacqmin, Thibo Rosemplatt

NLP M2 IDMC, Université de Lorraine

September 15th, 2020





Grammatical error

Modern approaches to grammatical error correction

Grammarly

Action plan





Grammatical error

Modern approaches to grammatical error correction





What is a grammatical error?

Grammatical error is an umbrella term used in prescriptive grammar to describe an instance of faulty, unconventional, or controversial usage of a given language.





Native vs non native

Errors made by learners are different from those made by native speakers. [2] and [3] have studied errors made by college students in the United States and ESL learners. Results showed that half of the ten most frequent error types made by native speakers were 'negligible' in ESL writings [6].





Different kind of errors

Errors are usually identified at five levels:

- lexical level (*tipe vs. type)
- syntactic level (the church *is/was rebuilt in 1948)
- semantic level (*big/long conversation)
- discourse structure level (some think [...], others believe [...]; I cannot agree with *them)
- pragmatic level (* The author rejects the weather for it is his father)

[5], [1], [4]





Modern approaches to grammatical error correction





Early approaches

- Rule-based systems
 - hand-coded rules based on syntactic analysis
 - easy to implement and still widely used
 - highly productive nature of language, avoided as a general solution
- Advent of large-scale resources: data-driven approaches
 - machine learning classifiers for closed-class errors
 - best pick among set of all possible correction candidates
 - possible features: POS tags, neighbouring words, dependency trees





Machine translation and error correction

- translate grammatically incorrect sentences into correct ones
- learn correction mappings from parallel examples
- Statistical machine translation
 - transform well-formed sentences into ungrammatical ones
- Candidate re-ranking
 - SMT designed to capture translation regularities
 - n-best list of candidate translations re-ranked using re-ranking algorithms
- Neural machine translation
 - encoder-decoder framework
 - word embeddings help correct previously unseen errors







Modern approaches to grammatical error correction

Grammarly





What is Grammarly?

Proprietary digital writing tool Using Deep Learning technologies

- Grammar checking
- Spell checking
- Plagiarism detection
- Suggestions about vocabulary, tone, etc.



Figure: Logo of Grammarly.

The code of some functionalities is open source





Functionalities of Grammarly

- Spelling (lexical & semantic errors)
- Fluency
- Concise writing
- Plagiarism
- Vocabulary diversity (synonymy suggestion)
- Tone detection

- Adapted to your native language
- Complexity and clarity of sentences
- Detect non-engaging sentences
- Inclusive language
- Typography
- Consistency (11/08/99 vs 12th Nov. 2005)







Functionalities continued



Firemen are the heroes of the nation.

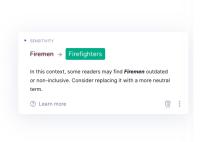


Figure: Inclusive writing.





Functionalities continued

possible from a wide range of speakers. The technique used to achieve this lexicon is called Grapheme-to-Phoneme (G2P) conversion. It is the task of predicting the pronunciation of a word given its graphemic or written form [34], It can be done by hand or using different algorithms.

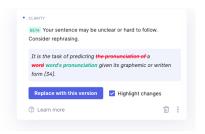


Figure: Clarity.





Possible improvements

- Implement a new language
- Add discourse analysis and/or coherence/anaphora correction
- Pragmatics?





Possible improvements continued

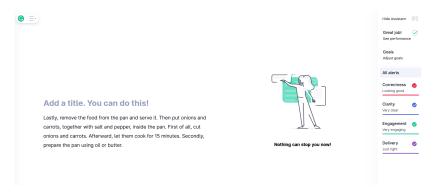


Figure: Discourse fallacy.





Modern approaches to grammatical error correction

Action plan





Action plan

- Objective
 - deploy an end-to-end GEC system on a web-application
- Technologies
 - NLP: Python (PyTorch, spaCy, OpenNMT)
 - Web-app: HTML, CSS, SQL, php, JS
- Data
 - use existing parallel corpora (NUCLE, CLC)
 - generate synthetic data?
- Evaluation
 - no single best evaluation metric, usefulness of a metric depends on the application
 - BEA Shared Task 2019, restricted/low resource track





Milestones



Figure: Milestones





References

X. Bao.

A discourse error analysis of english majors' english writing: Based on micro-level. *Studies in Literature and Language*, 10(5):29–35, 2015.

R. J. Connors and A. A. Lunsford.

Frequency of formal errors in current college writing, or ma and pa kettle do research. *College composition and communication*, 39(4):395–409, 1988.

S. Donahue.

Formal errors: Mainstream and esl students. In Conference of the Two-Year College Association (TYCA), 2001.

E. Kochmar.

Error detection in content word combinations.

Technical report, University of Cambridge, Computer Laboratory, 2016.

K. Kukich.

Techniques for automatically correcting words in text.

Acm Computing Surveys (CSUR), 24(4):377-439, 1992.

Z. Yuan.

Grammatical error correction in non-native english.

Technical report, University of Cambridge, Computer Laboratory, 2017.



