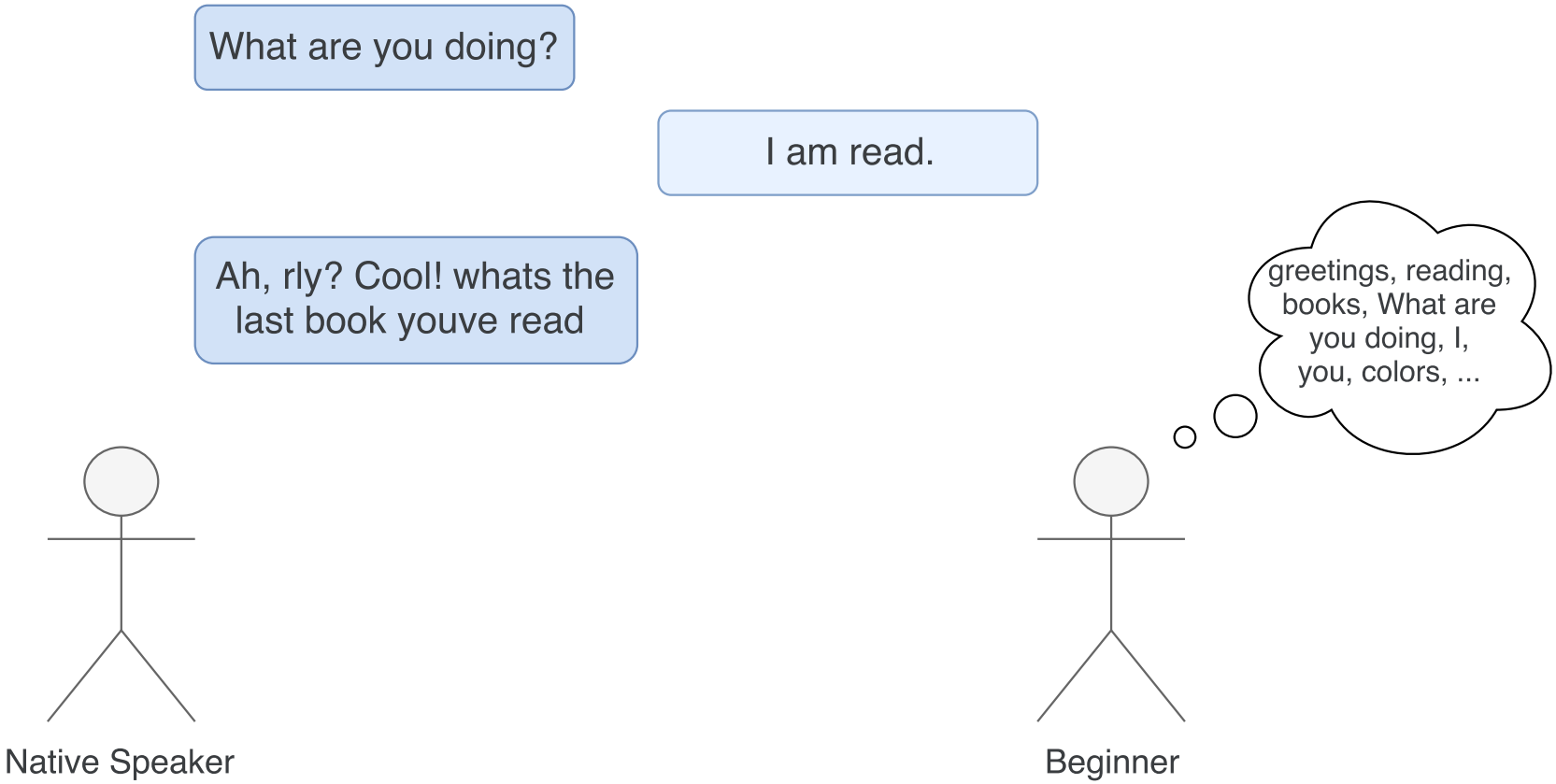
Chatbots for Language Learning

Anja Reusch

Technische Universität Dresden

Analyse eines Forschungsthemas (INF-D-960)

18.05.2018



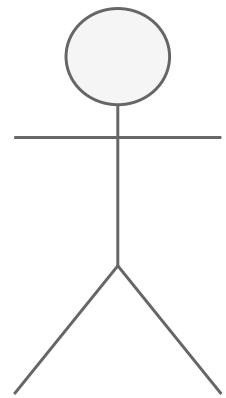
- Beginner talks to native speaker
- Beginner has limited knowledge of vocabulary and grammar

What are you doing?

I am read reading.

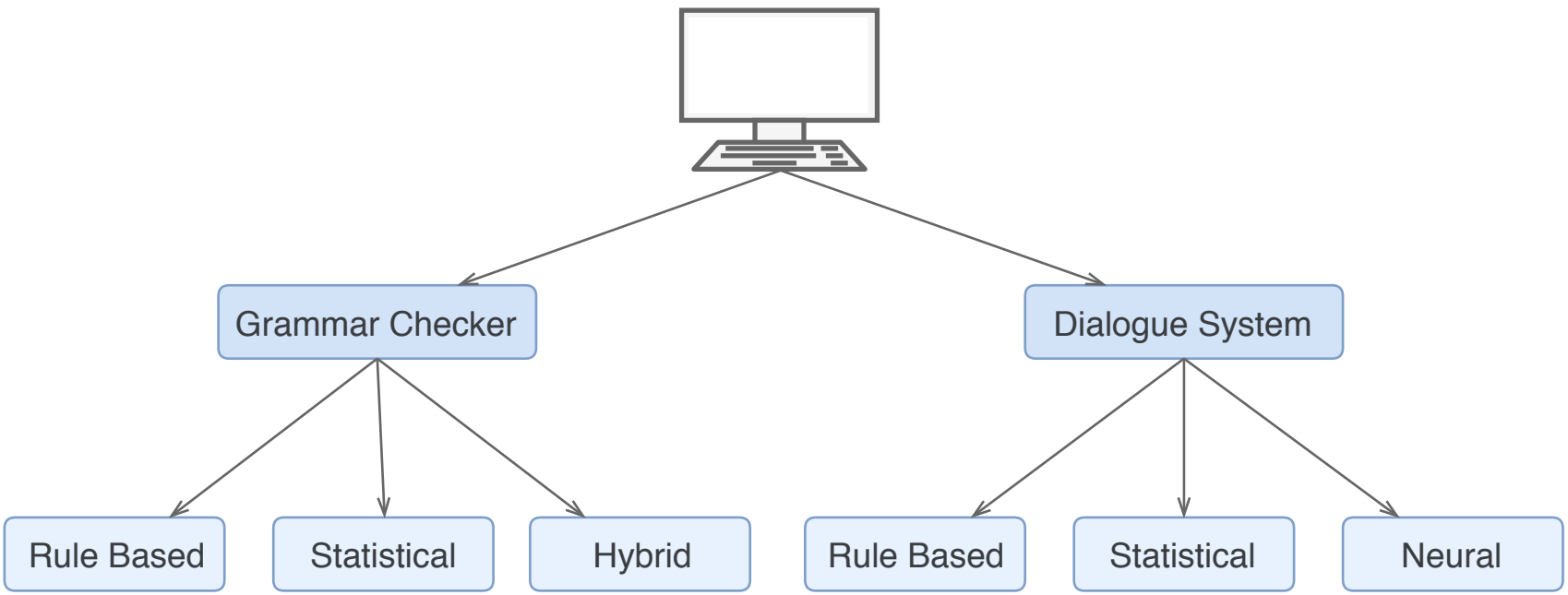
Interesting. What are you reading?

System



Beginner

Overview


Grammar Checker
Dialogue System
Rule Based
Statistical
Hybrid
Rule Based
Statistical
Neural

Overview

Problem

- User...
  - has limited knowledge (vocabulary, grammar)
  - does not understand native speakers
  - needs to practice to get better
- Grammar checker...
  - checks the users input for grammar mistakes
- Dialogue system...
  - communicates with user
  - can handle a small set of topics (e.g., shopping, asking for directions)
  - uses vocabulary and grammar the user knows

Dialogue System

- System $C = (\Sigma, \Gamma, \mathit{reply})$
- Input alphabet $\Sigma = \{I, you, go, dog, ...\}$, output alphabet Γ
- Dialog $D = (u_0, u_1), (u_2, u_3), ..., (u_{2i}, u_{2i+1})$ with $i \in \mathbb{N}$ and utterances $u_{2j} \in \Sigma^*, u_{2j+1} \in \Gamma^*, 0 \leq j \leq i$
- $\mathcal{D}$ set of all dialogues
- $\mathit{reply} : \Sigma^* \times \mathcal{D} \times \Theta \rightarrow \Gamma^*$
  - $\mathit{reply}(u, D, \theta) = u'$ defines the reply of the system
  - User input u
  - Dialog history D
  - Context parameters θ

Rule Based Systems [Dr.00]

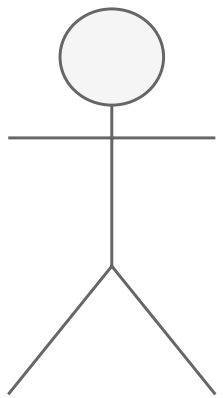
- Matches input with rules

Rule:

Hello *!
Hi, how are you?

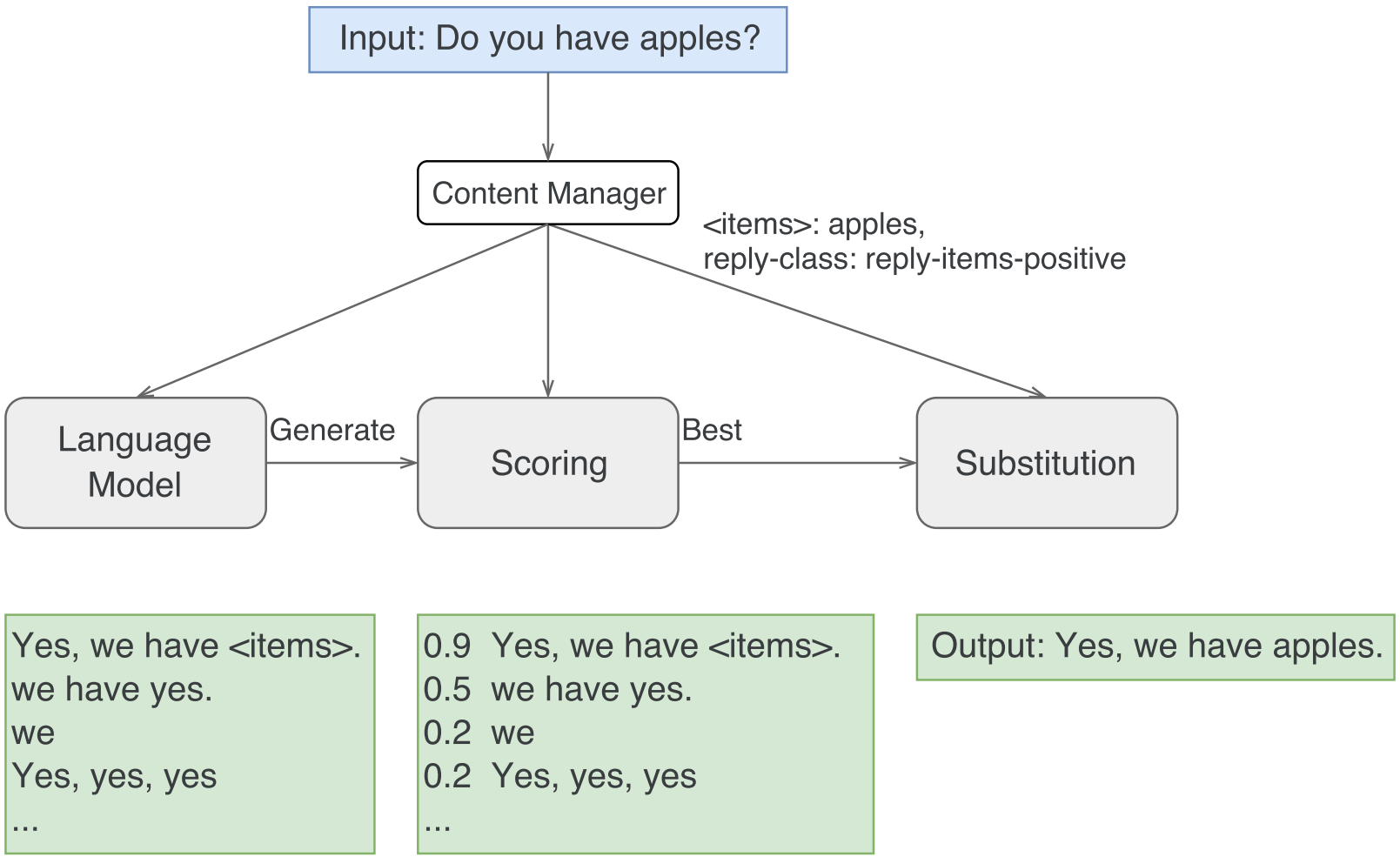
Hello Bot!

Hi, how are you?



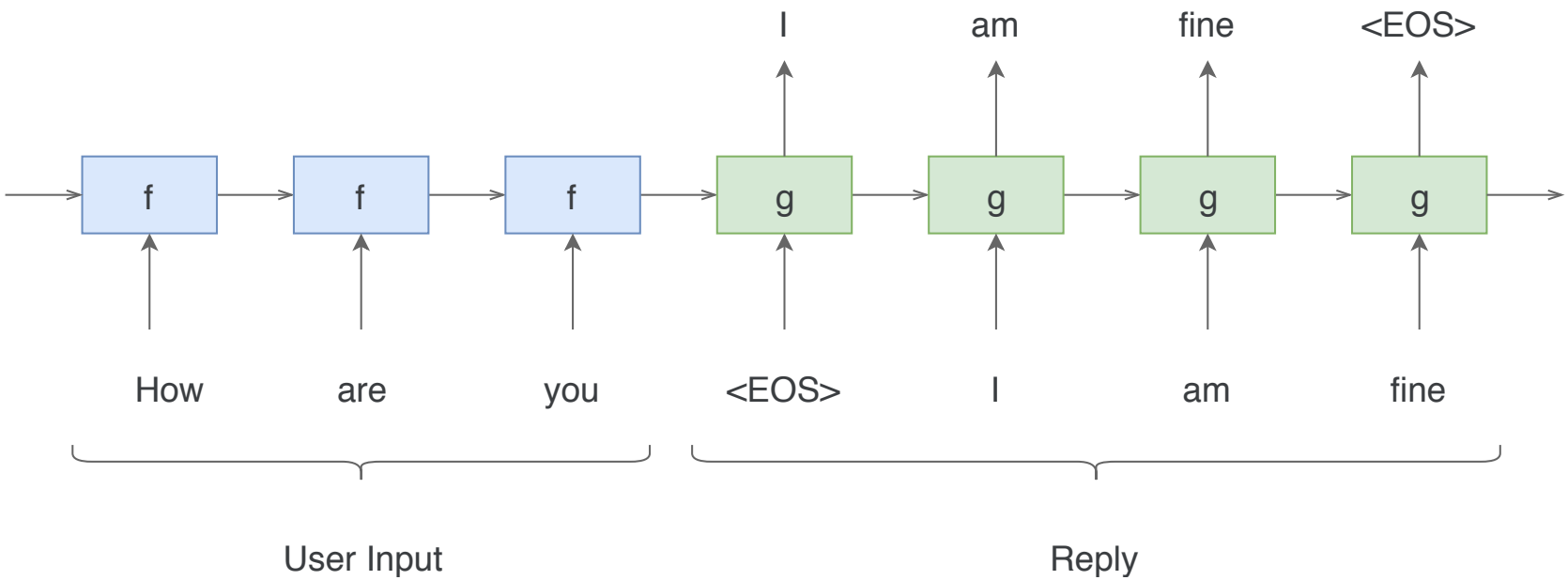
Statistical Systems [OR00]

- Analyzes input and generates output using a language model
- word class: set of words, e.g. {pens, apples, computers}
- reply class: type of the output, e.g. reply-items-positive



Neural Systems [VL15]

- end-to-end approach using Long Short-Term Memory cells [HS97]
- Words are encoded using word embeddings (high dimensional vectors)



Qualitative Evaluation

Rule Based Systems

- Rules written by hand are time consuming
- Language dependent
- Do not consider dialogue history

Statistical Systems

- More flexible than rule based systems
- Do not consider dialogue history
- Need annotated corpus

Neural Systems

- Consider dialogue history
- Needs large data set for training

Corpora and Metrics

Corpora

- Ubuntu Dialogue Corpus, 1 million multi-turn dialogues [LPSP15]
- Twitter data set, [RCD10]

Metrics

- Hard to find suitable evaluation, [LLS+16]
- Human judges, e.g., in [Jia09]

Grammar Checkers

Grammar Checkers

Correcting mistakes in sentences

- $f : \Sigma^* \to \Sigma^*$ ungrammatical sentence $\mapsto$ correct sentence

Annotating mistakes

Example:

"This is you're house." $\mapsto\{(\{3\}$, "Maybe you confused your and you're")$\}$

- $g : \Sigma^* \to \mathcal{P}(\mathcal{P}([n]) \times \Sigma^*)$
- Input: potentially ungrammatical sentence
- Output: One tuple for each error over positions and annotations with helpful information

Rule Based Systems [Nab03]

- Match input with rules
- If rule matches, message is displayed
- Example rule:

```
<rule>
  <pattern>
    <token postag="SENT_START"/><marker>
    <token>Your</token></marker>
    <token regexp="yes">not|an?|the</token>
  </pattern>
  <message>Did you mean <suggestion>You're</suggestion>?
  </message>
</rule>
```

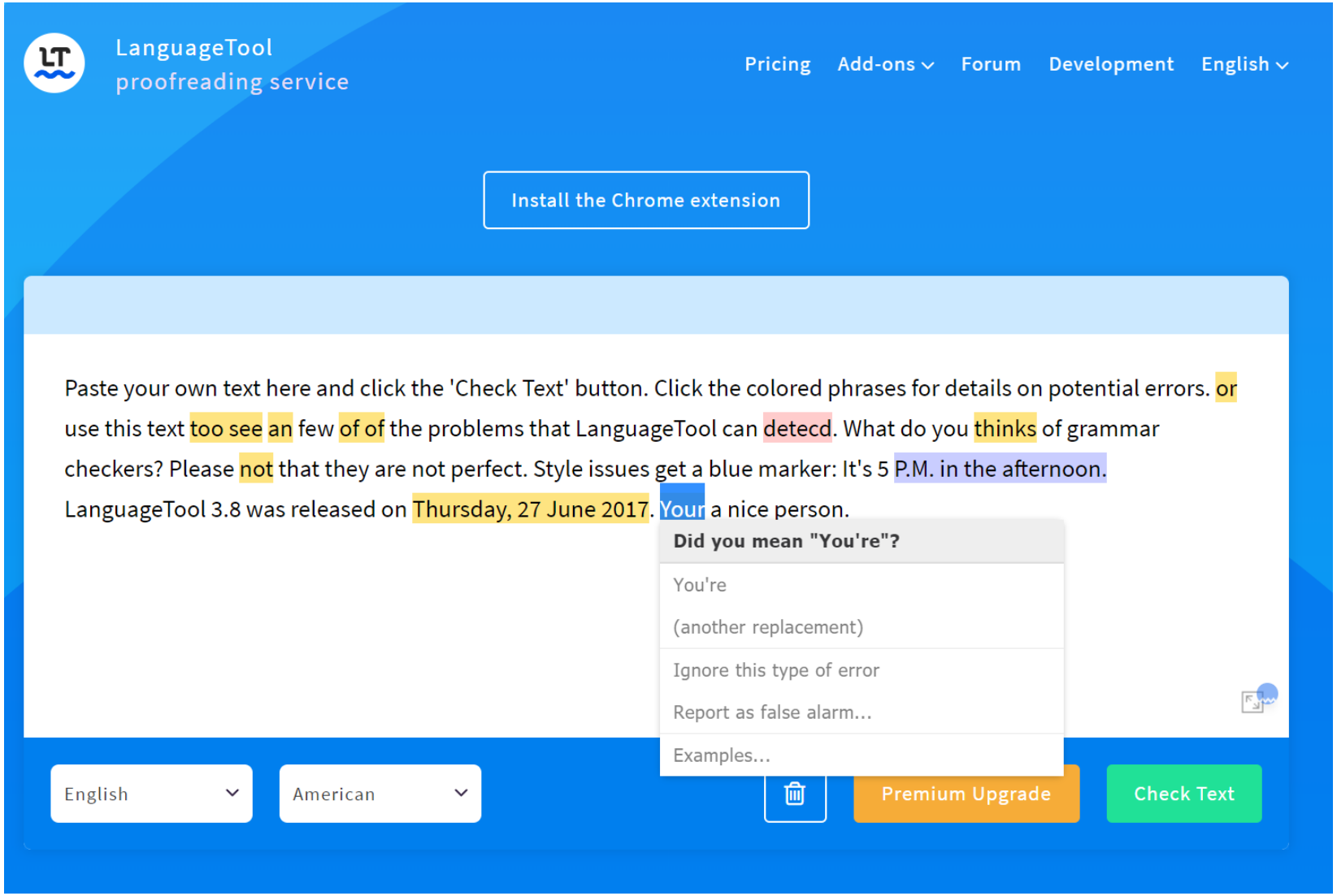
- Matches "Your a nice person."

Rule Based Systems

LanguageTool (https://languagetool.org/)

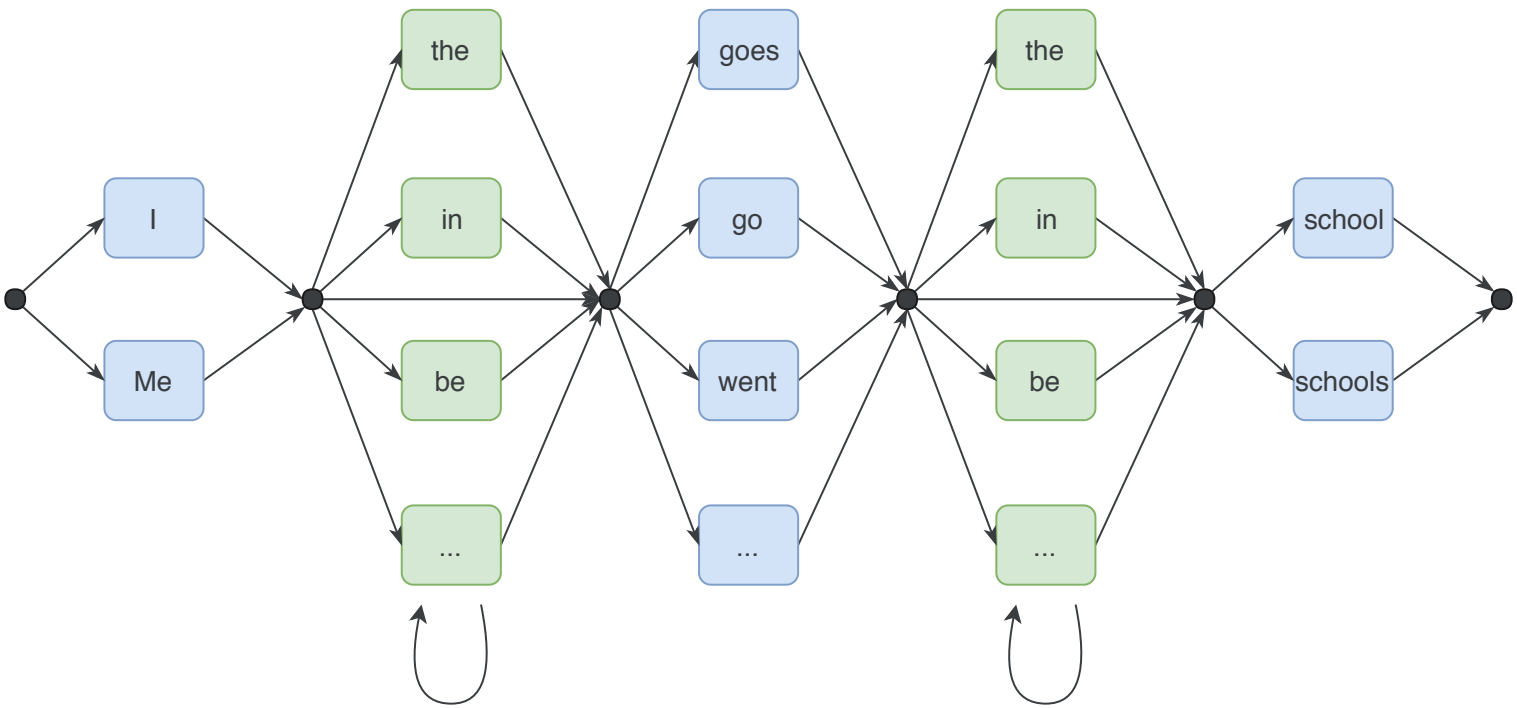
- Rule Based Grammar and Style Checker
- Community based
- Over 30 languages and dialects
- Over 2000 rules for European languages

LanguageTool



Statistical Systems [LS06]

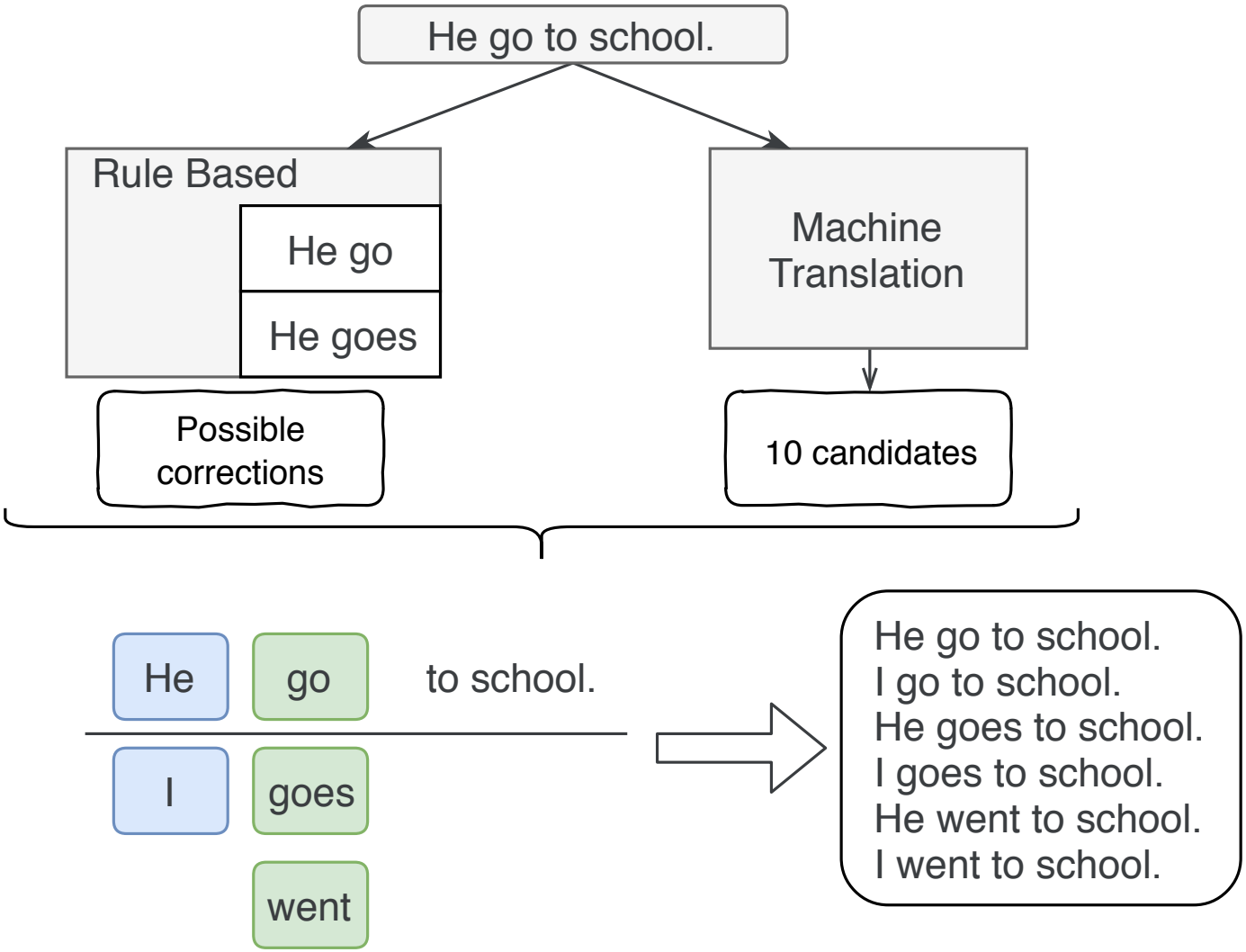
- Tries to correct ungrammatical sentence
- Creates normal form:
  - Remove articles, prepositions and auxiliaries as "can" or "would"
  - Transform nouns into singular and verbs into infinitive
- Builds a lattice by inserting all possible articles, prepositions,...
- Uses a trigram language model to score sentences



Hybrid Systems [FYA+14]

- Combines rules with statistical machine translation
- Rules:
  - Rules are obtained from unigram, bigrams and trigrams of a learner corpus
  - n-gram contains error if 90 percent of its occurrences are wrong
- Translates ungrammatical sentence into grammatical sentence
- 10 candidates are created using a 4-gram language model
- Combine both corrections

Hybrid Systems [FYA+14]



Qualitative Evaluation

Rule Based Systems

- Writing rules by hand is time consuming
- Language dependent
- Working open source system LanguageTool

Statistical Systems

- Good for Japanese English learners
- Not designed for other languages
- Does not consider context

Hybrid System

- Needs learner corpus
- Does not consider context

Metrics and Corpora

Metrics

- BLEU Score
- Precision, Recall, $F_{0.5}$-Score
- Human judges

Corpora

- Cambridge Learner Corpus (not open to the public) [Nic03]
- UD English-ESL / Treebank of Learner English [BKS+16]
- EAGLE: an Error-Annotated Corpus of Beginning Learner German [Boy10]

Conclusion

Dialog Systems:

- Statistic and neural systems are promising
- Not enough training data, especially in language learning topics
- Evaluation metrics not clear

Grammar Checkers:

- Good existing systems
- Largest corpus not open to the public

$\Rightarrow$ Due to these problems it is not possible to implement such a system quickly or easily.

The End

References I


Yevgeni Berzak, Jessica Kenney, Carolyn Spadine, Jing Xian Wang, Lucia Lam, Keiko Sophie Mori, Sebastian Garza, and Boris Katz, *Universal dependencies for learner english*, arXiv preprint arXiv:1605.04278 (2016).


Adriane Boyd, *Eagle: an error-annotated corpus of beginning learner german*, LREC, 2010.


Dr. Richard S. Wallace, *Aiml overview*, 2000.


Mariano Felice, Zheng Yuan, Øistein E. Andersen, Helen Yannakoudakis, and Ekaterina Kochmar, *Grammatical error correction using hybrid systems and type filtering*, Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task, 2014, pp. 15–24.

References II

- Sepp Hochreiter and Jürgen Schmidhuber, *Long short-term memory*, Neural computation **9** (1997), no. 8, 1735–1780.
- Jiyou Jia, *Csiec: A computer assisted english learning chatbot based on textual knowledge and reasoning*, Knowledge-Based Systems **22** (2009), no. 4, 249–255.
- Chia-Wei Liu, Ryan Lowe, Iulian Serban, Mike Noseworthy, Laurent Charlin, and Joelle Pineau, *How not to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation*, Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2016, pp. 2122–2132.
- Ryan Lowe, Nissan Pow, Iulian Serban, and Joelle Pineau, *The ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems*, 2015.

References III


John Lee and Stephanie Seneff, *Automatic grammar correction for second-language learners*, Ninth International Conference on Spoken Language Processing, 2006.


Daniel Naber, *A rule-based style and grammar checker*.


Diane Nicholls, *The cambridge learner corpus: Error coding and analysis for lexicography and elt*, Proceedings of the Corpus Linguistics 2003 conference, vol. 16, 2003, pp. 572–581.


Alice H. Oh and Alexander I. Rudnicky, *Stochastic language generation for spoken dialogue systems*, ANLP/NAACL 2000 Workshop on Conversational systems - (Morristown, NJ, USA) (Unknown, ed.), Association for Computational Linguistics, 2000, pp. 27–32.

References IV


Alan Ritter, Colin Cherry, and Bill Dolan, *Unsupervised modeling of twitter conversations*, Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2010, pp. 172–180.


Oriol Vinyals and Quoc Le, *A neural conversational model*, arXiv preprint arXiv:1506.05869 (2015).