Leveraging Known
Semantics for
Spelling Correction

Levi King &
Markus Dickinson

Motivation

Data & Method

Spelling correction

Leveraging Known Semantics for Spelling Correction

Levi King & Markus Dickinson
Indiana University

NLP4CALL 2014, November 13; Uppsala, Sweden

Motivation

Issue:

- ▶ Intelligent Computer-Assisted Language Learning (ICALL) / Intelligent Language Tutor (ILT) systems tend to focus on grammatical errors & feedback.
- ▶ Second Language Acquisition (SLA) research has established:
    - ▶ correcting a learner's grammar is often ineffective
    - ▶ a dispreference for explicit grammar instruction

Overarching Goal:

- ▶ See ICALL/ILT focus on interaction, with learners producing *more* target language rather than perfect target language.
- ▶ **Barrier:** NNS misspellings!

This work: Boost system performance with spelling correction & language model (LM).

Leveraging Known
Semantics for
Spelling Correction

Levi King &
Markus Dickinson

Motivation

Data & Method

Spelling correction

Data Collection

Picture Description Task (PDT) example

Response (L1)
He is droning his wife pitcher. (Arabic)
The artist is drawing a pretty women. (Chinese)
The artist is painting a portrait of a lady. (English)
The painter is painting a woman's paint. (Spanish)

Method

Leveraging Known
Semantics for
Spelling Correction

Levi King &
Markus Dickinson

Motivation

Data & Method

Spelling correction

1. Parse a sentence into a dependency representation
2. Extract a simple semantic form ($v(s,o)$) from this parse
   ▶ to compare to gold standard (NS) semantic forms

Spelling correction via language modeling

Leveraging Known
Semantics for
Spelling Correction

Levi King &
Markus Dickinson

Motivation

Data & Method

Spelling correction

- Consider the following NNS sentence:

  A women in straet ride bicycly.

- A proficient English speaker can use context and knowledge of the language to understand this
  - Our system has lacked such ability.
- To compensate for this, we implemented a preprocessing module incorporating:
  - spelling correction & trigram LM (linguistic knowledge)
  - NS responses (context)

Spelling correction process

1. Obtain spelling suggestions for all NNS words using Aspell/Enchant.
2. Keep any NNS words/suggestions matching NS words or stop words & combine remaining suggestions to get all possible sentences.
3. Evaluate candidate sentences with trigram LM (trained on newspaper text).
4. Select candidate sentence with lowest perplexity.

Spelling correction evaluation

The spelling correction module resulted in a 13.7% decrease in errors & 13.4% increase in coverage.

Performance is limited by the gold standard (14 NSs).

- ▶ Including false negatives results in a 25.9% decrease in errors.

Example problematic sentences

Good correction:

(1a) NNS: the old man shouts to beird. $\Rightarrow \emptyset$(shout,beird)
(1b) LM: the old man shoots to bird. $\Rightarrow$ shoot(man,bird)

(2a) NNS: a men is swapping the leaves. $\Rightarrow$ swap(man,leaf)
(2b) LM: a man is sweeping the leaves. $\Rightarrow$ sweep(man,leaf)

Bad correction:

(3a) NNS: a man shooted a bird. $\Rightarrow$ shoot(man,bird)
(3b) LM: a man shouted a bird. $\Rightarrow$ shout(man,bird)