

# Language and Computers

## LTS: Grammatical Error Detection

Based on slides from Ross Israel  
Indiana University  
Spring 2017

Language and Computers  
LTS: Grammatical Error Detection

Introduction  
NLP and (I)CALL  
Types of Errors  
Error Detection in Action  
Techniques  
A Quick Intro to Machine Learning  
English Prepositions  
References

1 / 16

## What Is Grammatical Error Detection?

We will be talking about errors made by learners in a second language acquisition context.

Language learners often make non-native-like mistakes when constructing sentences:

- ▶ We arrived *to* the station.
- ▶ There is *the* garden in my house.
- ▶ I *eat* rice, nikujaga and salad yesterday.

**Grammatical error detection** entails trying to find these mistakes automatically.

Recommended reading: Leacock et al 2014, *Automated Grammatical Error Detection for Language Learners*

Language and Computers  
LTS: Grammatical Error Detection

Introduction  
NLP and (I)CALL  
Types of Errors  
Error Detection in Action  
Techniques  
A Quick Intro to Machine Learning  
English Prepositions  
References

2 / 16

## Where Is This Useful?

- ▶ Automatic grading
  - ▶ Language teachers
  - ▶ Standardized testing
- ▶ Analysis and annotation of learner data for research
- ▶ Language learning software (ICALL)

Language and Computers  
LTS: Grammatical Error Detection

Introduction  
NLP and (I)CALL  
Types of Errors  
Error Detection in Action  
Techniques  
A Quick Intro to Machine Learning  
English Prepositions  
References

3 / 16

## Where Error Detection Fits In

(a bit of a review)

- ▶ CALL: Computer Assisted Language Learning
  - ▶ Using computers and media in language learning and teaching
  - ▶ e.g., Rosetta Stone, eLanguage
  - ▶ Exercises are typically very simple in design, and offer little feedback
- ▶ ICALL: Intelligent Computer Assisted Language Learning
  - ▶ Utilize computational linguistics tools, such as POS tagging & parsing along with statistical language modeling strategies (e.g., n-grams)
    - ▶ These tools often need to be altered to expect and diagnose errors, or at least handle learner data better
    - ▶ We can also build software for *specific* kinds of errors. (**today's discussion**)
  - ▶ Focus on precision; we don't want to tell a learner that they've made a mistake when they haven't!

Language and Computers  
LTS: Grammatical Error Detection

Introduction  
NLP and (I)CALL  
Types of Errors  
Error Detection in Action  
Techniques  
A Quick Intro to Machine Learning  
English Prepositions  
References

4 / 16

## Learner Errors

Learners typically make different kinds of mistakes than native speakers.

- ▶ Content Word Choice (19.9% of all errors in *CLC*)
  - ▶ We need to deliver the merchandise on a daily *\*base/basis*.
- ▶ Preposition Error (13.4%)
  - ▶ Our society is developing *\*in/at* high speed.
- ▶ Determiner Error (11.7%)
  - ▶ There is *\*the/a* garden in my house.

*CLC* = Cambridge Learner Corpus

Language and Computers  
LTS: Grammatical Error Detection

Introduction  
NLP and (I)CALL  
Types of Errors  
Error Detection in Action  
Techniques  
A Quick Intro to Machine Learning  
English Prepositions  
References

5 / 16

## Motivation

Some common areas of research in English error detection are articles, prepositions, and collocations. We'll look a little more in depth at prepositions.

- ▶ Because prepositions make up a large portion of errors commonly made by learners, there has been a good deal of research on how to find and diagnose preposition errors
- ▶ Also, *crucially*: prepositions are a closed set, so it's a problem that's easier to define than a more open error type like use of the wrong content word
  - ▶ Prepositions can be treated as a confusion set where we know that one is being substituted for another
  - ▶ This is not the case with many parts of the grammar

Language and Computers  
LTS: Grammatical Error Detection

Introduction  
NLP and (I)CALL  
Types of Errors  
Error Detection in Action  
Techniques  
A Quick Intro to Machine Learning  
English Prepositions  
References

6 / 16

# Cloze Test

- ▶ Choosing the correct preposition can be a tough task even for native speakers

There has been concern \_\_\_\_ syncing phone contacts \_\_\_\_ Facebook. "As long as you are aware of who is \_\_\_\_ the group it can be a great privacy tool. If it gets out \_\_\_\_ hand it could give you a sense \_\_\_\_ false security." The roll out \_\_\_\_ new products comes \_\_\_\_ reports that a syncing feature \_\_\_\_ the iPhone lets Facebook access contact data and share it \_\_\_\_ the site. "It's very possible that your private phone numbers - and those \_\_\_\_ lots \_\_\_\_ your and their friends - are \_\_\_\_ the site," said Charles Arthur \_\_\_\_ the Guardian newspaper.

Language and Computers  
LTS: Grammatical Error Detection

Introduction  
NLP and (I)CALL

Types of Errors

Error Detection in Action

Techniques  
A Quick Intro to Machine Learning  
English Prepositions

References

7 / 16

# Cloze Test

- ▶ Choosing the correct preposition can be a tough task even for native speakers

There has been concern **over** syncing phone contacts **with** Facebook. "As long as you are aware of who is **in** the group it can be a great privacy tool. If it gets out **of** hand it could give you a sense **of** false security." The roll out **of** new products comes **amid** reports that a syncing feature **on** the iPhone lets Facebook access contact data and share it **on** the site. "It's very possible that your private phone numbers - and those **of** lots **of** your and their friends - are **on** the site," said Charles Arthur **of** the Guardian newspaper.

Language and Computers  
LTS: Grammatical Error Detection

Introduction  
NLP and (I)CALL

Types of Errors

Error Detection in Action

Techniques  
A Quick Intro to Machine Learning  
English Prepositions

References

7 / 16

# Challenges with Prepositions

- ▶ Negative Transfer: *in the garden, at home, on (the) campus*: same preposition in Arabic
- ▶ Adjuncts: *on the beach* vs. *at the beach*
- ▶ Arguments: *The loaded the hay on the wagon* vs. *The loaded the wagon with hay*
- ▶ Phrasal Verbs: *add up the numbers, add the numbers up*
- ▶ Idioms: *on the house*
- ▶ PP Attachment: *I put the ring on the table in the safe*
- ▶ Lexical Ambiguity: *eat with a fork, view with anxiety, strike with fear, combine with others, furnish with supplies*

See section 3.3.1 of Leacock et al (2014)

Language and Computers  
LTS: Grammatical Error Detection

Introduction  
NLP and (I)CALL

Types of Errors

Error Detection in Action

Techniques  
A Quick Intro to Machine Learning  
English Prepositions

References

8 / 16

# Commonly Used Techniques

A sampling ...

- ▶ Language Model - Gamon et al. (2008)
  - ▶ Build *n*-grams of POS and/or parsing labels from native text and check if learner *n*-grams align with that model
- ▶ Web-based methods - Gamon and Leacock (2010)
  - ▶ Take a few words of context on either side of a preposition to generate a web query
  - ▶ Replace the preposition with neighbors from a confusion set and search those queries
  - ▶ The search with the greatest number of hits is selected as the right answer
- ▶ Heuristic-based systems - Eeg-Olofsson and Knutsson (2003)
  - ▶ Write linguistic rules designed to find errors in learner data
- ▶ Statistical - Tetreault and Chodorow (2008)
  - ▶ Statistical methods means building a classifier
  - ▶ So, what is a classifier?

Language and Computers  
LTS: Grammatical Error Detection

Introduction  
NLP and (I)CALL

Types of Errors

Error Detection in Action

Techniques  
A Quick Intro to Machine Learning  
English Prepositions

References

9 / 16

# Machine Learning

**Machine Learning**: give examples to a computer system & have it learn what the patterns are

- ▶ We will explore this topic in more detail when we get to the *document classification* unit

Example: based on your previous purchases, what coupons should you receive?

- ▶ Home & Gardening coupons? Music & Movie coupons? Grocery coupons? etc.

Language and Computers  
LTS: Grammatical Error Detection

Introduction  
NLP and (I)CALL

Types of Errors

Error Detection in Action

Techniques  
A Quick Intro to Machine Learning  
English Prepositions

References

10 / 16

# Understanding Classifiers

Machine learning is not as scary as it sounds!

- ▶ There are a number of algorithms for classification that we could talk about
  - ▶ Maximum Entropy, Support Vector Machines, Memory Based Learning
  - ▶ Each method requires different representations of information
  - ▶ These slides are indicative of Memory Based Learning

Language and Computers  
LTS: Grammatical Error Detection

Introduction  
NLP and (I)CALL

Types of Errors

Error Detection in Action

Techniques  
A Quick Intro to Machine Learning  
English Prepositions

References

11 / 16

# Running a Classifier

- ▶ We will need two sets of data:
  - ▶ Training Set - needs to be big
  - ▶ Testing Set - usually smaller
- ▶ The data sets are full of events (*instances*) that contain *features* that describe the circumstances of the event and a *class* that is the answer we are trying to guess
- ▶ Basic process:
  - ▶ Open a file
  - ▶ Extract bits of text (features) that you deem useful
  - ▶ Print those bits of text on a single line for each instance

**Challenge:** select appropriate features

Language and Computers  
LTS: Grammatical Error Detection

Introduction  
NLP and (I)CALL

Types of Errors

Error Detection in Action

Techniques

A Quick Intro to Machine Learning

English Prepositions

References

12 / 16

# Selecting Features

Let's consider a real-world example:

- ▶ The Task: We want to classify the weather as either **good** or **bad**.
- ▶ We would want features like
  - ▶ temperature
  - ▶ sunny?
  - ▶ cloudy?
  - ▶ windy?
  - ▶ humidity level
  - ▶ rain/snow/none

Language and Computers  
LTS: Grammatical Error Detection

Introduction  
NLP and (I)CALL

Types of Errors

Error Detection in Action

Techniques

A Quick Intro to Machine Learning

English Prepositions

References

13 / 16

# Running the Classifier

- ▶ Then, we would build **vectors** for every measurement we take and *label* them to build training data:
  - ▶ 75,yes,no,no,70%,none,**good**
  - ▶ 35,no,no,yes,50%,none,**bad**
  - ▶ 105,yes,no,no,98%,rain,**bad**
  - ▶ 68,yes,yes,no,75%,none,**good**
- ▶ Now, when we give the classifier an unknown feature vector, we hope that it makes a wise decision
  - ▶ 85,yes,no,no,65%,none - classifier's guess = **good** yay!
  - ▶ 15,no,no,yes,70%,snow - classifier's guess = **bad** yay!
  - ▶ 75,yes,no,yes,70%,none - classifier's guess = **bad** oops!

# Machine Learning for Prepositions

- ▶ Tetreault and Chodorow used a maximum entropy classifier to try to find preposition confusions and extraneous uses
- ▶ They extracted 25 features including:
  - ▶ words/POS tags in a 2 word window(+/-) around preposition
  - ▶ the head verb and noun of the preceding VP and NP
  - ▶ the head noun of the following NP
- ▶ *John went to the store this morning.*
  - ▶ word+POS bigrams: went.VBD, the.DET
  - ▶ head of previous VP = went
  - ▶ head of previous NP = John
- ▶ Their system achieved 84% precision and 19% recall.
  - ▶ This might sound low, but keep in mind, we want to get the best possible precision, even if it means losing recall.

Language and Computers  
LTS: Grammatical Error Detection

Introduction  
NLP and (I)CALL

Types of Errors

Error Detection in Action

Techniques

A Quick Intro to Machine Learning

English Prepositions

References

14 / 16

Language and Computers  
LTS: Grammatical Error Detection

Introduction  
NLP and (I)CALL

Types of Errors

Error Detection in Action

Techniques

A Quick Intro to Machine Learning

English Prepositions

References

15 / 16

# Types of systems

Systems differ in terms of:

1. the kinds of features they use
  - ▶ surface level features, syntactic features, L1 information, etc.
2. the training data they use
  - ▶ correct usage, artificially generated errors, real errors
3. the kinds of models (e.g., classifiers) they use
  - ▶ classifiers, language models, web counts, etc.

# References

Daelemans, Walter, Jakob Zavrel, Ko van der Sloot, Antal van den Bosch, Timbl Tilburg and Memory based Learner (2007). TIMBL: Tilburg Memory-Based Learner - version 6.1 - Reference Guide.

Eeg-Olofsson, Jens and Ola Knutsson (2003). Automatic Grammar Checking for Second Language Learners - the Use of Prepositions. In *Proceedings of Nodalida'03*. Reykjavik, Iceland.

Gamon, Michael, Jianfeng Gao, Chris Brockett, Alexander Klementiev, William Dolan, Dmitriy Belenko and Lucy Vanderwende (2008). Using Contextual Speller Techniques and Language Modeling for ESL Error Correction. In *Proceedings of IJCNLP-08*. Hyderabad, India.

Gamon, Michael and Claudia Leacock (2010). Search right and thou shalt find... Using Web Queries for Learner Error Detection. In *Proceedings of NAACL 2010*.

Tetreault, Joel and Martin Chodorow (2008). The ups and downs of preposition error detection in ESL writing. In *Proceedings of COLING-08*. Manchester.

Language and Computers  
LTS: Grammatical Error Detection

Introduction  
NLP and (I)CALL

Types of Errors

Error Detection in Action

Techniques

A Quick Intro to Machine Learning

English Prepositions

References

16 / 16

Language and Computers  
LTS: Grammatical Error Detection

Introduction  
NLP and (I)CALL

Types of Errors

Error Detection in Action

Techniques

A Quick Intro to Machine Learning

English Prepositions

References

16 / 16