Automating Second Language Acquisition Research: Integrating Information Visualisation and Machine Learning

Helen Yannakoudakis, Ted Briscoe, Theodora Alexopoulou

University of Cambridge

Visualisation of Linguistic Patterns

EACL 2012
Outline

1. Introduction
2. Dataset
3. System
4. Case study
5. Conclusions
Common European Framework of Reference for Languages (CEFR)

International benchmark of language attainment at different stages of learning
Introduction

Common European Framework of Reference for Languages (CEFR)

International benchmark of language attainment at different stages of learning
Common European Framework of Reference for Languages (CEFR)

Divides learners into three broad divisions:

A Basic User
   A1 Breakthrough or beginner
   A2 Waystage or elementary

B Independent User
   B1 Threshold or intermediate
   B2 Vantage or upper intermediate
      (e.g., can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options)

C Proficient User
   C1 Effective Operational Proficiency or advanced
   C2 Mastery or proficiency
Common European Framework of Reference for Languages (CEFR)

International benchmark of language attainment at different stages of learning

English Profile (EP) research programme

- Enhance the learning, teaching and assessment of English as an additional language
- Reference level descriptions of the language abilities expected at each learning stage
Common European Framework of Reference for Languages (CEFR)

- International benchmark of language attainment at different stages of learning

English Profile (EP) research programme

- Enhance the learning, teaching and assessment of English as an additional language
- Reference level descriptions of the language abilities expected at each learning stage

Goal

- Understand the linguistic abilities that characterise different levels of attainment and, more generally, developmental aspects of learner grammars
**Theory-driven approach**

**Approach**

- Theory-driven approach
- Linguistic intuition
- Literature on learner English
- Hypotheses that are well understood
  - Target language determiner systems cause problems for learners whose native language doesn't utilise determiners
Theory-driven approach

Approach

- Theory-driven approach
- Linguistic intuition
- Literature on learner English
- Hypotheses that are well understood
  - Target language determiner systems cause problems for learners whose native language doesn’t utilise determiners
- Risks ‘finding the obvious’
- Large-scale databases
  - How can we extract data efficiently and reliably to evaluate linguistic hypotheses?
  - How can we make ”observations” or extract patterns that may lead to new hypotheses?
## Data-driven approach

**Our approach**
- More empirical perspective for linguistic hypotheses on learner grammars
- Machine Learning

**Advantages**
- Partially automate the process of hypothesis creation
- Alternative route to learner grammars
- Useful adjunct to hypothesis-driven approach
- Powerful methodology for exploring a large hypothesis space
- Data-driven approaches quantitatively very powerful
First Certificate in English (FCE) exam

FCE Writing Component

- CEFR level: vantage or upper-intermediate (B2)
- Two tasks eliciting free-text answers, each one between 120 and 180 words (e.g. ‘write a short story commencing …’)
- Answers annotated with mark (in the range 1–40), fitted to a RASCH model (Fischer and Molenaar, 1995)
- Manually error-coded using a taxonomy of \( \sim 80 \) error types (Nicholls, 2003)

Meta-data

- Candidate’s grades
- Native language
- Age
FCE Writing Component

Manually error-coded using a taxonomy of ~80 error types (Nicholls, 2003)

Examples

It is a very beautiful place and the people there <NS type='AGV'> <i>is</i> <c>are</c> </NS> very kind and generous.

I will give you all <NS type='MD'> <c>the</c> </NS> information you need.
FCE Writing Component

Manually error-coded using a taxonomy of ∼80 error types (Nicholls, 2003)

Examples

*It is a very beautiful place and the people there* <NS type='AGV'> *is</i> <c>are</c> </NS> very kind and generous.*

*I will give you all* <NS type='MD'> *the</c> </NS> information you need.*
FCE Writing Component

Manually error-coded using a taxonomy of ~80 error types (Nicholls, 2003)

Examples

*It is a very beautiful place and the people there are very kind and generous.*

*I will give you all the information you need.*
Discriminative Learning

- Supervised discriminative machine learning methods to automate the assessment of the FCE exam (Briscoe et al., 2010)
  - Binary classifier that best discriminates passing from failing FCE scripts (trained on FCE scripts)
    - Linear Perceptron classifier
  - Feature set: lexical and part-of-speech (POS) ngrams (among other feature types)
### Highly Ranked Discriminative Feature Instances

<table>
<thead>
<tr>
<th>Feature</th>
<th>Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>VM_RR (+)</td>
<td>POS bigram</td>
<td>could clearly</td>
</tr>
<tr>
<td>_because (−)</td>
<td>word bigram</td>
<td>*, because</td>
</tr>
<tr>
<td>how_to (−)</td>
<td>word bigram</td>
<td>teach the others how to dance</td>
</tr>
<tr>
<td>necessary (+)</td>
<td>word unigram</td>
<td>it is necessary that</td>
</tr>
<tr>
<td>the_people (−)</td>
<td>word bigram</td>
<td>*the people are clever</td>
</tr>
<tr>
<td>probably (+)</td>
<td>word unigram</td>
<td>we are probably going</td>
</tr>
<tr>
<td>VVϕ_VVϕ (−)</td>
<td>POS bigram</td>
<td>*technology keep develop</td>
</tr>
<tr>
<td>NN2_VVG (+)</td>
<td>POS bigram</td>
<td>children smiling</td>
</tr>
<tr>
<td>II_VVN (−)</td>
<td>POS bigram</td>
<td>*I want to gone</td>
</tr>
</tbody>
</table>

Helen Yannakoudakis, Ted Briscoe, Theodora Alexopoulou

Automating Second Language Acquisition Research
Issues

- Hundreds of thousands of discriminative feature instances
- Proxies to aspects of the grammar and need interpretation
- Evaluate higher-level, more general and comprehensible hypotheses
Information Visualisation

**Appeal**

- Gain a deeper understanding of important phenomena that are represented in a database
- Navigate large amounts of data faster, intuitively, and with relative ease
- No need to learn query language syntax
- Identify the most productive paths to pursue
Feature relations

Given \( S = \{s_1, s_2, \ldots, s_N\} \) and \( F = \{f_1, f_2, \ldots, f_M\} \), a feature \( f_i \in F \) is associated with a feature \( f_j \in F \), where \( i \neq j \) and \( 1 \leq i, j \leq M \), if their relative co-occurrence score is within a predefined range:

\[
\text{score}(f_j, f_i) = \frac{\sum_{k=1}^{N} \text{exists}(f_j, f_i, s_k)}{\sum_{k=1}^{N} \text{exists}(f_i, s_k)}
\]  

(1)

where \( s_k \in S \), \( 1 \leq k \leq N \), \( \text{exists}() \) is a binary function that returns 1 if the input features occur in \( s_k \), and \( 0 \leq \text{score}(f_j, f_i) \leq 1 \).

Two features are connected by an edge if their score is within a user-defined range.

Outgoing edges of \( f_i \): \( \text{score}(f_j, f_i) \), incoming edges of \( f_i \): \( \text{score}(f_i, f_j) \).
The functionality, usability and tractability of graphs is severely limited when the number of nodes and edges grows by more than a few dozen (Fry, 2007)
Dynamic creation of graphs – cont.

Graph of the 5 most frequent negative features using a score range of 0.8–1
107374.0 I could also feed them, teach them and sing them some songs or tell them how to make their beds or hoover.

97401.0 I'd like to apply more information about these courses and how to join the club.

81760.0 The best thing would be to use always always to use a bicycle because nowadays people are sitting spending most of the daytime in offices, schools, at home in front of the TV, in cars, and in some cases there are people whose don't know...
Interpreting discriminative features: a case study

RG_JJ_NN1 (—): 18th most discriminative (negative) feature

- Degree adverb followed by an adjective and a singular noun (e.g., very good boy)
Interpreting discriminative features: a case study

**RG_JJ_NN1 (—): 18th most discriminative (negative) feature**

- Degree adverb followed by an adjective and a singular noun (e.g., very good boy)
Interpreting discriminative features: a case study

**RG_JJ_NN1 (−): 18th most discriminative (negative) feature**

- Degree adverb followed by an adjective and a singular noun (e.g., *very good boy*)
- **Why negative?**
Interpreting discriminative features: a case study

**RG_JJ_NN1 (−): 18th most discriminative (negative) feature**

- Degree adverb followed by an adjective and a singular noun (e.g., *very good boy*)
- Why negative?
- Related to:
Interpreting discriminative features: a case study

RG_JJ_NN1 (−): 18th most discriminative (negative) feature

- Degree adverb followed by an adjective and a singular noun (e.g., very good boy)
- Why negative?
- Related to:
  - very_good (−)
Interpreting discriminative features: a case study

RG_JJ_NN1 (−): 18th most discriminative (negative) feature

- Degree adverb followed by an adjective and a singular noun (e.g., very good boy)
- Why negative?
- Related to:
  - very_good (−)
  - JJ_NN1_NN (−) (e.g., difficult sport at)

Examples

1a It might seem to be very difficult sport at the beginning.
1b We know a lot about very difficult situation in your country.
Interpreting discriminative features: a case study

**RG_JJ>NN1 (−):** 18th most discriminative (negative) feature

- Degree adverb followed by an adjective and a singular noun (e.g., very good boy)
- Why negative?
- Related to:
  - very good (−)
  - JJ>NN1>II (−) (e.g., difficult sport at)
  - VBZ_RG (−) (e.g., is very)

Examples

1a *It might seem to be very difficult sport at the beginning.*
1b *We know a lot about very difficult situation in your country.*
1c *I think it’s very good idea to spending vacation together.*
1d *Unix is very powerful system but there is one thing against it.*
Interpreting discriminative features: a case study

**RG_JJ_NN1** – related to article omission errors?

- 23% of sentences that contain RG_JJ_NN1 also have a MD error
Interpreting discriminative features: a case study

**RG_JJ_NN1** – related to article omission errors?

- 23% of sentences that contain **RG_JJ_NN1** also have a MD error
- **very_good** (12%), **JJ_NN1_II** (14%), **VBZ_RG** (15%)
Interpreting discriminative features: a case study

**RG_JJ_NN1 – related to article omission errors?**

- 23% of sentences that contain RG_JJ_NN1 also have a MD error
- *very_good* (12%), *JJ_NN1_II* (14%), *VBZ_RG* (15%)
- **MD:doc**
Interpreting discriminative features: a case study

RG_JJ_NN1 – related to article omission errors?

- 23% of sentences that contain RG_JJ_NN1 also have a MD error
- very_good (12%), JJ_NN1_Il (14%), VBZ_RG (15%)
- MD:doc
  - Across all scripts: 2.18
Interpreting discriminative features: a case study

**RG JJ NN1** – related to article omission errors?

- 23% of sentences that contain RG JJ NN1 also have a MD error
- very good (12%), JJ NN1 II (14%), VBZ RG (15%)
- MD:doc
  - Across all scripts: 2.18
  - **RG JJ NN1**: 2.75
Interpreting discriminative features: a case study

**RG_JJ_NN1 – related to article omission errors?**

- 23% of sentences that contain RG_JJ_NN1 also have a MD error
- `very_good` (12%), `JJ_NN1_Il` (14%), `VBZ_RG` (15%)
- MD:doc
  - Across all scripts: 2.18
  - RG_JJ_NN1: 2.75
  - `VBZ_RG`: 2.68
  - `JJ_NN1_Il`: 2.48
  - `very_good`: 2.32
  - `VBZ_RG_JJ`: 2.73
  - `VBZ_RG_JJ & RG_JJ_NN1`: 3.68
Interpreting discriminative features: a case study

**RG_JJ_NN1** – related to article omission errors?

- 23% of sentences that contain RG_JJ_NN1 also have a MD error
- very_good (12%), JJ_NN1_II (14%), VBZ_RG (15%)
- MD:doc
  - Across all scripts: 2.18
  - RG_JJ_NN1: 2.75
  - VBZ_RG: 2.68
  - JJ_NN1_II: 2.48
  - very_good: 2.32
  - VBZ_RG_JJ: 2.73
  - VBZ_RG_JJ & RG_JJ_NN1: 3.68

- The error mostly involves the indefinite article
Interpreting discriminative features: a case study

**RG_JJ_NN1 – related to article omission errors?**

- 23% of sentences that contain RG_JJ_NN1 also have a MD error
- `very_good` (12%), `JJ_NN1_Ii` (14%), `VBZ_RG` (15%)
- MD:doc
  - Across all scripts: 2.18
  - RG_JJ_NN1: 2.75
  - VBZ_RG: 2.68
  - JJ_NN1_Ii: 2.48
  - `very_good`: 2.32
  - VBZ_RG_JJ: 2.73
  - VBZ_RG_JJ & RG_JJ_NN1: 3.68

- The error mostly involves the indefinite article
- **Why these richer nominals should associate with article omission?**
Interpreting discriminative features: a case study

<table>
<thead>
<tr>
<th>RG_JJ_NN1 – related to article omission errors?</th>
</tr>
</thead>
<tbody>
<tr>
<td>- 23% of sentences that contain RG_JJ_NN1 also have a MD error</td>
</tr>
<tr>
<td>- very_good (12%), JJ_NN1_Ii (14%), VBZ_RG (15%)</td>
</tr>
<tr>
<td>- MD:doc</td>
</tr>
<tr>
<td>- Across all scripts: 2.18</td>
</tr>
<tr>
<td>- RG_JJ_NN1: 2.75</td>
</tr>
<tr>
<td>- VBZ_RG: 2.68</td>
</tr>
<tr>
<td>- JJ_NN1_Ii: 2.48</td>
</tr>
<tr>
<td>- very_good: 2.32</td>
</tr>
<tr>
<td>- VBZ_RG_JJ: 2.73</td>
</tr>
<tr>
<td>- VBZ_RG_JJ &amp; RG_JJ_NN1: 3.68</td>
</tr>
<tr>
<td>- The error mostly involves the indefinite article</td>
</tr>
<tr>
<td>- Why these richer nominals should associate with article omission?</td>
</tr>
<tr>
<td>- <strong>Why only singular nouns are implicated in this feature?</strong></td>
</tr>
</tbody>
</table>

Helen Yannakoudakis, Ted Briscoe, Theodora Alexopoulou

Automating Second Language Acquisition Research
Why these richer nominals should associate with article omission?

- Typical of learners coming from L1s lacking an article system (Robertson, 2000; Ionin and Montrul, 2010; Hawkins and Buttery, 2010)

Helen Yannakoudakis, Ted Briscoe, Theodora Alexopoulou

Automating Second Language Acquisition Research
Why these richer nominals should associate with article omission?

- Typical of learners coming from L1s lacking an article system (Robertson, 2000; Ionin and Montrul, 2010; Hawkins and Buttery, 2010)

- Learners analyse articles as adjectival modifiers rather than as a separate category of determiners or articles (Trenkic, 2008)
Interpreting discriminative features: a case study

Why these richer nominals should associate with article omission?

- Typical of learners coming from L1s lacking an article system (Robertson, 2000; Ionin and Montrul, 2010; Hawkins and Buttery, 2010)

- Learners analyse articles as adjectival modifiers rather than as a separate category of determiners or articles (Trenkic, 2008)

- **Hypothesis**: with complex adjectival phrases, learners may omit the article
Interpreting discriminative features: a case study

Why these richer nominals should associate with article omission?

- Typical of learners coming from L1s lacking an article system (Robertson, 2000; Ionin and Montrul, 2010; Hawkins and Buttery, 2010)
- Learners analyse articles as adjectival modifiers rather than as a separate category of determiners or articles (Trenkic, 2008)
- Hypothesis: with complex adjectival phrases, learners may omit the article
  - Is article omission more pronounced with more complex adjectival phrases?
Interpreting discriminative features: a case study

Why these richer nominals should associate with article omission?

- Typical of learners coming from L1s lacking an article system (Robertson, 2000; Ionin and Montrul, 2010; Hawkins and Buttery, 2010)
- Learners analyse articles as adjectival modifiers rather than as a separate category of determiners or articles (Trenkic, 2008)
- Hypothesis: with complex adjectival phrases, learners may omit the article
  - Is article omission more pronounced with more complex adjectival phrases?
  - Is this primarily the case for learners from L1s lacking articles?
Interpreting discriminative features: a case study

Is article omission more pronounced with more complex adjectival phrases?

- MD:doc
  - Across all scripts: 2.18
  - RG_JJ_NN1: 2.75
  - VBZ_RG: 2.68
  - JJ_NN1_II: 2.48
  - very_good: 2.32
  - VBZ_RG_JJ: 2.73
  - VBZ_RG_JJ & RG_JJ_NN1: 3.68
  - JJ_NN1: 2.20
Interpreting discriminative features

Is this primarily the case for learners from L1s lacking articles?

<table>
<thead>
<tr>
<th>Language</th>
<th>sentences%</th>
<th></th>
<th></th>
<th>MD:doc</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RG_JJ_NN1</td>
<td>VBZ_RG JJ</td>
<td>RG_JJ_NN1</td>
<td>VBZ_RG JJ</td>
</tr>
<tr>
<td>all</td>
<td>23.0</td>
<td>15.6</td>
<td>2.75</td>
<td>2.73</td>
</tr>
<tr>
<td>Turkish</td>
<td>45.2</td>
<td>29.0</td>
<td>5.81</td>
<td>5.82</td>
</tr>
<tr>
<td>Japanese</td>
<td>44.4</td>
<td>22.3</td>
<td>4.48</td>
<td>3.98</td>
</tr>
<tr>
<td>Korean</td>
<td>46.7</td>
<td>35.0</td>
<td>5.48</td>
<td>5.31</td>
</tr>
<tr>
<td>Russian</td>
<td>46.7</td>
<td>23.4</td>
<td>5.42</td>
<td>4.59</td>
</tr>
<tr>
<td>Chinese</td>
<td>23.4</td>
<td>13.5</td>
<td>3.58</td>
<td>3.25</td>
</tr>
<tr>
<td>French</td>
<td>6.9</td>
<td>6.7</td>
<td>1.32</td>
<td>1.49</td>
</tr>
<tr>
<td>German</td>
<td>2.1</td>
<td>3.0</td>
<td>0.91</td>
<td>0.92</td>
</tr>
<tr>
<td>Spanish</td>
<td>10.0</td>
<td>9.6</td>
<td>1.18</td>
<td>1.35</td>
</tr>
<tr>
<td>Greek</td>
<td>15.5</td>
<td>12.9</td>
<td>1.60</td>
<td>1.70</td>
</tr>
</tbody>
</table>

Table: sentences%: proportion of sentences containing $f_i$ that also contain a MD.
Why only singular nouns are implicated in this feature?

The association with predicative contexts may provide a clue. Such contexts select nominals which require the indefinite article only in the singular case.

Example: *Unix is (a) very powerful system* vs. *Macs are very elegant machines.*
Conclusions & Future Work

- Formed initial interpretations for why a particular feature is negatively discriminative
- Nominals with complex adjectival phrases appear particularly susceptible to article omission errors by learners of English with L1s lacking articles
- Usefulness of visualisation techniques for navigating and interpreting large amounts of data
- Relevance of features weighted by discriminative classifiers
Conclusions & Future Work

- Formed initial interpretations for why a particular feature is negatively discriminative.
- Nominals with complex adjectival phrases appear particularly susceptible to article omission errors by learners of English with L1s lacking articles.
- Usefulness of visualisation techniques for navigating and interpreting large amounts of data.
- Relevance of features weighted by discriminative classifiers.
- More rigorous evaluation techniques, such as longitudinal case studies (Shneiderman and Plaisant, 2006; Munzner, 2009).
- Investigation and evaluation of different visualisation techniques of machine learned or extracted features that support hypothesis formation about learner grammars.
Conclusions & Future Work

- Formed initial interpretations for why a particular feature is negatively discriminative
- Nominals with complex adjectival phrases appear particularly susceptible to article omission errors by learners of English with L1s lacking articles
- Usefulness of visualisation techniques for navigating and interpreting large amounts of data
- Relevance of features weighted by discriminative classifiers
- More rigorous evaluation techniques, such as longitudinal case studies (Shneiderman and Plaisant, 2006; Munzner, 2009)
- Investigation and evaluation of different visualisation techniques of machine learned or extracted features that support hypothesis formation about learner grammars

**Available upon request as a web service on FCE scripts**
Acknowledgments: we are grateful to Cambridge ESOL for supporting this research. The third author acknowledges support from Education First. We would like to thank Marek Rei, Øistein Andersen, Tim Parish, Paula Buttery, Angeliki Salamoura as well as the anonymous reviewers for their valuable comments and suggestions.