TELL @ NTU: Intelligent Tutoring and Error Detection Systems

Bringing Language Teaching and Learning to a Technological Enhanced Reality

by

Luis Morgado da Costa
Automated Writing Support for NTU Engineering Students

Interactive Error Diagnosis and Coaching in L2 Learning
Automated Writing Support for NTU Engineering Students
Origins and Goals

- collab. with Language and Communication Centre, NTU
- cohorts of around 2000 engineering students per semester
- pedagogical challenges: correction, feedback, timing
- provide timely, high quality feedback to students

Goals:
- strictly Error Detection (no correction!)
- encourage independence and critical thinking
- exploring and evaluating possible solutions
- deciding on best corrections on their own
A New Learner Corpus

- NTU Corpus of Learner English (NTUCLE)
- 180 human tagged documents
  - course assignments
  - 60 documents were double-tagged
  - \( \sim 9.50k \) words, \( \sim 120k \) words
- new tagging schema
  - based on other schemas (e.g. NUS, Cambridge)
  - 53 error tags divided in 15 categories
- 6 annotators, course lecturers
A Learner Corpus Annotation Tool

- extends IMI (Bond et al., 2015)

- integrated with all of our other semantic annotated layers (sense, sentiment, etc.)

- online, concurrent annotation, open source

- ‘full-featured’ to our needs:
  - tagged at the level of word tokens
  - contiguous and non-contiguous word spans
  - allowing comments and/or corrections
  - multiple tags, total/partial overlap
A Learner Corpus Annotation Tool

Screenshot of the IMI extension (annotation layer)
Relative error distribution, by error category (not individual tags)
### Annotation Results (Distribution II)

#### Automated Writing Support for NTU Engineering Students

<table>
<thead>
<tr>
<th>A#</th>
<th>Most Common Error</th>
<th>2nd Most Common Error</th>
<th>3rd Most Common Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>awkward expression (21%)</td>
<td>word choice (11%)</td>
<td>unclear expression (10%)</td>
</tr>
<tr>
<td>A2</td>
<td>singular/plural forms (22%)</td>
<td>word choice (7%)</td>
<td>missing article/det. (6%)</td>
</tr>
<tr>
<td>A3</td>
<td>singular/plural forms (12%)</td>
<td>missing article/det. (10%)</td>
<td>word choice (8%)</td>
</tr>
<tr>
<td>A4</td>
<td>missing article/det. (21%)</td>
<td>singular/plural forms (11%)</td>
<td>verb form (9%)</td>
</tr>
<tr>
<td>A5</td>
<td>unclear expression (12%)</td>
<td>awkward expression (11%)</td>
<td>word choice (7%)</td>
</tr>
<tr>
<td>A6</td>
<td>singular/plural forms (11%)</td>
<td>word choice (9%)</td>
<td>missing article/det. (9%)</td>
</tr>
<tr>
<td>Total</td>
<td>singular/plural forms (10%)</td>
<td>missing article/det. (8%)</td>
<td>word choice (8%)</td>
</tr>
</tbody>
</table>

Top errors by annotator (before harmonisation)
Annotation Results (Frequency)

Annotator 1: 1183 errors
Annotator 2: 1101 errors
Annotator 3: 972 errors
Annotator 4: 908 errors
Annotator 5: 556 errors
Annotator 6: 380 errors

Number of errors per annotator (similar samples)
The cost of non-standardisation (i.e., naturalistic tagging)

- error analysis (mismatches)
  - misunderstanding or misapplying a tag
  - multiple tags could apply
  - error span selection
    (e.g., selecting only heads vs. full phrases)

- individual differences
  - different sensibilities to error spotting
  - different correction techniques
  - strictness
Corpus Release

<table>
<thead>
<tr>
<th>DB</th>
<th>Docs.</th>
<th>Overlapped Docs.</th>
<th>Sents.</th>
<th>Words</th>
<th>Sents. w/Errors</th>
<th>Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>40</td>
<td>10 (A6) + 10 (A2)</td>
<td>2,051</td>
<td>26,176</td>
<td>812</td>
<td>1108</td>
</tr>
<tr>
<td>A2</td>
<td>40</td>
<td>10 (A1) + 10 (A3)</td>
<td>2,144</td>
<td>26,764</td>
<td>372</td>
<td>390</td>
</tr>
<tr>
<td>A3</td>
<td>40</td>
<td>10 (A2) + 10 (A4)</td>
<td>2,269</td>
<td>27,603</td>
<td>625</td>
<td>1193</td>
</tr>
<tr>
<td>A4</td>
<td>40</td>
<td>10 (A3) + 10 (A5)</td>
<td>2,223</td>
<td>27,246</td>
<td>361</td>
<td>575</td>
</tr>
<tr>
<td>A5</td>
<td>40</td>
<td>10 (A4) + 10 (A6)</td>
<td>2,093</td>
<td>26,654</td>
<td>579</td>
<td>908</td>
</tr>
<tr>
<td>A6</td>
<td>40</td>
<td>10 (A5) + 10 (A1)</td>
<td>2,024</td>
<td>26,103</td>
<td>564</td>
<td>972</td>
</tr>
<tr>
<td>Tagged</td>
<td>180</td>
<td>n.a.</td>
<td>9,571</td>
<td>119,727</td>
<td>2,751</td>
<td>4,860</td>
</tr>
<tr>
<td>Untagged</td>
<td>93</td>
<td>n.a.</td>
<td>5,174</td>
<td>64,462</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>All</td>
<td>273</td>
<td>n.a.</td>
<td>14,745</td>
<td>184,189</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

Corpus Statistics

- Still in process of data anonymization
- Everything released under a CC-0
- Expanding (i.e. +393 documents last semester)
  - with some automatic tagging (!?)
Welcome, user.

Dear Students
Welcome to the error detection system. Please pay attention to the following important instructions before you upload your assignment for checking by the system.

- Make sure the student who uploaded your assignment to Turnitin is the same student who uploads the assignment to the error detection system.
- Please sign in using your NTU user name as both user name and password (both in UPPER CASE).
- Only upload your assignment in docx – NO PDFs.
- Once you have uploaded your assignment, the system will take about 2-3 minutes to flag down sentences with possible errors. In some cases, the system may flag some items as possible errors only. In such cases, you may decide not to take any action.
- Errors highlighted in RED indicate major errors that almost certainly require changes, while YELLOW highlighted errors indicate possible errors which may require changes.
- You can use this system to upload your assignment as many times as you want, progressively addressing any errors flagged by the system.
- If you have any technical problems, please wait a couple hours and try again since too many concurrent uploads might slow down the system to the point of rejection. If the problem persists, email Mr. Kuribayashi at takayuki.k@ntu.edu.sg, including your NTU username and the docx file that you had problems with.

User Panel
Upload a new assignment

Online Error Detection System - Welcome Page
Welcome, user.

Upload Assignments

Welcome to the upload section. Before we proceed, we would like to remind you that upload only files with the extension '.docx'. These files can be produced from a variety of word processors including Microsoft Word, Google Documents, and Pages.

License

Before you upload your assignment, we would like to request your permission to use your assignment for research purposes. Systems like this are only possible to be improved if there is relevant data available. As NTU students, your privacy is very important to us, and any data you agree to release will be hand-checked and anonymised before being used for research. Anonymised data will be released under a Creative Commons CC0 Licence and used to improve this system. This choice will not affect your grade in any way, and lecturers won't know your choice before they grade your assignment.

Choose your file to upload:

- Choose File
- No file chosen

Online Error Detection System - Upload
Online Error Detection System - Loading/Checking
The Error Detection System (IV)

Online Error Detection System - Feedback

The Error Detection System (IV)

Automated Writing Support for NTU Engineering Students
Behind the Curtains

- online, Flask-based (Python), open source
- scalable, but resource hungry
- 59 checks in total
  - English Resource Grammar (w/ selected grammar checks)
  - extra checks (mostly on style and content)
  - different levels of confidence/severity (color marked)
- learner corpus compilation (automatic tagging)
- still under development (⚠️)

Automated Writing Support for NTU Engineering Students
Feedback Examples

- This sentence may have a verb which does not agree in person (e.g. 'I', 'you', 's/he') and number (singular/plural) with its subject: {{placeholder}}. Please check the sentence and ensure that the verb agrees with its subject.

- You may be using an indefinite article, ‘a’ or ‘an’, before an uncountable noun (such as ‘research’): {{placeholder}}. Indefinite articles should only precede singular countable nouns. Please check your sentence for uncountable nouns and remove any indefinite articles that precede them.

- This sentence may contain subjective or informal words or expressions: {{placeholder}}. You may want to replace these words and expressions with more formal and objective alternatives.

- This sentence is much longer than the average sentence. It may be difficult for readers to read the sentence and understand it after reading it once. There is also a higher risk of making grammar mistakes in such a long sentence. You may want to consider breaking up the sentence to make it easier for the reader to follow the text.

- You have used ‘there’ in this sentence. Please check if it should be ‘their’ instead and make the change if necessary.
Learning Experiment

- 1000 groups of students (paired assignment)
- one extra week after the original deadline
- free use of the tool (i.e. multiple uploads)
- (ongoing) comparision between submissions
  - over 90% participation
  - greatly asymmetric usage
    (i.e. no. of submissions, number of errors detected)
- generally positive qualitative feedback
  - some false positives
  - feedback not enough to correct the errors
  - general ‘something wrong’
Qualitative Results (I)

'I found the online error detection tool useful.' (n=236)
'I would like to use the online error detection tool for other courses and assignments.' (n=236)
Qualitative Feedback

- Additional feature: Reference checking.
- At times, there were errors identified which essentially said “There seems to be an error here but we are not sure what it exactly is”. Could improve on clarifying such errors. / Could be more detailed on the mistakes.
- Releasing it earlier and allowing access to it out of campus way better.
- Could be more specific in diagnosing the error of the assignment. / Feedbacks could be more specific / Could improve on the “could not detect error in sentence”.
- Provide better error messages.
- I think that it might give some students confusion as it does not show what to improve on the phrase/sentence.
- Some errors didn’t gave a clear explanation.
- It is basically not functioning in the sense to detect grammar errors. / Not very accurate.
- Just quite funny.
What Lies Ahead

- finish corpus anonymization and harmonization
- finish (quantitative) evaluation of the experiment
- expand error checks
- optimize feedback messages (i.e. useful to student learning, not to students’ contentment)
- enable individual sentence checks
- rinse / repeat / publish results (experimental design)
Practice is important in mastering a language:
- Problems have a difficulty optimum
- Risk of reinforcing errors

Precise, timely feedback is essential:
- Costly and time-consuming
- Individual feedback is practically impossible

Knowing a classroom is important:
- General trends must be addressed quickly
- Individual trends can be addressed on a case-by-case basis
And what about TELL?

**Education and Technology:**
Technology Enhanced Learning is an increasingly important part of higher education! (MOOCS, E-Learning Platforms, Blended-Learning Classrooms, etc.)

**What we want does not exist yet:**
- to know and help individual student’s weaknesses
- to drill student’s weaknesses with exercises
- to provide precise, informed feedback on how to improve
- to adapt itself to the student proficiency level
- to help assess students’ level of proficiency
- to scale to 100s or 1000s of students
Go Mal-Rules?

- extend prescriptive grammars
- model ungrammatical sentences
- identify specific language errors
- used for student feedback
Mal-Rule (Inflectional Rule)

* These students buyed many books.

```
S
  NP
    DP    N
      these students
  VP
    V   NP
    *buyed    DP    N
              many books
```
* These students buys many books.

```
*S

NP
  \-- VP
    \-- V
        \-- NP
            \-- N
                \-- DP
                    \-- these
                    \-- students
                \-- many
            \-- books
```
what is the correct form of ‘this students’?
‘this student’? ‘these students’?

*NP

DP      N
this    students
we can force the singular meaning

```plaintext
NP
  [NUM singular]
  /
  DP
  [NUM singular]  *N_{mal_{pl_noun_as_sg}}
  /
  this
  /
  N
  [NUM plural]
  /
  students
```
we can force the plural meaning
and then we have a problem...

```
NP
[ NUM singular ]

  DP
[ NUM singular ]
  [ NUM singular ]
  this

  *N_{mal\_pl\_noun\_as\_sg}

  N

  [ NUM plural ]
  this

students

NP
[ NUM plural ]

  *DP_{mal\_this\_as\_these}

  N

  [ NUM plural ]
  this

students
```
This students sleep.

*S_{head\_subj\_mal\_rule\_subj}

[meaning: This student sleeps.]

NP  

[ NUM singular ]  

VP

[ NUM plural ]

DP

[ NUM singular ]  

|  

* N_{mal\_pl\_noun\_as\_sg}  

|  

sleep

[ meaning: These students sleep. ]

NP

[ NUM plural ]  

VP

[ NUM plural ]

* DP_{mal\_this\_as\_these}  

N

[ NUM plural ]  

|  

this

|  

students

[ meaning: These students sleep. ]
Introducing LIT (working title)

Ungrammatical Ambiguity for Chinese speakers learning English

<table>
<thead>
<tr>
<th>Student:</th>
<th>That dog like the cat happy.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hmm... something is wrong with your sentence. Did you mean any of these?</td>
</tr>
<tr>
<td>A.</td>
<td>那只狗和猫一样高兴。</td>
</tr>
<tr>
<td></td>
<td>[That dog, like the cat, is happy.]*</td>
</tr>
<tr>
<td>B.</td>
<td>那只狗喜欢猫高兴。</td>
</tr>
<tr>
<td></td>
<td>[That dog likes the cat happy.]*</td>
</tr>
<tr>
<td>C.</td>
<td>那只狗喜欢高兴的猫。</td>
</tr>
<tr>
<td></td>
<td>[That dog likes the happy cat.]*</td>
</tr>
<tr>
<td>Student:</td>
<td>C. 那只狗喜欢高兴的猫。</td>
</tr>
<tr>
<td></td>
<td>Ok! Then I believe you forgot to conjugate the verb ‘to like’.</td>
</tr>
<tr>
<td></td>
<td>Also, remember that an adjective must come before the noun it’s modifying.</td>
</tr>
<tr>
<td></td>
<td>Please try again!</td>
</tr>
</tbody>
</table>

* The English translation is what the system thinks is correct, but it is not shown.
The System

- LIT is a Language-Aware Intelligent Tutor

  - goals:
    - provide drill and feedback
    - randomized examples with controlled vocabulary
    - gradual increase in difficulty
    - adapt to students based on their responses

  - semantic-based (i.e. MRS + wordnets) translation / disambiguation to pinpoint errors

- targeting university level L2 learners

- multilingual (Chinese, Japanese and English)
If there is mal-rule activation

- no ambiguity about which error was made – feedback can be generated automatically
- ambiguity in the student’s intended meaning (i.e. multiple semantic reconstructions) — *semantic disambiguation is required*

- basic dialog system powered by a MT system using the semantics reconstructed by mal-rules
- can use parse ranking algorithms to help select the most probable set of intended meanings
- provides feedback based solely on the student’s intended meaning - *never guessing*
The Basic Bits

- **Language Curricula**: what students need to know

- **Graded Parsers**: syntactic knowledge bases using HPGS grammars
  - step-by-step grammar introduction

- **Graded Wordnets**: graded lexical knowledge bases
  - step-by-step lexicon introduction

- **Learner Corpus**: what are the common errors
  - the source of mal-rule design

- **Semantic MT Environment**: high-quality, closed coverage
The Fancy Bits

Rich Student Models:
- knows each and every student (individually & by class)
- their strengths and weaknesses
- their progress through the syllabus
- topicalize examples based on student interests

automatic collection of Learner Corpora
- tree-banked ungrammatical input (useful for mal-rule ranking)
- corrected sentence
- sense tagged

Gamified Interaction
- fun language games
- competitive social games (e.g. within the classroom, against other classrooms)
Interactive Error Diagnosis and Coaching in L2 Learning

- **Syllabi Corpus** with dialogues, texts, examples, drills and vocabulary from text-books (≈ 800 sentences)
  - sense annotated (COW, ≈ 1000 senses, + new senses)
  - treebanked with Zhong (Chinese HPSG)

- **Learner Corpus** from past-exams (≈ 5,000 sentences)
  - annotated for error types (≈ 1,600 errors)
  - 21 error-types based on teaching experience

- Started the theoretical **design of the mal-rules**
  - some already implemented in Zhong
Most is yet to come! (PhD)

- scale-up: higher proficiency levels and other languages (English, Japanese, ...?)
- research on the impact of different kinds feedback
- research on models of error prediction
- research on student modelling

**Evaluation:**

- **intrinsic:** ability to diagnose and correct naturally occurring grammatical mistakes by learners (i.e. corpus study)
- **extrinsic:** controlled experiments for blended and/or lab learning environments
language is always a complex matter, even ‘ungrammatical language’ is ambiguous!

TELL is not simple but it is possible! (and hopefully worth it…)

hand-made rich lexical (e.g. wordnets) and syntactic (e.g. computational grammars) resources can shine in tasks where precision should be the primary measure

we hope to empower students and teachers, taking a step forward in TELL

if you have comments or ideas, I will be working on this for the next 3 years!
Acknowledgments

- Thanks to the LCC Team and Takayuki Kuribayashi.
- Thanks to Dan Flickinger (CSLI, Stanford) and the Deep Linguistic Processing with HPSG Initiative (DELPH-IN).
- Thanks to the MOE for their support through TRF, and NTU through EdEx. This project builds on earlier work on multilingual understanding: That’s what you meant: A Rich Representation for Manipulating Meaning (Tier 2) and Shifted in Translation — An Empirical Study of Meaning Change across Languages (Tier 1).
- And, of course, Francis Bond for entertaining my lunacy!