TELL @ NTU:
Intelligent Tutoring and Error Detection Systems

Bringing Language Teaching and Learning to a Technological Enhanced Reality

by

Luis Morgado da Costa

Projects

Automated Writing Support for NTU Engineering Students

Winder et al. (2017). NTUCLE: Developing a corpus of learner English to provide writing support for engineering students. Proceedings of the 4th Workshop on NLP Techniques for Educational Applications (NLPTEA 2017). Taipei, Taiwan.

Interactive Error Diagnosis and Coaching in L2 Learning

Morgado da Costa et al. (2016) Syntactic Well-Formedness Diagnosis and Error-Based Coaching in Computer Assisted Language Learning using Machine Translation. Proceedings of the 3rd Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA2016). Osaka, Japan.

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 chalenges: 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

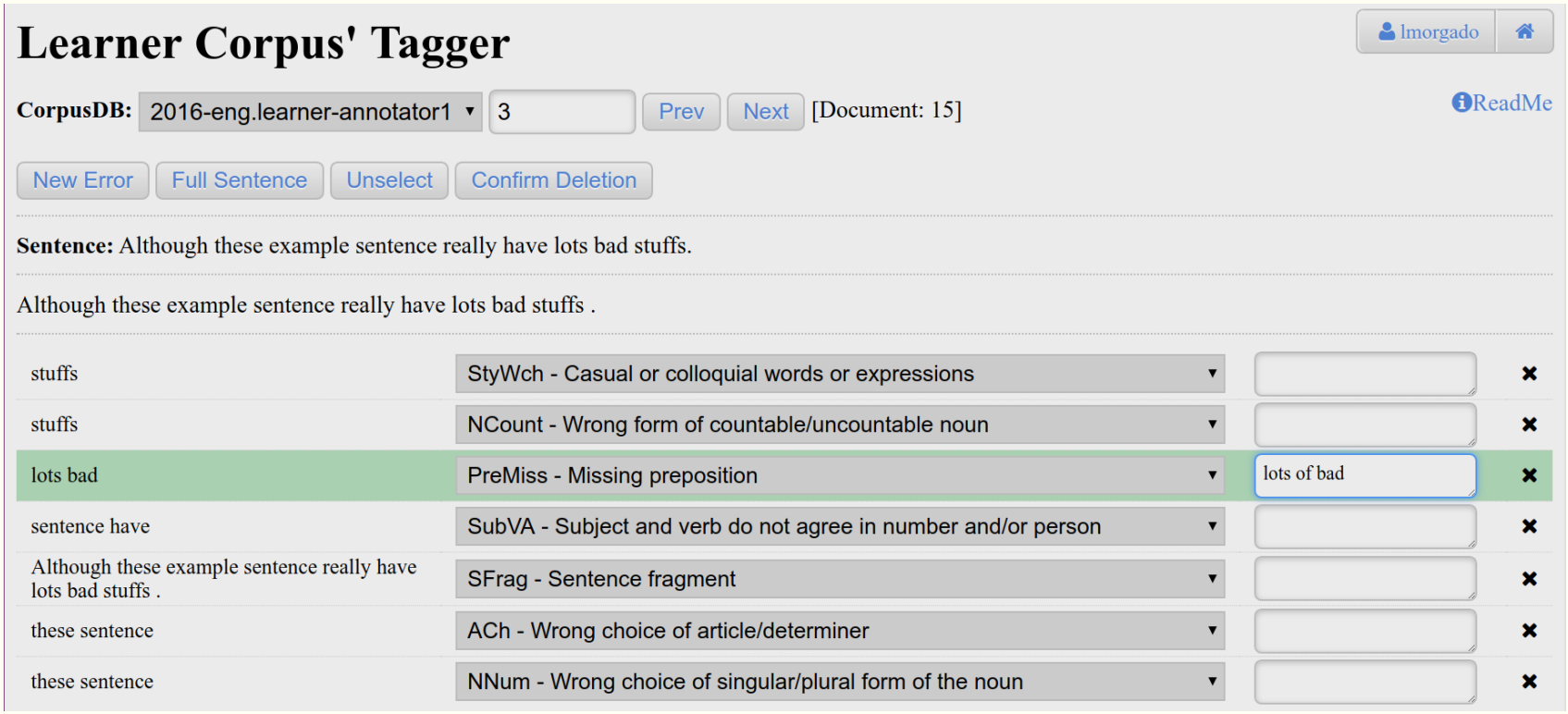
A New Learner Corpus

- NTU Corpus of Learner English (NTUCLE)
- 180 human tagged documents
  - course assignments
  - 60 documents were double-tagged
  - ∼9.50k words, ∼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

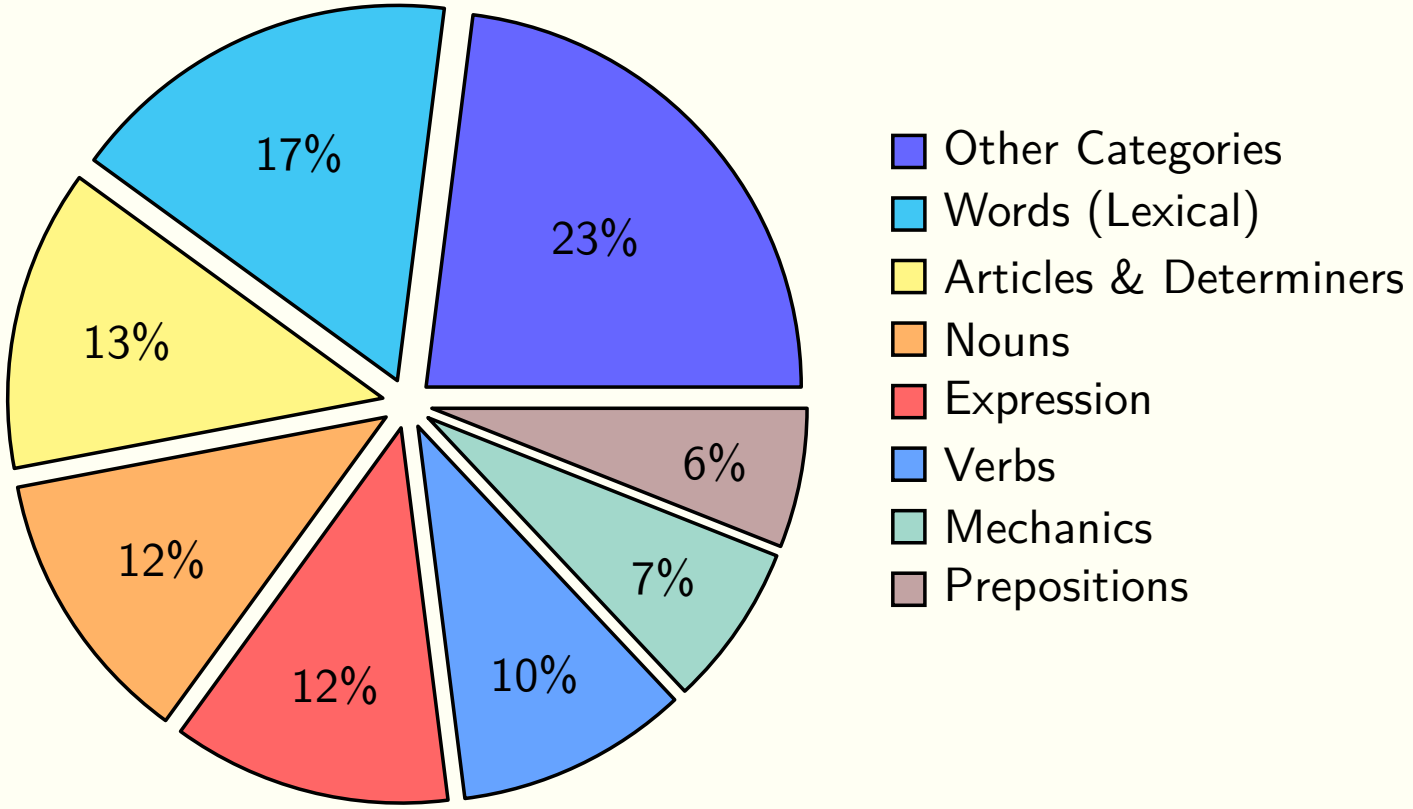
- extends IMI (Bond et al., 2015)
- integrated with all of our other semantic annotated layers (sense, sentiment, etc.)
- online, concurrent annotation, open source
- 'full-featured' to our needs:
  - tagged at the level of word tokens
  - contiguous and non-contiguous word spans
  - allowing comments and/or corrections
  - multiple tags, total/partial overlap

A Learner Corpus Annotation Tool



Screenshot of the IMI extension (annnotation layer)

Annotation Results (Distribution)



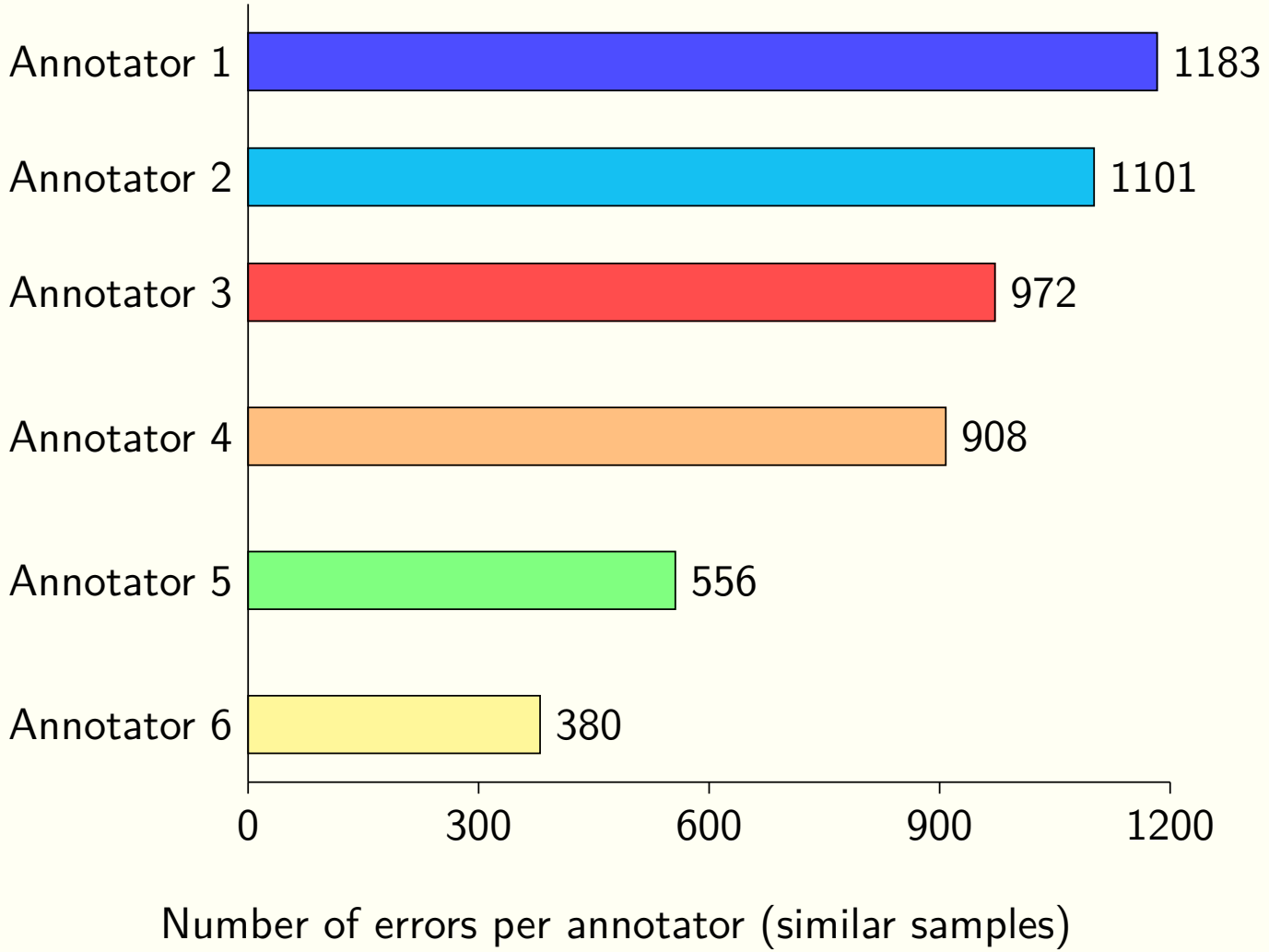
Relative error distribution, by error category (not indivual tags)

Annotation Results (Distribution II)

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

Top errors by annotator (before harmonisation)

Annotation Results (Frequency)



Annotation Results (Discussion)

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

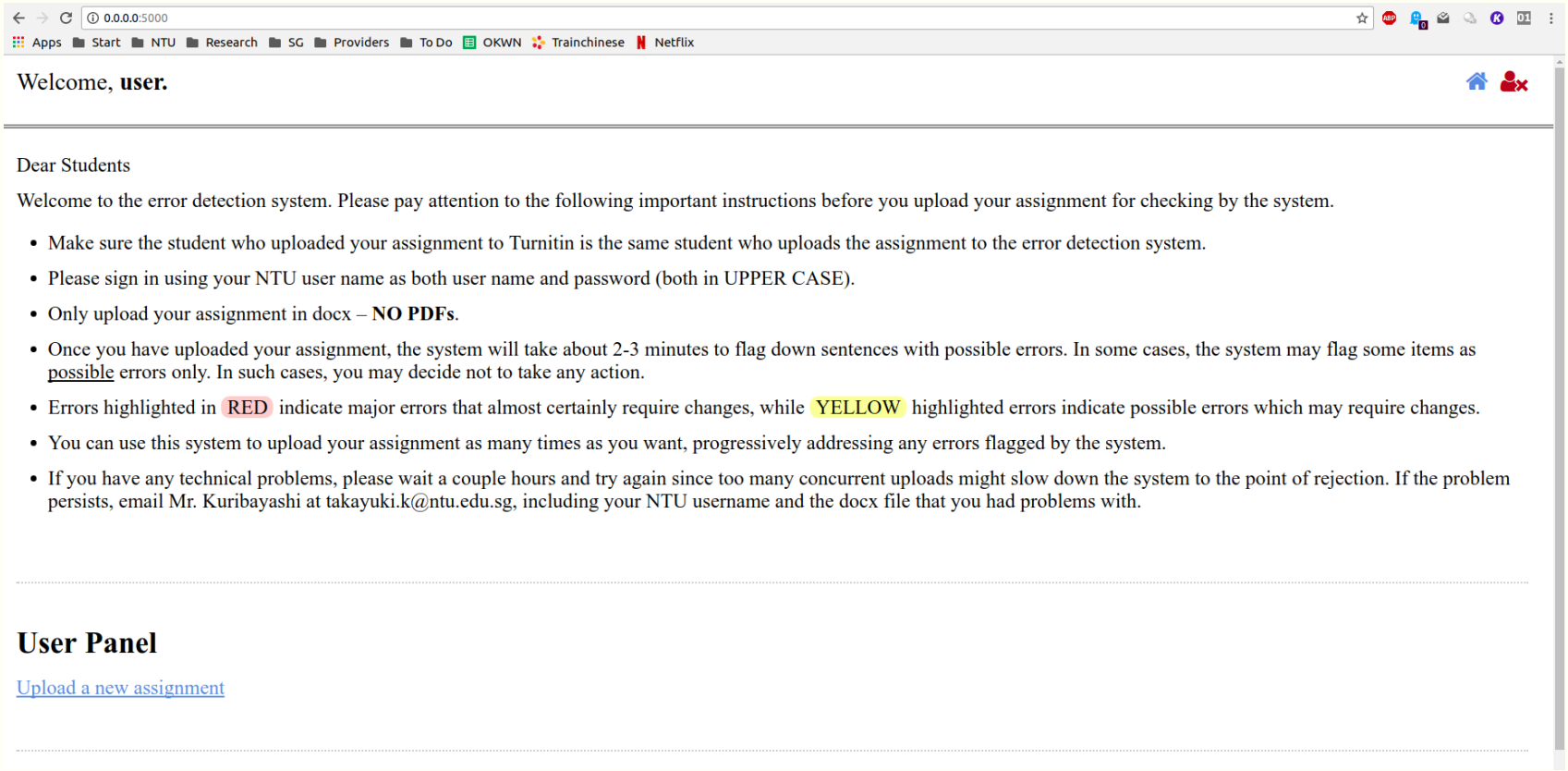
Corpus Release

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

Corpus Statistics

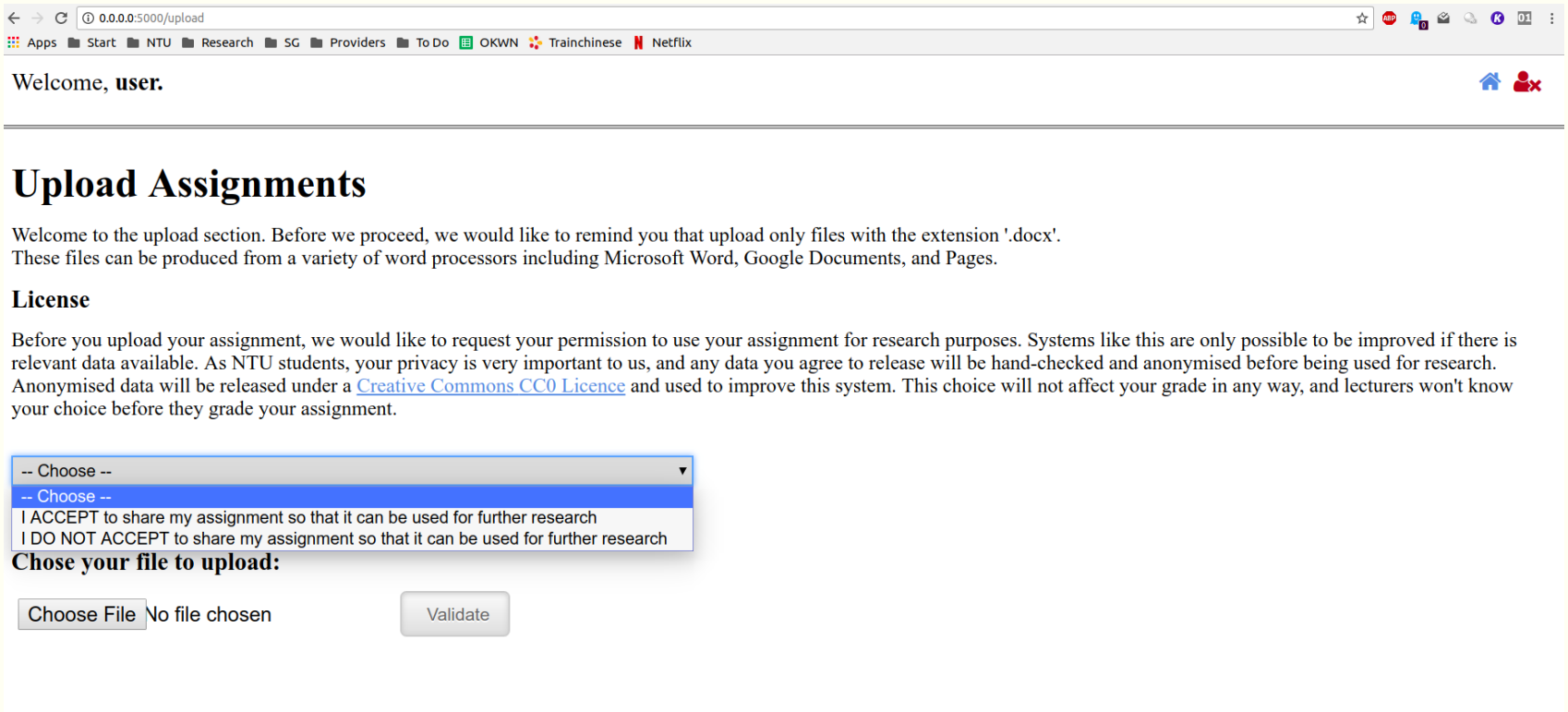
- Still in process of data anonymization
- Everything released under a CC-0
- Expanding (i.e. +393 documents last semester)
  - with some automatic tagging (!?)

The Error Detection System (I)



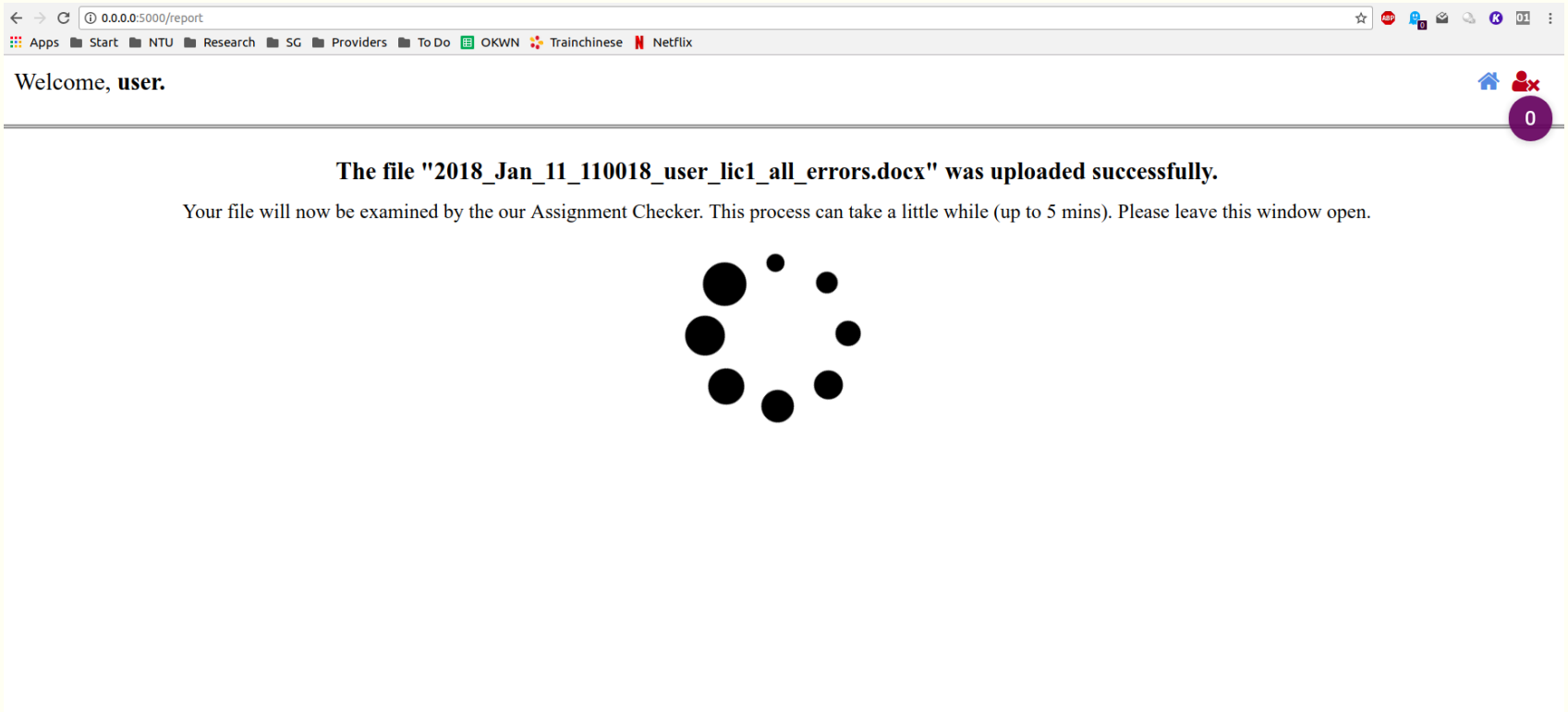
Online Error Detection System - Welcome Page

The Error Detection System (II)



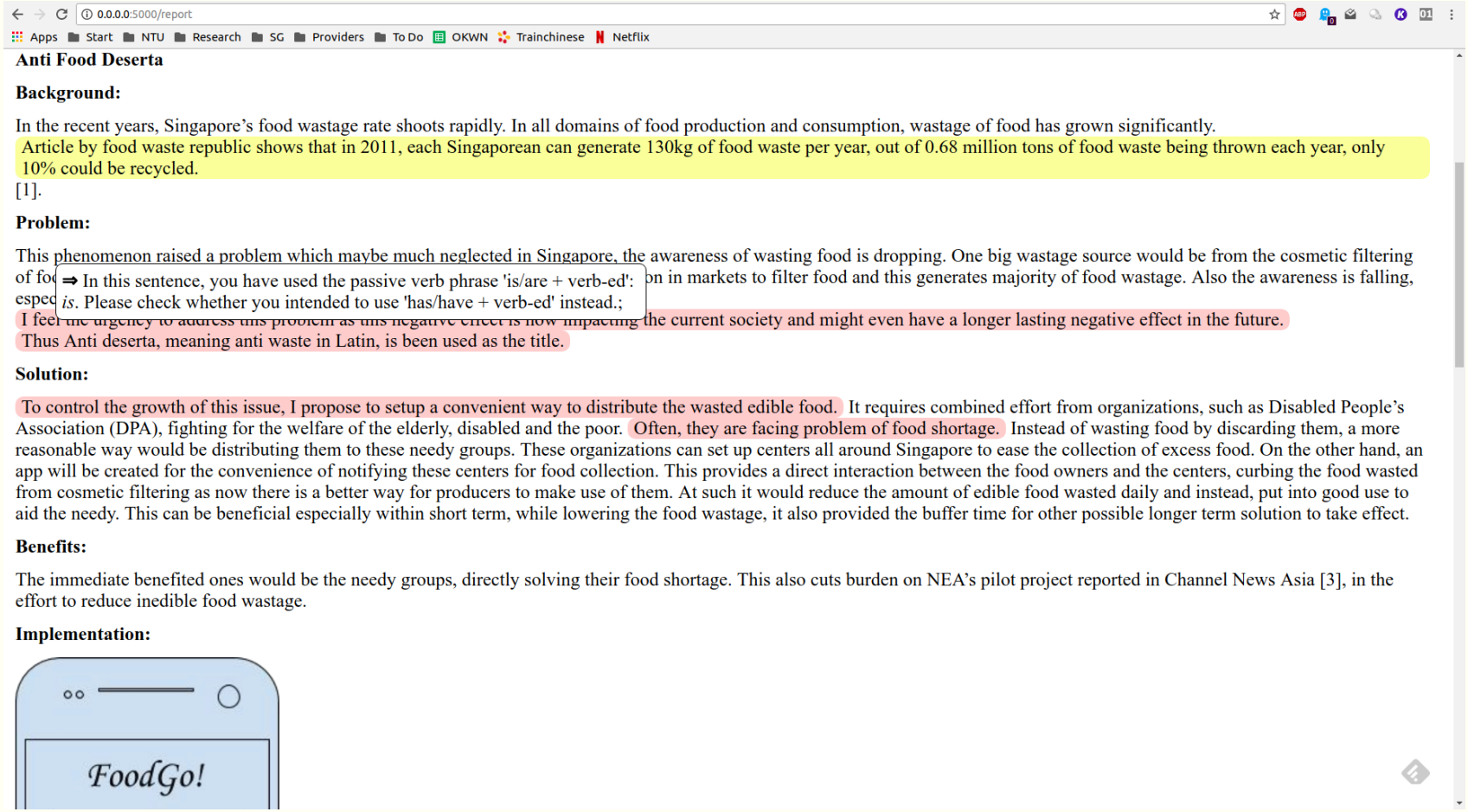
Online Error Detection System - Upload

The Error Detection System (III)



Online Error Detection System - Loading/Checking

The Error Detection System (IV)



Online Error Detection System - Feedback

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 (⚠)

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.*

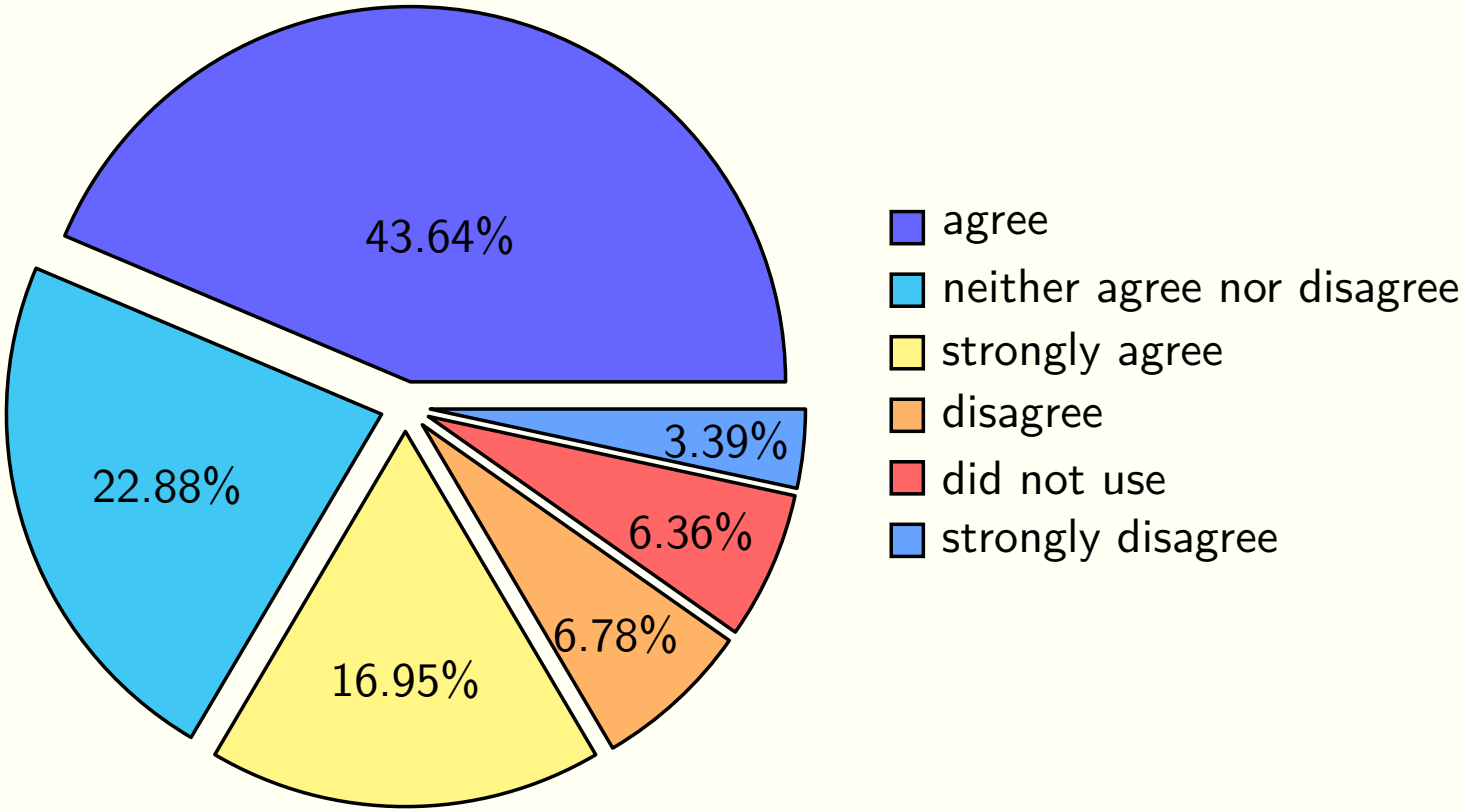
- *This sentence is much **longer than the average sentence**. It may be difficult for readers to read the sentence and understand it after reading it once. There is also a higher risk of making grammar mistakes in such a long sentence. You may want to consider breaking up the sentence to make it easier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

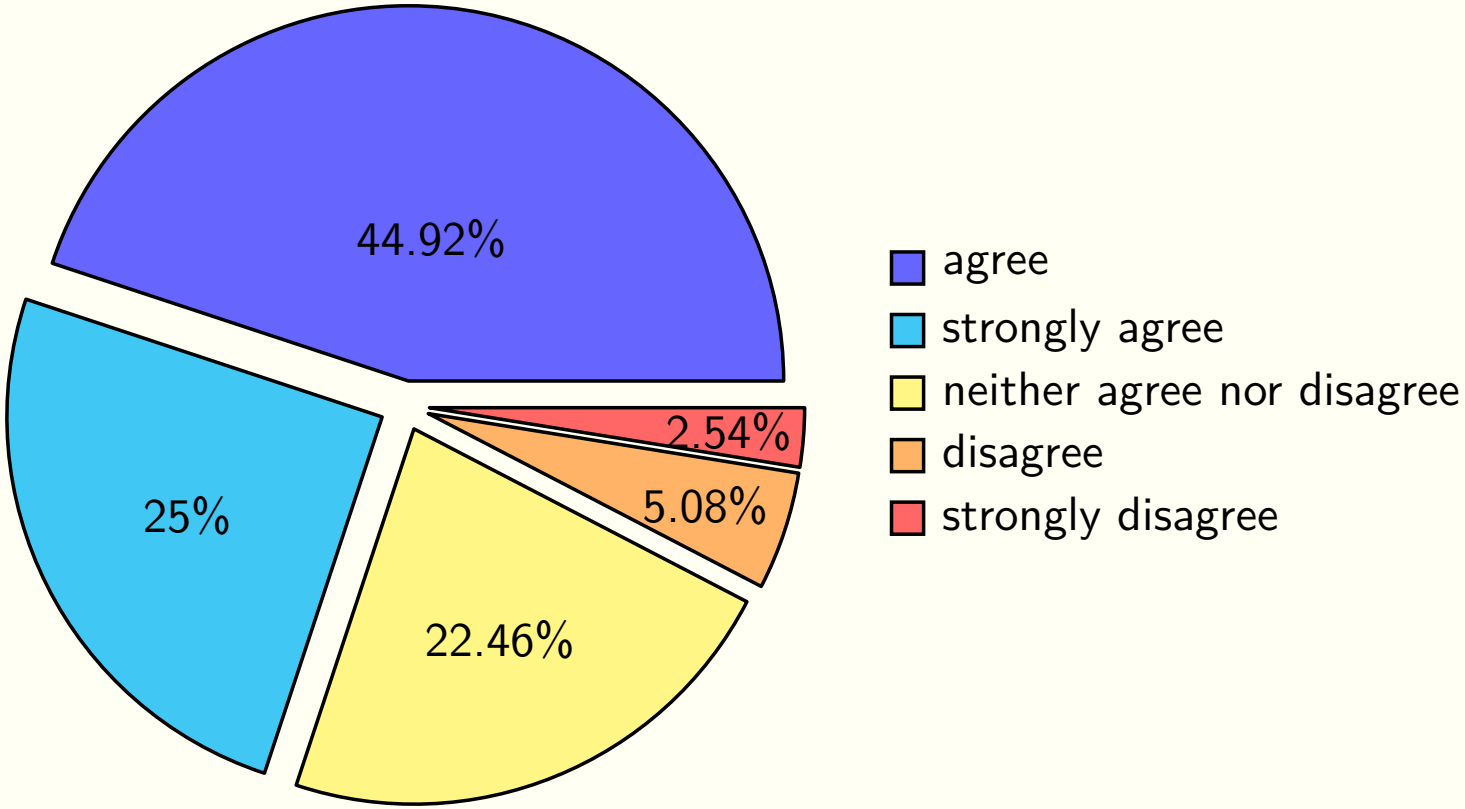
- 1000 groups of students (paired assignment)
- one extra week after the original deadline
- free use of the tool (i.e. multiple uploads)
- (ongoing) comparision between submissions
  - over 90% participation
  - greatly asymmetric usage
    (i.e. no. of submissions, number of errors detected)
- generally 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)

Qualitative Results (II)



'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)

Interactive Error Diagnosis and Coaching in L2 Learning

Background

- 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:

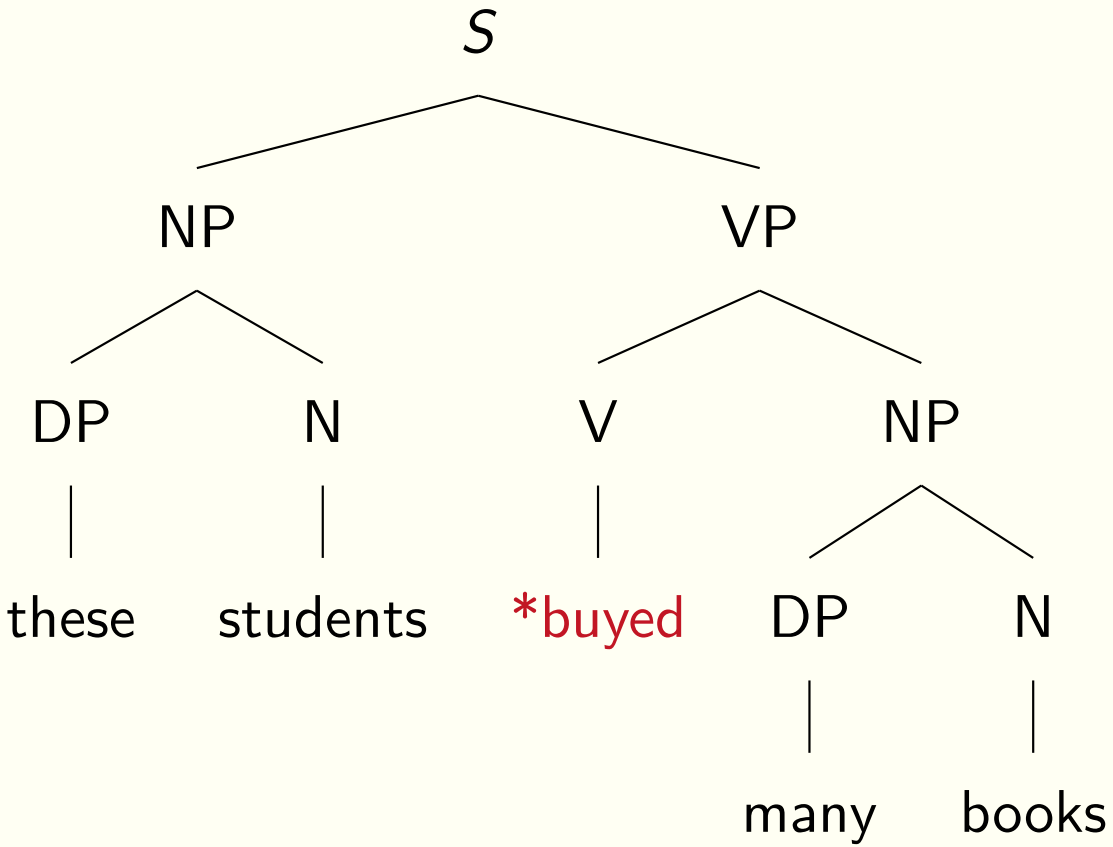
- to know and help individual student's weaknesses
- to drill student's weaknesses with exercises
- to provide precise, informed feedback on how to improve
- to adapt itselt to the student proficiency level
- to help assess students' level of proficiency
- to scale to 100s or 1000s of students

Go Mal-Rules?

- extend prescriptive grammars
- model ungrammatical sentences
- identify specific language errors
- used for student feedback

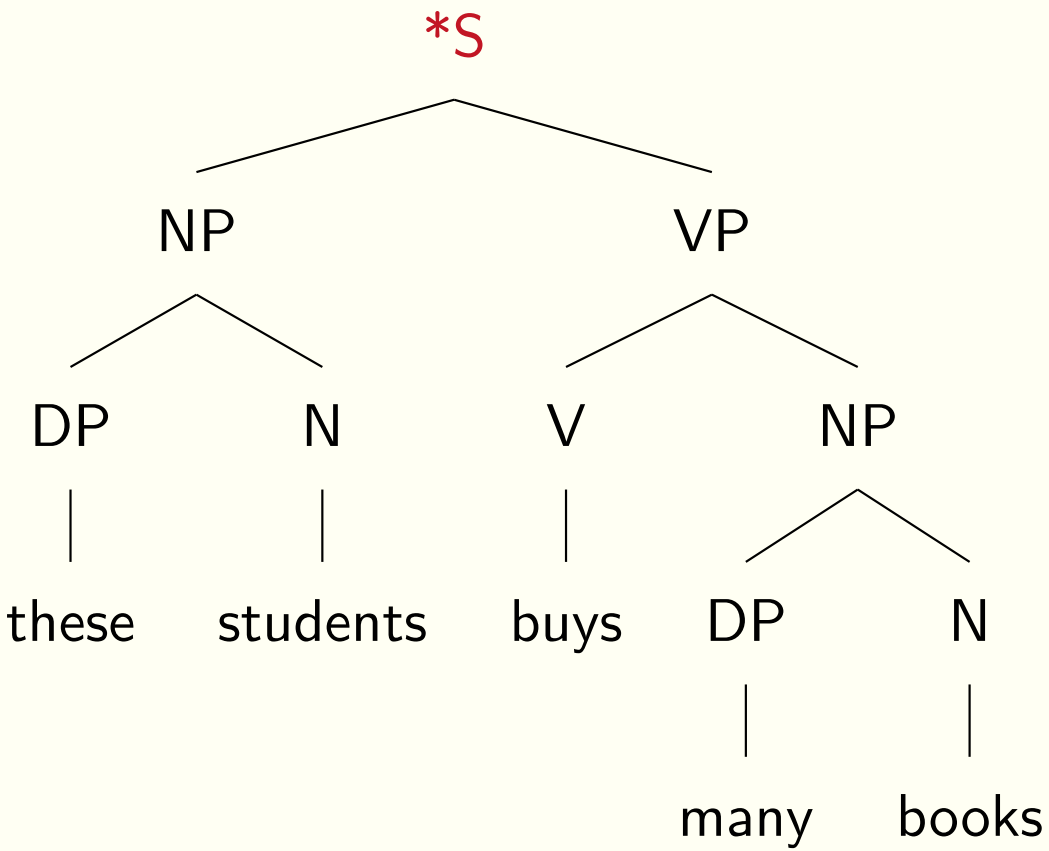
Mal-Rule (Inflectional Rule)

* These students buyed many books.



Mal-Rule (Subj-Verb Agreement)

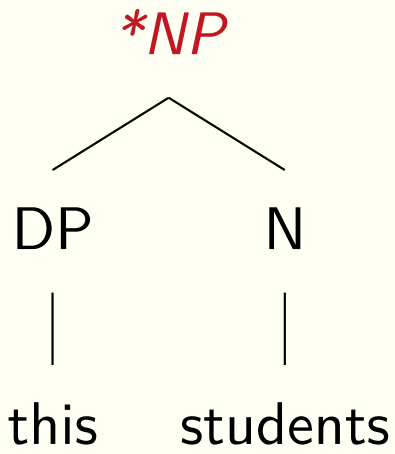
* These students buys many books.



Ungrammatical Ambiguity

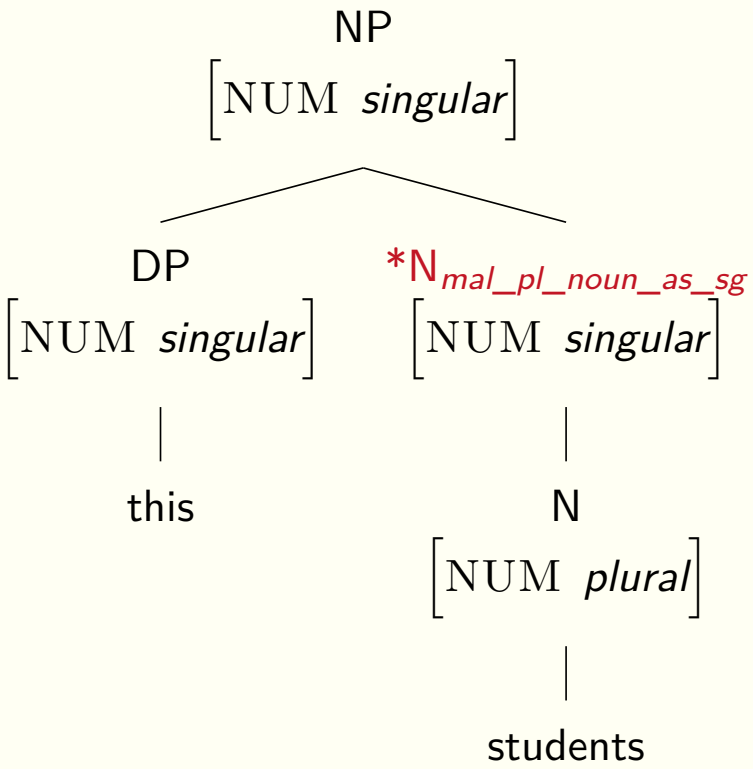
- what is the correct form of ‘this students’?

  ‘this student’? ‘these students’?



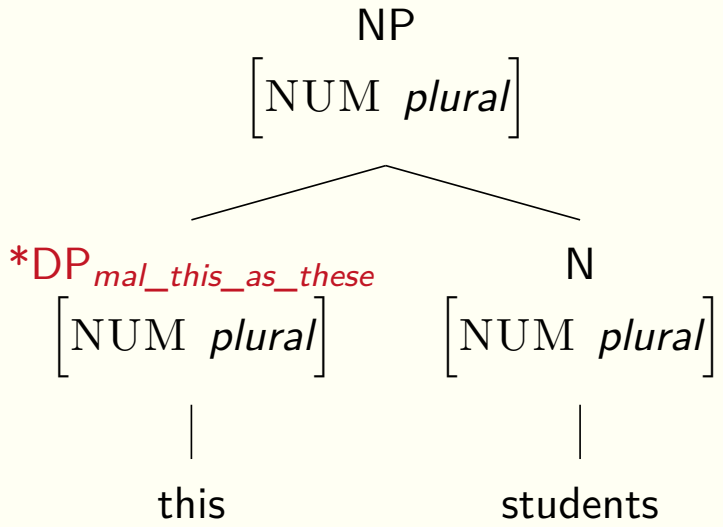
Mal-Lexical-Rule

- we can force the singular meaning



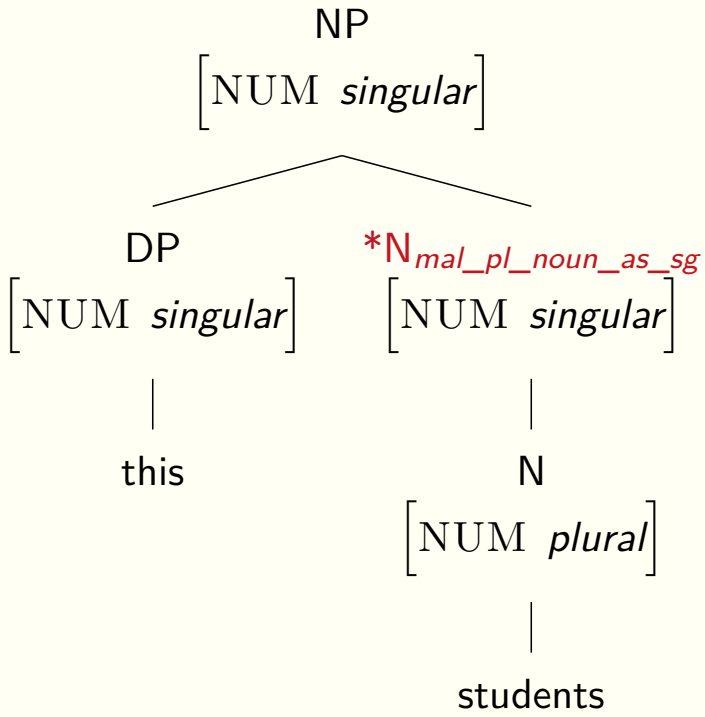
Mal Lexical Entry

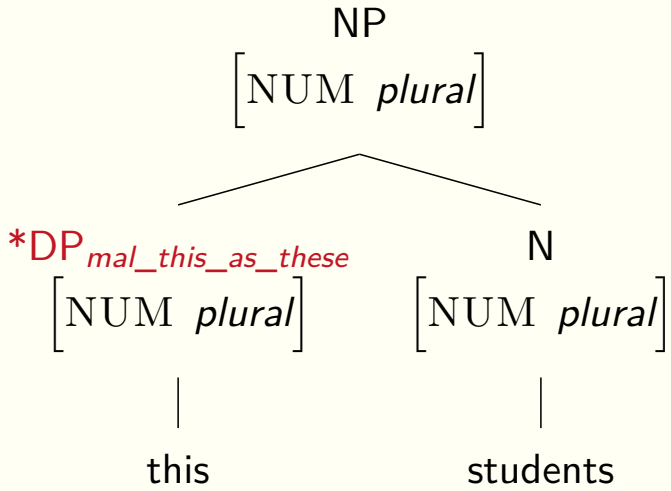
- we can force the plural meaning



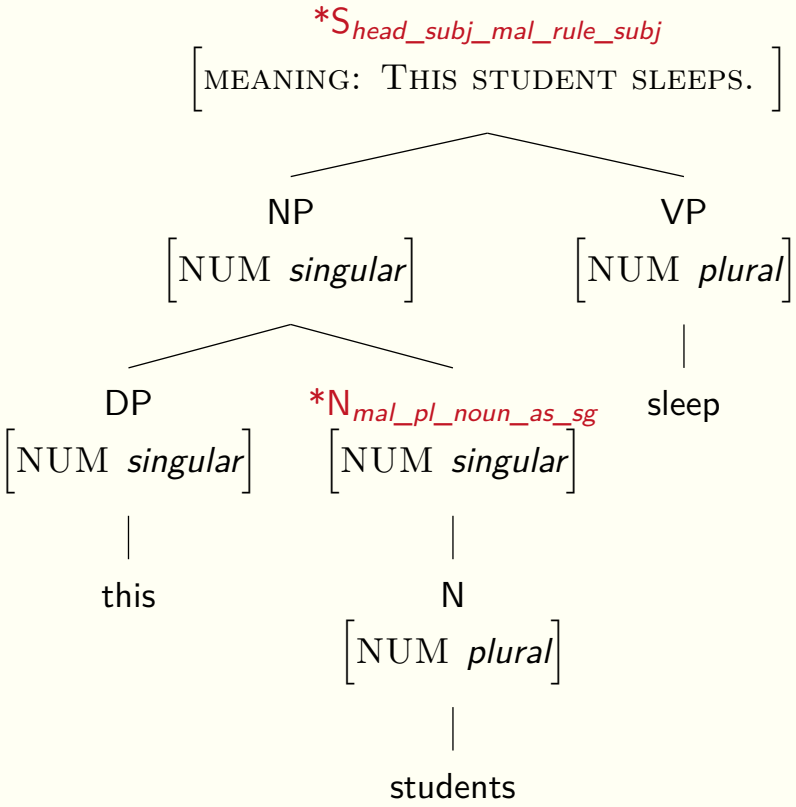
Combining Approaches

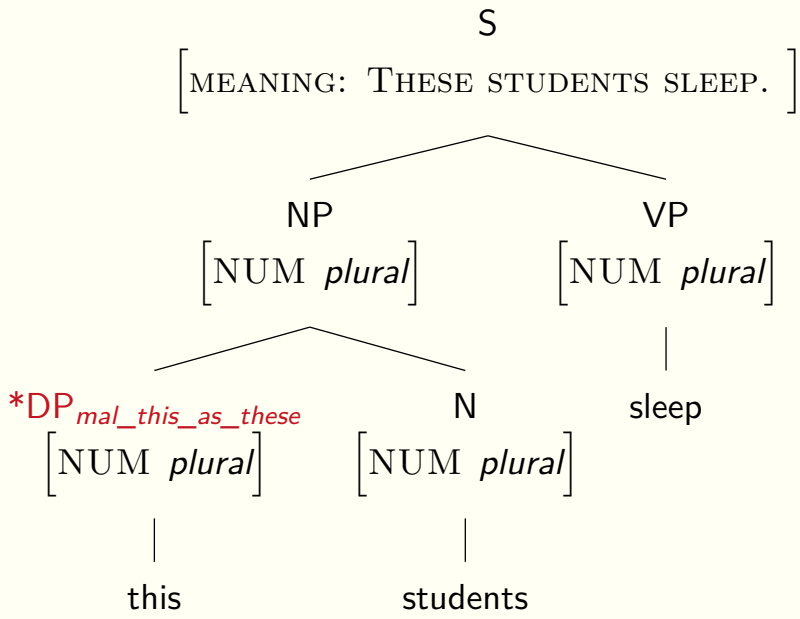
- and then we have a problem...





This students sleep.





Introducing LIT (working title)

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

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

Ungrammatical Ambiguity for Chinese speakers learning English

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)

Grammar Feedback Through Translation

- 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)

Preliminary Results (Mandarin)

- **Syllabi Corpus** with dialogues, texts, examples, drills and vocabulary from text-books ($\approx$ 800 sentences)
  - sense annotated (COW, $\approx$ 1000 senses, + new senses)
  - treebanked with Zhong (Chinese HPSG)
- **Learner Corpus** from past-exams ($\approx$ 5,000 sentences)
  - annotated for error types ($\approx$ 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

Conclusions

Closing Thoughts

- 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!