

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 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
 - ❖ ~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

Learner Corpus' Tagger Imorgado

CorpusDB: 2016-eng_learner-annotator1 3 Prev Next [Document: 15] ReadMe

New Error Full Sentence Unselect Confirm Deletion

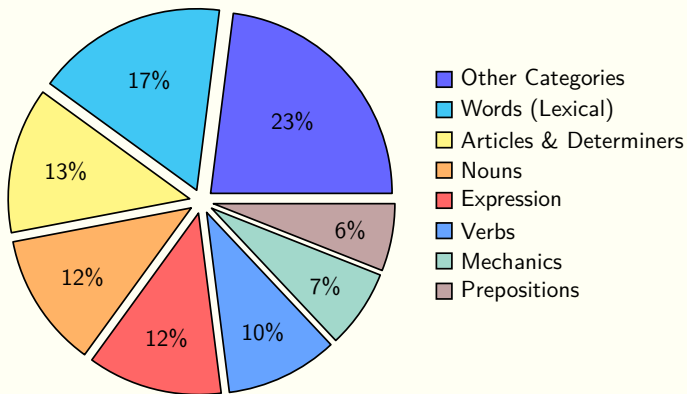
Sentence: Although these example sentence really have lots bad stuffs.

Although these example sentence really have lots bad stuffs .

stuffs	StyWch - Casual or colloquial words or expressions	<input type="text"/>	✘
stuffs	NCount - Wrong form of countable/uncountable noun	<input type="text"/>	✘
lots bad	PreMiss - Missing preposition	lots of bad	✘
sentence have	SubVA - Subject and verb do not agree in number and/or person	<input type="text"/>	✘
Although these example sentence really have lots bad stuffs .	SFrag - Sentence fragment	<input type="text"/>	✘
these sentence	ACh - Wrong choice of article/determiner	<input type="text"/>	✘
these sentence	NNum - Wrong choice of singular/plural form of the noun	<input type="text"/>	✘

Screenshot of the IMI extension (annotation layer)

Annotation Results (Distribution)



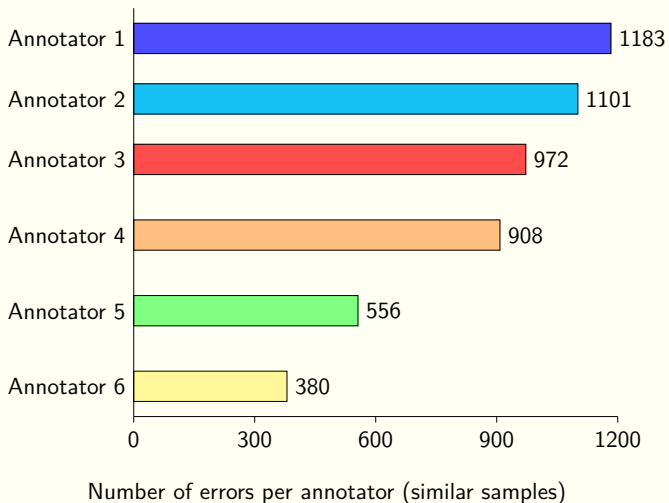
Relative error distribution, by error category (not individual tags)

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

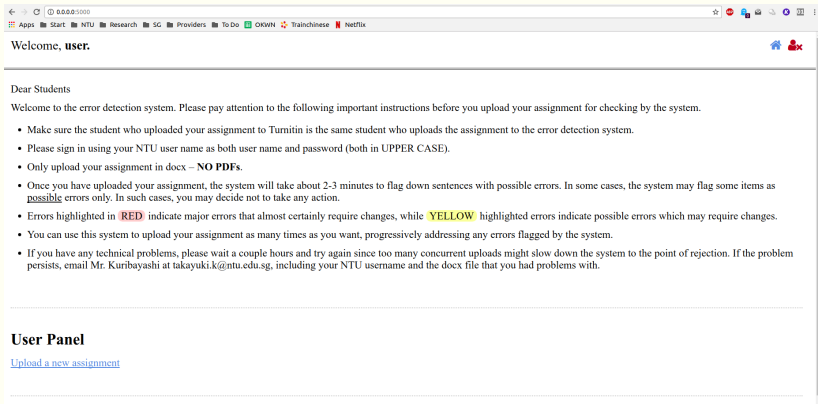
Corpus Release

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)



0.0.0.0:5000

Apps Start NTU Research SG Providers To Do OKWN Trainchinese Netflix

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.

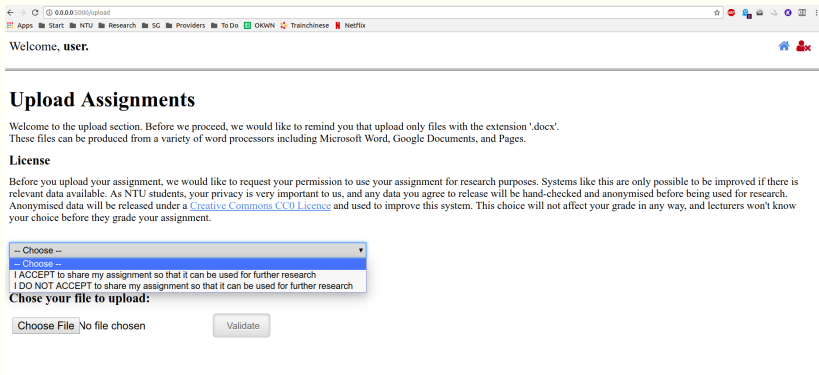
- Make sure the student who uploaded your assignment to Turnitin is the same student who uploads the assignment to the error detection system.
- Please sign in using your NTU user name as both user name and password (both in UPPER CASE).
- Only upload your assignment in docx – **NO PDFs**.
- Once you 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

The Error Detection System (II)



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.

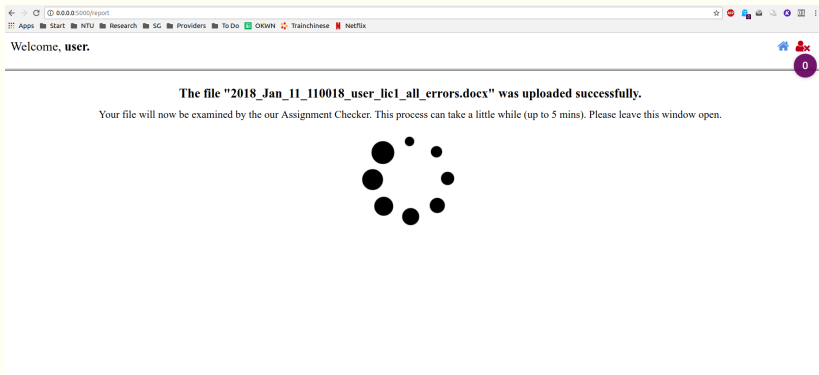
-- Choose --
-- Choose --
I ACCEPT to share my assignment so that it can be used for further research
I DO NOT ACCEPT to share my assignment so that it can be used for further research

Choose your file to upload:

Choose File No file chosen Validate

Online Error Detection System - Upload

The Error Detection System (III)



A screenshot of a web browser window. The address bar shows the URL "0.0.0.0:5000/report". The browser's tab bar includes "Apps", "Start", "NTU", "Research", "SG", "Providers", "To Do", "OKWN", "Trainchinese", and "Netflix". The page content starts with "Welcome, user." in the top left and a notification icon with the number "0" in the top right. The main message reads: "The file '2018_Jan_11_110018_user_lic1_all_errors.docx' was uploaded successfully." Below this, it says: "Your file will now be examined by the our Assignment Checker. This process can take a little while (up to 5 mins). Please leave this window open." In the center of the page is a loading spinner consisting of seven black circles of varying sizes arranged in a circle.

Online Error Detection System - Loading/Checking

The Error Detection System (IV)

0.0.0.0:5000/report

Apps Start NTU Research SG Providers To Do OKWV Trainchinese Netflix

Anti Food Deserta

Background:

In the recent years, Singapore's food wastage rate shoots rapidly. In all domains of food production and consumption, wastage of food has grown significantly. Article by food waste republic shows that in 2011, each Singaporean can generate 130kg of food waste per year, out of 0.68 million tons of food waste being thrown each year, only 10% could be recycled.

[1].

Problem:

This phenomenon raised a problem which maybe much neglected in Singapore, the awareness of wasting food is dropping. One big wastage source would be from the cosmetic filtering of food in markets to filter food and this generates majority of food wastage. Also the awareness is falling, espec in. Please check whether you intended to use 'has/have + verb-ed' instead; I feel the agency to address this problem as this negative effect is now impacting the current society and might even have a longer lasting negative effect in the future. Thus Anti deserta, meaning anti waste in Latin, is been used as the title.

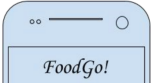
Solution:

To control the growth of this issue, I propose to setup a convenient way to distribute the wasted edible food. It requires combined effort from organizations, such as Disabled People's Association (DPA), fighting for the welfare of the elderly, disabled and the poor. Often, they are facing problem of food shortage. Instead of wasting food by discarding them, a more reasonable way would be distributing them to these needy groups. These organizations can set up centers all around Singapore to ease the collection of excess food. On the other hand, an app will be created for the convenience of notifying these centers for food collection. This provides a direct interaction between the food owners and the centers, curbing the food wasted from cosmetic filtering as now there is a better way for producers to make use of them. At such it would reduce the amount of edible food wasted daily and instead, put into good use to aid the needy. This can be beneficial especially within short term, while lowering the food wastage, it also provided the buffer time for other possible longer term solution to take effect.

Benefits:

The immediate benefited ones would be the needy groups, directly solving their food shortage. This also cuts burden on NEA's pilot project reported in Channel News Asia [3], in the effort to reduce inedible food wastage.

Implementation:



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 (⚠)

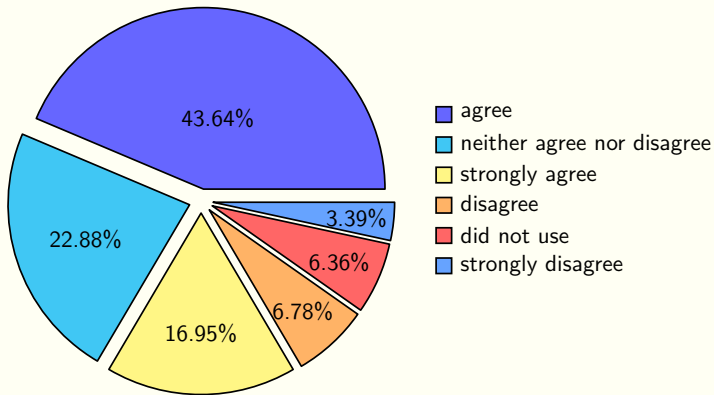
Feedback Examples

- ❑ *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

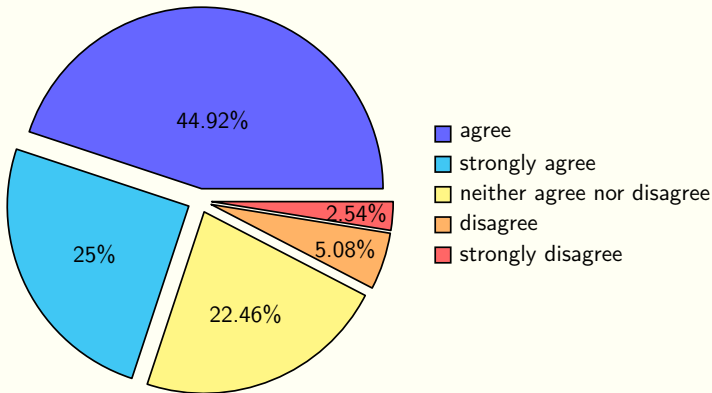
- ❖ 1000 groups of students (paired assignment)
- ❖ one extra week after the original deadline
- ❖ free use of the tool (i.e. multiple uploads)
- ❖ (ongoing) comparison between submissions
 - ❖ over 90% participation
 - ❖ greatly asymmetric usage
(i.e. no. of submissions, number of errors detected)
-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

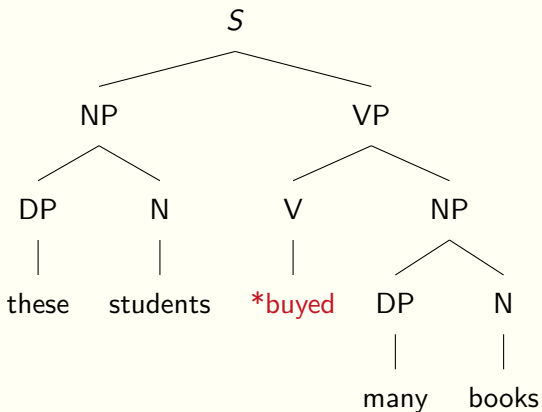
- ❑ to know and help individual student's weaknesses
- ❑ to drill student's weaknesses with exercises
- ❑ to provide precise, informed feedback on how to improve
- ❑ to adapt itself to the student proficiency level
- ❑ to help assess students' level of proficiency
- ❑ to scale to 100s or 1000s of students

Go Mal-Rules?

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

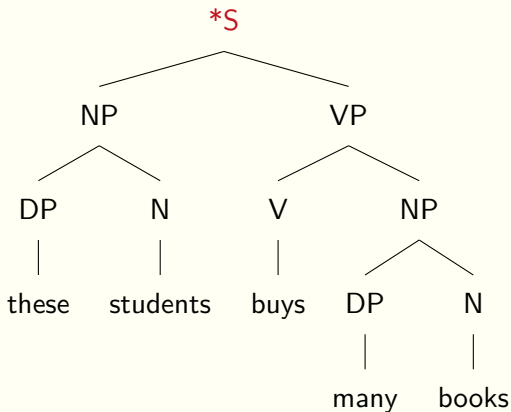
Mal-Rule (Inflectional Rule)

* These students **buyed** many books.



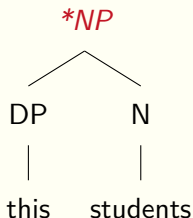
Mal-Rule (Subj-Verb Agreement)

* These students buys many books.



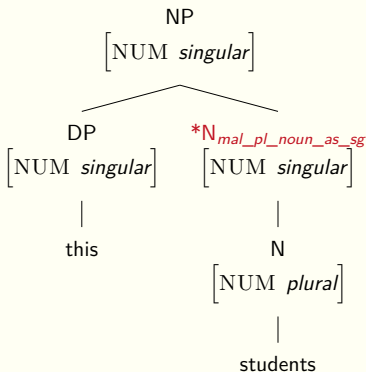
Ungrammatical Ambiguity

- what is the correct form of 'this students'?
'this student'? 'these students'?



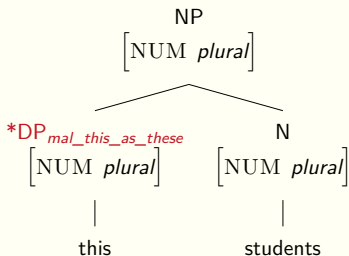
Mal-Lexical-Rule

- we can force the singular meaning



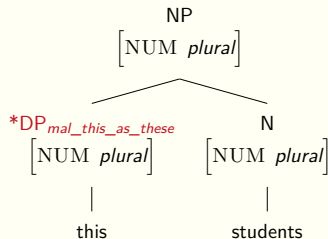
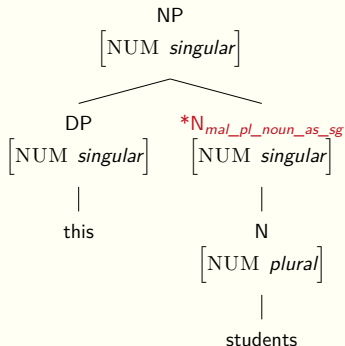
Mal Lexical Entry

- ❖ we can force the plural meaning

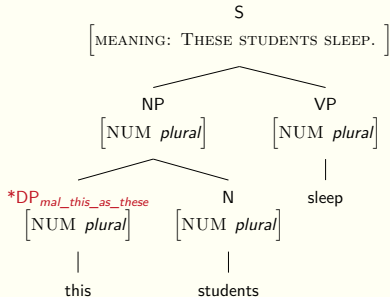
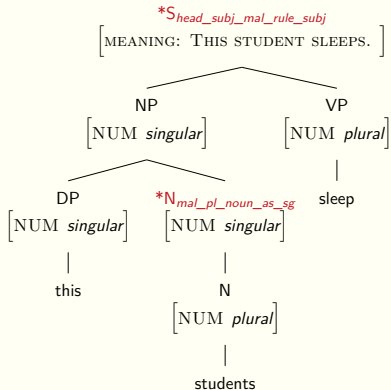


Combining Approaches



❖ 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!