NTUCLE: Developing a Corpus of Learner English to Provide Writing Support for Engineering Students

Roger V. P. Winder,[♡] Joe MacKinnon,[♡] Shu Yun Li,[♡] Benedict Lin,[♡]
Carmel Heah,[♡] Luís Morgado da Costa,[♦] Takayuki Kuribayashi,[♠] and Francis Bond[♠]
[♡]Language and Communication Centre, [♠]Linguistics and Multilingual Studies
[♡♠]School of Humanities,
[♦]Global Asia, Interdisciplinary Graduate School
Nanyang Technological University, Singapore
rogerwinder@ntu.edu.sg

Abstract

This paper describes the creation of a new annotated learner corpus. The aim is to use this corpus to develop an automated system for corrective feedback on students' writing. With this system, students will be able to receive timely feedback on language errors before they submit their assignments for grading. A corpus of assignments submitted by first year engineering students was compiled, and a new error tag set for the NTU Corpus of Learner English (NTUCLE) was developed based on that of the NUS Corpus of Learner English (NUCLE), as well as marking rubrics used at NTU. After a description of the corpus, error tag set and annotation process, the paper presents the results of the annotation exercise as well as follow up actions. The final error tag set, which is significantly larger than that for the NUCLE error categories, is then presented before a brief conclusion summarising our experience and future plans.

1 Introduction

In this paper, we report on a new project which involves the creation of a new annotated Learner Corpus (LC), and which aims to develop an automated system for corrective feedback at Nanyang Technological University (NTU), Singapore. The goal of this system is to provide immediate feedback to students on possible errors in syntax, grammar and lexis, as well as possible style problems, in their assignment drafts.

In this project, we follow studies such as Nagata (1996), which shows that it is not the medium itself (e.g. a computer, a book, a lecturer, etc.) that determines success in learning, it is the quality of the feedback produced by that medium that affects the results. This is why a language teacher is likely to be a better medium than a book, and the same reason why a properly designed Computer Assisted Language Learning (CALL) system can also be a better medium than writing guidelines, assuming that such systems can provide timely and constructive feedback to the learner.

Given our current course design and manpower constraints, students are much more likely to learn from the system's automated feedback than from receiving the same feedback from tutors, which will take longer, after an assignment has been submitted and graded. The immediate feedback through the automated system will enable students to address the possible errors before submitting the final versions for assessment. Consequently, students are more likely to take the feedback seriously because it can be used to improve the quality of the assignment before it is submitted (Price et al., 2010). Furthermore, this automated system will enable tutors to focus more attention on areas that require human judgement in their feedback, such as content, organization and use of rhetorical strategies.

To develop the system, we have tagged an LC of 180 written assignments for a course entitled *Engineering Communication I*, taught at NTU. We then developed an error coding system based on the 27 labels used in the NUS Corpus of Learner English (NUCLE, Dahlmeier et al., 2013) because of similarities in the demographic profile of the participating learners. However, we removed some categories and expanded others so the final list consists of 53 labels. Part of this was to include categories that are not purely grammatical, but pertain to matters of writing style which we are concerned about, some of which can be automatically detected. These include not only obvious style issues such as the use of contractions and collo-

Proceedings of the 4th Workshop on Natural Language Processing Techniques for Educational Applications, pages 1–11,
Taipei, Taiwan, December 1, 2017

quial words or expressions but also more subtle ones such as overly long and convoluted sentences and missing parallel clause structures. In this, our corpus distinguishes itself from the Cambridge Learner Corpus (CLC, Nicholls, 2003), NUCLE and other corpora which focus solely on grammar.

Our primary motivation for assembling the NTU Corpus of Learner English (NTUCLE), in other words, is to help individual students to identify their language and style problems, and to rectify these on their own. This is unlike the broader intentions of the CLC, whose error coding and analysis is intended to provide "lexicographers, researchers, ELT [English Language Teaching] authors and examiners with easy, direct information which they can interpret and use for widely varying purposes" (Nicholls, 2003). Similarly, our initial motivation differs from that of the NUCLE, whose goal is to provide a large data resource for research purposes, and for development of grammatical error correction systems (Ng et al., 2014).

Our ultimate goal is also different from many current Natural Language Processing (NLP) projects, which appear to focus on building automated grammatical correction tools, with the holy grail of a "complete end-to-end application" that can identify and correct mistakes for the writers, with a high degree of precision (see Ng et al., 2014). Instead, the goal for NTUCLE was to develop a system that will be able to prompt students to review possible mistakes in their writing drafts and correct them on their own. This will allow learners to participate more meaningfully in the error correction process and to actively identify and choose from multiple options which are often available and would be considered acceptable by different annotators (Rozovskaya and Roth, 2010).

Finally, NTUCLE differs from other similar corpora in its narrower focus on a specific genre (i.e. technical proposals) and target students (i.e. Singaporean engineering undergraduates). Nevertheless, we foresee that our project might be expanded to include other genres and groups of learners, though sub categorisation of specific groups of learners and genres will be ensured.

We have now completed annotation of the corpus, and are currently using this to develop the system for providing feedback to students. This system will detect and tag potential errors in drafts submitted by students, and identify likely errors using our categories. It will not correct any error, but will prompt, with different degrees of confidence, students to consider whether corrections are needed. In this way, we hope to encourage students to adopt a more independent and critical approach to error correction. We also hope to enable a pedagogy focused on timely, high quality feedback to students.

This paper discusses the completed phases of our new LC primarily from the perspective of professional English instructors. In Section 2, we describe the compilation of the corpus and the establishment of initial error tag set. We then describe in Section 3 the annotation process, before presenting in Section 4 the outcomes of our initial annotation, including findings on the most frequent errors identified, inter-annotator differences in tagging and how we resolved them. Section 5 highlights our revised error tag set. We conclude with a brief note on the corpus release, followed by a summary of our experience and our future plans for the corpus.

2 Corpus and Error Tag Set

2.1 Corpus Compilation

Approval was obtained from the university's Institutional Review Board for the research protocol and the use of students' written assignments, subject to the students' consent. Over three semesters (from 2015 to 2016), 349 students gave written consent, and their assignments were retrieved for the corpus.

Of the assignments retrieved, we selected only files in doc/docx format, because it would be difficult to automate text extraction, while preserving headings, paragraphs, style and sentence boundaries for the other formats (e.g. pdf). We ended up with 273 documents from which we tagged only a random sample of 180 documents, due to time and manpower constraints. The 93 untagged documents were kept to test the error-detecting system under development.

The documents are assignments from a communication skills course taught at NTU for first-year engineering students. These authors are predominantly Singaporean (about 80%), with many likely to have native speaker proficiency in English, male (70%), and between 18 and 22 years of age. The assignments consist of a 500-word technical proposal that offers an engineering solution to a real life problem. The solution could be a new product,

service or process, or an improvement of an existing one. The instructions for the assignment specify a structure for the proposal consisting of seven sections: background, problem, solution, benefits, implementation, costs/budget and conclusion.

2.2 Initial annotation schema

We next developed a preliminary error tag set by referring to NUCLE (Dahlmeier et al., 2013) as well as marking rubrics used at NTU. Six annotators, all professional English instructors, then tagged the same selected paper from the data set using this tag set. After conferring and reviewing the error tags and agreeing on what constituted an error in the student's paper, we created a modified tag set with 15 broad categories covering 50 error labels. This is much larger than the NUCLE tag set, which has 13 broad categories and 27 error labels, though we were conscious of how excessive granularity could lead to greater difficulty in applying the annotation schema to the documents (Nagata et al., 2011). Below are the ways in which we modified the NUCLE tag set:

Removed two broad categories:

(a) 'Redundancy' because tags created in other categories dealt with this issue more specifically, and

(b) 'Word Choice' replaced with 'Words (lexical)' to reflect a broadening of the category.

Created three additional categories:

(c) 'Expression', covering two tags, 'Awkward expression' (not used in NUCLE) and 'Unclear expression' (similar to 'Unclear meaning' under 'Others' in NUCLE);

(d) 'Prepositions', with three tags (NUCLE covers prepositions under a single tag for 'Wrong collocation/idioms/prepositions'); and

(e) 'Style', with two tags unique to NTUCLE ('StyF' for overly formal words or expressions and 'StySh' for inappropriate shifts in style and formality), and one other tag ('StyC' for inappropriate use of casual or colloquial words or expressions) similar to 'Wtone' for 'Tone' under 'Word Choice' in NUCLE.

Added tags in most categories, through:

(f) specifying whether an error involved something missing, unnecessary or inappropriate (for 'Articles, determiners', 'Prepositions', 'Pronouns', 'Verbs' and 'Words'), similar to the use of 'insertion'/'missing', 'deletion'/'unnecessary' and 'replacement' tags in other projects (see Bryant et al., 2017; Rozovskaya and Roth, 2010);

(g) expanding tags that were collapsed in NUCLE (e.g. two separate tags for run-on sentences and comma splices instead of one, and more specific tags for case, punctuation, spacing and spelling instead of the generic 'Mechanics'); and

(h) creating new tags such as 'VVoice' (for wrong choices of active or passive voice), 'NCount' (for wrong forms of countable/uncountable nouns), and 'SMMod' (for misplaced modifiers) based on errors we have found from experience to be common in our students' writing.

Reduced the error tags in 'Others':

(i) replacing the tag 'Unclear meaning' with the tag 'ExpUC' (for 'Unclear expression') in our new 'Expressions' category.

3 Annotation Process

From the 180 documents collected (see 2.1), each of the 6 annotators was randomly assigned 40 documents, ensuring that 20 of these 40 documents were overlapped evenly with two other annotators (i.e. 10 documents overlapped with another annotator, and another 10 documents overlapped with a second annotator). Each annotator tagged the assigned scripts independently, and the identities of the other annotators tagging the same documents were not revealed. Annotators were also not aware which samples were being double tagged with other annotators. The double tagging was done to check accuracy and inter-annotator agreement.

A total of 60 documents were double annotated. Annotators were instructed to tag every error identified as specifically as possible, and to use more than one tag for the same set of words if there were multiple ways of tagging the error. While we acknowledge that it would also have been useful to correct the errors identified, this was not done because of the complexity of the task, especially in identifying all possible options for correcting each error while preserving the student's intended meaning (Sakaguchi et al., 2017). Unfortunately, this would have required more time and resources than were available.

3.1 Annotation Tool

The annotation process was done on an expanded version of IMI – A Multilingual Semantic Anno-

tation Environment (Bond et al., 2015). We used the open source platform to build an extra layer to the annotation environment, allowing us to tag the documents with our own tag-set (discussed in 2.2 and presented in 5).

The annotators used this new system to tag each document by sentence, in ascending order. Although the system currently only allows tagging sentence by sentence, (i.e. annotators could work on only one sentence at a time on-screen), annotators had access to the full text of each document so that they could identify errors in context (e.g. errors in pronouns with referents in earlier sentences).

To tag each error, annotators could select a single word, a contiguous word-string (e.g. a phrase), a set of non-contiguous words (e.g. a pronoun and its referent earlier or later in the sentence), or the entire sentence. Multiple errors could be tagged for each sentence. Total and partial overlap of errors within the same sentence were allowed and encouraged. This happened, for example, when the same span of words could be corrected in more than one way (i.e. two error tags were assigned to the same span of words), or when a smaller error occurred within a larger error (e.g. an agreement error inside an overly long sentence). Errors were tagged at the level of word tokens, which means that sub-word units could not be selected. Missing words were indicated by tagging the words surrounding the location of the hypothesized missing word. A text-box was provided for each error, which could be used to correct it or to leave comments (e.g. to flag referents that should be anonymized). A screen-shot of the annotation environment is shown in Fig 1.

4 Results and Annotation Issues

The results of the annotation exercise revealed a wide range in the number of errors tagged by each annotator, from 380 (Annotator 2) to 1,183 (Annotator 3), as shown in Table 1. This is not unusual and similar differences have been observed in other annotation exercises (see, for example, Bryant and Ng, 2015). Further discussion suggested that the differences are likely to be due to different levels of sensitivity to particular errors and different tagging practices, including decisions about which particular word, phrase, clause or even sentence to tag with a single label. Annotators also differed in the frequency with which they tagged the same word or word string with different tags to acknowledge different ways of identifying errors. It is also possible, but unlikely, that particular annotators may have received assignments from weaker students – since the assignments were distributed randomly across annotators.

As Table 1 also indicates, three annotators (A2, A3 and A6) were highly similar in their tagging patterns in relation to the three main error categories tagged, namely 'singular/plural forms', 'missing article/determiner' and 'word choice' – which were also the top three error categories overall. Two others (A1 and A5) also had similar top three error categories ('word choice', 'awkward expression' and 'unclear expression') but these were the third, fourth and fifth most common error categories tagged overall.

The five most common errors, distributed by annotators, are shown in Figure 2: errors in using singular or plural forms, omitting articles or determiners, choosing inappropriate words, using awkward expressions and using unclear expressions. However, the annotators appeared to have had different emphases in their annotation. While overall, strictly grammatical errors (i.e. use of singular/plural forms and omission of articles and determiners) were the most commonly identified, annotators 1 and 5 identified far more errors in 'expression' (unclear/awkward), which may relate more to issues in semantics or idiomaticity.

4.1 Double Tagging

As has been mentioned earlier, 60 documents were double tagged. In many cases, both annotators tagged the same errors in the same sentences, either for exactly the same word strings or for word strings with some overlapping words. However, there were also significant differences, such as different word strings tagged for the same error type, or the same word string tagged for different error types.

Interestingly, although two pairs of annotators (A3+A4, and A5+A6) had relatively high degrees of overlap in using the same error tags, they also had relatively high degrees of discrepancy in assigning error tags to the same word strings. This suggests that while they had a strong common understanding of some error tags, they quite possibly also had rather different interpretations of others, or that they had quite different foci where the word strings may have more than one error type.

All the annotators met to review every instance

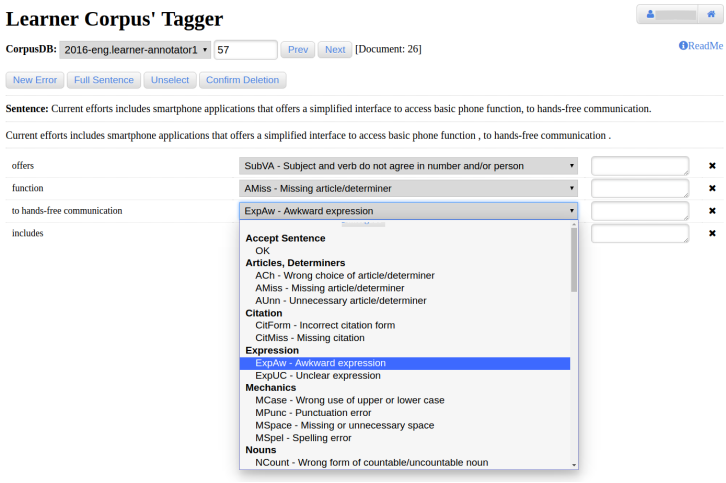

Figure 1: Annotation tool developed for the corpus annotation, as an extension of IMI

A#	No. Errors	Most Common Error	2nd Most Common Error	3rd Most Common Error
A1	1,101	awkward expression (21%)	word choice (11%)	unclear expression (10%)
A2	380	singular/plural forms (22%)	word choice (7%)	missing article/det. (6%)
A3	1,183	singular/plural forms (12%)	missing article/det. (10%)	word choice (8%)
A4	556	missing article/det. (21%)	singular/plural forms (11%)	verb form (9%)
A5	908	unclear expression (12%)	awkward expression (11%)	word choice (7%)
A6	972	singular/plural forms (11%)	word choice (9%)	missing article/det. (9%)
Total	**5,100**	**singular/plural forms (10%)**	**missing article/det. (8%)**	**word choice (8%)**

Table 1: Top errors by annotator (before harmonisation)

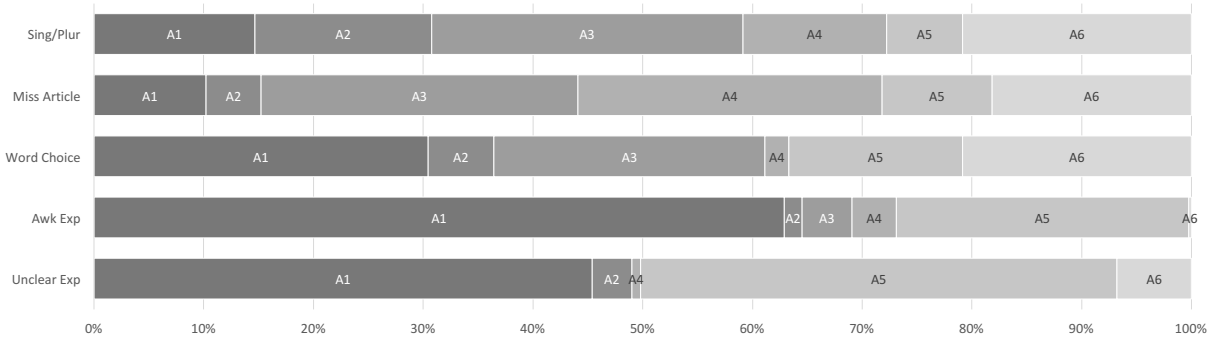

Figure 2: Contributions of annotators to top five errors tagged

of different error tags assigned to the same exact sentence spans (words, expressions, etc.). In most cases, it was agreed that one or both of the annotators had made a mistake, either unintentionally, or through misunderstanding or misapplying a tag. The relevant word strings were then re-tagged with the correct tags. In a few instances, either tag could apply, e.g. *However, trolley has its own limitations*, which can be construed as either 'NNum' (*However, trolleys have their own limitations*) or 'AMiss' (*However, a trolley has its own limitations*). In yet a few others, both tags apply, i.e. there are two errors conflated in the same word set (e.g. *For example, individual seats for individual cubical will be installed with motion sensor.*, where there is both an 'MSpel' (*cubical* to *cubicle*) and 'NNum' (*cubical* to *cubicles*) error). In both these kinds of cases, we agreed that both tags should remain, in the first instance because there are two ways of correcting the error, and in the second, because there are two overlapping errors.

Some differences arose because the annotators involved found it difficult to use existing tags. In many of these instances, one annotator tagged the error under 'Others' and provided his or her own labels or comments. From these, we identified new categories for tagging, namely 'StyMood' for the inappropriate use of imperatives or interrogatives, 'SLong' for overly long sentences, and 'SConv' for convoluted sentences.

We understand that such differences could have been avoided, and the tagging process made more efficient had the annotators been given more detailed guidelines or met for a more extensive standardisation exercise prior to annotation. However, as an exercise to test natural discrepancies in human tagging, this was a useful exercise. Given that our final goal is to emulate human feedback, while providing constructive feedback on issues lecturers usually highlight in student assignments, it was an important part of our experiment to allow this kind of naturalistic tagging, which captures differences in grading expectations, editing experience and perceptions of acceptable or exemplary language use (Daudaravicius et al., 2016; Rozovskaya and Roth, 2010). Consequently, to be able to create a useful error-feedback system, we wanted to restrain ourselves from creating a highly mechanical process to assign tags – even at the cost of inter-annotator agreement.

The annotators also discussed their own 'pet peeves' in the texts they annotated. Among those most commonly shared was the problem of overly long sentences that made comprehension difficult. This reinforced the need for the category 'SLong'. Another commonly shared 'pet peeve' was the inappropriate over-use of certain colloquial words and informal clichés, *tackle* (to mean *study*, *address* or *solve* a problem) and *hassle* (to mean *inconvenience* or the like) being two of the most common. Another new category 'StyWch' for the use of casual or colloquial words and expressions was created to tag such words.

Our observations of instances tagged for inappropriate style also helped us to identify the specific ways in which this problem was realized in linguistic form, leading to a further two new categories – 'StyContr' for the use of contractions and 'StyPron' for the use of first and second person pronouns.

5 Revised Error Tag Set

The review of the annotation exercise resulted in an amended error tag set with the same 15 categories but with 53 tags. After the amendments, the tag discrepancies mentioned in 4.1 were resolved. Based on the discussion above and the results of the initial tagging, errors that had been tagged under the 'Style' category were re-tagged with one of the tags available in the final tag set.

Table 3 presents our final error tag set, with an indication of the frequency of each error type in the corpus after re-tagging. The 'Source' column indicates how the tags were created:

- 'Sub-divided': broader NUCLE tags that were sub-divided to be more specific
- 'Modified': NUCLE tags that were modified slightly to be more specific
- 'Moved': NUCLE tags that were moved to other categories
- 'NUCLE': NUCLE tags that were not changed
- 'Re-named': NUCLE tags that were re-named to fit the NTUCLE schema
- 'NTUCLE': tags created for NTUCLE

6 Corpus Release

The corpus described above will soon be available at the following url: http://compling.hss.ntu.edu.sg/ntucle.

The corpus includes eight databases, all of them following the database schema used in IMI (Bond

et al., 2015). All anonymised data will be released under an Attribution 4.0 International license (CC BY 4.0),[1] in conformity with our IRB and the students' consent.

Table 2 provides a quick overview of the corpus to be released: number of documents, overlaps, number of sentences, number of word tokens, number of sentences that contain at least one error label, and the total number of errors included in each database.

We will release the six individual databases, each tagged by a different professional English instructor, along with a compiled database of the 180 documents tagged, merging documents that were double tagged. While the compiled database has more traditional usages, we believe the individual databases can be used to further analyse and discuss individual differences between annotators. Lastly, we will also release a database with the remaining untagged documents.

7 Conclusion

Based on the NUCLE, we have started the NTU Corpus of Learner English using written assignments submitted by first year engineering students. This corpus will be used to develop an automated system for corrective feedback which is expected to cultivate greater student autonomy and critical awareness in error correction when writing. Our system will be piloted and tested with the next round of submissions for the same writing assignment used to develop the corpus. We plan to add these submissions to the corpus, and keep expanding it.

For our corpus, we have developed a new learner error tag set with 53 tags, which is significantly larger than NUCLE's. This is to meet the specific needs and goals of this corpus, the development of the online tool for corrective feedback without automated correction. As expected, there were significant differences among the annotators in applying the initial tag set, with some annotators being more or less sensitive to particular errors than others. In samples that were double tagged, there were both overlaps and differences in the words tagged and the error tags used. We discussed and resolved all instances where the same word strings were tagged differently, and retagged the word strings. We agreed unanimously that the annotation process could have been improved if more detailed guidelines or brief training had been provided to the annotators prior to annotation. At the same time, the goal of building an automated system for corrective feedback of student's writing, as mentioned above, invited us to firstly acknowledge the low inter-annotator agreement and different foci of professional instructors when correcting student assignments.

We believe that our annotated corpus can be a useful new learner corpus, which can complement and advance on the purposes of corpora such as NUCLE, and we hope to expand it with other genres and learners in the near future. We would like to have the opportunity to further revise and harmonize the annotations in the corpus, and we also acknowledge that it would be beneficial to provide corrections for the identified errors. Unfortunately, this will be dependent on the availability of resources.

All compiled, we are releasing, under an open license, 273 anonymised student assignments, comprising over 14,700 sentences. Roughly 65% of this corpus has been tagged using our newly proposed tagset (available in Table 3).

[1] https://creativecommons.org/licenses/by/4.0/

Acknowledgments

This research is supported by Nanyang Technological University's EdeX Teaching and Learning Grant administered through the Teaching, Learning and Pedagogy Division (TLPD) and the MOE TRF grant *Syntactic Well-Formedness Diagnosis and Error-Based Coaching in Computer Assisted Language Learning using Machine Translation Technology*. We thank the reviewers for their comprehensive comments. We would also like to acknowledge the valuable contributions of Boon Tien Lim in the earlier phases of the project.

References

Francis Bond, Luís Morgado da Costa, and Tuan Anh Le. 2015. IMI – A multilingual semantic annotation environment. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, System Demonstrations (ACL 2015)*, pages 7–12, Beijing, China.

Christopher Bryant, Mariano Felice, and Ted Briscoe. 2017. Automatic annotation and evaluation of error types for grammatical error correction. In *Proceedings of the 55th Annual Meeting of the Association*

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.

Table 2: Corpus Statistics

for Computational Linguistics, pages 793–805, Vancouver, Canada. Association for Computational Linguistics.

Christopher Bryant and Hwee Tou Ng. 2015. How far are we from fully automatic high quality grammatical error correction? In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 697–707, Beijing, China. Association for Computational Linguistics.

Daniel Dahlmeier, Hwee Tou Ng, and Siew Mei Wu. 2013. Building a large annotated corpus of learner English: The NUS corpus of learner English. In *BEA@ NAACL-HLT*, pages 22–31.

Vidas Daudaravicius, Rafael E. Banchs, Elena Volodina, and Courtney Napoles. 2016. A report on the automatic evaluation of scientific writing shared task. In *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 53–62, San Diego, CA. Association for Computational Linguistics.

Noriko Nagata. 1996. Computer vs. workbook instruction in second language acquisition. *CALICO journal*, 14(1):53–75.

Ryo Nagata, Edward Whittaker, and Vera Sheinman. 2011. Creating a manually error-tagged and shallow-parsed learner corpus. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 1210–1219, Portland, Oregon, USA. Association for Computational Linguistics.

Hwee Tou Ng, Siew Mei Wu, Ted Briscoe, Christian Hadiwinoto, Raymond Hendy Susanto, and Christopher Bryant. 2014. The CoNLL-2014 shared task on grammatical error correction. In *CoNLL Shared Task*, pages 1–14.

Diane Nicholls. 2003. The Cambridge learner corpus: Error coding and analysis for lexicography and elt. In *Proceedings of the Corpus Linguistics 2003 conference*, volume 16, pages 572–581.

Margaret Price, Karen Handley, Jill Millar, and Berry O'Donovan. 2010. Feedback: all that effort, but what is the effect? *Assessment & Evaluation in Higher Education*, 35(3):277–289.

Alla Rozovskaya and Dan Roth. 2010. Annotating esl errors: Challenges and rewards. In *Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 28–36, Los Angeles, California. Association for Computational Linguistics.

Keisuke Sakaguchi, Courtney Napoles, and Joel Tetreault. 2017. Gec into the future: Where are we going and how do we get there? In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 180–187, Copenhagen, Denmark. Association for Computational Linguistics.

Categories	Tags	Explanation	Freq.	Source
Articles, determiners	ACh	Wrong choice of article/determiner *A development of a new product is required*	69	Expanded
	AMiss	Missing article/determiner *a stall with [a] shorter queue*	449	Expanded
	AUnn	Unnecessary article/determiner *two holes in the two of the sides*	144	Expanded
Citations	CitForm	Incorrect citation form *(Sim, R. 2013)*	100	Expanded
	CitMiss	Missing citation *According to a study [citation], Singaporean students ...*	6	Expanded
Expression	ExpAw	Awkward expression (meaning is clear) *paths are of high human traffic*	366	NTUCLE
	ExpUC	Unclear expression (meaning is unclear) *A rubbish bin to test our idea as well as human resources from the companies*	249	Moved
Mechanics	MCase	Wrong use of upper or lower case *The Rubbish bin is a common object*	98	Expanded
	MPunc	Punctuation error *This[,] in turn[,] would create an orderly environment*	190	Expanded
	MSpace	Missing or unnecessary space *They can not be used in open areas*	27	Expanded
	MSpel	Spelling error *a cold and quite environment*	58	Expanded
Nouns	NCount	Wrong form of countable/uncountable noun *Users can exchange notes and advices*	77	NTUCLE
	NNum	Wrong choice of singular/plural form of the noun *one of his speech*	525	NUCLE
	NPoss	Wrong choice of possessive form *the timers can be adjusted to workers['] feedback*	22	NUCLE
Prepositions	PreCh	Wrong choice of preposition *at the comfort of his home*	227	Expanded
	PreMiss	Missing preposition *EasyGrip will be a great addition [to] every household*	53	Expanded
	PreUnn	Unnecessary preposition *video tutorials can be played to teach users on how to use the mouse*	54	Expanded
Pronouns	ProAgr	Pronoun and reference do not agree in number/person/gender *An electrostatic precipitator works by absorbing dirty air, passing them through ionising electrodes*	88	Re-named
	ProCh	Wrong choice of pronoun *they things tend to slip off their mind easily*	32	Expanded
	ProMiss	Missing pronoun *5 'X's will identify owners as irresponsible and deny [them] a pet.*	21	Expanded
	ProRef	Unclear reference for pronoun *The components can be mounted onto a circuit board, which is covered with a plastic housing once it is completed.*	92	Modified
	ProUnn	Unnecessary pronoun *Death then follows if the victim he is been left untreated within minutes*	8	Expanded

Categories	Tags	Explanation	Freq.	Source
Sentence structure	SComS	Comma splice *The wobbling table can cause food and drinks to be spilled out of their containers, writing can become messy.*	40	Expanded
	SConv	Convoluted sentence *Rubbish bins are facing one problem in crowded areas where bins fill up quickly that cleaners have hard time discerning as there are too many bins, and only come at fixed timings to clear the rubbish currently.*	-	NTUCLE
	SDMod	Dangling modifier *Looking at the bigger picture, a canteen can efficiently accommodate more diners in a given time.*	16	Expanded
	SFrag	Sentence fragment *Thus, showing that our students have a huge desire to always learn something new.*	58	NUCLE
	SLong	Overly long sentence *However, they would not be able to do the required printing if they possess an EZ-link card that has insufficient stored monetary value and hence may require the assistance of friends by borrowing their EZ-link cards, or make their way back to (...) [+38 words]*	14	NTUCLE
	SMMod	Misplaced modifier *An ideal conducive learning environment is essential as it facilitates effective teaching and learning process coupled with a well-equipped lecture theatre*	11	NTUCLE
	SPar	Parallelism missing *students will find it a hassle to go through emails and calling to find out more*	37	NUCLE
	SRun	Run-on sentence *there is an increase in commuters for public transport[;] this leads to higher congestion in public transport*	26	Expanded
	SSub	Problematic subordinate clause *The immediate benefited [sic] ones would be the needy groups, directly solving their food shortage.*	25	NUCLE
Style	StyContr	Contractions *It's a rectangular device*	25	NTUCLE
	StyF	Overly formal words or expressions *To solve the aforementioned problems*	1	NTUCLE
	StyMood	Inappropriate use of interrogatives and imperatives *Establish a collaboration with an existing music-streaming app.*	13	NTUCLE
	StyPron	Inappropriate use of first and second person pronouns *I could not manage to find the cost of one EZ link top up machine*	9	NTUCLE
	StyWch	Casual or colloquial words or expressions *some find it a hassle to search for an available power socket*	92	NTUCLE
Subject-verb agreement	SubVA	Subject and verb do not agree in number and/or person *The portable charger are basically portable*	148	NUCLE

Categories	Tags	Explanation	Freq.	Source
Transitions	TCh	Wrong choice of link words/phrases *Hence users will also be able to purchase a UV light, where they can use it to identify areas which were not cleaned properly*	50	Expanded
	TMiss	Missing link words/phrases *The food owners select the nearest food centre, [and] fill in their address and contact number.*	26	Expanded
	TUnn	Unnecessary link words/phrases *Skipping lunch can cause students to be distracted by hunger and thus affecting academic performance.*	34	Expanded
Verbs	VForm	Wrong form of the verb *NTU is rank 13th in the world*	231	NUCLE
	VMiss	Missing verb *The files they need [?] directly streamed to their computer.*	23	NUCLE
	VMod	Missing, inappropriate or unnecessary modal *To produce the application, the following steps are taken:*	138	NUCLE
	VTense	Verb tense *Each year Nanyang Technological University (NTU) welcomed approximately 4,500 students into their freshmen year*	121	NUCLE
	VVoice	Wrong choice of active or passive voice *The phenomenon of overcrowding of Canteen B has been existed for a long time.*	27	NTUCLE
Word order	PosAd	Wrong position of adjective/adverb *vacuum cleaners can be used to clean narrow spaces also*	3	Re-named
	PosW	Incorrect word order *the problem of dropping things off the desk*	13	Re-named
Words (lexical)	WCh	Wrong choice of word *The air conditioner is an electric appliance that alternates the surrounding temperature.*	411	NTUCLE
	WColloc	Words do not collocate *Find assistance from Sistic to sell tickets*	73	NTUCLE
	WForm	Wrong form of the word *Rentascoot™ is environmental friendly*	96	NUCLE
	WMiss	Missing words *This system can simplify [?] and reduce the time of packing away [?].*	95	NTUCLE
	WUnn	Unnecessary words *... which poses severe risks to nature as well as human health issues*	195	NTUCLE
Others	Oth	Other errors requiring correction	140	NUCLE

Table 3: Final list of error tags. Examples for each error are provided below the explanation of each tag, with the words selected for each error underlined. Possible corrections are provided in brackets when deemed necessary.