# An Error Analysis of the English Resource Grammar with English Learner Data, and Application

**Sara Ng**
University of Washington
sbng@uw.edu

## Abstract

My intent is to test the efficacy of syntactic and semantic parsing of an off-the-shelf grammar (the English Resource Grammar) on academic writing from language learners, using the TECCL (Ten-thousand English Compositions of Chinese Learners) corpus. Through error analysis, I show the strengths of the grammar in parsing learner data, the common ways that the English Resource Grammar fails to accommodate learner language, and in addition offer possible applications of this analysis and others likef it.

## 1 Introduction

### 1.1 Vision

It is my goal to test an off-the-shelf syntactic and semantic parser on corpus data sampled from language learners, in the hopes of validating functionality of the system as well as being able to distributionally describe the types of errors that English language learners make that could contribute to their intelligibility. In this section, I will describe the resources I use to conduct such an error analysis. In the next section, I present similar work in related fields. In Section 3 I describe my methodology, and in the subsequent two sections I present results of the analysis, and a brief discussion.

### 1.2 ERG

The English Resource Grammar (ERG) is a precision grammar of standard English, that boasts high coverage (83% on well-edited text) and integrative potential with various grammar engineering environment (Copestake & Flickinger, 2000). The grammar features a hand-crafted lexicon and a meticulously-detailed set of grammatical rules.

The ERG is a daughter project of the original Verbmobil project, an NLP task in machine translation using transcribed speech data from travel inquiries (Copestake & Flickinger, 2000). The grammar was used within VerbmobilT to create off-the-cuff translations between English, German, and Japanese. The "meat" of the grammar is contained in excess of 15,000 lines of code, with an additional 30,000 of hand-curated lexical entries (Copestake & Flickinger, 2000). Via a grammar engineering toolkit, it is capable of producing syntactic parses in tree and nested forms, as well as a few different semantic representation forms using Minimal Recursion Semantics (mrs).

There is not detailed information available about the speakers who created the grammar (unless one were to deep dive the list of lab members), and so the ERG does not have ostensible priors on any target demographics. Baldwin et al. (2004) test the system on the British National Corpus (BNC). But like the other data that has been used to test the ERG, the BNC is largely well-edited text, with the bulk of its body coming from journalistic publications (Asthon & Burnard, 1998).

I am of the opinion that the ERG is an ideal test subject for the type of analysis I perform. It's (mostly) intuitive parsing structure, and ability to run with parsing and visualization tools like the Answer Constraint Engine (ACE), make it easy for novel users like myself to quickly retrieve parses and error statistics. In addition, its linguistic precision allows it to serve as proxy for the kinds of grammar rules that second language learners of English may be exposed to. It is for these reasons that I use the ERG as the grammar baseline for this error analysis.

### 1.3 TECCL

Ten-thousand English Compositions of Chinese Learners (the TECCL corpus) is a representative

corpus of the writings of current Chinese-national foreign language learners of English. There are more than 10,000 individual stories, each containing about a paragraph or two of text, that correspond to an assignment or written exam from a student in an English language classroom.

The corpus boasts a diverse background of speakers for the data. Samples can be arranged by age, education level, and English level, as well as provincial region within China. The designers of the corpus sought to present equal representation across these metrics in the data (Xu, 2015). While it is not recorded in the documentation for the data, Xu also notes that the data come from a variety of activity types, such as a timed essay activity or a partnered take-home journal entry.

Despite its robustness, there are few existing applications of the data. Song & Wolter (2017) use the data to compare First language(L1)/Second language (L2) effects cross-linguistically. Huang et al. (2017) use the data as a test set for their computational semantic entity grid model. Through validation from the corpus data, they show how their model improves automatic assessment of student essay submissions in the English as a Foreign Language (EFL) context. LI & LIU (2017) use the corpus as an informant of tendencies of Chinese EFL learners with respect to synonyms.

The TECCL poses an interesting data source to the English Resource Grammar for a number of reasons. First, it is unabashedly text data composed from writing samples; there's no confounding of speech processing. This is in contrast to the English Resource Grammar, which was originally tested on transcribed speech from the Verbmobil data set (Copestake & Flickinger, 2000). The previous testing set also shows temporal variance. Verbmobil transcriptions were collected in the 1980s (Wahlster, 1993). In contrast, the designers of the TECCL corpus specifically attempted to collect as modern of a data set as possible, and therefore much of the data was sampled around 2015. It is also interesting to note that the English Resource Grammar has a hand-curated lexicon, which may also show its age when faced with newer data (Copestake & Flickinger, 2000). Finally, and most markedly, the ERG is optimized for "well-formed" or standard grammatical English utterances. How does TECCL stack up to this description? Xu (2015) notes that in general, Learner corpora can perform well with off-the-shelf parsers like the ERG, because learner grammar tends to rely on simplistic and repetitive forms. However, the learner corpora are also victim to grammatical errors, unusual lexical choices, and other problems such as co-reference confusion, etc. A problem that Xu (2015) identifies as specific to Chinese learner corpora is widespread, incorrect punctuation. Xu notes that "Chinese learners have a notorious habit of typing words immediately after the commas and full stops without a space" 2015. I also found in using this data that Chinese learners incorporate Chinese punctuation of titles into English writing, so *Great Expectations* is rendered as << Great Expectations >>, which is not a standard English form. Other interesting challenges of this sort will be discussed in Section 3.

## 2 Related Work

Learner corpora hold a significant place in the computational study of natural language. A learner corpus contains a statistical picture of L2 acquisition, and offer the researcher a unique demographic on which to test existing and developing systems. They can also be used to inform L2 pedagogy.

### 2.1 Informing Pedagogy

Of interest to language teachers and designers of curricula is the types of errors language learners make, and what kind of mistakes are important to correct and which can be ignored. Learner corpora can aid this task in two ways. First, they show what kinds of errors students make and with what frequency. For example, Dickinson & Ragheb (2009) are able to classify and tag learner corpora by error type. This kind of information shows what are the errors that should be considered for correction. Running learner data against existing NLP machines, especially when data is tagged in this way, can inform instructors what kinds of production errors are most likely to cause failure in comprehension of the listener.

### 2.2 Understanding L2 Language

Learner corpora also inform what we know about Second Language Acquisition (SLA) and the grammar of a language learner. For examples **?** use learner corpora to establish categories of language learner levels, and to assign grammatical *criteria* to each of these levels. Learner corpora

snapshot the errors that learners are likely to make. In addition, they record the canonical grammar forms utilized by language learners.

Some learner corpora also contain meta-linguistic judgments from learners, or can be used to analyze language comprehension (Granger, 2002). In these cases, the corpora not only model production, but also language processing. This could be integral information for grammar models that are said to be based on neurological reality; should such grammars be expected to accommodate learner language, they must also mirror the processing skills of a second/foreign language learner.

### 2.3 Broad-Brush NLP

It is not surprising that learner corpora do not have a place of prestige as a testing tool for the state-of-the-art in NLP. To my knowledge, there is no parser or ASR system trained on explicit learner data, although some like the Stanford parser have been applied to Learner Corpora. This seems like a glaring oversight to the field at large. In an era of increasing globalization, the proportion of the global population who are L2 learners of administrative/prestige languages is increasing dramatically. Especially since our language tools are only optimized for these languages, they need to be able to accommodate such users.

## 3 Methodology

In this section, I describe how I take the raw TECCL data and tokenize and process it to be usable in the ACE interfacing for parsing with the ERG.

### 3.1 Tokenization

The organization of the TECCL corpus preserves student formatting of writing samples, which mandates that before parsing, things like titles and dates be removed, and sentences tokenized. I have chosen to use the nltk function sent_tokenize(). Samples with index 01410, 04338, 01030, 03438, 00144, 02427, 08621, 07449, and 06832 were unable to be tokenized by nltk's sentence tokenizer because of Unicode errors. They are excluded from all analysis.

After tokenization, I further edited the samples in the following way: First, I made two copies of the data. In one copy, which I call the Uncorrected Version, I only remove titles and salutations (e.g. "Dr Mr. Halifax"), and angle braces (e.g. << The Pearl >>). The removal of salutations does not change the efficacy of the parser, but ensures that the data being fed in is intended to be a complete sentence. While I didn't want to remove any punctuation from the uncorrected set, the presence of angle braces yielded an error in ACE, and thus had to be excluded.

The second copy of the data I call the Hand-Corrected Version. In this copy, I attempted to correct the punctuation issues noted by Xu, as well as some errors made by the tokenizer (such as splitting abbreviations). This process came in two parts. First , I ran a script to split appropriate punctuation. This adds space between utterances like "time.The" but maintains spacing in things like "1.80 meters tall." The script also removed leading numbering or bullets from sentences. The second phase of correction was true hand editing. I took what the script gave me and deleted punctuation and line returns that would create a parsing error (e.g. Mr.\nWang) when it was clear that that was the speakers' intent. I did not delete or change misused punctuation, such as periods used in the place of commas. I also deleted lines containing a single word, where the word showed no inflectional morphology. It is assumed that there are titles to writing prompts. For example, "matrix." was deleted, but "others." was retained. Any words which could be sentences in isolation were retained as well (e.g. 'Finally.'). Where the corrector failed to split words (e.g. At8:30), I inserted space ('At 8:30'). However I did not split misspellings ('iwant' instead of 'I want'). All hand-editing is optimistic, as the corrected version had 38,791 lines and I lacked the resources to fine-comb all the data.

### 3.2 Parser and Sub-Sampling

For this task, I chose to interact with the ERG via the ACE parser. This is a command line parser that takes a grammar as input, and combinatorialy produces corresponding syntactic parses and mrs's for test sentences. At test time, each of the large samples was divided into mini-batches of size 1000 for computation convenience. The random subsample of size 200 was collected from the hand-corrected sample, and was processed as a whole. While I would have preferred to use the capabilities of the [[incr tsdb()]] GUI on all the sets to better visualize details of the parse and semantics

and evaluate for correctness, it proved too computationally expensive for samples of this size.

### 3.3 Evaluation Metrics

For the purposes of evaluation, I propose the following heuristic: A sentence is said to be parsed correctly if the ERG can produce a reasonable syntactic structure, with a corresponding semantic representation (read: a semantic reading that exists).

In addition, I will evaluate errors along the following as classification designations:

- Is generally ungrammatical to a native speaker.
- Lexical items being used incorrectly.
- Morphological or conjugation problems. Misspellings.
- Tokenization error.
- Punctuation being used incorrectly.
- Unknown vocabulary.
- Incorrect use of function words.
- No errors obvious

For each of the 2 sentences in the sampled set that returns no parses, I assign it the most salient category from this list.

All files, including the tokenized text at various time steps, is available at www.github/com/SaraBlalockNg/LING-575.

## 4 Results

Table 1 gives the percentage of parses yielded by the ERG when testing the three sets of data outlined in Section 3. The hand-corrected set accounts for 38,730 lines, and the uncorrected set has 88,446, of which 39,446 were sampled for parsing. As stated before the Random Sample set is 200 lines from the hand-corrected set. Considering that the coverage for the ERG was attested at 83%, the results from the hand-corrected and random samples is inline with expected behavior of the ERG in general.

The uncorrected test set also performed surprisingly well, with coverage at 73.15%. Impressionistically, parsing failures were common when function words and lexemes were concatenated by punctuation, leaving gaps in both the syntactical and semantic combinatorics. In the next section, I break down the 34 errors thrown for the Random Sample.

Table 1: Coverage Across TECCL Data

| **Data Portion Name** | **% Sent. Parsing** |
|---|---|
| Hand-Corrected | 79.8270 |
| Uncorrected | 73.1538 |
| Random Sample | 83 |

## 5 Discussion

Table 2 divides the retrieved error sentences from the Random Sample according to the error types proposed in Section 3. A complete list of the sentences yielding error can be found in Appendix A.

Table 2: Error Classification

| **Error Type** | **# of Samples** | **% of Total** |
|---|---|---|
| Is ungrammatical to a native speaker. | 6 | 17.65 |
| Lexical items being used incorrectly. | 3 | 8.82 |
| Conjugation/Morphology error | 11 | 32.35 |
| Misspellings | 2 | 5.88 |
| Tokenization error. | 1 | 2.94 |
| Punctuation being used incorrectly. | 4 | 11.76 |
| Unknown vocabulary. | 1 | 2.94 |
| Incorrect use of function words. | 6 | 17.65 |
| None | 3 | 8.82 |

The types of errors rendered un-parseable in the random sub-sample illuminate the problems that a generic parser cannot handle. First, in just over 17% of cases, the sample sentence had so many compounding issues that no meaning could be derived by the scorer. In these cases, it is predictable that the grammar would not accommodate.

The highest frequency among errors were errors containing conjugation or morphological issues. We know that pedagogically, much classroom time in the EFL environment is spent learning tense and aspect. This is especially true in a Chinese to English environment, where the systems of tense and agreement are vastly different between the target and source languages.

Also high frequency in the non-parsing set were sentences with omitted essential function words. Language learners confuse words without strongly-salient semantic forms, and this bears out in the data: about 17% of non-parsing sentence were missing an essential function word.

As expected, punctuation and sympathetic tokenization issued persisted even in the hand-corrected data. In the tokenization error, a name was split at the title, leaving the surname (in small

case) in isolation. This didn't produce an error on its own, but impacts the readability of the sentence from a native perspective. Punctuation was also an issue; however, in the cases represented in the sample, the punctuation issues were caused by speaker misuse altogether, and not by accompanying kerning issues.

The cases which are cause for the most concern are the three samples that contained no strong errors, but that still received no parses. While it must be acquiesced that an error rate of 1.5% over the sample for genuine grammatical sentences is very high-performing, they still have fairly high frequency amongst the frequency of errors in general. The grammatical failure to parse these sentences shows a systematic failure within the grammar, either in lexicon or rules, that will prevent any native or learning utterances of similar form to be rejected by the grammar.

Because of the size of the samples, it is unfeasible to thoroughly question what over-generation of parses is incurred in the set. However in this case, over-generation may not be a bad thing. If the semantic representations in parsed ungrammatical sentences is reasonable, then it speaks to the intelligibility of non-fluent learners. Speakers of all levels experience at least the occasional successful conversation, and we may say that an over-generating grammar is the computational analogue of a speaker's ability to ignore errors.

It may be tempting to say that for all cases except the 8.82% where there was no genuine error, we would not want the grammar to produce a parse. However, if employed in a commercial product, we would hope that the system would be accommodating to learners and the kinds of errors they make. This is beneficial for all users. For example, even the most literat of native speakers make okaisional spelling error, or use non-prescribed morphology. Since we know that native speakers make and can parse these types of errors without difficultly, we should expect that a grammar based on speaker knowledge should also have these capabilities.

### 5.1 Implications for ERG

The high performance of the ERG in parsing learner sentences is a testament to the power of the ERG's parsing power.

The coverage is particularly impressive given that the TECCL data does contain beginner data, and writing samples where students have to write about complicated topics with limited language knowledge; in this situation, one would expect that grammar and lexical errors will have high frequency. However, the ERG's capability to accommodate OOVs, maintain many different syntactic environments, and provide sympathetic semantic mapping allows it to handle well what might otherwise be considered ungrammatical by other frameworks.

However, the success of the system may not be purely due to superior design. As noted in Xu (2015), parsers can perform well on learner data by the merit that early learners.

One note for system improvement is the issue previously discussed on punctuation. There are merits apart from cross-linguistic writing in including punctuation in language processing. Punctuation serves as the written correlate of prosody (Steinhauer, 2003), and may be able to inform parsing as well. For example, quotations around titles are a cue that the words contained therein serve as a single syntactic and semantic unit. If the grammar could handle the punctuation issues in learner data, it may bring the system closer to using the cues more broadly to the benefit of its precision.

### 5.2 Implications for TECCL

TECCL exists as a record of Chinese EFL students in various regions of China and from varied ages and abilities. What can be said of the findings of its interaction with the ERG, that could inform its own existence?

As previously stated, the TECCL exists in three forms: raw text, tagged text, and parsed text. Tagging for the data set was completed using the CLAWS POS tagger for English (Garside, 1987). Because the syntactic component of the parse gives provides label information (albeit in a slightly different format), it can be used to either validate or dispute the parses given by the elder parser, without the need for laborious hand-annotation.

The parsed version of the TECCL corpus has been completed by its creators via the Stanford parser. It stands to question why then, a parse from the ERG is necessary at all. The information has already been seen by a highly developed parser, and the data honestly looks better than what the ERG has provided (read: there are actually

parses for the reasonable-yet-not-prescriptively-grammatical sentences). What is the merit to the system of interacting with another parser?

Xu (2015) notes that the Stanford parses have never been validated for the data set. It is arguable, then, that future comparison of the outputs from each parser could serve as a proxy to manual validation in this case, as well. If both parsers, developed with different machinery at different institutions, can return similar parses of the same data, it could boost the confidence of those concordant parses. While it is obvious that this isn't a complete replacement for old-fashioned validation, given the likelihood of validation comparison is a reasonable substitution.

### 5.3 Potential Applications

It seems curious to think that there could be application outside of insular improvement for such an analysis. Indeed, no English as a Foreign Language instructors I know would be willing to download a corpus at set it at their students' data. Many of the errors that prohibited the parser from outputting a result are probably better-suited to a grammatical explanation rather than a distributional one. However, some of the error types found in this study could inform the focus of troubleshooting writing in the EFL classroom. We assume that grammar-based parsers are a proxy or rough estimator of the kinds of writing that a native speaker would be able to naturally parse. It stands to reason then, that the grammatical features that make an utterance un-parseable to a machine parser should be features on which language learners should apply most focus; the grammatical errors that render a parse can respectively given less effort to learn.

For example, EFL curricula often emphasize the importance of learning tense and agreement. Learners are told that an error in verb tense can change the interpretability of a sentence. However, this doesn't seem to be an issue as far as the ERG is concerned. In an utterance like "I will walk home yesterday," the ERG correctly renders that the tense of the VP is future, and it assigns a temporal value of YESTERDAY in the semantics. While the two features are practically incompatible, the system does not have a problem. The intuition of the native speaker also shows relatively little confusion: one may assume that the speaker walked home yesterday, or that they will walk home in the future. There are two valid interpretations. Compare this with the TECCL sentence, "Therefore, the most striking conclusion is obvious that network keep use the unreal-name system will keep the freedom." This sentence, unlike the former, does not yield a parse with the ERG. For this utterance, there are too many possible interpretations to allow the native speaker any confidence: it could be that the speaker is saying the conclusion other people found striking is actually obvious ("that network keep use the unreal-name system will keep the freedom"), it could be that the network will continue to use the "unreal-name system," thereby maintaining it's algorithmic freedom. From the greater context of the writing sample, what I believe to be the most realistic interpretation is that maintaining online anonymity (the "unreal-name system") enables online users to speak honestly online. This reading requires context to know what the novel vocabulary usages (e.g. *unreal-name system*, *freedom*, and *network*) mean, and what can be semantically ignored from the contradiction of an obvious, striking conclusion.

While it must be acquiesced that this sentence suffers from the compounding issues of grammatical errors, it can also be said that many of the issues in isolation are not subjects receiving devoted lessons in the EFL classroom. The phrase *unreal-name system* is symptomatic of the speakers 1) ignorance of appropriate vocabulary, and 2) inability to successfully concatenate novel noun compounds and morphological forms.

## 6 Conclusion

In general, I am heartened by the results presented in this paper. The TECCL corpus was an easy-to-use resource with very good documentation and demographic detail that I am looking forward to re-using in the future. The off-the-shelf parser performed well on learner data, and the errors seemed consistent with what would be considered difficult to interpret by a native speaker.

This work also highlights some of the potential pitfalls in working with standard programs and secondary source detail. First, the issue with punctuation shows the utility of of providing prose descriptions from the collector about data resources; without notes from the collector, I would have been wading in un-meaningful errors. Moving forward, I am looking forward to also composing

a better work flow architecture for error analysis with the selected grammar; I would like to expand the error set and include a more fine-grained categorization for the grammatical errors.

## 7 Acknowledgments

The word contained in this paper would not be possible without the helpful input of many people. I would like to extend my thanks to my peer editor, Amina Venton for her helpful grammar notes, and vital comments about the the methodology and background sections. Your insights have been extremely helpful.

I'd like to thank Xu Jiajin at the National Research Centre for Foreign Language Education Beijing Foreign Studies University, who created and maintains the information about the TECCL corpus on the corpus website. I am especially appreciative of Dr. Xu's helpful hints about the tendencies of Chinese ELLs; I certainly would not have understood many of the punctuation issues had I not had access to Dr. Xu's writing on the matter.

In addition, I would like to thank Dr. Emily M. Bender, for maintaining a spirit of public information in keeping her course resources with respect to the ERG public and dense. While I unfortunately ran into a slew of problems processing the data, it at least evoked fond (scare-quotes) memories of LING 567.

Finally, I'd like to thank the Ryan and the class at large for creating an inquisitive atmosphere in which we are allowed to explore the minutia of ethics (and error analysis). I had not previously considered an error analysis in learner data to be a personal interest. However, this course has allowed me the opportunity to see who NLP is failing, and why it is important to dig through the data and see why. Thank you.

## References

Aston, G., & Burnard, L. (1998). The BNC handbook: exploring the British National Corpus with SARA. Capstone.

Baldwin, T., Bender, E. M., Flickinger, D., Kim, A., & Oepen, S. (2004, May). Road-testing the English Resource Grammar Over the British National Corpus. In LREC.

Copestake, A. & Flickinger, D. (2000). An open-source grammar development environment and broad-coverage English grammar using HPSG In Proceedings of the Second conference on Language Resources and Evaluation (LREC-2000), Athens, Greece.

Dickinson, M., & Ragheb, M. (2009). Dependency annotation for learner corpora. In Proceedings of the Eighth Workshop on Treebanks and Linguistic Theories (TLT-8) (pp. 59-70).

Garside, R. (1987). The CLAWS Word-tagging System. In: R. Garside, G. Leech and G. Sampson (eds), The Computational Analysis of English: A Corpus-based Approach. London: Longman.

Granger, S. (2002). A birds-eye view of learner corpus research. Computer learner corpora, second language acquisition and foreign language teaching, 6, 3-33.

Hawkins, J. A., & Buttery, P. (2010). Criterial features in learner corpora: Theory and illustrations. English Profile Journal, 1.

Huang, G., Tan, M., Huang, S., Mo, R., & Zhou, Y. (2017, December). A discourse coherence model for analyzing Chinese students' essay. In Progress in Informatics and Computing (PIC), 2017 International Conference on Progress in Informatics and Computing (PIC) (pp. 430-434). IEEE.

LI, X., & LIU, J. (2017). A Corpus-based Contrastive Study on the Acquisition of Synonyms of Chinese EFL Learners. Journal of Literature and Art Studies, 7(7), 925-934.

Song, L., & Wolter, B. (2017). Effects of L1 Transfer on L2 Learners VN Collocational Use: A Corpus-based Study from Semantic Preference and Semantic Prosody Perspective.

Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. Brain and language, 86(1), 142-164.

Wahlster, W. (1993). Verbmobil. In Grundlagen und anwendungen der knstlichen intelligenz (pp. 393-402). Springer, Berlin, Heidelberg.

Xu, Jiajin. (2015). Ten-thousand English Compositions of Chinese Learners (the TECCL corpus) (Version 1.1). Retrieved at http://corpus.bfsu.edu.cn/content/teccl-corpus.

## Appendix A

| Sentence | Error Type |
|---|---|
| The advantage of living in the suburban is obvious: living in the suburban is suitable for fear of noisy people. | lexical |
| Our life also requires us to be wise, lasting study makes our life wonderful and our mind improved, everyone admires wisdom, but not everyone can achieve that level, because learning always be a tough things, it requires people to hold on. | ungrammatical |
| Thirdly, they don't know what to do when they out. | function |
| At last, they should kind and patient. | none |
| besides, she not only remembering the details of happy old days but also understanding better the strongly feelings of pains and joys. | morphology/conjugation |
| Though it's hard and tired, We need to get up early in the morning and go back to the bed late in the night, our parents keep nagging us and we are always worried about our grade, but everyday is full and meaningful. | punctuation |
| The houses problem is also secious. | oov |
| If you need to go there by train or plane, you should go to ticket office to by the ticket or pay for it on the internet. | misspelling |
| I be good at work and get along with my workmates , we having fun every day. | punctuation |
| We can use mobile phone to do a lot of things online. | morphology |
| One can have a lot of dream. | morphology |
| Even though my hometown is smal , it has a quantity of earth where grows all kinds of crops. | misspelling/function |
| The TV Program become more and more colourful. | conjugation |
| wang teach me computer. | tokenization/conjugation |
| He is a man which one of the leads. | umgrammatical |
| You know, the Double Ninth Festival is Chinese traditional festival. | function |
| For esample, with the improvement of people's livng stangard, the pursuit of people to higher and higher, more and more people like pop music. | ungrammatical |
| He puts away the things on half past eight in afternoon . | function |
| More job opening is good for city too, which lead more person to city. | morphology/conjugation |
| They usually hit me use their shoes. | lexical |
| Manypeople who like reading books use e-book. | punctuation (spacing) |
| I'm more easily understand the teacher's meaning and Mmake me in class has a high efficiency filter so as to achieve better results. | ungrammatical |
| The country which are next to the Switzerland are Austria, France, Italy and Germany. | morphology |
| First, it is easily to buy and easily to get. | morphology |
| when I had to come home, she send me a dictionary to encourage me to work hard. | conjugation |
| I'm sure you unforgettable these foods. | lexical |
| It is adjacent to Bohai sea. | none |
| so each other deeply, you will find society each piece of people things are so beautiful, light, further inspired deep inside you touched, exerting their own expertise to go and work for the society, the country and the world service; after all, I door of all long, if people no longer foreigners. | ungrammatical |
| On the other hand, there are some people tihnk the romance can inspire two students to work hard, and two people can talk with each other mind to mind, they will have a good time on Campus. | function |
| Besause of haven't money . | ungrammatical |
| But we are dare to challenge, rich creative idea and full of youthful spirit. | function/conjugation |
| He said, Histories make men wise; poets witty; the mathenatics subtle; natural philosophy deep; moral grave; logic and rhetoric able to contend. | punctuation |
| Learn that life abound in bumps down the road. | none |
| We should help they, because all of us are the member of earth. | conjugation |