Errors Versus Mistakes: A False Dichotomy?

SIMON PHILIP BOTLEY
Academy of Language Studies, Universiti Teknologi MARA Sarawak, Malaysia.

ABSTRACT

In this paper, the dichotomy between errors and mistakes in applied linguistics will be discussed, with a view to arriving at a clearer understanding of how to describe learner language from an empirical viewpoint. Errors are usually defined as systematic deviations from the rules of a target language. They may occur because a learner does not know a rule, such as Subject-Verb Agreement in English. Mistakes, on the other hand, are seen as unintentional, accidental slips resulting from simple laziness or forgetting. Distinguishing between errors and mistakes has always been fraught with problems. Using examples from a corpus of written learner English constructed in Malaysia, this paper argues that it is not empirically feasible, or even desirable, to maintain a dichotomy between errors and mistakes, especially from a corpus-based empirical perspective on learner language. It is argued that empirical analysis of learner language cannot practically access the knowledge of a learner to determine if an error or a mistake has been produced. Furthermore, many existing objective tests of 'errorness' or 'mistakeness' are not reliable. Instead, it is proposed that the phenomena currently known as 'errors' or 'mistakes' should be termed 'interlanguage features', a term which is relatively value-free.

KEYWORDS: **Corpus, error, mistake, interlanguage feature, learner English**

Introduction

For those who set out to describe and analyse language in a rigorous and scientific manner, linguistics terminology provides a definitional minefield. Linguistic concepts are notorious for the difficulty with which they can be defined and delineated clearly. This is especially true in the field of Applied Linguistics, where two terms are commonly used, but are extremely difficult to define and distinguish from one another. These terms are 'error' and 'mistake'.

Errors can be defined as systematic deviations from the rules of a target language, as they are believed to occur because a learner does not know a given rule or feature, such as Subject-Verb Agreement or Noun Plurality in English. Errors might arise from little or no input on a given language feature during language lessons. Mistakes, on the other hand, are usually seen as unintentional, accidental slips resulting from simple laziness or forgetting, or insufficiently internalised rules. If a learner makes a mistake, he or she will normally be capable of correcting it, because of awareness of the rule that has been broken. On the other hand, errors cannot be self-corrected by the learner (James, 1998) because the rule being broken is not part of the learner's knowledge of the target language.

Despite this apparently clear dichotomy between 'error' and 'mistake', it is argued in this paper that errors and mistakes are not only challenging to distinguish, but they are not always easily defined. This is a problem for applied linguists, for two main reasons. Firstly, the difficulty in clearly defining and delineating errors and mistakes provides a challenge for those who aim to investigate both Second Language Acquisition (SLA), and how effectively language educators teach a new language. This is because understanding what goes wrong when learners communicate in a target language reflects not only their nascent knowledge of the language's rules, but also offers direct evidence of how well they have learned, and how well they have been taught.

The second reason why the error/mistake dichotomy is problematic is that it provides a distraction for those who, like the author of this paper, aim to rigorously describe the language performance of learners using empirical methodologies such as Corpus Linguistics. Two reasons for this are that 'error' and 'mistake' are commonly conflated into the term 'error' (e.g. Error Analysis), and that using terms like 'error' and 'mistake' is effectively an evaluative act that negates much of the objective character from the practice of describing learner language.

Therefore, this paper argues that it is time to abandon the error/mistake dichotomy, or at least replace it with something clearer and more scientific which can be more helpful and less problematic in describing the language produced by learners. This paper will discuss the challenges of using such a contentious pair of terms to empirically investigate learner language using large computerised text collections (or 'corpora'). However, before moving on to this discussion, it is necessary to briefly survey the issues involved in distinguishing between mistakes and errors.

Errors, mistakes and error analysis

Defining 'error'

Let us begin the discussion by briefly stating what it is that terms like 'error' and 'mistake' are supposed to describe. It was James (1998, p. 1) who gave a somewhat preliminary definition of an error as "an unsuccessful bit of language". We can assume that he also used this definition to cover mistakes. It has to be said that most writers on this subject, perhaps bound by the limitations of the English language, will tend to use the term 'error' to describe any unsuccessful use of language, whether or not it is in fact an error or a mistake. For the sake of brevity, this paper will not take issue

with this strategy, and indeed this writer has been guilty of using 'error' in an over-inclusive manner on many occasions.

In Chapter 3 of his book, James (1998) explores a number of reasons why defining 'error' in a clear and unambiguous manner is challenging. Firstly, it is not clear whether *grammaticality* or *acceptability* should be used as the defining criterion for an error. As reiterated by Ellis and Burkhuizen (2005, p. 56), errors defined by the grammaticality criterion constitute what Corder (1971) states as breaches of "the rules of the code". This begs the question of which code is being referred to – effectively which variety of the language. This issue will be discussed in depth later.

James points out that errors defined by grammaticality need to be divided into *overt* and *covert* errors, where overt errors are detectable in the sentence in which they occur, while covert errors require analysis of larger stretches of discourse in order to be detected. On the other hand, the *acceptability* criterion for defining errors is, as James states, more subjective and unreliable than *grammaticality*, and will often require consideration of the context of utterance to determine whether or not a given expression is acceptable.

One more definition of 'error' worth mentioning was provided by Lennon (1991, p. 182), who said that an error is "a linguistic form or combination of forms which, in the same context and under similar conditions of production, would, in all likelihood, not be produced by the speakers' native speaker counterparts". Lennon's definition, while not perfect and perhaps too inclusive, does in some ways side-step the problems with grammaticality and acceptability and provides a starting point for rigorous analysis. Essentially, then, an error or mistake is unsuccessful language use that would not be found in the production of a native speaker.

In order to come to terms with a definition of error and mistake, it is necessary to expand the discussion beyond the Second Language Acquisition (SLA) perspective and take on board a more sociological view of the issues. This can be seen in the recent English as a Lingua Franca (ELF) paradigm associated with such scholars as Kirkpatrick (2011), Seidlhofer (2005) and Mauranen (2003). From an ELF perspective, English is widely spoken in communities where speakers may not share a common language, such as South East Asia (Kirkpatrick, 2011) and the European Union. Rather than needing to look to British and American native speakers as custodians of the rules of the code, English speakers and learners from an ELF perspective can now draw their linguistic models from localised examples of successful ELF speakers, rather than the standardised norms used and promoted by native speakers.

The corollary of this view is that those 'successful ELF speakers' may now come from countries such as Malaysia, Singapore or Thailand rather than the traditional countries in the Inner Circle of the World English model proposed by Kachru (1992). As a result, learners no longer need to follow linguistic norms set by the 'native speakers'. This means that the concept of error or mistake needs to be re-thought if a more inclusive, progressive, sociologically-plausible definition of the learner is to be arrived at. Therefore, if an error or a mistake is a breach of the code or violates some linguistic norm, these codes and norms may no longer be those imposed by or drawn from Standard British English, American English or other prestigious genres such as Academic English or Business English. They may in fact be found in the linguistic and discursive practices of non-native speakers of English as a Lingua Franca.

The importance of 'errors'

Errors, or unsuccessful bits of language, are important because they provide an indicator of problems that a learner may have in learning a language, in terms of grammar patterns or pronunciation features. From the traditional cognitively-oriented Second Language Acquisition (SLA) perspective, Corder (1967) argued that errors serve three important functions: they inform teachers about what has been learned and what has not been learned, they help researchers to understand how languages are learned, and finally they help learners to discover the rules of the language they are trying to master, by means of feedback.

In particular, errors can provide us with detailed information about a learner's Interlanguage (Selinker, 1972). The Interlanguage is the evolving and transitional 'mental grammar' (Ellis & Burkhuizen, 2005, p. 54), which is built up by a learner at different stages of the process of learning a new language. The Interlanguage can be said to consist of grammatical knowledge, akin to Chomsky's Competence, as well as learning strategies. It is essentially a way-point between total ignorance of a target language, and total mastery of it.

From within the current ELF paradigm, errors and mistakes are viewed differently. According to Kirkpatrick (2011), the goals of language learners in an ELF setting are vastly different from those of the older SLA paradigm. Under the older paradigm, languages such as English are acquired so that learners can attain native speaker proficiency, where linguistic best practice is derived from culturally prestigious native models. However, from the ELF point of view, the goal of learning languages like English is to use them successfully as lingua francas in particular places in the world.

This means that rather than thinking of 'errors' or 'mistakes' as deviating from a common standard model, it is necessary to see them in terms of non-standard forms, which may be derived from localised varieties of English (such as Malaysian English), and preferred by ELF speakers over standard forms. Examples of these would be such standard bugbears of Malaysian EFL practitioners as missing or misused articles, non-standard plurals or auxiliary constructions such as 'also can'. Effectively, features of language which may have been considered errors or mistakes in the past will instead gain a wider acceptance, as long as they do not negatively affect clear communication.

'Error' Analysis

Whatever paradigm one subscribes to – SLA or ELF – if one aims to investigate the Interlanguage of learners systematically, it is necessary to carry out Error Analysis. Error Analysis, according to Ellis and Burkhuizen (2005, p. 51), "consists of a set of procedures for identifying, describing and explaining learner errors." James (1998, p. 1) also defines Error Analysis as "the process of determining the incidence, nature, causes and consequences of unsuccessful language". Corder (1974) described five stages in conducting an Error Analysis, namely Collection, Identification, Description, Explanation and Evaluation. In this paper, we will largely be focusing on the second, third and fourth of these stages.

Error Analysis has traditionally been a 'pen-and-paper' enterprise. It has involved the close scrutiny of relatively small samples of productions collected from learners (see Ellis & Burkhuizen, 2005, p. 56-69, as well as James, 1998 for detailed information on how to carry out Error Analysis). In recent years, computer technology has given Error Analysis a new lease of life. Granger (2002, p.11-14) identified the emerging methodology of Computer-Aided Error Analysis (CEA). CEA enables errors to be studied on a much larger scale than before, using data collected in a Learner Corpus. A Learner

Corpus is a computerized text database containing written or spoken productions by language learners, including essays, oral examinations or story tasks.

Many Learner Corpora exist, such as the International Corpus of Learner English (ICLE) (Granger, Hung & Petch-Tyson, 2002) based in Europe, and the EMAS (English of Malaysian School Students) corpus (Abdul Samad et al., 2002) in Malaysia. From the ELF perspective, some specialised ELF corpora already exist, such as the English as a Lingua Franca in Academic Settings (ELFA) corpus (Mauranen, 2003) and the Vienna-Oxford International Corpus of English (VOICE) (Seidlhofer, 2004).

Using such corpora, it is possible to carry out large CEA studies of learner Interlanguage as well as Contrastive Interlanguage Analysis (CIA) studies (Granger, 2002, p.11-14) where different populations of learners can be compared, such as learners from different first-language backgrounds. Thanks to rapid advancement in technology, errors are now amenable to analysis and investigation on a scale, and at a level of complexity which was never before possible.

Errors versus mistakes

It was Corder (1974) who reminded us that Errors and Mistakes are different, even though these two terms are often used as if they are synonymous. Errors, on the one hand, reflect deviations on the part of the learner from the grammatical rules of a native speaker, and reflect the incomplete 'Competence' (Chomsky, 1965) of the learner. Errors are systematic and indicative of incorrect knowledge of the target language. Mistakes, on the other hand, are reflective of Chomsky's 'Performance' – un-systematic slips where a learner knows the correct grammatical system but for some reason fails to use it properly.

Because Errors tell us more about the learner's internalised grammatical knowledge of the target language, Corder (1974) argued that Error Analysis should only concentrate on Errors, and should ignore Mistakes. Furthermore, in order to identify, collect and describe errors we need to know if they are in fact Errors or Mistakes. However, the crucial question is: how do we tell whether a given example of 'unsuccessful language' is in practice an Error or a Mistake? Here is a sentence from a Learner Corpus, with two Errors/Mistakes highlighted:

This **become** a serious **issues**...

Here, one could say that there are two grammatical rules being broken, or deviated from. Firstly, 'become' reflects a problem with English verb tenses, while 'issues' represents a difficulty with the English Number system.

A learner corpus might contain hundreds or indeed thousands of examples like this, all showing examples of James' 'unsuccessful language', and all representing observations made in the field from the Performance of a set of learners. Are they Errors or are they Mistakes? The argument of this paper is that it is not clear how in practice one can tell whether any given deviation is an Error or a Mistake. Despite this, a number of strategies have been put forward, which are discussed below (Ellis & Burkhuizen, 2005).

The first strategy might involve simply asking the learners if they are aware of the rule they have deviated from, and if they can correct the error themselves. Presumably, this would involve taking an essay written by a learner and going through each issue one by one. If the learner can self-correct the deviant form, then it can be said to be a Mistake (Ellis & Burkhuizen, 2005, p. 64). The problem with

this strategy, apart from the practical challenge of doing this for a class of thirty or more learners, is that the learners are rarely available for consultation, especially with a learner corpus. The only time the learners are present to interrogate is usually during class time when, hopefully, they are too busy with other activities. A learner corpus might contain thousands of examples of putative errors or mistakes, but it is not currently clear how learner awareness of each error or mistake can be captured and encoded in a corpus.

Furthermore, it is not clear whether problems that arise from asking the learner arise from the learners themselves, or with weaknesses of the whole empirical approach underpinning corpus-based linguistics. It could be the case that useful data cannot easily be obtained by asking the learners because they lack the vocabulary or the linguistic competence to provide any useful answers, or, it could be that corpus linguistics lacks the capacity to capture the knowledge of learners. Corpus linguistics is admittedly a descriptivist methodology that tends to avoid taking human intuitions into account, instead relying on collecting large amounts of data which could be used to inform existing intuitions.

Another strategy for distinguishing an error from a mistake could be to ask teachers which aspects of the target language *should* have been learned in a given class. In theory, deviation from a given rule or structure in the target language would reflect a learner's not being exposed to that topic in class. However, just because a teacher thinks a given form should have been taught does not mean that it has actually been internalised by the learner, making it even more difficult to tell an Error from a Mistake in this way.

A third strategy might be to look in the learner's essays to see if the student varies or alternates systematically between the wrong form and the right one (Ellis & Burkhuizen, 2005, p. 64). This can be done during Error Analysis, without any recourse to the faulty intuitions of learners or educators. If a given form, such as simple past tense, is incorrectly used once or twice in an essay, but all other instances of simple past tense in the essay are correct, then the deviant example is a mistake, because it is an isolated example.

However, as pointed out by Ellis and Burkhuizen (2005, p. 64), this argument hinges on what is meant by a 'form' in this case. If the form is the grammatical feature in general, and there is one instance of this form being incorrectly used, then this instance is a mistake. However, if 'form' is meant to refer to a specific instance of an incorrect form, such as the past tense of a particular verb, the situation is not clear because there may not be any other correct examples of that form present in the essay.

Finally, it might be possible to identify an error if there is a significantly frequent number of examples in a given essay, or in a given corpus sample. Therefore, the most frequent deviations are error candidates, and the least frequent ones are mistake candidates. Still, even in these circumstances, it is not absolutely clear if these frequent examples are really errors or mistakes, in accordance with the reasons given earlier.

Looking inside a learner corpus

In order to investigate the Error/Mistake dichotomy from an empirical viewpoint, this paper makes use of a sample of student essays collected in a learner corpus here in Malaysia. The corpus in question is an experimental learner corpus collected in Sarawak and Sabah, called CALES (Corpus Archive of Learner English in Sabah/Sarawak) (Botley & Metom, 2011). The CALES corpus contains almost 500,000 words of argumentative essay writing produced by Malaysian undergraduates in a number of higher education institutions in East Malaysia.

A learner corpus was used because of the advantages that learner corpora bring to the empirical analysis of learner language. According to McEnery and Wilson (2001) and Granger (2002), learner corpora allow for large scale, high speed studies of language in use, and also allow for the collection of sizable and representative samples of language productions that can be used in further research. A corpus, because it is stored on a computer, allows for rapid search and retrieval of information, especially where linguistic information has been added to the corpus in the form of annotation, such as error tagging (Botley & Metom 2011). If a corpus is enriched with extra information, it is possible to gain detailed statistical data such as frequency of occurrence of certain words or indeed errors/mistakes.

A subcorpus of about 32,000 words (160 essays) had been manually error-tagged using the UCL Error-tagging taxonomy (Dagneaux, Denness & Granger, 1998) as part of an earlier study (Botley & Metom, 2011). Once this annotated subcorpus was established, frequency statistics for the distribution of a range of error categories identified were derived using the ANTCONC concordance software (Anthony, 2013). ANTCONC enables the user to extract keywords from context lists and word frequency lists from a corpus, as well as other statistical reports.

The following table shows the most frequent error categories in the sample as extracted using ANTCONC. Due to the constraint of space, the table contains only the most frequent (n>=100) categories. The total number of errors tagged was 12,830.

Table 1. Error Frequencies in a Learner Corpus

Error Tag	**N**	**%**
GNN	1806	14.08
LS	1787	13.93
GA	1329	10.36
WM	1293	10.07
FS	1209	9.42
GVT	1083	8.44
WR	998	7.78
GP	724	5.64
GVN	497	3.87
GWC	397	3.09
S	363	2.83
SU	208	1.62
GVM	111	0.86
WO	106	0.83
OVERALL TOTAL	12,830	1

The table above shows that the most frequent categories of error/mistake in this sub-corpus include GNN (noun number), LS (incorrect or inappropriate lexical choice), GA (articles), WM (missing word), FS (spelling), GVT (verb tense), WR (redundant word), GP (Pronouns), GVN (Subject-Verb Number Agreement), GWC (wrong word class), S (Style), SU (Style Unclear), GVM (Verb Morphology) and WO (Word Order).

In order to establish whether any of these categories represent examples of errors or mistakes, it is necessary to address the following research questions:

1. For a given high-frequency error category, do students systematically alternate between correct or incorrect forms, or do they get them wrong all the time?
2. Is there a significantly large number of examples of a given category in a given essay or a large number of essays?
3. Does the overall distribution of a category throughout all the essays suggest that all students are having trouble getting this form right?

For the sake of space, we will only focus on the most frequent error category from Table 1, GNN (noun number), which comprises just over 14% of the total errors identified in the sub-corpus. Firstly, if we use ANTCONC to sort all examples of GNN according to the essays they occurred in, we can see clearly if the learners systematically alternate between correct forms or not, thereby addressing question 1 above.

For reasons of brevity, three error-tagged files were selected to discuss the issue of whether or not the GNN errors contained within were in fact errors or mistakes. Table 2 below shows how many nouns were tagged GNN in these files (meaning they were wrong) along with the number of times the student writers produced correctly pluralised nouns overall.

Table 2. Three Selected Essays with Correct and Incorrectly Pluralised Nouns

Essay	Incorrect Nouns (GNN)	Correct Nouns
1	15	51
2	36	81
3	3	48

On the face of it, Table 2 tells us that for the three essays shown, the learners seemed to have much less trouble with the English noun plural, because they effectively got it right far more than they got it wrong, even where, as with Essay 2, there were quite a lot of cases where they got it wrong.

This would provide evidence for the argument that systematic variation between the correct and incorrect form shows that the incorrect one is a mistake – the learner knows the rule, but sometimes gets it wrong for some reason, and gets it right more times than not. This might be the case in Essay 3, where there were only three incorrect examples.

However, as Ellis and Burkhuizen (2005) point out, things are not as simple as that, and the data analysed so far seem to support the importance of what is meant by 'form'. For instance, the three Essays 1, 2 and 3 include many examples where particular nouns are incorrectly pluralised or not

pluralised more than once in the same essay, which suggests an error – perhaps the writer did not know the word or how to pluralise it. Also, there are plenty of examples where particular noun plurals are wrong only once in the essay, which makes it unclear if the learner knew them or not.

For example, Table 3 shows those words in all three essays where the learner has sometimes been correct and sometimes incorrect.

Table 3. Correct and Incorrect Noun Uses in Three Essays

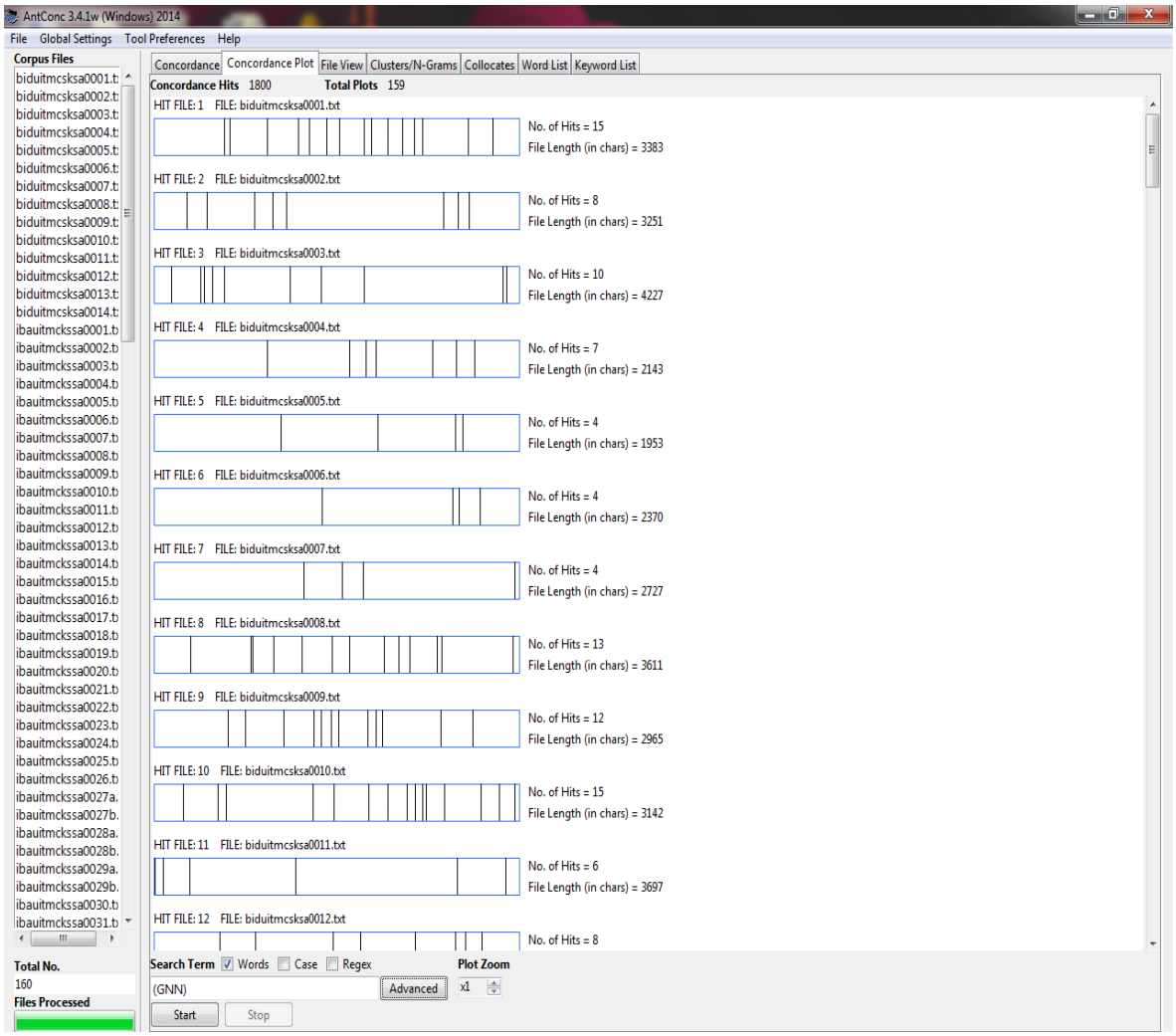
Essay	Word	No. Times Correct	No. Times Incorrect
1	Evil	1	2
1	Effect	1	1
1	(human) being	0	3
2	Life	1	1
3	Course	1	3
1	Salary	1	1
1	University	3	2
1	Job	1	2
1	Problem	5	3

This table shows that there is some evidence of variation between correct and incorrect uses of the plural rule for English Nouns in this data. However, it is not clear whether the variation shown here is systematic, or even significant enough to present clear examples of errors. There is not enough data to make a definitive statement. Even the example of 'problem', where the learner has correctly pluralised the word more times than not, does not give a clear answer. This raises the question whether the three incorrect uses of this word are mistakes just because the learner has got it right more often. The only strong claim that can be made at this stage is that the learners have got their English plurals wrong sometimes. Any evidence for why they did so for specific nouns is inconclusive.

Next, let us consider the second research question, which seeks to determine whether there is a significantly large number of examples of the GNN category in a large number of essays. As we can see in Table 1, the GNN category comprises the largest category of incorrect uses of language found in the 32,000 words of data under analysis. There were 1,806 examples of GNN errors, making 14% of all examples tagged. Therefore, the GNN category is frequent, suggesting that the 160 learners who contributed their essays to this study may all have problems with the English plural. There is clearly some issue with this aspect of the Interlanguage of this group of learners.

It needs to be established how widely are the GNN errors distributed throughout the subcorpus. This can be addressed once again using the ANTCONC software. ANTCONC allows us to see how the GNN cases are distributed among different essays in the subcorpus. This will show if all 160 learners are having problems with the English plural, or just a few. Figure 1 shows a fragment of a list of distribution plots for some of the essays included in the subcorpus under analysis. The horizontal columns represent essays, and the vertical lines within represent individual GNN cases.

Figure 1. Distribution Plot for GNN Cases (partial)



Although this only shows a partial picture, due to lack of space, it can be seen that GNN cases are scattered widely throughout this portion of the corpus. There are no essays where a single GNN case does not occur, and on average, there were 11.29 GNN cases per essay, with 56.44 GNN cases per 1,000 words of the whole 32,000 word subcorpus. Whether or not these cases are all errors or mistakes, there is clearly an issue with the Interlanguage of this group of students, as they are getting English plurals wrong most of the time, even though they are getting some of them right, and clearly, some students are producing more GNN cases than others.

Conclusion: errors or mistakes?

Some conclusions can be drawn from this small study of errors and mistakes. Firstly, we can say that unsuccessful language, or errors or mistakes, are highly frequent in the corpus sample analysed. This tells us that the students whose essays made up the sample have clear problems with various aspects of English. In this way, Research Question 2 has been answered in the positive.

However, despite the frequency of such deviant productions, it remains unclear from the data if the productions represent errors or mistakes, even if a given learner has only got a particular form wrong

and never right (e.g. *being*) or if the learners have produced more correct versions of a particular form than incorrect ones. This makes it difficult to provide conclusive answers to Research Questions 1 and 3.

One of the main reasons why we cannot clearly distinguish between errors and mistakes in the subcorpus is that there is insufficient data on specific forms in individual essays, though there are plenty of examples of deviance over the whole subcorpus. Furthermore, we do not have access to the knowledge of the learners nor their lecturers; knowledge which would in itself most likely be unreliable, subjective and therefore unhelpful.

In order to answer our research questions fully using corpus data, there clearly needs to be a great deal more annotated essays, to enable statistically valid claims to be made. The challenge in producing such data is that the process of annotating a learner corpus for language errors is highly time-consuming and labour intensive. However, given enough effort and time, it is hoped that future projects may enable us to throw more light on the error/mistake distinction, if indeed it can still be empirically supported.

In the meantime, this paper has to conclude by arguing that the error/mistake dichotomy as it has been stated in the SLA literature appears to be false as an empirical construct. It would appear to be too challenging at present to arrive at hard and fast practical methods to distinguish errors from mistakes in written productions by EFL learners. At best, it may be possible to deal with this need to distinguish errors and mistakes by side-stepping the problem somewhat. One way of doing this would be to change the focus of Error Analysis so that it does not just focus on what is wrong in a learner's productions, but also considers what the learner has got right.

Such a focus would allow us to move away from the current somewhat restrictive practice of identifying and distinguishing 'errors' and 'mistakes'. Once we are free from this restriction, we can move toward to take on board the totality of features of the learner's Interlanguage, whether they be grammatical, lexical or phonological. This would, for example, reveal variation between correct and incorrect productions not merely as evidence of something that has gone wrong, but evidence that a learner is avoiding difficult forms by using them rarely.

It is the contention of this paper, then, that the analysis of learner language productions should aim to avoid unreliable, value-laden, negative and unclear terms like 'error' and 'mistake'. Instead, we should be adopting terminology that avoids values and which allows for a clearer, more scientific analysis of 'unsuccessful language' in the productions of learners. In this way, Error Analysis would be following the philosophy of corpus-based linguistics as laid down by Sinclair (1991, 2004) as well as others such as McEnery and Wilson (2001). This philosophy is underpinned by a strictly empirical, statistically-oriented analysis of text, lexis and grammar patterns and phonological patterns, which aims to put description and observation at the centre of linguistic theorising. This stands opposed to the Chomskyan reliance on human intuition as a method for accounting for language universals and syntactic structures.

In order to confer some descriptive objectivity on the analysis of learner language, this paper proposes the adoption of a new term: 'Interlanguage Feature' (IF). This new term is intended to cover what has previously been thought of as 'error' and 'mistake'. This new term frees the analyst to consider 'unsuccessful language' as features of the transitional linguistic knowledge of learners – their Interlanguage – rather than having to worry about whether or not the learner has made an error or a mistake. This means that we should start to use the term 'Interlanguage Feature Analysis' (IFA) instead of 'Error Analysis'.

Admittedly, we may still be in danger of the stigma of values, even if we adopt a new supposedly value-free term such as IF or IFA. It remains uncertain whether or not it is possible or even desirable to be completely value-free when exploring the language of learners. Everything may depend on which side of the divide between cognitivist SLA theory and the sociological ELF paradigm the analyst is sitting on. If one is of a more SLA-based persuasion, learners, for instance, might be treated theoretically as 'failed native speakers' whose productions are forever negatively compared with so called native speaker language norms which may in fact be unattainable. Or, if one is to take the ELF paradigm as a starting point, the learner is more of a free agent whose non-standard linguistic productions can be privileged because they follow localised norms.

Whichever viewpoint we adopt, it is the contention of this paper that, if we are true to the descriptivist tenets of corpus linguistics, we can and should be value-free or neutral in our analysis of learner productions. Anything less than this should be viewed as bad science. Furthermore, if it is possible to operationalise a neutral terminology to analyse learner language or English as a Lingua Franca, we will avoid many of the definitional issues encountered whenever we try to decide whether a given case represents failure of knowledge or a momentary slip on the part of the learner. These issues, along with the terms 'error' and 'mistake', should be retained for the Error Explanation stage of Corder's steps. Meanwhile, Interlanguage Feature Analysis could, along with more advances in computer technology, become a standard scientific tool for investigating the language of English learners and speakers of ELF.

Acknowledgements

An earlier version of this paper was presented at the 12th International Asia TEFL Conference and 23rd MELTA International Conference, held in Kuching, Sarawak, Malaysia in August 2014, and published in the conference proceedings.

References

Abdul Samad, A., Hassan, F., Mukundan, J., Kamarudin, G., Syed Abdul Rahman, S. Z., Mohd Rashid, J., & Vethamani, M. E. (2002). *The English of Malaysian School Students (EMAS) Corpus*. Serdang: Universiti Putra Malaysia.

Anthony, L. (2013). Developing AntConc for a new generation of corpus linguists. Proceedings of the Corpus Linguistics Conference (CL 2013), July 22-26, 2013. Lancaster University, UK, 14-16.

Botley, S. P., & Metom, L. (2011). *A Corpus-Based Investigation of the Interlanguage of University Students in East Malaysia*. Final Project Report, Research Management Institute, Universiti Teknologi MARA.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics*, *5*, 161-70.

Corder, S. P. (1971). Idiosyncratic Dialects and Error Analysis. *International Review of Applied Linguistics*, *9*(2), 147-160.

Corder, S. P. (1974). Error analysis. In J. Allen & S. Corder (eds.): *The Edinburgh Course in Applied Linguistics Volume 3: Techniques in Applied Linguistics* (pp. 122-154). Oxford: Oxford University Press.

Dagneaux, E., Denness, S., & Granger, S. (1998). Computer-aided error analysis. *System: An International Journal of Educational Technology and Applied Linguistics*. *26*(2), 163-174.

Ellis, R., & Barkhuizen, G. (2005). *Analysing Learner Language*. Oxford: Oxford University Press.

Granger, S., Hung, J., & Petch-Tyson, S. (Eds.) (2002). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins.

Granger, S. (2002). A Bird's-eye view of computer learner corpus research in S. Granger, J. Jung & S. Petch-Tyson (eds.) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 3-33). Amsterdam: John Benjamins.

James, C. (1998). *Errors in Language Learning and Use: Exploring Error Analysis*. Harlow, Essex: Addison-Wesley Longman.

Kachru, B. (ed.) (1992). *The Other Tongue* (Second Edition). Urbana and Chicago: University of Illinois Press.

Kirkpatrick, A. (2011). English as an Asian lingua franca and the multilingual model of ELT. *Language Teaching*, *44*(2), 212-224.

Lennon, P. (1991). Error: some problems of definition, identification and distinction, *Applied Linguistics*, *12*(2), 180-96.

Mauranen, A. (2003) Academic English as lingua franca—a corpus approach. *TESOL Quarterly*, *37*, 513–27.

McEnery, A. M., & Wilson, A. (2001). *Corpus Linguistics* (Second Edition). Edinburgh: Edinburgh University Press.

Seidlhofer, B. (2004). Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics*, *24*, pp. 209–39.

Seidlhofer, B. (2005). English as a lingua franca. *ELT Journal*, *59*(4) (October), 339-341. doi:10.1093/elt/cci064.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics,10*, 209-31.

Sinclair, J. M. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, J. M. (2004). *Trust the Text: Language, Corpus and Discourse*. London: Routledge.