Ventola, E. and Mauranen, A. (1991) Non-native writing and native revising of scientific articles. In E. Ventola (ed.) *Functional and Systemic Linguistics: Approaches and Uses* (pp. 457–492). Berlin: Mouton de Gruyter.

Zwaan, R.A., Magliano, J.P. and Graesser, A.C. (1995) Dimensions of situation-model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21, 386–397.

5 Error Patterns and Automatic L1 Identification

Yves Bestgen, Sylviane Granger and Jennifer Thewissen

Introduction

The strong version of the Contrastive Analysis Hypothesis rested on the 'difference equals difficulty principle'. That is to say, learner errors can be predicted on the basis of differences between their mother tongue (L1) and the target language (L2). Although this strong version has been abandoned, research in error analysis and more recently, computer-aided error analysis has demonstrated that a sizable proportion of errors are L1-induced (see results of studies by Chuang & Nesi, 2006; Díez-Bedmar & Papp, 2008; Hawkins & Buttery, 2010; Neff & Bunce, 2005). It would thus seem reasonable to use error patterns as a clue to L1 identification. However, a survey of the literature shows that errors have rarely been used as a profiling criterion for this purpose, although they are one of the standard variables used in other types of stylometric studies, notably those that aim to determine learners' proficiency level (see Dikli, 2006 for a survey of automated scoring systems). The aim of this chapter is to investigate the extent to which error patterns can be relied on for automatic L1 identification.

Two recent studies have used errors as a sole or additional feature for the purposes of L1 identification. Koppel *et al.* (2005) based their investigation on five sub-corpora of the first version of the *International Corpus of Learner English* (L1 Russian, Czech, Bulgarian, French and Spanish). Using a mixture of features (several error types as well as function words, letter *n*-grams, and rare POS bigrams), they achieved an automatic identification rate of 80%. In their conclusion, the authors add an important caveat, which is that they may have taken 'unfair advantage of differences in overall proficiency among the different sub-corpora' (Koppel *et al.*, 2005: 627). They give as an example Bulgarian learners who were considerably less error prone than the Spanish learners. However, their investigation brings out a number of distinctive patterns that can be exploited for identifying the L1 of particular learner populations, such

as the unnecessary doubling of consonants by Spanish learners or the overuse and misuse of *indeed*, which is characteristic of French learners. In a subsequent investigation, Wong and Dras (2009) used the same five sub-corpora as Koppel *et al.*, to which they added two native languages (i.e. Chinese and Japanese) from the second version of the *International Corpus of Learner English*. They used three types of syntactic errors (subject–verb disagreement, noun–number disagreement and misuse of determiners) and obtained accuracy rates that were slightly though significantly higher than the 14% majority class baseline.[1] However, when combined with other features (function words, character *n*-grams and POS *n*-grams), syntactic errors did not demonstrate any improvement in classification accuracy. ANOVA tests were run for each error type. The results show significant differences in both raw and relative frequencies for only one error category: determiner misuse. Overall, Czech and Chinese learners display significantly more such errors than French and Spanish learners, a fact that the authors attribute to the absence or less extended use of articles in their mother tongue.

Although the results of those two studies are mixed, they illustrate the potential of errors as L1 identification features, thus providing an impetus for further research. The two studies suffer from a number of weaknesses that need to be addressed, in particular as regards the actual selection of error types and the methods used to identify them. Research so far has been based on a highly restricted number of error types (often formal and syntactic) selected on rather weak grounds. Koppel *et al.* give a very practical justification for this, namely that 'they can be identified with relative ease' (Koppel *et al.*, 2005: 626). However, automatic identification of errors with spell- and grammar-checkers is far from fool-proof. This is acknowledged by Wong and Dras, who report 'a relatively high false positive rate of 48.2% in determiner misuse errors' (Wong & Dras, 2009: 58). They give a more theoretical explanation for the selection of their three syntactic error types, claiming to have chosen them because they are 'amongst the frequently observed syntactic error types in nonnative English which it has been argued are attributable to language transfer' (Wong & Dras, 2009: 53). This claim is not very convincing in the case of disagreement errors which are at least partly developmental (Osborne, 2008), a factor which may explain the lack of significant differences between the three L1 groups for these error types. In their conclusion, the authors mention two desiderata for future research in the field: 'using more error types' and designing 'a method for more accurately identifying them' (Wong & Dras, 2009: 60).

In the current study we try to address some of the shortcomings of earlier studies by making use of manually error-tagged samples from the *International Corpus of Learner English* (Granger *et al.*, 2009), annotated on the basis of the *Louvain Error Tagging System* (Dagneaux *et al.*, 1998) which covers a wide range of error types. In the section 'Data and Methodology', we describe the corpus data and error annotating system used. The following section 'Automatic Identification of the L1 Groups' gives an outline of the statistical methods

employed for the automatic L1 identification and presents the classification results. Next, in the section 'Bringing Proficiency into the Equation' we tackle the possible impact of proficiency level differences on the results. The section 'Fine-Grained Qualitative Analysis' then provides a fine-grained analysis of the most discriminatory error patterns, while the 'Conclusion' section concludes our chapter and suggests avenues for future research.

Data and Methodology

The learner corpus data used in this study were taken from the *International Corpus of Learner English (ICLEv2)*, which contains essays written by learners of English from 16 mother tongue backgrounds (Granger *et al.*, 2009). The texts selected for the present study were all written by French-, German- and Spanish-speaking EFL learners (henceforth FR, GE, SP), are argumentative in nature, and vary between 500 and 900 words in length. A total of 223 learner essays were selected, amounting to c. 50,000 tokens per L1 group. The detailed breakdown of the learner corpus sample used here is presented in Table 5.1.

The rather limited size of the corpus sample can be explained by the fact that each of the texts was exhaustively error tagged (i.e. annotated for errors, in accordance with the latest version of the *Louvain Error Tagging Manual*; Dagneaux *et al.*, 1998).[2] The Louvain error tagging system is hierarchical: it includes a total of seven main error domains,[3] the majority of which are broken down into further error subcategories. The main error domains are briefly described in Table 5.2.

Examples (1) and (2) below target error subcategories, more specifically spelling errors (FS) which are part of the formal error domain, and verbs used with an erroneous dependent preposition, which belong to the lexico-grammatical domain (XVPR). As can be seen from the examples, the error tag is written between brackets and placed in front of the erroneous element while the correction is presented between dollar signs after the error:[4]

(1) *The fast spread of television can transform it into a double-edged* **(FS)** ***wheapon $weapon$****. (SP)*
(2) *The classroom must often have* **(XVPR)** ***resembled to $resembled$*** *a 'Chamber of Horrors'. (GE)*

Table 5.1 *ICLE* corpus sample

L1 background	Number of learner essays	Total tokens
FR	74	50,195
GE	71	49,856
SP	78	51,397
Total	223	151,448

Table 5.2 The Louvain error tagging system

Error tags	Definition	Description
F	Formal errors	Spelling or morphological errors that result in a nonexistent English word (+ homophones)
G	Grammatical errors	Errors that break the general rules of English grammar
L	Lexical errors	Errors involving the semantic properties of words or phrases (conceptual, collocational or connotative)
X	Lexico-grammatical errors	Errors that violate the lexico-grammatical properties of words, that is, erroneous dependent prepositions, complementation patterns or countable/uncountable noun confusion.
Q	Punctuation errors	Errors that target punctuation problems, for example, confusion between punctuation markers, missing or redundant markers
W	Word redundant/ missing/order errors	Unnecessary use of words, missing necessary words, or misordered words
S	Style errors	Sentence fragments and incomprehensible sentences

Along with its seven main error domains, the Louvain system comprises a total of 54 error subtags. A number of these tags (six in total) were excluded from the present analysis, however, because they occurred with too low a frequency (i.e. in fewer than 10 different texts). A full list of the error tags used in our study is included in the appendix (the six excluded error tags are not represented).

Automatic Identification of the L1 Groups

In what follows, we first present the statistical measures that were chosen to best tackle our research question, namely the possibility of automatically identifying the learners' L1 on the basis of the errors found in the *ICLE* learner corpus sample. The success rate with which the L1s were identified along with the error types which best discriminate the L1 groups will be described in a second stage.

Statistical Methods

In keeping with the work carried out in this book, we have used the statistical method known as Discriminant Analysis (DA) to find out whether errors can be relied upon to automatically classify a text according to its L1.[5] This classification technique, which aims at selecting the variables (in our case errors) that best discriminate between given groups, is detailed in Chapter 1 and in the chapter preceding ours. Two elements that are specific to our study are in need of additional attention, however.

First, having a large number of potential (L1) predictors (i.e. 46 error types in total) in relation to the number of observations (i.e. 223 learner texts) may lead to the problem known as overfitting (e.g. Hawkins, 2004; Schiavo & Hand, 2000). Overfitting occurs when a statistical model (which is formed by the variables included in the prediction of group membership) is excessively specialized to account for the particular sample at hand, therefore lowering the predictive power of the discriminant analysis. In order to circumvent the problem of overfitting, researchers are advised to first carry out a procedure to select the best group of predictors (Schiavo & Hand, 2000: 302). It is noteworthy that such a procedure does not prevent overfitting, but only reduces it by limiting the number of variables used in the model. However, it brings other benefits which, in themselves, serve to justify its use: it facilitates the identification of the variables that will help discriminate between the groups and should ensure that the discriminant analysis is more successful than if carried out on the basis of all the variables (Louw & Steel, 2006). In order to do so, we have chosen an automatic selection procedure known as Stepwise analysis. To start off, the model does not contain any predictor. At each step, the algorithm adds the variable with the highest discriminatory power. However, because adding variables to the model modifies the discriminatory power of the other variables present, the algorithm each time checks that the variable with the lowest discrimination still significantly contributes to the analysis. If not, the variable is taken out of the model. The procedure ends when none of the variables that are not yet in the model are significant enough to be included. The significance level to enter the model or to be removed from the model is 0.05 so as to avoid the selection of too many variables in comparison with the number of texts.

Second, in order to evaluate the effectiveness of DA in the selection of new variables, we used the Leave-one-out cross-validation technique (LOOCV) which classifies each text on the basis of the model derived from the DA for the other 222 texts (Lachenbruch & Mickey, 1968). This step results in an unbiased estimate, ensuring that if the model were to be applied to new texts taken from the same learner population, the rate of accurately classified texts would be similar to the one reported by the LOOCV procedure.

Combining the stepwise and LOOCV methods is not without problems, however. The statistical packages available apply the stepwise procedure

and thus select the best set of variables on the basis of the dataset as a whole. It is only afterwards that the LOOCV is performed. Thus, when the LOOCV is performed, the model that is used to classify each text has partially been constructed on the basis of this text. A consequence of this is that the LOOCV is favorably biased; that is, it provides an estimation of the classification accuracy that is overly optimistic. To find a way around this problem, we used the procedure advocated by Schulerud and Albregtsen (2004), which involves excluding each text one by one *before* actually selecting the variables with the stepwise analysis. In other words, 223 stepwise analyses were carried out, each time leaving one text out. The variables selected on this basis were used to build the best model for the 222 texts. We then determined how this model classified the 223rd text. This ensures an unbiased LOOCV measure despite the use of the stepwise procedure.

Results

L1 Group Identification

Table 5.3 presents the results for the L1 group identification obtained by combining the stepwise and LOOCV techniques. In this analysis the stepwise procedure was set to iterate through no more than 22 steps. In this way, a maximum of 22 error types could be incorporated into the prediction model and, consequently, the ratio between the number of variables in the model and the number of cases (or texts) in the training set was limited to 1 variable per 10 cases. The bolded figures presented diagonally in the table give the number of texts that were accurately classified in each L1 group: 48 FR texts, 51 GE texts and 47 SP texts.

It appears from Table 5.3 that error types can indeed lead to successful L1 group identification. The overall score for accurate classification is 65% (i.e. 146 out of the 223 texts) which is a very satisfactory result in view of the fact that chance would have set the score at approximately 33%. Looking at the individual L1 backgrounds, we see that the number of texts which have been accurately classified is quite similar in the three groups. However, what also stands out is that the discriminant analysis classified very few FR and

Table 5.3 L1 group identification scores for FR, GE and SP texts

	Predicted L1			
Actual L1	FR	GE	SP	Total
FR	**48**	21	5	74
GE	13	**51**	7	71
SP	15	16	**47**	78
Total	76	88	59	223

GE texts in the SP group (i.e. 5 FR and 7 GE texts were classified as SP), while quite a high number of GE and SP texts were classified as FR (i.e. 13 and 15 respectively); similarly a number of FR and SP texts were identified as belonging to the GE group (i.e. 21 and 16 respectively). What this indicates is that when DA classifies a text as SP, its decision is in general more accurate than when it classifies a text as being either FR or GE. This may also be interpreted as meaning that while error profiles generally enable us to clearly distinguish the Spanish group from its counterparts, it may be more difficult to tease apart a French text from a German text.

Identifying L1 discriminatory error types

To determine which specific error types are the most useful to differentiate texts written by speakers of the three L1s, we followed Schulerud and Albregtsen (2004) and relied on the number of times that a given variable is selected across the folds of a cross-validation with the internal stepwise feature selection procedure. We assumed that any error type that is selected in more than half of these stepwise DAs is potentially useful. Each of these error types was then submitted to a Student–Newman–Keuls post-hoc test to determine whether their mean relative frequency (out of 100 tokens) differed significantly across the three L1 groups. Table 5.4 provides all the error

Table 5.4 Discriminatory error types for the three L1 backgrounds

		Mean relative frequency		
Sign. diff.	Error category	FR	GE	SP
SP>FR>GE	LS	1.79614	1.48447	2.69455
SP>GE>FR	LP	0.63539	0.79587	1.10214
SP>(FR=GE)	GA	0.50257	0.33697	1.14883
SP>(FR=GE)	FS	0.62141	0.79184	1.61050
SP>(FR=GE)	GPU	0.09966	0.06123	0.24312
SP>(FR=GE)	XVPR	0.10129	0.09632	0.27498
SP>(FR=GE)	GDD	0.02048	0.01416	0.06117
GE>(FR=SP)	GADJO	0.00775	0.03733	0.00943
GE<(FR=SP)	GNN	0.22731	0.11789	0.23780
GE<FR	LCLS	0.17842	0.08718	0.12460
FR>(GE=SP)	QL	0.16064	0.03337	0.04836
FR<(GE=SP)	LCS	0.06320	0.10906	0.11787

Note: The sign. diff. column presents the statistically significant differences between the means according to the Student–Newman–Keuls procedure. For example, GE>(FR=SP) indicates that the mean for the FR group does not differ significantly from that of the SP group, while both of these means are significantly lower than the GE mean. The numbers in the FR, GE and SP columns represent the respective group means.

types that have passed these two selection tests as well as the statistical results of the mean comparison tests.

Table 5.4 shows seven error types to be strongly associated with the SP group: spelling errors (FS), lexical errors on single words (LS) and phrases (LP), as well as erroneous article use (GA), erroneous-dependent prepositions used with verbs (XVPR), unclear pronominal reference (GPU)[6] and erroneous demonstrative determiners (GDD). Errors involving the order of adjectives (GADJO) are more typical of the GE group. Strongly associated with the FR population are punctuation errors of the QL type (i.e. use of a conjunction of coordination instead of a punctuation marker, or of a punctuation marker instead of a conjunction of coordination, as in *he took the books* (QL) *and $,$ the records and the computers*). Some of the other error categories found in Table 5.4 are less frequent in one of the L1 groups than in the other two (cf. noun number disagreement errors (GNN) and subordinating conjunction errors (LCS)). A last error category (LCLS) distinguishes between just two of the three groups: single logical connectors are significantly fewer in the GE than in the FR group, neither of which show significant differences from the SP group. These results show that the error types that play a part in L1 identification are extremely diverse, which justifies widening the error base beyond the traditional spelling and agreement errors.

The above analysis suggests that it is possible to successfully discriminate the learners' L1 backgrounds on the basis of error types. However, like Koppel *et al.* (2005), we acknowledge that proficiency level differences between the three L1 groups may have played a part in the discriminant analysis. In the following section we investigate the possible impact of proficiency differences on the good L1 identification rates obtained above.

Bringing Proficiency into the Equation

Because the learners whose writing is included in the *ICLE* corpus were in their third or fourth year of undergraduate university studies, the data were initially believed to be representative of advanced learner writing. However, following a pilot study carried out by the Centre for English Corpus Linguistics in which a professional rater was asked to assign a Common European Framework score (henceforth CEF) to 20 texts randomly selected from each of the 16 mother tongue backgrounds, it was revealed that the corpus as a whole actually 'falls in the intermediate-advanced range' (Granger, 2004: 130).[7] In view of this finding, it is important to establish whether the reasonably successful automatic identification of the three L1 groups presented in Table 5.3 might be due to possible differences in proficiency level between the mother tongue populations. Two questions will be investigated in relation to this issue, the first being whether there are indeed significant differences in the level of English proficiency between the FR, GE and SP groups. If so, we will determine whether error profiles can nevertheless be used to successfully discriminate between subgroups of learners with the same level of proficiency but with a different L1.

CEF Rating Procedure

Because the *ICLE* does not include any built-in information on *individual* learners' level of proficiency in English, we had each of the 223 learner essays assessed by two professional raters who were both involved in the assessment of writing at the University of Cambridge Local Examinations Syndicate (UCLES).[8] A consequence of this methodology is that what we have termed 'proficiency level' is in fact a description of the quality of one single essay produced by each individual learner, rather than a more independent assessment of the learners' overall proficiency in English as would have been obtained if they had taken a standardized proficiency test such as TOEFL, for instance.[9] Importantly also, errors were a feature that the raters took into account when assessing the quality of the essays, as accuracy was part of the criteria they had to rate, along with elements pertaining to complexity and coherence and cohesion. Table 5.5 details the specific linguistic competences that were assessed for each text.

Both raters were asked to assign a CEF grade to each of the linguistic components in Table 5.5. The scores they could choose from ranged from the lower intermediate to the most advanced CEF levels, that is, B1, B2, C1 or C2.[10] In addition, the raters also attributed a holistic score to each text so as to convey their overall impression of the quality of the 223 essays. It was made possible for raters to use + or – signs to further specify quality within each proficiency level (e.g. vocabulary control = B2–, grammatical accuracy = C1+ etc).[11] The scores thus obtained were recorded using an 11-point numerical scale (i.e. B1– = 0.67, B1 = 1, B1+ =1.33 etc until C2 = 4,[12] cf. Table 5.6) so as to calculate the degree of correlation between the grades given by the two raters. The correlation between both raters on this 11-point scale reached $r = 0.69$. Overall,

Table 5.5 Linguistic competences considered in the rating procedure

Linguistic competence	Linguistic domain represented
Vocabulary control	Accuracy in lexical choice
Grammatical accuracy	Accuracy in grammatical use
Orthographic control	Accuracy in spelling and punctuation
Vocabulary range	Complexity, that is, range and richness of lexical knowledge
Coherence and cohesion	Accuracy of the chosen coherence and cohesion devices and complexity, that is, range of coherence and cohesion devices used

Table 5.6 11-point numerical scale

Holistic CEF score	B1–	B1	B1+	B2–	B2	B2+	C1–	C1	C1+	C2–	C2
Numerical value	0.67	1	1.33	1.67	2	2.33	2.67	3	3.33	3.67	4

Table 5.7 Interpreting mean scores to assign final CEF score

Mean score	Final CEF score assigned
Lower than 1.5	B1
Exactly 1.5	B1
Between 1.51 and 2.5	B2
Exactly 2.5	B2
Between 2.51 and 3.5	C1
Exactly 3.5	C1
Between 3.51 and 4	C2

Table 5.8 Converting mean numerical scores into the final CEF score

	Txt	Holistic score Rater 1	Holistic score Rater 2	Mean score	Final CEF score
Initial CEF scores	SP49	C1+	B2		
Numerical values	SP49	3,33	2	**2,67**	**C1**

both raters completely agreed on a total of 102 texts (46%); they reached near agreement (i.e. a maximal difference of one band score as in B2–C1 or C1–C2) on 87 texts (39%) and disagreed by more than one band score (e.g. B1–C1) on 34 texts (15%). In accordance with the language testing guidelines described in Alderson *et al.* (2001), the 34 texts on which the two raters substantially disagreed were submitted to a third rater, who was given the same descriptors and guidelines as her first two colleagues.

In order to attribute one final CEF score to each text and so as to be able to stratify the corpus sample accordingly, we computed the mean of the holistic scores given by the two (or three) raters (each holistic score had been given a numerical value, as shown in Table 5.6). The mean was then reinterpreted in terms of CEF score, as shown in Table 5.7.

A concrete example of how this procedure was conducted is given in Table 5.8 for one of the SP texts whose final CEF score was calculated to be C1.

Impact of Proficiency Level on the Discriminant Analysis Results

A one-way between-groups analysis of variance (ANOVA) was carried out to determine whether the mean CEF scores in the three L1 groups (see Table 5.9) differed significantly or not. The ANOVA indicated that there

Table 5.9 Mean CEF scores in the three L1 groups

	N	Mean	Std Dev	Minimum	Maximum
SP	78	**1.44**	0.64	0.67	3.00
FR	74	**2.64**	0.67	1.44	4.00
GE	71	**3.03**	0.87	0.83	4.00

are highly significant differences in mean CEF scores between the groups ($F(2, 220) = 96.14$; $p < 0.0001$). The Student–Newman–Keuls post-hoc test revealed that all three group means actually differ significantly from each other, namely all three learner populations display a different level of proficiency in English. As can be noted, however, the difference between the SP group and the FR/GE groups is much more marked than that between the FR and GE groups themselves.

Table 5.10 gives the breakdown of the number of texts per proficiency in each of the L1 sub-corpora. The SP group is overwhelmingly representative of the lower intermediate level B1 (i.e. 68% of all SP essays are B1); the FR learners are mainly situated between B2 (45%) and C1 (41%), while the GE population is well within the advanced C1 (42%) and C2 (31%) range. It is difficult to determine whether the differences in the mean proficiency scores between the L1 groups are specific to the *ICLE* sample used here or whether they can be generalized to these three populations as a whole. The latter option is plausible given that these results were confirmed by the rating of ICLE as a whole (Granger *et al.*, 2009), which also highlighted the closeness of the FR and GE data in comparison with the SP group, which was found to be at a lower level of proficiency. The differences in proficiency level can partly explain the effectiveness of the discriminant analysis in distinguishing between the groups on the basis of errors as there is a very strong correlation ($r = -0.81$; $p < 0.0001$) between the total number of errors (out of 100 tokens) and the CEF score (as an 11-point scale). The horizontal x-axis in Figure 5.1 represents the total number of errors (out of 100 tokens) while the vertical y-axis stands for the CEF score (in numerical values). As can be seen, the SP population (represented by the triangles) is clustered on the right-hand side of the x-axis with a high number of errors and toward the lower

Table 5.10 Text breakdown per L1 and proficiency level

	B1	B2	C1	C2	Total
SP	**53**	18	7	0	78
FR	5	**33**	**30**	6	74
GE	8	11	**30**	**22**	71
Total	66	62	67	28	223
	29.60	27.80	30.04	12.56	100.00

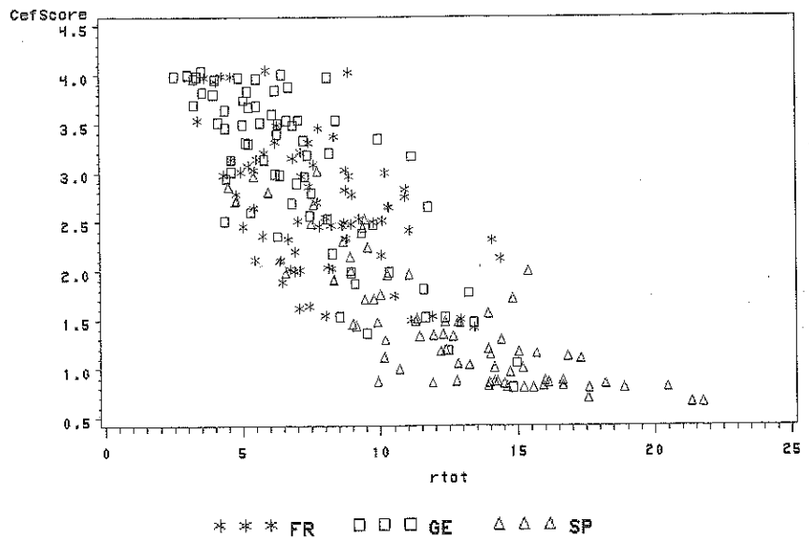

Figure 5.1 Scatterplot for the correlation between the total errors and the CEF score

end of the *y*-axis for low CEF proficiency scores. On the other hand, the FR and GE groups (represented by the stars and squares, respectively) are more difficult to tease apart. It follows from this observation that, in order to distinguish the SP learners from their FR and GE counterparts, the discriminant analysis can rely on this clear difference in proficiency levels. What can also be concluded, therefore, is that one cannot exclude the possibility that automatic L1 group identification may at some level be based on proficiency-level differences.

Controlling for Proficiency Level

In order to control for the possible impact of proficiency-level differences on the 65% accurate classification score obtained for the data as a whole (cf. Table 5.3), we selected a group of learners with the same proficiency level but with a different L1. The same analyses as were carried out earlier on the FR, GE and SP data as a whole (i.e. Stepwise and LOOCV) were conducted on the 30 FR and 30 GE texts that were assessed as C1. As indicated in Table 5.10 above, this is the CEF category for which we have the highest number of texts in two different L1 groups. However, because the analysis is carried out on a small data sample, the results should be cautiously interpreted and seen as preliminary.

The LOOCV results are given in Table 5.11. In total, 75% of texts (i.e. 22 FR and 23 GE) were accurately classified, with chance setting the score at 50%.

Table 5.11 L1 group identification for C1 FR and GE texts

	To L1		
From L1	FR	GE	Total
FR	**22**	8	30
GE	7	**23**	30
Total	29	31	60

Table 5.12 Discriminatory error types for the C1 FR and GE texts

Sign. diff.	Error types	FR	GE
FR>GE	QL	0.158	0.035
FR>GE	GNN	0.231	0.083
GE>FR	FS	0.530	0.808

Note: The sign. diff. column presents the statistically significant differences between the relative error means as reported by the Student–Newman–Keuls procedure.

To determine which specific error types are the most useful to differentiate texts written by speakers of the two L1s, we relied on the method that was outlined in the section 'Identifying L1 discriminatory error types'. Table 5.12 provides the three error types that have passed the two selection tests as well as the statistical results of the mean comparison tests. Strongly associated with the FR C1 group are punctuation errors of the QL type (i.e. use of 'and' instead of a punctuation marker or of a punctuation marker instead of 'and') and noun number disagreement errors (GNN) (these two error types were also more typical of the FR texts than of the GE texts in Table 5.4, which presented the discriminatory errors for the corpus sample as a whole). As for spelling errors (FS), these are more tightly associated with the GE than the FR C1 population.

This small-scale analysis, which would need to be replicated on a larger corpus, suggests that errors can be used for successful automatic L1 group identification even for populations that have been controlled for proficiency.

Fine-Grained Qualitative Analysis

An important point we wish to put forward here is that successfully identifying L1 groups on the basis of errors is not tantamount to saying that the discriminatory error types all necessarily result from transfer from the L1. Many other factors can be at play here to explain why certain L1 populations commit more errors than other L1 groups, such as differences in the level of English proficiency reached by the L1 populations or different English

teaching methods (some of which may be more accuracy oriented while others may be more communication-oriented). Each error category may thus include errors that are the result of many different processes, including though not restricted to transfer. In order to identify potential transfer errors among the errors committed, we carried out a more fine-grained analysis which involved analyzing a number of 'error couples', that is, errors together with their suggested corrections. Examples include the *'his-its'* couple, which pertains to the GDO category (grammatical errors on possessive determiners) or *'increase of-increase in'* which belongs to the XNPR error category (noun used with an erroneous dependent preposition). An ANOVA test was used to compare the mean frequencies of the error couples in the three L1 sub-corpora. This enabled us to identify the couples that were significantly more frequent in one corpus than in the other two. In spite of the many comparisons being carried out and because this is an exploratory qualitative study, a rather liberal 0.05 significance value was adopted. Importantly, this procedure can only be applied if a couple occurs frequently enough in the data. We set the frequency threshold at seven occurrences in the whole corpus sample.

Figure 5.2 shows the number of error couples that are significantly more frequent in one population than in the other two. As can be seen, the SP sub-corpus clearly stands out as containing the most L1-salient couples.

Table 5.13 lists the couples that occur significantly more frequently in the SP, FR and GE sub-corpora, respectively. The error couples presented below

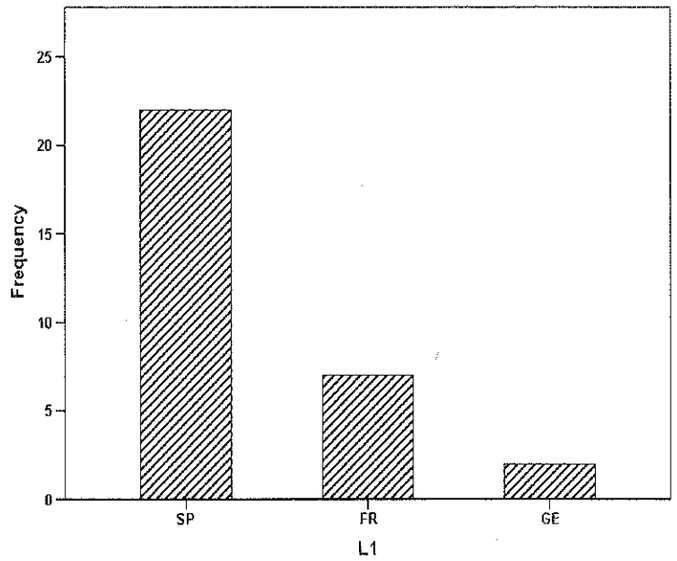

Figure 5.2 Frequency of L1-discriminatory error/correction couples in the three sub-corpora

Table 5.13 Error couples

Dependent	*F* Value	Prob *F*	Mean	*N*	Level
SP > [FR = GE]					
r_theX0	23.61	<0.0001	.	.	.
r_theX0	23.61	.	0.660	78	SP
r_theX0	23.61	.	0.249	74	FR
r_theX0	23.61	.	0.120	71	GE
r_0Xit	13.29	<0.0001	.	.	.
r_0Xit	13.29	.	0.038	78	SP
r_0Xit	13.29	.	0.003	71	GE
r_0Xit	13.29	.	0.000	74	FR
r_inXon	8.42	0.0003	.	.	.
r_inXon	8.42	.	0.035	78	SP
r_inXon	8.42	.	0.004	71	GE
r_inXon	8.42	.	0.002	74	FR
r_thisXthese	8.42	0.0003	.	.	.
r_thisXthese	8.42	.	0.027	78	SP
r_thisXthese	8.42	.	0.004	71	GE
r_thisXthese	8.42	.	0.000	74	FR
r_0Xthe	8.39	0.0003	.	.	.
r_0Xthe	8.39	.	0.271	78	SP
r_0Xthe	8.39	.	0.135	74	FR
r_0Xthe	8.39	.	0.092	71	GE
r_haveXhas	7.65	0.0006	.	.	.
r_haveXhas	7.65	.	0.025	78	SP
r_haveXhas	7.65	.	0.004	74	FR
r_haveXhas	7.65	.	0.000	71	GE
r_0Xby	7.65	0.0006	.	.	.
r_0Xby	7.65	.	0.026	78	SP
r_0Xby	7.65	.	0.004	74	FR
r_0Xby	7.65	.	0.000	71	GE
r_commaXfull stop	7.47	0.0007	.	.	.
r_commaXfull stop	7.47	.	0.097	78	SP
r_commaXfull stop	7.47	.	0.035	74	FR
r_commaXfull stop	7.47	.	0.030	71	GE
r_it_sXit_is	7.24	0.0009	.	.	.
r_it_sXit_is	7.24	.	0.059	78	SP
r_it_sXit_is	7.24	.	0.010	71	GE

(*continued*)

Table 5.13 (continued)

Dependent	*F* Value	Prob *F*	Mean	*N*	Level
SP > [FR = GE]					
r_it_sXit_is	7.24	.	0.005	74	FR
r_enterpriseXcompany	6.92	0.0012	.	.	.
r_enterpriseXcompany	6.92	.	0.013	78	SP
r_enterpriseXcompany	6.92	.	0.000	71	GE
r_enterpriseXcompany	6.92	.	0.000	74	FR
r_otherXanother	6.84	0.0013	.	.	.
r_otherXanother	6.84	.	0.014	78	SP
r_otherXanother	6.84	.	0.000	71	GE
r_otherXanother	6.84	.	0.000	74	FR
r_commaXcolon	5.54	0.0045	.	.	.
r_commaXcolon	5.54	.	0.050	78	SP
r_commaXcolon	5.54	.	0.019	74	FR
r_commaXcolon	5.54	.	0.012	71	GE
r_itXquestion mark	5.46	0.0048	.	.	.
r_itXquestion mark	5.46	.	0.015	78	SP
r_itXquestion mark	5.46	.	0.000	71	GE
r_itXquestion mark	5.46	.	0.000	74	FR
r_full stopX0	5.38	0.0052	.	.	.
r_full stopX0	5.38	.	0.094	78	SP
r_full stopX0	5.38	.	0.004	71	GE
r_full stopX0	5.38	.	0.002	74	FR
r_0Xa	5.19	0.0063	.	.	.
r_0Xa	5.19	.	0.082	78	SP
r_0Xa	5.19	.	0.037	71	GE
r_0Xa	5.19	.	0.027	74	FR
r_speciallyXespecially	5.19	0.0063	.	.	.
r_speciallyXespecially	5.19	.	0.016	78	SP
r_speciallyXespecially	5.19	.	0.000	71	GE
r_speciallyXespecially	5.19	.	0.000	74	FR
r_0Xthey	4.82	0.0089	.	.	.
r_0Xthey	4.82	.	0.015	78	SP
r_0Xthey	4.82	.	0.000	71	GE
r_0Xthey	4.82	.	0.000	74	FR
r_lifesXlives	4.79	0.0092	.	.	.
r_lifesXlives	4.79	.	0.025	78	SP
r_lifesXlives	4.79	.	0.005	71	GE

Dependent	*F* Value	Prob *F*	Mean	*N*	Level
SP > [FR = GE]					
r_lifesXlives	4.79	.	0.000	74	FR
r_areXis	4.59	0.0112	.	.	.
r_areXis	4.59	.	0.022	78	SP
r_areXis	4.59	.	0.005	74	FR
r_areXis	4.59	.	0.002	71	GE
r_notXno	4.47	0.0125	.	.	.
r_notXno	4.47	.	0.011	78	SP
r_notXno	4.47	.	0.002	74	FR
r_notXno	4.47	.	0.000	71	GE
r_increase_ofXincrease_in	4.31	0.0145	.	.	.
r_increase_ofXincrease_in	4.31	.	0.017	78	SP
r_increase_ofXincrease_in	4.31	.	0.004	74	FR
r_increase_ofXincrease_in	4.31	.	0.000	71	GE
r_don_tXdo_not	4.30	0.0147	.	.	.
r_don_tXdo_not	4.30	.	0.060	78	SP
r_don_tXdo_not	4.30	.	0.019	74	FR
r_don_tXdo_not	4.30	.	0.018	71	GE
FR > [SP = GE]					
r_indeedX0	21.62	<.0001	.	.	.
r_indeedX0	21.62	.	0.075	74	FR
r_indeedX0	21.62	.	0.008	71	GE
r_indeedX0	21.62	.	0.003	78	SP
r_ellipsis(...)Xfull stop	9.12	0.0002	.	.	.
r_ellipsis(...)Xfull stop	9.12	.	0.077	74	FR
r_ellipsis(...)Xfull stop	9.12	.	0.033	78	SP
r_ellipsis(...)Xfull stop	9.12	.	0.002	71	GE
r_ellipsis(...)Xfull stop	7.78	0.0005	.	.	.
r_ellipsis(...)Xfull stop	7.78	.	0.021	74	FR
r_ellipsis(...)Xfull stop	7.78	.	0.000	71	GE
r_ellipsis(...)Xfull stop	7.78	.	0.000	78	SP
r_commaXand	7.66	0.0006	.	.	.
r_commaXand	7.66	.	0.083	74	FR
r_commaXand	7.66	.	0.029	78	SP
r_commaXand	7.66	.	0.019	71	GE
r_butXhowever	6.80	0.0014	.	.	.
r_butXhowever	6.80	.	0.141	74	FR
r_butXhowever	6.80	.	0.078	71	GE

(*continued*)

Table 5.13 (continued)

Dependent	*F* Value	Prob *F*	Mean	*N*	Level
FR > [SP = GE]					
r_butXhowever	6.80	.	0.073	78	SP
r_communityXeu	6.24	0.0023	.	.	.
r_communityXeu	6.24	.	0.036	74	FR
r_communityXeu	6.24	.	0.002	78	SP
r_communityXeu	6.24	.	0.000	71	GE
r_hisXits	4.11	0.0176	.	.	.
r_hisXits	4.11	.	0.016	74	FR
r_hisXits	4.11	.	0.003	78	SP
r_hisXits	4.11	.	0.002	71	GE
GE > [SP = FR]					
r_full stopXquestion mark	5.09	0.0069	.	.	.
r_full stopXquestion mark	5.09	.	0.025	71	GE
r_full stopXquestion mark	5.09	.	0.009	74	FR
r_full stopXquestion mark	5.09	.	0.000	78	SP
r_kidsXchildren	4.56	0.0115	.	.	.
r_kidsXchildren	4.56	.	0.037	71	GE
r_kidsXchildren	4.56	.	0.005	78	SP
r_kidsXchildren	4.56	.	0.002	74	FR

should be interpreted as follows: 'r_theX0', for instance, stands for the relative frequency of the couple involving the use of the definite article *the* where the zero article should have been used instead, for example, *they haven't access to* **(GA) *the 0*** *education and probably most of them will dye before they were 15 or 16, while the children of the manager who bought them, go to* **(GA) *the 0*** *university in Europe* (SP).

A close look at the SP error couples shows that a number of them can safely be assumed to be transfer related. The '0Xit' and '0Xthey' couples, illustrated by examples 3 through 5 below, can be explained by the fact that Spanish nearly always omits personal pronouns in subject position, a possibility that does not exist in French or German. It is symptomatic that all '0Xthey' errors and all but one '0Xit' errors are found in the SP sub-corpus.

(3) *Finally* **0** *must be added that in our days it is necessary for a country to be (...)*
(4) *by the other hand* **0** *is very difficult for a person who have hurt any harm damage to prove it*
(5) *particular and daily objects become so artificial that* **0 <they>** *lose their primitive charm.*

Similarly, the 'otherXanother' couple (cf. example 6) is most probably due to the fact that Spanish uses the same word *otro* for both *other* and *another* (cf. also Cowan & Leeser, 2007, who report similar results for this feature). Similarly, 'inXon' errors (cf. example 7) can be assumed to be due to the fact that both *in* and *on* are translated into Spanish by the same preposition *en*.

(6) *it is clear that a person who is born in Africa or in any other Third World country will have more problems to develop than* **other <another>** *who is born in Europe.*
(7) *All words, tones, images which appear* **in <on>** *television are controlled by someone.*

Determiner errors are complex and would require a more detailed investigation than is possible within the framework of this study. It is clear, however, that transfer may play a part in several of the error couples involving articles, for example in 0Xa errors, as Spanish often uses the zero article where English as well as French and German use the indefinite article.[13] Several of the FR couples also appear to be transfer-related, such as the overuse and misuse of *indeed*, already noted by Granger and Tyson (1996) and Koppel *et al.* (2005).

A sizable proportion of the SP error couples, however, are more likely to be developmental than transfer related. This is the case with certain number errors such as 'lifesXlives' or 'areXis' and punctuation errors where learners have used a comma instead of a full stop ('commaXfull stop'), thus producing a run-on sentence as illustrated in example 8:

(8) *Dreams are the instruments of imagination,* **<.>** *through them we begin to give shape to all the formless ideas than wander in (...)*

The inappropriate use of contractions in academic texts, apparent from the 'it'sXit is' and 'don'tXdo not' pairs are unlikely to be linked to the learners' L1 but might rather be a sign of novice writing. It is symptomatic that most of the errors are also found in the other two sub-corpora, though in much smaller numbers, most likely because the Spanish texts are scored lowest.

The fine-grained analysis has shown that certain L1 discriminatory error couples do indeed appear to result from transfer from the L1 but other errors that also contribute to L1 group discrimination do not necessarily find their source in transfer processes. Importantly, this indicates that discriminant analysis (as well as any other supervised classification technique such as Support Vector Machines) has the ability to rely on any type of information that will help toward group identification, whatever the theoretical relevance of the information in question may be (in our case, the fact that L1 discrimination is due to transfer).

Conclusion

This chapter has shown that the error patterns found in error-tagged learner corpora can indeed be used as a reliable source of information in L1 authorship profiling studies. The good L1 identification scores obtained here may be due to the fact that, as advised by Wong and Dras (2009), we drew on a wide array of errors ranging from the more traditional spelling and agreement errors to the more covert lexical, lexico-grammatical and punctuation error types, all of which were manually identified. Besides considering the valuable evidence yielded by error frequencies, we also felt it necessary to dig deeper into the qualitative aspect of the errors committed and to concretely identify a number of L1-salient error couples. Such a fine-grained approach is also advocated by Baroni and Bernardini in translation studies to automatically distinguish translated texts from original texts: 'the analysis should be carried out at a more detailed level, studying the impact not only of broad categories but also of specific elements (e.g. a certain pronoun or a certain punctuation mark)' (Baroni & Bernardini, 2006: 271).

Because of its exploratory nature, our study also displays a number of caveats, however. First, the learner corpus sample is admittedly limited in size. Additionally, the fact that only three L1 backgrounds were considered is very likely to have had an influence on the high percentage of accurately classified texts (see Chapters 2 and 3 in this volume for more on this issue). However, the complexity of the analyses carried out would not have easily allowed for the inclusion of more language backgrounds. Another question that has not been broached in this chapter has to do with whether errors actually contribute in a significant way to authorship studies when compared with other features such as function words, *n*-grams or POS patterns. This question is tackled in Chapter 6.

An important point highlighted by our study is the importance of controlling for other phenomena such as differences in proficiency levels that influence error frequency and, consequently, the subsequent discriminant analysis results. As shown in our analysis, the discriminant analysis was most successful in distinguishing the SP group—the group with the lowest level of proficiency—from the FR and GE groups, whose proficiency profiles are more homogeneous. What the fine-grained analysis further revealed is that the SP sub-corpus was also the one to include the most traces of potential transfer-related errors. These observations can be linked to a claim often made in SLA studies according to which reliance on transfer is inversely proportional to proficiency level; that is, the lower the proficiency level, the more learners rely on transfer from their L1 (cf. e.g. Taylor, 1975; Zhang, 2003).[14] A more extensive study would need to be carried out on a wider number of L1 backgrounds and range of proficiency levels for this claim to be further substantiated.

Besides SLA, the evidence gathered from error patterns may be of relevance for a number of other fields. Pedagogically, the identification of L1-salient error types can help teachers zoom in on learner-corpus-attested areas of difficulty. This is a valuable practice across the proficiency range as it can benefit beginner and intermediate learners, who frequently commit errors, as well as advanced learners, to try and erase the more visible signs of their nonnativeness. Other applications that derive from the study of L1 patterns include improved scoring algorithms for e-rating tools (Granfeldt *et al.*, 2005) as well as heightened security by helping toward the identification of phishing texts, for example. To this end, however, it remains to be seen whether high classification accuracy can be obtained from errors that have been automatically rather than manually 'identified'.

Notes

(1) The majority class baseline is the proportion of correctly classified texts when all the texts are classified in the most frequent class in the data.

(2) The error-tagging procedure was carried out within the framework of a PhD dissertation that is under completion at the Centre for English Corpus Linguistics (Thewissen, 2011).

(3) The Louvain taxonomy includes an additional main category, namely infelicities (tagged Z), which target odd-sounding yet nonerroneous language, and which were excluded from the discriminant analysis results reported below.

(4) The error tags and corrections were inserted into the learner texts with the help of the *Université catholique de Louvain Error Editor (UCLEE)*,that is, a computer software program which includes (a) an error tag menu from which the analyst can select the appropriate tag and (b) an error correction box which speeds up the inclusion of the corrections in the texts (for more details on this, see Dagneaux *et al.*, 1998).

(5) The statistical analyses were all carried out using SAS v9.1 (SAS Institute, Cary, NC).

(6) Although we have included unclear pronominal reference in the grammatical error category, this is an error that actually targets the discourse dimension of learner writing, for example, *But there are also imprisoned people waiting for their execution who are innocent. They never had a fair trial and a real chance to get out of* (GPU) *it $jail€$. These people often do not have enough money to get their own attorney.*

(7) See Granger *et al.* (2009) for a detailed account of the proficiency-level breakdown for each of the 16 ICLE mother tongue backgrounds.

(8) The rating procedure carried out for the 223 texts investigated here (described in Thewissen, 2011) is distinct from the pilot study reported in Granger *et al.* (2009) for each of the 16 ICLE L1 backgrounds (20 texts per L1 background were given a CEF score).

(9) Recent studies (i.e. Wulff & Römer, 2009; Carlsen, forthcoming; Thewissen, 2011) have shown that *individual* learners' proficiency level is a variable that learner corpus research has paid little attention to so far and that it would need to be more tightly controlled for in future corpus compilation work.

(10) The CEF includes a total of six proficiency levels which can be broken down into three groups of two, namely A1 and A2 (=basic users; elementary proficiency learners), B1 and B2 (=independent users; intermediate proficiency learners), C1 and C2

(=proficient users; advanced proficiency learners). We started the rating procedure at level B1 because the CEF specifies that learners need a B1 proficiency level to attempt essay writing.

(11) The essays were presented to the raters in a random order (the same random order for both raters). The judges were given clear rating guidelines and carried out the rating procedure completely independently, that is, they never met to discuss results, and were not told how many L1 backgrounds were represented in the data.

(12) The scale is an 11-point rather than 12-point scale as the data did not include any C2+ score.

(13) Díez-Bedmar and Papp (2008: 155) give the following example: *Maria tiene Ø coche* vs **Mary has Ø car.*

(14) Note that some scholars have made the opposite claim (cf. e.g. Håkansson *et al.*, 2002).

References

Alderson, J.C., Clapham, C. and Wall, D. (2001) *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.

Baroni, M. and Bernardini, S. (2006) A new approach to the study of translationese: Machine-learning the difference between original and translated text. *Literary and Linguistic Computing* 21, 259–274.

Carlsen, C. (forthcoming) Proficiency level: A fuzzy variable in computer learner corpora.

Chuang, F-Y. and Nesi, H. (2006) An analysis of formal errors in a corpus of Chinese student writing. *Corpora* 1, 251–271.

Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.

Cowan, R. and Leeser, M. (2007) The structure of corpora in SLA research. In R. Facchinetti (ed.) *Corpus Linguistics 25 Years on* (pp. 289–304). Amsterdam, NY: Rodopi.

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

Díez-Bedmar, M.B. and Papp, S. (2008) The use of the English article system by Chinese and Spanish learners. In G.S. Gilquin, S. Papp and M.B. Díez-Bedmar (eds) *Linking Up: Contrastive and Learner Corpus Research* (pp. 147–175). Amsterdam: Rodopi.

Dikli, S. (2006) An overview of automated scoring of essays. *The Journal of Technology, Learning, and Assessment* 5, accessed 22 July 2010. http://www.jtla.org.

Granfeldt, J., Nugues, P., Persson, E., Persson, L., Kostadinov, F., Ågren, M. and Schlyter, S. (2005) Direkt profil: A system for evaluating texts of second language learners of French based on developmental sequences. In *Proceedings of the Second Workshop on Building Educational Applications Using Natural Language Processing* (pp. 53–60). Ann Arbor, MI.

Granger, S. (2004) Computer learner corpus research: Current status and future prospects. In U. Connor and T.A. Upton (eds) *Applied Corpus Linguistics: A Multidimensional Perspective* (pp. 291–231). Amsterdam: Rodopi.

Granger, S. and Tyson, S. (1996) Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes* 15, 19–29.

Granger, S., Dagneaux, E., Meunier, F. and Paquot, M. (2009) *The International Corpus of Learner English. Handbook and CD-ROM* (2nd edn). Louvain-la-Neuve: Presses Universitaires de Louvain.

Håkansson, G., Pienemann, M. and Sayehli, S. (2002) Transfer and typological proximity in the context of L2 processing. *Second Language Research* 18, 250–273.

Hawkins, D.M. (2004) The problem of overfitting. *Journal of Chemical Information and Computer Sciences* 44, 1–12.

Hawkins, J. and Buttery, P. (2010) Criterial features in learner corpora: Theory and illustrations. *English Profile Journal* 1, 1–23.

Koppel, M., Schler, J. and Zigdon, K. (2005) Determining an author's native language by mining a text for errors. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining* (pp. 624–628). Chicago: Association for Computing Machinery.

Lachenbruch, P.A. and Mickey, R.M. (1968) Estimation of error rates in discriminant analysis. *Technometrics* 10, 1–11.

Louw, N. and Steel, S.J. (2006) Variable selection in kernel Fisher discriminant analysis by means of recursive feature elimination. *Computational Statistics and Data Analysis* 51, 2043–2055.

Neff van Aertselaer, J. and Bunce, C. (2005) An account of lexico-grammatical errors in Spanish EFL writers' argumentative texts: Data from the International Corpus of Learner English Error Tagging Project. In M. Genis and E. Orduna (eds) *Actas del III Congreso de ACLES* (pp. 697–705). Madrid: Universidad Antonio de Nebrija.

Osborne, J. (2008) Phraseology effects as a trigger for errors in L2 English: The case of more advanced learners. In F. Meunier and S. Granger (eds) *Phraseology in Foreign Language Learning and Teaching* (pp. 67–83). Amsterdam: Benjamins.

Schiavo, R.A. and Hand, D.J. (2000) Ten more years of error rate research. *International Statistical Review* 68, 295–310.

Schulerud, H. and Albregtsen, F. (2004) Many are called, but few are chosen: Feature selection and error estimation in high dimensional spaces. *Computer Methods and Programs in Biomedicine* 73, 91–99.

Taylor, B.P. (1975) The use of overgeneralisation and transfer learning strategies by elementary and intermediate students of ESL. *Language Learning* 25, 73–107.

Thewissen, J. (2011) Accuracy across proficiency levels: Insights from an error-tagged EFL learner corpus. PhD thesis, Centre for English Corpus Linguistics, Université catholique de Louvain.

Wong, S-M.J. and Dras, M. (2009) Contrastive analysis and native language identification. In *Proceedings of the Australasian Language Technology Association* (pp. 53–61). Cambridge, MA: The Association for Computational Linguistics.

Wulff, S. and Römer, U. (2009) Becoming a proficient academic writer: Shifting lexical preferences in the use of the progressive. *Corpora* 4, 115–133.

Zhang, W. (2003) Error analysis of college students' writing. *US-China Foreign Language* December, 52–60.

Appendix

Error tags used in the present study

Error tags	*Definition*	*Learner-corpus-attested examples*
F	**Form errors**	
FS	Form spelling	*The fast spread of television can transform it into a double-edged* (FS) *wheapon $weapon$.**
FM	Form morphology	*It is* (FM) *unpossible $impossible$.*
G	**Grammatical errors**	
GDD	Grammar determiner demonstrative	(GDD) *This $These$ elements cannot be separated.*
GDO	Grammar determiner possessive	*People accept jobs according to how much they get paid but not according to* (GDO) *his $their$ preferences.*
GDI	Grammar determiner indefinite	*He does not have* (GDI) *some any expectations.*
GPP	Grammar pronoun personal	*The big majority of children have a computer or a video-game, with which* (GPP) 0 *$they$ spend (waste, in my opinion) a great number of hours.*
GPI	Grammar pronoun indefinite	*A lower-class man is not on an equal footing with his middle- or upper-class* (GPI) *one $counterpart$.*
GPF	Grammar pronoun reflexive and reciprocal	*They didn't need to communicate* (GPF) *themselves 0 outside their homes.*
GPR	Grammar relative and interrogative pronouns	The government took several measures to stop the strikes, *(GPR)* that $which$ was not effective.
GPU	Grammar pronoun unclear reference	*But there are also imprisoned people waiting for their execution who are innocent. They never had a fair trial and a real chance to get out of* (GPU) *it $jail$. These people often do not have enough money to get their own attorney.*

Error tags	*Definition*	*Learner-corpus-attested examples*
GA	Grammar article	(GA) *The 0 life is beautiful.*
GNN	Grammar noun number	*Bearing in mind that sex equality is one of the great* (GNN) *reason $reasons$ for fights in most places around the world (. . .)*
GNC	Grammar noun case	*Behind the* (GNC) *Berlin's wall $Berlin wall$.*
GADJCS	Grammar adjective comparative/ superlative	*The role that women should play in a* (GADJCS) *more fair $fairer$present-day society.*
GADJN	Grammar adjective number	The last sentences have been *(GADJN)* favourables $favourable$ to women.
GADJO	Grammar adjective order	*A* (GADJO) *leather black small $small black leather$ handbag.*
GADVO	Grammar adverb order	*They* (GADVO) *see only $only see$ other criminals.*
GVAUX	Grammar verb auxiliary	*So if there is an army it* (GVAUX) *might $should$ be professional, and formed by people who believe in that and want to dedicate their lives to it.*
GVM	Grammar verb morphology	*It is generally* (GVM) *agree $agreed$ today that we live in a world where television plays an important part.*
GVN	Grammar verb number	*How do you think that a man* (GVN) *react $reacts$ when he hears that a woman is shocked when she receives a letter beginning with 'Dear Sirs'?*
GVNF	Grammar verb nonfinite/finite	(GVNF) *To travel $Travelling$ by public transport is recommended.*
GVT	Grammar verb tense	*He learned a profession in the prisson, and now he* (GVT) *wrote $writes$ poetry and* (GVT) *took $takes$ part in the publication of a prisson's journal.*
GVV	Grammar verb voice	*This seems impossible to* (GVV) *be achieved $achieve$.*
GWC	Grammar word class	*We are going to review the following subjects: Labour discrimination, the right to vote, the fight against male* (GWC) *chauvinist $chauvinistic$ behaviours.*

(*continued*)

Error tags	*Definition*	*Learner-corpus-attested examples*
L	**Lexical errors**	
LCC	Lexical conjunction coordination	*Life is not only work* (LCC) *or and study.*
		(*continued*)
LCS	Lexical conjunction subordination	*There is also the ethical question* (LCS) *if $whether$ people generally have the right to manipulate living beings.*
LCLS	Lexis logical connector single	*The second important point is the notion of 'risks' taken by the worker.* (LCLS) *Indeed 0 we often hear that 'in the public service you earn less money than in the private'.*
LCLC	Lexis logical connector complex	(LCLC) *In consequence to this $Consequently$, people are very busy and tired.*
LS	Lexical error on single words	*Resorting to violence might be somehow a* (LS) *comprehensible $understandable$ reaction.*
LP	Lexical phrase errors	*I was riding my bike through the village when I met a Turk, who was* (LP) *of middle age $middle-aged$.*
X	**Lexico-grammatical errors**	
XADJPR	Adjectives used with wrong/ missing dependent preposition	*How many public places are easily* (XADJPR) *accessible for $accessible to$ wheelchairs?*
XNCO	Nouns used with wrong complementation	*Students have the* (XNCO) *possibility to leave $possibility of leaving$.*
XNPR	Nouns used with wrong/missing dependent preposition	*He has a* (XNPR) *thirst of $thirst for$ knowledge.*
XNUC	Errors on countable/ uncountable nouns	*The tremendous* (XNUC) *progresses $progress$ realized by science have disrupted our habits and our way of living.*
XVCO	Verbs used with wrong complementation	*What about the people who cannot* (XVCO) *afford going $afford to go$ to these kind of centres?*

Error tags	*Definition*	*Learner-corpus-attested examples*
XVPR	Verbs used with wrong/ missing dependent preposition	*The classroom must often have* (XVPR) *resembled to $resembled$ a 'Chamber of Horrors'.*
W	**Word missing/redundant/order errors**	
WRS	Word redundant single	*Actual life is very complicate and* (WRS) *extremely 0 full of worries.*
WRM	Word redundant multiple	*Others comment that* (WRM) *the fact is that 0 once you are inside, if you like it, you can even re-enlist.*
WM	Word missing	*The future soldiers make an strict physical training and* (WM) *0 sit some exams.*
WO	Word order	*Think about* (WO) *how would be your house $how your house would be$ without the last century's inventions.*
S	**Style errors**	
SU	Sentence unclear	(SU) *Beggars of reflection have power and very often they use it in such a wrong way that make of imagination become slaves or just disappear$?$.*
SI	Sentence incomplete	(SI) *Another example. $Another example is:$ Yesterday we spoke about the Gulf War (...)*
Q	**Punctuation errors**	
QC	Punctuation confusion	*Some creative people make the most of their spare time by imagining or even building things* (QC), *$.$ others create works of art which are fruits of their inmost beings.*
QM	Punctuation missing	*Physically* (QM) *0 $,$ you do not run any risks but it is very dangerous for your mind.*
QR	Punctuation redundant	*Women were seen conducting affairs, and bringing negotiations to satisfactory conclusions in what men always claimed to be* (QR) *: 0 'a man's world'.*
QL	Punctuation lexical	*He took the books* (QL) *and $,$ the records and the computers.*

*The examples are replicated as in the original. In cases where the examples contain multiple errors, only the error that represents the tag being illustrated is highlighted.