Contribution à ouvrage collectif (Book Chapter)

"From EFL to ESL: Evidence from the International Corpus of Learner English"

Gilquin, Gaëtanelle ; Granger, Sylviane

Abstract

This chapter revisits the dichotomy that is traditionally made in Second Language Acquisition (SLA) research between English as a Foreign Language (EFL) and English as a Second Language (ESL) and argues, on the basis of data from the International Corpus of Learner English, that it should be viewed as a continuum instead, with many in-between categories corresponding to a variety of learning contexts. Using the case of the preposition 'into' as an illustration, we show that the different environments in which Spanish-, French-, Dutch- and Tswana-speaking students learn English are reflected in their syntactic, semantic and lexical use of the preposition. More precisely, it appears that the Spanish-, French- and Dutch-speaking learners, who represent a cline in terms of exposure to the target language, from little exposure for the Spanish learners to considerable exposure for the Dutch learners, also form a cline in their use of 'into', from most distant to most similar to native (Bri...

Référence bibliographique

From EFL to ESL: Evidence from the *International Corpus of Learner English*

Gaëtanelle Gilquin & Sylviane Granger
Centre for English Corpus Linguistics
Université catholique de Louvain

Abstract

This chapter revisits the dichotomy that is traditionally made in Second Language Acquisition (SLA) research between English as a Foreign Language (EFL) and English as a Second Language (ESL) and argues, on the basis of data from the *International Corpus of Learner English*, that it should be viewed as a continuum instead, with many in-between categories corresponding to a variety of learning contexts. Using the case of the preposition *into* as an illustration, we show that the different environments in which Spanish-, French-, Dutch- and Tswana-speaking students learn English are reflected in their syntactic, semantic and lexical use of the preposition. More precisely, it appears that the Spanish-, French- and Dutch-speaking learners, who represent a cline in terms of exposure to the target language, from little exposure for the Spanish learners to considerable exposure for the Dutch learners, also form a cline in their use of *into*, from most distant to most similar to native (British) English. As for the Tswana variety, which clearly displays characteristics of both EFL and ESL, it occupies different positions along the cline, being sometimes closest to native English and sometimes most dissimilar, depending on the features of the use of *into* that are considered.

1. Introduction

The distinction between English as a Foreign Language (EFL) and English as a Second Language (ESL) is a long-established one in Second Language Acquisition research. Like many other distinctions (e.g. nativeness vs non-nativeness, grammar vs lexicon), the distinction was initially presented as a dichotomy, but gradual recognition of the complexity of the language learning process and its many contextual determinants has led to a more qualified view. In this paper, we argue that, far from being clear-cut, the distinction between EFL and ESL should be viewed as a continuum with many in-between categories. We demonstrate this on the basis of data from the *International Corpus of Learner English* (ICLE). While this corpus was collected in such a way as to represent EFL, some components of it contain data produced by learners who studied English in a context closer to ESL. We may expect such differences in the learning environment to be reflected in language itself. This hypothesis is tested by means of a study of a notoriously difficult preposition, the preposition *into*, in four components of ICLE: the Spanish, French, Dutch and Tswana components. These data are compared with each other and with a reference corpus of British newspaper editorials (MULT-ED).

The chapter is organised as follows: in Section 2, we define EFL and ESL and argue for considering them as two extremes on a cline rather than as a dichotomy. Next, we justify the choice of prepositions, and in particular the preposition *into*, as a subject for studying variation among several learner varieties. Section 4 consists in the corpus analysis proper, with results for the frequency of *into*, its syntactic, lexical and semantic behaviour, as well as its phraseological and non-standard uses. In Section 5, we introduce another distinction, that between novice and expert writing and, using data from LOCNESS (*Louvain Corpus of*

Native English Essays), we briefly consider the role of the degree of expertise, and its relation with the nativeness/non-nativeness distinction. Section 6 concludes the paper.

2. From EFL to ESL

The general framework within which this chapter is situated is that of Second Language Acquisition (SLA), i.e. the learning of a language after the first language has been learned. Within that context, we follow Gass and Selinker (2001: 5) in using the term Foreign Language to refer to "the learning of a nonnative language in the environment of one's native language" and Second Language to "the learning of a nonnative language in the environment in which that language is spoken". Like them, however, we also recognise that the picture is more complex as the degree and type of exposure may vary considerably in the two learning contexts. The *International Corpus of Learner English* (ICLE), which we have used for this study, is essentially a corpus of writing by learners of English as a Foreign Language rather than Second Language (cf. Granger et al. 2009). The corpus contains argumentative essays produced by higher intermediate to advanced learners from 16 different mother tongue backgrounds (Bulgarian, Chinese, Czech, Dutch, Finnish, French, German, Italian, Japanese, Norwegian, Polish, Russian, Spanish, Swedish, Tswana, Turkish). It has been used to investigate a wide range of lexical, grammatical and discourse features of L2 writing. Researchers have either focused exclusively on one learner population (e.g. Neff van Aertselaer's [2008] study of interpersonal discourse phrases in Spanish learner writing) or compared two or more of them (e.g. Lozano & Mendikoetxea's [2008] study of postverbal subjects in Italian and Spanish learner writing). Another option consists in using several ICLE subcorpora (or all of them) and treating them as an aggregate, irrespective of the learners' mother tongue. Nesselhauf (2009), for example, uses a corpus, called ICLE-4L1, which contains data from German, French, Finnish and Polish learners. This corpus, however, is analysed with no distinction between the four L1 components and serves as a basis to draw conclusions about "learner English" in general.

Although ICLE is essentially an EFL corpus, it is important to bear in mind that there are a number of factors that blur the line between the two situations, amongst them the presence or absence of language instruction (in the case of ESL), the number of years of instruction, the focus of language lessons (focus on form and/or communication), the use of the target language for some or all of the non-language subjects (for EFL), the quality of teacher talk, the type and amount of exposure to the target language outside the classroom, in particular access to English-speaking media and in the case of EFL learners, the amount of time spent in a country where English is spoken. For our study, we selected four populations – Spanish, French, Dutch and Tswana – that occupy different points along the EFL-ESL cline with regard to two main factors: amount of exposure to the target language and focus of language instruction. In order to ensure comparability of the data, we controlled for one other factor, namely the number of months spent in a country where the target language is spoken. Using the ICLE interface, we only selected texts produced by learners who had spent a maximum of three months in an English-speaking country.

As shown in Table 1, the Spanish- and French-speaking learners are characterised by a relatively low amount of exposure to English. As regards the media, films and television serials are dubbed and the English in TV programmes, notably the news, is voiced over. Although nowadays the Internet is a source of potential contact with English, the ICLE data were collected before it became a major issue. The amount of exposure received by the French-speaking students represented in ICLE is arguably somewhat higher than that of the

Spanish learners as all the linguistics and literature courses of their English philology degree are taught via the medium of English, while practices on this point vary in Spain. By contrast, the Dutch and Tswana learner populations benefit from a much higher degree of exposure. The Dutch learners get to hear a lot of English as all the films, TV shows and soaps are subtitled rather than dubbed and English speakers on TV programmes are not voiced over. According to Koolstra & Beentjes (1999: 16), Dutch children spend about half their TV time watching programmes with English language sound. Ginsburgh & Weber (2006) attribute the much higher proficiency level of Dutch-speaking compared to French-speaking Belgians largely to that factor. The positive effect of undubbed TV programmes has led Van Parijs (2004) to launch a 'ban dubbing' campaign, which has unfortunately not had much impact so far in French-speaking Belgium or France. Besides the potential benefit gained from access to English-speaking media, the Tswana learners have an additional advantage as classes are taught through the medium of English from the fifth grade in primary school. While in primary school, code-switching between Setswana and English is the norm, in high school English instruction is dominant (cf. Van Rooy 2009: 199).

Table 1. ICLE subcorpora: EFL-ESL cline

ICLE subcorpus	Exposure	Focus on form
SP	-	+
FR	+/-	+
DU	+	+
TSW	+	-

As regards focus on form in the language classroom, the Spanish, the French and the Dutch cluster together in having their attention directed to morphological, grammatical and lexical accuracy. By contrast, the immersion education received by the Tswana learners leads to a high level of functional proficiency in English but, as demonstrated in numerous studies of immersion programmes, this advantage is counterbalanced by a much lower degree of (especially grammatical) accuracy and "endemic fossilization" (Sheen 2006: 828). A factor that further complicates the issue in the case of the Tswana learners is that they may be influenced by the emerging variety of Black South African English (Van Rooy 2009), notably via their own teachers who are predominantly Setswana speakers, not English speakers.

Applying the Contrastive Interlanguage Analysis (CIA) methodology (Granger 1996, Gilquin 2000/2001), we carried out two types of comparison: a comparison between learner corpora and a comparison between learner corpora and reference corpora. Table 1 gives the breakdown of the corpora used.

Table 2. Breakdown of learner and reference corpora

Corpus		No. of words
Learner corpora	ICLE-SP	156,840
	ICLE-FR	182,328
	ICLE-DU	192,771
	ICLE-TSW	199,380
	TOTAL	*731,319*
Reference corpora	BrE editorials	152,123
	SAE editorials	150,401
	LOCNESS	150,590
	TOTAL	*453,114*
TOTAL		*1,184,433*

The four learner corpora, which are extracted from the second version of ICLE (Granger et al. 2009), contain essays written by higher intermediate to advanced learners with Spanish, French, Dutch or Tswana as a mother tongue. They were compared to three reference corpora. Two of these contain editorials from British and South African English newspapers respectively and are part of a larger corpus collected at Louvain, the *Multilingual Editorials Corpus* (MULT-ED).[1] The third reference corpus is the *Louvain Corpus of Native English Essays* (LOCNESS), a corpus of essays written by American English students. As the analysis of the data hardly showed any difference between the British and South African English corpora,[2] we will only report the results for the British component (cf. Section 4). The comparison with LOCNESS will be the subject of Section 5.

Based on the configuration represented in Table 1, we hypothesised that the Dutch learner population would be closest to the reference corpora, followed by the French and the Spanish. In view of the mixed configuration displayed by the Tswana group, no hypothesis was formulated for that learner population.

3. The preposition *into*

In this study, we focus on the use of prepositions, and more precisely the preposition *into*. Prepositions have been shown to be problematic for non-native speakers of English. Kao (2001), for example, has demonstrated that communicatively redundant prepositions are likely to be omitted by learners. Many SLA specialists have also underlined learners' tendency to avoid prepositions (e.g. Hulstijn & Marchena 1989, Sjöholm 1995, Liao & Fukuya 2004, Siyanova & Schmitt 2007). In fact, prepositions are often considered as the *bête noire* of both teachers and learners, being impossible to teach and impossible to learn. Prepositions also have a special status in indigenised varieties of English (World Englishes), where they can be described both as a 'mutating species' and an 'endangered species': 'mutating species' because they are likely to lead to innovations in World Englishes (see, e.g., Mukherjee 2009 on new prepositional verbs in Indian English), and 'endangered species' because some prepositions tend to vanish (see Mwangi 2003, 2004 on Kenyan English).

Into is a particularly interesting preposition from an SLA perspective, because of its obvious link with (and hence possible confusion with) the preposition *in*. Thus, it is common for learners to use *in* instead of *into*, especially due to the lack of a similar contrast in the learner's mother tongue, cf. Swan's (2005: 244) example: *The ball rolled slowly *in the goal*. The distinction between *in* and *into* also seems to be gradually disappearing from some indigenised varieties of English, e.g. *There are so many people just coming **in** the country* (Mwangi 2004: 28).

In what follows, we will investigate several aspects of the use of *into* in ICLE (and the reference corpus of British English), namely its frequency, the syntactic structures in which it appears, the lexical variation it displays, the senses in which it is used, its phraseological uses, and its non-standard uses.

4. Quantitative and qualitative analysis of *into* in ICLE

4.1. *Frequency*

The relative frequency of *into* per 100,000 words ranges from 146 (in the corpus of British editorials, BrE) to 71 (in ICLE-TSW). If we examine the full range of results, shown in Figure 1, we notice that three groups seem to emerge: one with the native corpus, one with ICLE-DU and ICLE-FR, and one with ICLE-SP and ICLE-TSW.[3] These results confirm the general underuse of the preposition *into* in the non-native varieties of English. However, they also reveal a great disparity among the ICLE varieties, with a mild underuse in the Dutch and French subcorpora, and a marked underuse in the Spanish and Tswana subcorpora. With respect to our initial hypothesis, we see that the cline between ICLE-DU, ICLE-FR and ICLE-SP is confirmed, and that ICLE-TSW, which shows a mixed configuration in terms of learning context, comes last, after the Spanish learners.
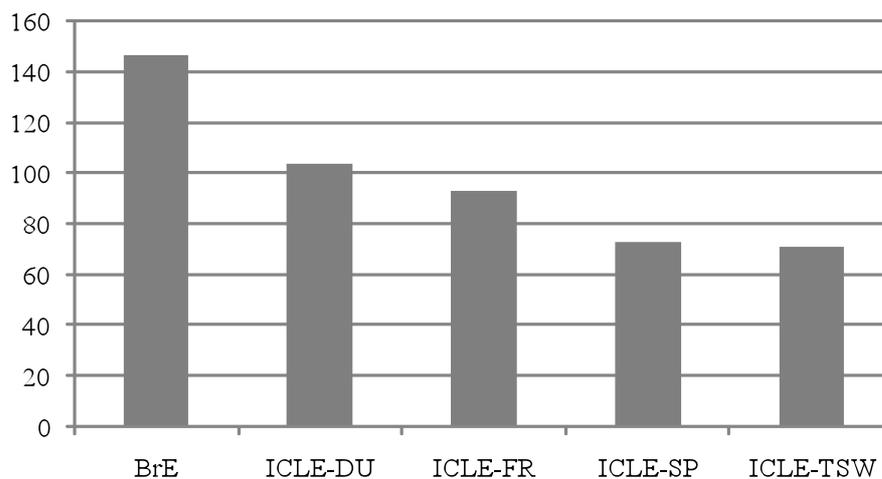

Figure 1. Relative frequency of *into* per 100,000 words

4.2. *Syntactic structures*

From a syntactic point of view, a distinction may be drawn between three main structures: N + *into*, $V_{intrans}$ + *into* and V_{trans} + *into* (+ NP or V_{ing}), as illustrated by the following sentences:

(1) It does not need much research **into** Labour Party history to see that pounds and pence have been the downfall of many previous Prime Ministers of the Left. <BrE>

(2) But on many issues his generally admirable resoluteness has descended **into** pig-headed obstinacy. <BrE>

(3) a. Nothing looked more certain than that he would lead his party **into** the next general election, due in 2005 or, at the latest, in 2006. <BrE>

   b. Others have ended because both sides were exhausted, or because outsiders cajoled them **into** putting down their weapons and starting to talk. <BrE>

Not too surprisingly, the three structures display an underuse among learners (although to different degrees), as appears from Table 3. This underuse is particularly pronounced in the Spanish and Tswana subcorpora and less strong in the French and (especially) Dutch subcorpora, which corroborates the results for the overall frequency of *into*.

Table 3. **R**elative frequencies (per 100,000 words) and raw frequencies of syntactic structures with *into**

	BrE	**ICLE-DU**	**ICLE-FR**	**ICLE-SP**	**ICLE-TSW**
N + *into*	12.49 (19)	6.23 (12)	5.48[(*)](10)	2.55[***] (4)	1.50[***] (3)
$V_{intrans}$ + *into*	49.96 (76)	43.06 (83)	24.13[***] (44)	19.77[***] (31)	29.09[**] (58)
V_{trans} + *into*	82.17 (125)	54.46[**] (105)	63.08[(*)] (115)	49.73[***] (78)	39.62[***] (79)

* The asterisks indicate the degree of statistical significance of the log-likelihood test (BrE vs ICLE): (*) for $p<0.05$, * for $p<0.01$, ** for $p<0.005$ and *** for $p<0.001$.

The figures for the transitive use of *into* (V_{trans} + *into*), however, hide an interesting variation. If we consider the causative use of *into*, i.e. V_{trans} + *into* + V_{ing}, as exemplified by (3b), we notice that, while this structure is generally rare among learners, it is used more often by the Dutch and Tswana learners, less often by the French-speaking learners, and it never occurs in the Spanish component of ICLE. The Tswana learners, who present the lowest frequency of *into*, thus appear to do comparatively well when it comes to the causative use of *into*. These results also show that, while the Dutch, French and Spanish learners seem to have a relatively fixed position along the cline predicted on the basis of the learning context, the Tswana learners appear to occupy varying positions depending on the feature that is investigated: further down the cline if we consider the overall frequency of *into* (see Figure 1), but closer to the native speakers if we examine the causative use of *into*.
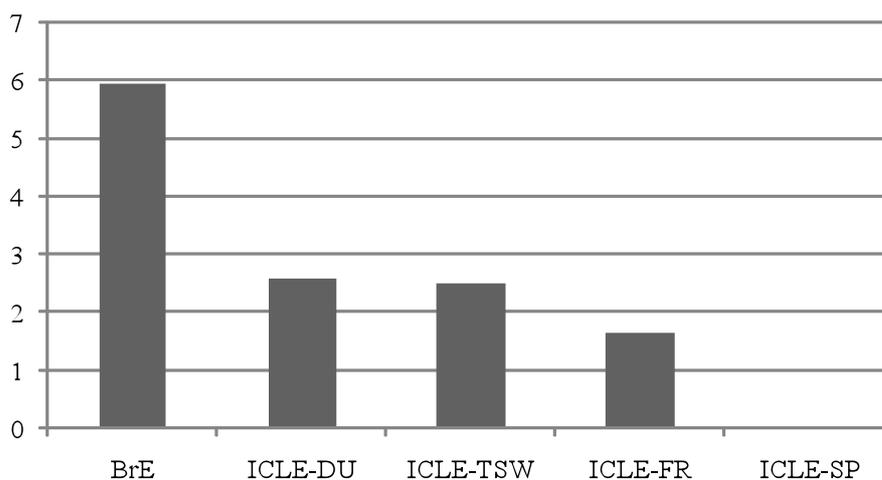
Figure 2. Relative frequency of causative structures with *into* per 100,000 words

4.3. *Lexical variation*

Limiting ourselves to the verbal structures with *into* ($V_{intrans}$ + *into* and V_{trans} + *into*), we then examined the lexical variation displayed by the verb among the four learner populations. In order to do so, we calculated the number of different lemmas per 100,000 words in the ICLE subcorpora, and compared these results with the results for the British English reference corpus. As appears from Figure 3, the learners use significantly fewer lemmas than the native speakers (log-likelihood value = 16.40, $p<0.001$), which suggests repetition of a limited repertoire of verbs with *into* – a phenomenon which, incidentally, has been brought to light for other aspects of the learner's lexicon (see, e.g., Hasselgren 1994). While the Dutch learners exhibit slightly better results, there is, generally speaking, little variation among the

four groups of learners (no significant differences according to the log-likelihood test, except for a marginally significant one between ICLE-DU and ICLE-SP, $p<0.05$).
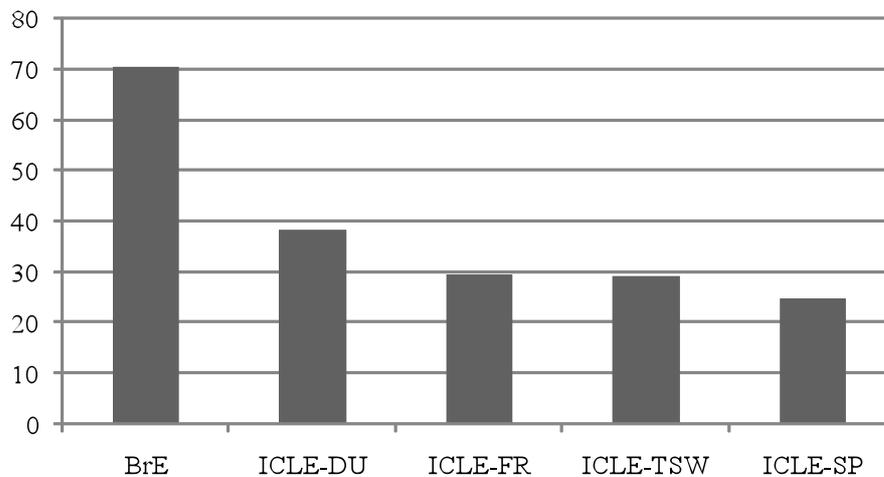

Figure 3. Relative frequency of verb lemmas with *into* per 100,000 words

Table 4 lists the most frequent verb lemmas in the five corpora.[4] It will be noticed that, despite the higher frequency of verbal structures with *into* in the British corpus, the list of lemmas occurring five times or more in this corpus is shorter than the list for ICLE-DU and ICLE-FR, which points to a lower degree of repetition in native English. In ICLE-SP and ICLE-TSW, the list of frequently recurring lemmas is not so long, but the top lemmas appear to be extremely common (*take* occurs 23 times in ICLE-TSW and as many as 27 times in ICLE-SP). Another phenomenon worth underlining is the higher degree of recurrence of certain verbs in the ICLE varieties as compared to the reference corpus, in particular *take* and *put* (which have been highlighted in the table). Some examples are given in (5) to (8).[5]

Table 4. Top verb lemmas (5 occurrences or more)

BrE	ICLE-DU	ICLE-FR	ICLE-SP	ICLE-TSW
turn (21)	turn (19)	**take (22)**	**take (27)**	**take (23)**
bring (9)	go (15)	turn (16)	***put (12)***	come (9)
pour (9)	**take (15)**	***put (12)***	divide (10)	look (8)
go (7)	***put (14)***	divide (10)	turn (8)	turn (7)
break (5)	come (12)	bring (9)	fall (6)	***put (6)***
fall (5)	get (11)	come (9)		
get (5)	bring (6)	get (8)		
	change (5)	transform (8)		
	force (5)			

(5)   Another important fact to *take* **into** account is the bad treatment received while doing the military service. <ICLE-SP>

(6)   My view is that the economy of the country should be *taken* **into** consideration. <ICLE-TSW>

(7)   To conclude money is useful, as long as it helps people to respect true human values and to *put* them **into** practice. <ICLE-FR>

(8)   All this must be *put* **into** perspective, of course. Religion as well as television can be very positive and valuable in life. <ICLE-DU>

4.4. *Semantic analysis*

In order to study the semantic behaviour of *into*, we distinguished between eight senses and categories of use of the preposition, using a dictionary- and corpus-derived methodology similar to that described in De Cock & Granger (2004).[6] These are listed in Table 5 and illustrated with examples from the corpus of British newspaper editorials.

Table 5. Semantic classification of *into*

Sense/use	Example
Movement	Max Hastings *marched* **into** Port Stanley at the head of a column of British troops.
Abstract movement	And they are not only *moving* **into** manufacturing – they are increasingly competing in services, too.
Transformation	*Turning* the entire country **into** a focus group won't solve any real problems of government.
Causation	The insurers hope that taking a hard line will *prod* a capricious Government **into** taking flood protection more seriously.
Division	The Eskimos *compartmentalise* their flakes **into** fine, fresh, drifting, clinging or crusted.
Other meanings	The new editors have declined to extend the story **into** the 20th century.
Phrasal verbs	Law-abiding citizens should have a greater entitlement to take action against burglars who *break* **into** their homes.
(Semi-)fixed expressions	A graduate tax *takes* **into** *account* the level of an individual's earnings.

The semantic analysis of the corpora (see Table 6) reveals that the prototypical sense of concrete movement is never predominant. Instead, the most frequent sense in the reference corpus is abstract movement, as in (9), closely followed by the sense of transformation, whereas in the four ICLE components it is the (semi-)fixed expressions that are most common. Several senses are significantly underused by all four groups of learners, namely abstract movement, transformation and causation (which may be related to the underuse of the causative structure, since most of the uses of *into* expressing causation are of this type). Most of the other results point to a lower frequency among the learners too, even though they do not always reach the threshold of statistical significance. This widespread underuse stands in stark contrast to the overuse of (semi-)fixed expressions, which are used with a relative frequency ranging from 19.56 to 27.42 per 100,000 words in ICLE, as against 18.41 in the reference corpus. The only other sense that is found more often among some of the learners is division, significantly overused by the French- and Spanish-speaking learners, e.g. (10).

Table 6. Relative frequency of senses/uses of *into* per 100,000 words (and percentages)

	BrE	ICLE-DU	ICLE-FR	ICLE-SP	ICLE-TSW
Movement	15.12	11.93	9.32	7.01[(*)]	6.52[(*)]
	(11.4%)	(12.2%)	(10.7%)	(10.1%)	(9.5%)
Abstract	36.15	22.31[(*)]	9.87[***]	7.01[***]	18.06[***]
movement	(27.4%)	(22.9%)	(11.3%)	(10.1%)	(26.3%)

Transfor-mation	34.84	22.31[(*)]	23.04[(*)]	15.3[***]	8.53[***]
	(26.4%)	(22.9%)	(26.4%)	(22.0%)	(12.4%)
Causation	14.46	4.67[**]	3.29[***]	0[***]	3.51[***]
	(10.9%)	(4.8%)	(3.8%)	(0.0%)	(5.1%)
Division	1.97	2.59	8.23[(*)]	8.29[(*)]	0.5
	(1.5%)	(2.7%)	(9.4%)	(11.9%)	(0.7%)
Other meanings	3.29	1.04	2.74	4.46	6.02
	(2.5%)	(1.1%)	(3.1%)	(6.4%)	(8.7%)
Phrasal verbs	7.89	7.78	3.29	1.28[*]	6.02
	(6.0%)	(8.0%)	(3.8%)	(1.8%)	(8.7%)
(Semi-)fixed expressions	18.41	24.9	27.42	26.14	19.56
	(13.9%)	(25.5%)	(31.5%)	(37.6%)	(28.5%)

(9) Instead of *falling* **into** the easy temptation to also posture grandly, Mr Blair should seek a constructive relationship with the unions while standing firm on his policy agenda. <BrE>

(10) In this case, the believers are *divided* **into** two groups: catholics and protestans. <ICLE-SP>

4.5. *Phraseological uses*

In the preceding section, we saw that phraseological usage plays a prominent role in the learners' use of *into*. In this section, we zoom in on two types of phraseological uses of the preposition, namely its use in (semi-)fixed expressions and its use in phrasal verbs. We use the term '(semi-)fixed expression' to refer to those expressions that are described as such in the *Oxford Advanced Learner's Dictionary* (e.g. *take into account, burst into the open, play into the hands of*), whereas phrasal verbs are defined here as non-compositional prepositional verbs (e.g. *look into, feed into*). Interestingly, these two phraseological uses show divergent patterns of use in ICLE, with a high frequency of (semi-)fixed expressions and a relatively low frequency of phrasal verbs.

Figure 4 shows the relative frequency of (semi-)fixed expressions in native English and learner English (see also Table 4). Although the results are not statistically significant and we would clearly need more data on which to base the analysis, it is interesting to notice that the four learner populations make a greater use of (semi-)fixed expressions than native speakers, especially the French- and Spanish-speaking learners and, to a lesser extent, the Dutch-speaking learners. The Tswana learners come closer to the standard set by the native speakers (as was the case for the causative use of *into*). This seems to contradict the common claim that a great deal of exposure is necessary in order to acquire formulaic expressions, since the learners with the least exposure (French- and Spanish-speaking learners) show the highest frequency of (semi-)fixed expressions. There are at least two possible explanations for this apparent contradiction.[7] One is that such expressions often have a direct equivalent in the learners' mother tongues. A detailed contrastive analysis would be needed in order to support this hypothesis, but the equivalence between, e.g., *take into account* and French *prendre en compte* or Spanish *tener en cuenta* seems to point in this direction.
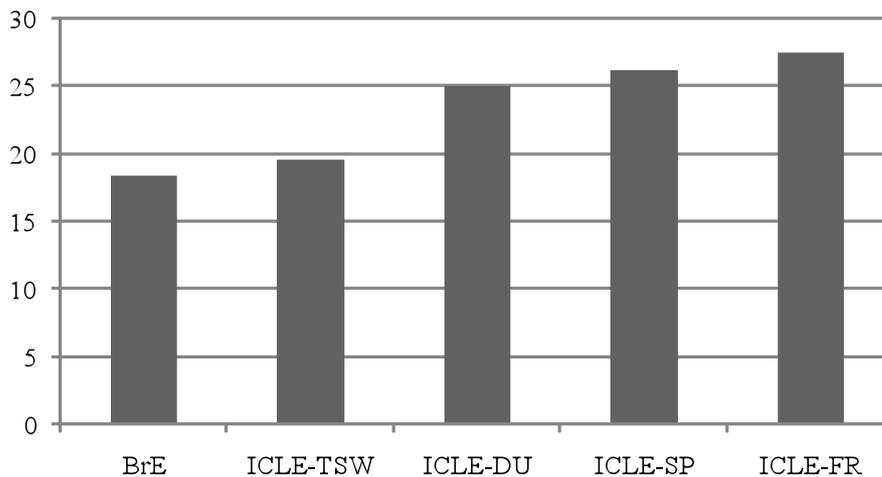
Figure 4. Relative frequency of (semi-)fixed expressions with *into* per 100,000 words

Another possible explanation for our results is that the learners are likely to repeat expressions that are familiar to them and appear to be safe. This explanation is confirmed by Table 7, which lists the most frequent (semi-)fixed expressions with *into* in the four ICLE subcorpora, together with their absolute frequencies and the percentage they represent in the total of (semi-)fixed expressions. It turns out, among other things, that the use of *take into account* accounts for 50% of all (semi-)fixed expressions with *into* in ICLE-FR, and that in ICLE-SP this proportion reaches 54%. In native British English, by contrast, *take into account* occurs only three times, which represents a percentage of some 10%, and none of the (semi-)fixed expressions with *into* is repeated more than three times. The strong preference for certain expressions in learner English is reflected in the results for the type/token ratio of (semi-)fixed expressions with *into* (Table 8) – although the results also reveal a continuum among the learner populations, with the type/token ratio being lower, and hence repetition being more likely in ICLE-FR and ICLE-SP than in ICLE-TSW and ICLE-DU. This continuum, incidentally, corresponds to the continuum predicted on the basis on the learning context, with the Dutch learners coming closer to the native speakers, and the French and Spanish learners lagging behind. As was the case with some of the other features investigated (but not all of them), the Tswana learners turn out to be relatively high on the continuum, coming just after the Dutch learners.

Table 7. Most frequent (semi-)fixed expressions with *into*

Corpus	Expression	Frequency
ICLE-DU	*take into account*	9 (19%)
	come into being	4 (8%)
	take into consideration	4 (8%)
ICLE-FR	*take into account*	25 (50%)
	put into practice	7 (14%)
ICLE-SP	*take into account*	22 (54%)
	put into practice	6 (15%)
ICLE-TSW	*take into consideration*	16 (41%)

Table 8. Type/token ratio of (semi-)fixed expressions with *into*

Corpus	TTR

BrE	0.79
ICLE-DU	0.50
ICLE-TSW	0.41
ICLE-FR	0.28
ICLE-SP	0.22

It is noteworthy that the learners' preference for certain (semi-)fixed expressions may vary from one group of learners to the other. Table 9 compares the frequency of two synonymous expressions, *take into account* and *take into consideration*, in the four ICLE subcorpora under study. While the French- and Spanish-speaking learners show a marked preference for *take into account*, the Tswana learners clearly prefer the alternative expression *take into consideration* (in the Dutch subcorpus the frequency of these two expressions is closer to the reference corpus). This finding partially confirms the tendency, already noted by Sand (2005) for World Englishes, to reduce functionally equivalent variants to "a small number of choices or a single preferred variant" – with the additional caveat that the preferred variant may vary depending on the learner's mother tongue. At the same time, our results contradict Nesselhauf's (2009) conclusion that *take into consideration* is the preferred option for learners (in general), and hence underline the danger of treating several learner populations as an aggregate.[8]

Table 9. Frequency of *take into account* and *take into consideration*

	BrE	**ICLE-DU**	**ICLE-FR**	**ICLE-SP**	**ICLE-TSW**
take into account	3	9	25	22	1
take into consideration	0	4	2	5	16

In contrast to (semi-)fixed expressions with *into*, which are more frequent in learner English than in native English, phrasal verbs with *into* tend to be underused by the learners. Figure 5 shows that this is the case in ICLE-SP and ICLE-FR and, to a lesser extent, ICLE-TSW (although, again, the differences are not statistically significant, except for the difference between BrE and ICLE-SP, significant at the 0.005 level); the Dutch learners use approximately the same number of phrasal verbs as native speakers (see Table 6 for the exact figures). The amount of exposure may explain the difference observed between the Dutch and Tswana learners on the one hand and the French and Spanish learners on the other, as a high degree of exposure to the target language is said to be necessary in order to acquire phrasal verbs (Sjöholm 1995). The influence of the mother tongue may also be at work and account for the particularly good results of the Dutch learners, who have phrasal verbs in their mother tongue, unlike the other three groups of learners (see Waibel [2007] on the influence of the mother-tongue background on the use of phrasal verbs). Whatever the reason(s) for these results, however, it is remarkable that the Dutch, French and Spanish ICLE subcorpora, once again, are ordered as predicted in Section 2. As for the Tswana subcorpus, it occupies an intermediate position, being situated in-between ICLE-DU and ICLE-FR.
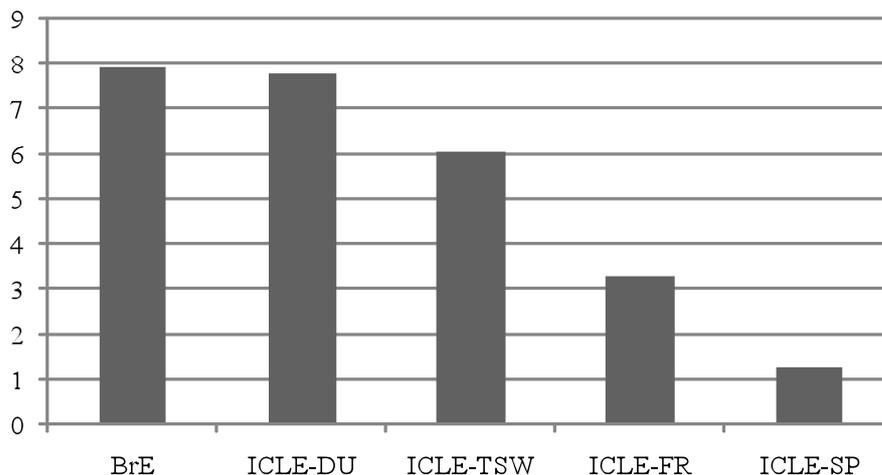
Figure 5. Relative frequency of phrasal verbs with *into* per 100,000 words

As is the case with (semi-)fixed expressions, we notice a tendency among the learners to repeat a small number of different phrasal verbs. Table 10 displays the type/token ratio of phrasal verbs with *into* in the four ICLE subcorpora and in the reference corpus. The type/token ratios for ICLE-FR and ICLE-SP have been put between brackets, as they correspond to only six and two tokens respectively (by contrast, there are over ten tokens in the other subcorpora). The type/token ratio in ICLE-TSW turns out to be particularly low (0.33). In the data, this translates into a very high degree of repetition of the phrasal verb *look into* (67% of all the phrasal verbs found in ICLE-TSW).

Table 10. Type/token ratio of phrasal verbs with *into*

Corpus	TTR
ICLE-FR	(1)
ICLE-SP	(1)
BrE	0.58
ICLE-DU	0.47
ICLE-TSW	0.33

4.6. *Non-standard uses*

We also examined the non-standard uses of *into*. Although the results should be seen as merely indicative, since they rely on the judgement of one native speaker only,[9] they still reveal interesting findings. As appears from Table 11, two ICLE varieties stand out: ICLE-SP with almost 16% of non-standard uses and ICLE-TSW with over 30%. In ICLE-FR and ICLE-DU, the non-standard rate stays well under 10%.

Table 11. Proportion of non-standard uses of *into*

Corpus	Non-standard uses
ICLE-FR	4.7%
ICLE-DU	6.5%
ICLE-SP	15.8%
ICLE-TSW	30.5%

The low proportion of non-standard uses among the French- and Dutch-speaking learners could be due to a "play-it-safe" strategy (Hulstijn & Marchena 1989): *into* is only used when the learners feel confident that they can use it. It could also be a reflection of a higher proficiency level and/or greater attention to form/accuracy during instruction (see Table 1). As for the many non-standard uses found in ICLE-SP and ICLE-TSW, they seem to have several origins. One of them is the confusion between static *in* and directional *into*, as illustrated by (11) and (12). In the Tswana subcorpus, this also happens with more fixed uses, e.g. (13).[10]

(11)   The great problem of the prisioning is located **into** the jails or cells. <ICLE-SP>

(12)   People on the continent of Africa find themselves fallen or trapped **into** the net of HIV/AIDS due to the fact that, most africans are very poor, hence they cannot affort a living. <ICLE-TSW>

(13)   This resulted **into** one scarverging for employment in other the cope with advanced life in the city. At the end it encourages prostitution. <ICLE-TSW>

Another source for non-standard usage is interference from the mother tongue. (14) and (15) are two instances of transfer of phraseological expressions from Spanish. The literal translation of *fall into account* in Spanish, *caer en la cuenta*, means 'to realise', and *put into relevance* also has a word-for-word equivalent in Spanish, *poner en relevancia*, meaning 'to highlight'.

(14)   Both men feel very bad, because they fall **into** account that they have treated very badly Mr. Hardcastle. <ICLE-SP>

(15)   She represents just another human being who has died because she wanted a change, and she dies because her world was not prepared for that change, this is put **into** relevance in the epilogue and in the final sentence of the play put in her mouth. <ICLE-SP>

In addition, there are a number of instances, especially common in the Tswana subcorpus, where the non-standard use seems to be the result of creativity on the part of the learner. This is the case in the following sentences, all taken from ICLE-TSW:

(16)   I plea to South African football association to take soccer **into** a serious consideration. <ICLE-TSW>

(17)   In Uganda the government has tried to fight HIV/AIDS and this has come **into** fruition. <ICLE-TSW>

(18)   Africa is by and by moving towards its last grave, this is due to the following unoticed facts, yet not taken **into** seriusness: Poverty is the ambrella "word" and it has other contributary factors which include the following, unemployment, wars and language. <ICLE-TSW>

(19)   Safa should arrange with companies to request them to assist the clubs or sponsor them, therefore the attracting force at European teams must come **into** fiasco. <ICLE-TSW>

Although they do not belong to the repertoire of expressions with *into* in standard English, these expressions are perfectly understandable and thus enable the speaker to get his/her message across. Often, they seem to result from the extension of existing patterns (e.g. *come*

into fruition [17] and *come into fiasco* [19] seem to be built by analogy with expressions like *come into being* or *come into contact*) and/or blends (e.g. *take into seriousness* in example [18] could be interpreted as a blend of *take into consideration* and *take seriously*). In fact, we may wonder whether such creative uses should be considered as real errors, or rather as new prepositional verbs. To further illustrate this, consider the two examples below:

> (20)   The most important novels written by women were written by people without any experience of life that could enter **into** the house of a respectable clergyman. <ICLE-SP>
>
> (21)   Soccer players don't have to rely on soccer only they can open up their businesses, enter **into** the corporate world. <ICLE-TSW>

In both cases, the verb *enter* is followed by a noun phrase representing a place (*the house of a respectable clergyman, the corporate world*), a use which normally does not require the preposition *into*, but which seems to be licensed by the existence of expressions like *enter into partnership* or *enter into discussion*s, through a process of "semantico-structural analogy" (Mukherjhee & Hoffmann 2006: 166). The fact that this expression occurs several times in the Spanish and Tswana subcorpora forces us to reconsider its exact status, as does the presence of the expression in other ICLE subcorpora (the German component in particular), as well as in corpora representing indigenised varieties of English (including Singapore and Kenyan English, cf. Nesselhauf 2009). The line is thin between errors and creative uses (see also Rimmer 2008). Yet, one must recognise that non-native speakers are often denied the right to creativity. As Bamgbose (1998: 1) aptly puts it, "[i]nnovations in non-native Englishes are often judged not for what they are or their function within the varieties in which they occur, but rather according to how they stand in relation to the norms of native Englishes. To this extent, it is no exaggeration to say that these innovations are torn between two sets of norms". Mukherjee (2009) recommends upholding "the distinction between 'norm-developing' L2 speakers and 'norm-dependent' foreign-language learners of English", which amounts to interpreting departures from native standards as errors in the case of learner English and as creative innovations in the case of institutionalised L2 varieties. While descriptive studies such as this one or Nesselhauf's do not solve the problem of how to treat this type of usage, they at least have the merit of drawing attention to this crucial issue by highlighting the commonalities across several varieties of English.

5. Novice vs expert writing

The control corpus used in our study is a corpus of expert native writing. Some linguists, among others Hyland & Milton (1997), Lorenz (1999) and McCrostie (2008), have criticised this type of reference variety on the basis that it sets too high a standard for EFL learners and suggested using a corpus of native student writing instead. To assess the impact of the native variety on the results, we revisited the analysis of *into* using the *Louvain Corpus of Native English Essays* (LOCNESS) as comparable data.[11] If, as demonstrated in several studies (cf. Hyland & Milton 1997 and Neff van Aertselaer 2008), native and non-native students share a large number of novice writer characteristics, many of the differences highlighted in Section 4 might disappear.

The results paint a varied picture. For a number of features there is no difference between the two native varieties. For example, novice native writers display the same frequency of use of *into* as expert native writers (cf. Figure 6) and a high degree of similarity

in the use of syntactic structures. As regards lexical variation, however, the lemma frequency displayed by novice native writers stands midway between expert writers and EFL learners (see Figure 7). This in-between status is confirmed by the results of the semantic analysis. On the one hand, LOCNESS is similar to BrE (and differs from the ICLE subcorpora) in having abstract movement as the most frequent sense. On the other, it is closer to the ICLE subcorpora in having (semi-)fixed expressions as the second most frequent sense, which suggests that "chunkiness" might be a transient feature in the acquisition of literacy.
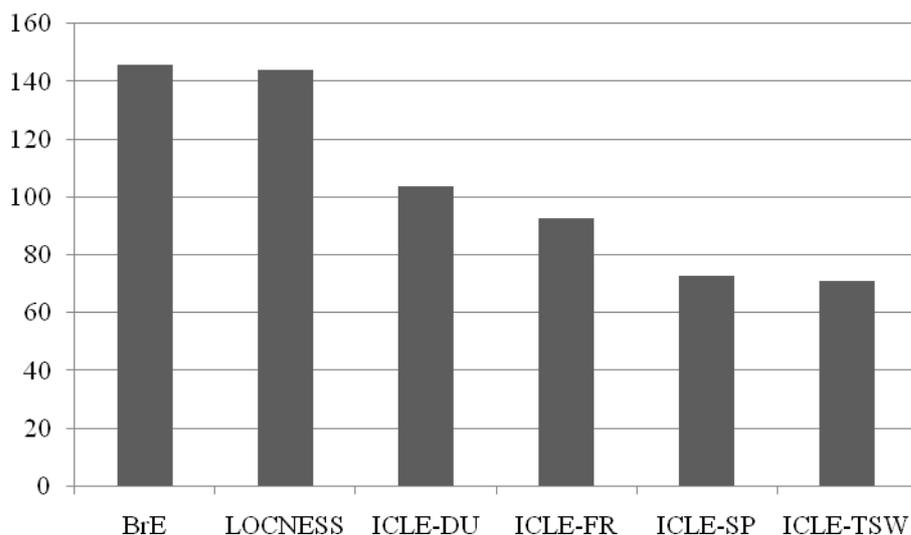

Figure 6. Relative frequency of *into* per 100,000 words (with LOCNESS)
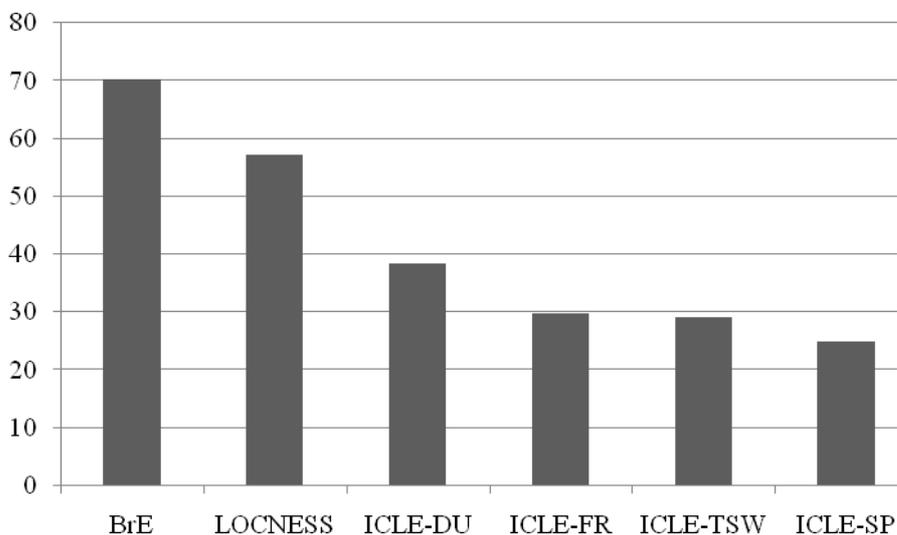

Figure 7. Relative frequency of verb lemmas with *into* per 100,000 words (with LOCNESS)

While confirming the fuzzy nature of the native/non-native distinction, our results show that the distinction cannot simply be abandoned in favour of one undifferentiated category of 'novice writers'. EFL learners prove to display a number of unique characteristics

that are not found in novice native writing and require dedicated pedagogical attention (see Gilquin et al. 2007 for further discussion of this issue).

6. Conclusion

Our study shows that the concept of 'learner English' needs to be broken down. Depending on a series of factors, notably the amount of exposure to the target language and the focus of language teaching, learner varieties display different degrees of similarity with the reference corpora. The term 'learner English**es**' reflects this diversity and is therefore more appropriate than the cover term 'learner English'.

Table 12 summarises the main results of the corpus analysis by showing how the different learner varieties are related to each other and to the reference variety with respect to a number of syntactic, semantic and lexical features. A mere glimpse at the table is enough to make obvious a number of striking similarities and differences. First, the expert native writing reference corpus (BrE) clearly stands out from all the learner varieties. Second, the Dutch, French and Spanish learner corpora display a high degree of consistency while the Tswana variety occupies a range of different positions. Third, our hypothesis for the Dutch, French and Spanish varieties is largely confirmed: the Dutch learners are the closest to the reference corpus, followed by the French and the Spanish. Fourth, Tswana learner English, for which we found it hard to make any predictions, presents both similarities and differences with the other ICLE varieties. One particularly striking finding is the closeness between ICLE-TSW and ICLE-DU, the two learner populations that have benefited from a high degree of exposure to the target language, albeit of a different nature. This closeness was also established in a study of the passive (Granger 2009), which brought out a much more frequent use of the passive by the Tswana and Dutch learners than all the other learner populations in ICLE.

Table 12. Summary table

Frequency	TSW	SP		FR		DU		BrE
Causative structure		SP		FR	TSW	DU		BrE
Lexical variation		SP	TSW	FR		DU		BrE
Abstract movement	TSW	SP		FR		DU		BrE
Freq. expressions	FR	SP				DU	TSW	BrE
TTR expressions		SP		FR	TSW	DU		BrE
Freq. phrasal verbs		SP		FR	TSW	DU		BrE
TTR phrasal verbs		-		-	TSW	DU		BrE
Non-standard/ creative uses	TSW	SP		DU	FR			BrE

The Tswana variety clearly has a status of its own. Exactly what this status is is difficult to establish at this stage. In relation to ICLE-TSW, Van Rooy (2006: 62) claims that "[a] new outer circle variety of English is clearly emerging in South Africa". For Kasanga (2006: 76), "it is reasonable to theorize that BSAE [Black South African English] is not a 'learner language'". However, Kasanga (2006: 77) further qualifies this statement: "It is important to point out that the form of BSAE which qualifies as a distinct variety in its own right is the 'acrolang' form which has reached a certain degree of stability, spread and prominence and excludes the 'mesolang' and 'basilang' forms". The impression one gets from analysing ICLE-TSW is that it rather qualifies as a mesolang form of BSAE. As such, it shares features with both inner/outer circle varieties of English and 'mesolang' varieties of the

expanding circle, viz. learner English. As pointed out by Gilmour (2007), who describes a similar situation in Sri Lanka, extensive fieldwork is needed in order to identify the typical (i.e. stable, spread and prominent) linguistic features of the different varieties.

Another major finding of our study concerns the degree of expertise of the native speakers represented in the reference corpus. The results show that the novice native writers share features with both the expert native writers and the non-native writers. This suggests that, while the degree of expertise is an important factor to take into account when comparing learner English with native English, it does not make the nativeness/non-nativeness distinction redundant. Rather, it adds a layer to our understanding of the learner variety, which appears to be characterised by non-native as well as non-expert features.

SLA specialists have been aware for quite some time that the EFL/ESL distinction is not a clear-cut dichotomy but a continuum, with many factors pulling language varieties in one or the other direction. In spite of its limited scope, our investigation of the use of *into* by students learning English in different environments has brought out the power of corpus linguistic methods in substantiating this continuum. In particular, the striking contrast between the Dutch, French and Spanish learners and the Tswana learners has shed some light on the hazy border between the expanding and the outer circle. While the results are promising, however, the field is vast and complex and we can only claim to have lifted a very small corner of a much larger veil.

Endnotes

1. More information can be found at http://juppiter.fltr.ucl.ac.be/FLTR/GERM/ETAN/CECL/MULTED.html.

2. This, admittedly, may be partly due to the fact that newspaper editorials tend to be heavily edited by native speakers of English. That differences may nonetheless exist between the British and African varieties of English is suggested by Mwangi's (2003) study, which shows that *into* is significantly more frequent in ICE-GB, the British component of the *International Corpus of English*, than in ICE-K, the Kenyan component of the corpus. It should be emphasised, however, that using the written part of ICE-K as a reference corpus still results in a significant underuse of *into* among the four learner populations investigated here ($p<0.01$ with the log-likelihood test), as is the case with our corpus of British editorials (see Section 4.1).

3. The log-likelihood test reveals significant differences between BrE and ICLE-DU ($p<0.001$) and between ICLE-FR and ICLE-SP ($p<0.05$), but no significant differences between ICLE-DU and ICLE-FR, nor between ICLE-SP and ICLE-TSW.

4. Although these figures are likely to be influenced by corpus size, it should be noted that the corpora used in this study are relatively similar in size, varying between 150,000 and 200,000 words. The same remark applies to some other results in the following sections as well.

5. The examples are reproduced exactly as they appear in ICLE.

6. The dictionaries we used are the *Oxford Advanced Learner's Dictionary* (Wehmeier 2000) and the *Macmillan English Dictionary for Advanced Learners, Second Edition* (Rundell 2007).

7. As pointed out by one of the reviewers, teaching might also be an additional factor to consider when seeking to account for the overuse of (semi-)fixed expressions among learners.

8. Cf. Mollin (2006) for a similar warning within the framework of English as a Lingua Franca.

9. Experienced Cambridge ESOL rater, native speaker of British English.

10. The same problem of confusion between *in* and *into* is mentioned by Mwangi (2003: 105-106) for Kenyan English, but with a focus on cases where *in* is used instead of *into*.

11. While LOCNESS contains data produced by American students, we believe that this does not fundamentally affect the validity of the comparison, for preliminary analyses reveal that the frequency of *into* in the corpus of British editorials is not significantly different from its frequency in a comparable corpus of American English.

References

Bamgbose, A. 1998. Torn between the norms: Innovations in world Englishes. *World Englishes* 17(1): 1-14.

De Cock, S. & Granger, S. 2004. High frequency words: The bête noire of lexicographers and learners alike. A close look at the verb *make* in five monolingual learners' dictionaries of English. In *Proceedings of the 11ᵗʰ EURALEX International Congress*, G. Williams & S. Vessier (eds), 233-243. Lorient: Université de Bretagne-Sud.

Gass, S. & Selinker, L. 2001. *Second Language Acquisition. An Introductory Course*. Mahwah: Lawrence Erlbaum.

Gilmour, K. 2007. World Englishes and Sri Lanka. Available at http://www.kln.ac.lk/humanities/eltu/events/perkumpulan/kanegilmour.pdf.

Gilquin, G. 2000/2001. The Integrated Contrastive Model. Spicing up your data. *Languages in Contrast* 3(1): 95-123.

Gilquin, G., Granger, S. & Paquot, M. 2007. Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes* 6(4): 319-335.

Ginsburgh, V. & Weber, S. 2006. La dynamique des langues en Belgique. *Regards économiques* 42: 1-10.

Granger, S. 1996. From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In *Languages in Contrast. Text-based Cross-linguistic Studies*, K. Aijmer, B. Altenberg & M. Johansson (eds), 37-51. Lund: Lund University Press.

Granger, S. 2009. More lexis, less grammar? What does the learner corpus say? Keynote presentation at the *Third International Conference Grammar and Corpora*, Mannheim, 22-24 September 2009.

Granger, S., Dagneaux, E., Meunier, F. & Paquot, M. 2009. *International Corpus of Learner English. Handbook & CD-ROM. Version 2.* Louvain-la-Neuve: Presses universitaires de Louvain. Available from http://www.i6doc.com.

Hasselgren, A. 1994. Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics* 4(2): 237-258.

Hulstijn, J. & Marchena, E. 1989. Avoidance: Grammatical or semantic causes? *Studies in Second Language Acquisition* 11(3): 241-255.

Hyland, K. & Milton, J. 1997. Qualifications and certainty in L1 and L2 students' writing. *Journal of Second Language Writing* 6(2): 183-205.

Kao, R.-R. 2001. Where have the prepositions gone? A study of English prepositional verbs and input enhancement in instructed SLA. *IRAL* 39: 195-215.

Kasanga, L.A. 2006. Requests in a South African variety of English. *World Englishes* 25(1): 65-89.

Koolstra, C. & Beentjes, J. 1999. Children's vocabulary acquisition in a foreign language through watching subtitled TV programmes at home. In *Educational Technology Research and Development* 47(1): 51-60.

Liao, Y. & Fukuya, Y.J. 2004. Avoidance of phrasal verbs: The case of Chinese learners of English. *Language Learning* 54(2): 193-226.

Lorenz, G. 1999. *Adjective Intensification - Learners versus Native Speakers. A Corpus Study of Argumentative Writing.* Amsterdam & Atlanta: Rodopi.

Lozano, C. & Mendikoetxea, A. 2008. Postverbal subjects at the interfaces in Spanish and Italian learners of L2 English: A corpus analysis. In *Linking up Contrastive and Learner Corpus Research*, G. Gilquin, S. Papp & M.B. Díez-Bedmar (eds), 85-125. Amsterdam & New York: Rodopi.

McCrostie, J. 2008. Writer visibility in EFL learner academic writing: A corpus-based study. *ICAME Journal* 32: 97-114.

Mollin, S. 2006. *Euro-English: Assessing Variety Status.* Tübingen: Gunter Narr Verlag.

Mukherjee, J. 2009. Creative variation in New Englishes: Corpus-based insights into verb-complementational innovations in Indian English. In *Grammar between Norm and Variation*, A.N. Lenz & A. Plewnia (eds). Amsterdam: John Benjamins.

Mukherjee, J. & Hoffmann, S. 2006. Describing verb-complementational profiles of New Englishes. A pilot study of Indian English. *English World-Wide* 27(2): 147-173.

Mwangi, S. 2003. *Prepositions in Kenyan English. A Corpus-Based Study in Lexico-Grammatical Variation.* Aachen: Shaker Verlag.

Mwangi, S. 2004. Prepositions vanishing in Kenya. *English Today* 20(1): 27-32.

Neff van Aertselaer, J. 2008. Contrasting English-Spanish interpersonal phrases: A corpus study. In *Phraseology in Foreign Language Learning and Teaching*, F. Meunier & S. Granger (eds), 85-99. Amsterdam & Philadelphia: John Benjamins.

Nesselhauf, N. 2009. Co-selection phenomena across New Englishes: Parallels (and differences) to foreign learner varieties. *English World-Wide* 30(1): 1-25.

Rimmer, W. 2008. Grammatical creativity in learner corpora. *Humanising Language Teaching* 10(1). Available at http://www.hltmag.co.uk/jan08/idea.htm.

Rundell, M. (ed.) 2007. *Macmillan English Dictionary for Advanced Learners, Second Edition.* Oxford: Macmillan Education.

Sand, A. 2005. *Angloversals? Shared morpho-syntactic features in contact varieties.* Habiliations-thesis, University of Freiburg.

Sheen, R. 2006. Comments on R. Ellis's "Current Issues in the Teaching of Grammar: An SLA Perspective". *TESOL Quarterly* 40(4): 828-832.

Siyanova, A. & Schmitt, N. 2007. Native and nonnative use of multi-word vs. one-word verbs. *IRAL* 45: 119-139.

Sjöholm, K. 1995. The influence of crosslinguistic, semantic and input factors on the acquisition of English phrasal verbs. PhD dissertation, Abo Akademi University Press.

Swan, M. 2005. *Practical English Usage. Third Edition*. Oxford: Oxford University Press.

Van Parijs, P. 2004. Europe's linguistic challenge. *Archives européennes de sociologie* 45(1): 113-154.

Van Rooy, B. 2006. The extension of the progressive aspect in Black South African English. *World Englishes* 25(1): 37-64.

Van Rooy, B. 2009. The status of English in South Africa. In *International Corpus of Learner English. Version 2. Handbook and CD-ROM*, S. Granger, E. Dagneaux, F. Meunier & M. Paquot (eds), 198-204. Louvain-la-Neuve: Presses universitaires de Louvain.

Waibel, B. 2007. *Phrasal verbs in learner English: A corpus-based study of German and Italian learners*. PhD dissertation, Albert-Ludwigs-Universität Freiburg.

Wehmeier, S. (ed.) 2000. *Oxford Advanced Learner's Dictionary*. Compass CD-ROM. Oxford: Oxford University Press.

Book chapter. The contribution of self-reflection and inner dialogue to personality development. P. 155-171. Kostenko V., Leontiev D. Â The contribution of inner dialogue became visible from the Stage E5 (Selfawareness) on, while below this level self-reflection mechanisms were neither differentiated nor associated with inner dialogical activity. Language: English. Full text (PDF, 9.87 Mb). Keywords: personality development dialogical self inner dialogue ego development. In book. Dialogical Self: Inspirations, Considerations and Research. Edited by: M. Puchalska-Wasyl, P. OleÅ›, H. J. Hermans.