

At the interface of contact linguistics and second language acquisition research:

New Englishes and Learner Englishes compared*

Gaëtanelle Gilquin

FNRS – Université catholique de Louvain

This paper examines the possible interface between contact linguistics and second language acquisition research by comparing the institutionalized second-language varieties of English known as “New Englishes” and the foreign varieties of English called “Learner Englishes”. On the basis of corpus data representing several populations of various origins, it investigates four linguistic phenomena, ranging from syntax (embedded inversion) to lexis (phrasal verbs with *up*), through phraseology (word clusters) and pragmatics (discourse markers), with a view to identifying similarities and differences between the two types of varieties at several levels of the language. The paper also explores avenues for going beyond a descriptive account towards a more explanatory one, in an attempt to build the foundations of a theoretical rapprochement between contact linguistics and second language acquisition research.

Keywords: New Englishes; Learner Englishes; contact linguistics; second language acquisition; paradigm gap; embedded inversion; phrasal verbs; word clusters; discourse markers

* This paper is based on a plenary talk I gave at the *Crossling Symposium: Language Contacts at the Crossroads of Disciplines* (University of Eastern Finland – Joensuu, February 28 – March 1, 2013). I thank the organizers of the conference for the invitation and the audience at this event for the fruitful discussion. I also thank Marianne Hundt and the two anonymous reviewers for comments on an earlier version of this paper.

1. Introduction

With English having become a language used worldwide for international and intranational communication, linguists have taken an interest in the different varieties which have developed in various parts of the world, both native and non-native. One well-known model that seeks to describe and classify these varieties is Kachru's (1985) model of the three concentric circles, which distinguishes between the Inner Circle, the Outer Circle and the Expanding Circle. The Inner Circle includes countries where English is used as a native language (UK, USA, Australia, etc.). These native varieties are well established and are usually referred to under the cover term "native English" (or "English as a Native Language" [ENL]) – although differences between them have long been recognized, a fact which would probably justify the use of the plural form "native Englishes". The Outer Circle includes former British or American colonies like India, Malaysia or Singapore where English is an official or semi-official language, used with intranational functions, e.g. in administration, education, the media or literature. These varieties are variously called "English as a Second Language" (ESL), "New Englishes" (NEs), "indigenized varieties of English" or "institutionalized second-language varieties of English". Finally, the Expanding Circle includes countries such as France, China or Russia, where English has no official status but non-native speakers use it as an international lingua franca in such areas as business, politics, research or tourism. Besides "English as a Foreign Language" (EFL), the term "Learner English" is often employed to refer to these varieties, or indeed "Learner Englishes" (LEs) (using a plural form by analogy with New Englishes, see Gilquin and Granger 2011: 74). In this article, I will use the terms New Englishes and Learner Englishes to refer to the varieties corresponding to the Outer Circle and the Expanding Circle, respectively. The native varieties belonging to the Inner Circle will be called English as a Native Language.

Both the Outer and Expanding Circles have close links with the Inner Circle, if only because the concept of the native speaker occupies a central place in the study of varieties of English. Coulmas (1981: 1), in the introduction to his *Festschrift for Native Speaker*, claims that "there is no way of doing linguistics without taking account of him [sic]" and that "he [sic] can be conceived of as a common reference point for all branches of linguistics". Admittedly, there have been attempts in recent years to call into question the relevance of a native norm. Honna (2006: 118) explains that "[t]raditionally, non-native speakers were expected to conform to native speakers' norms of linguistic behavior [...]. In

view of world Englishes, however, these assumptions are increasingly questioned". At the same time, it must be recognized that both Outer and Expanding Circles still very much depend on the Inner Circle to define themselves. The learner varieties are clearly identified as "norm-dependent", i.e. dependent on native English for their norms (see Kachru 1985: 17). As for the description of NEs as "norm-developing" (Kachru 1985: 17), it suggests that these varieties are in the process of developing their own norms, but have not yet quite reached the point where their norms are fully developed and where they can do away with the norm-providing native English. Besides, it appears that the native norm is always lurking behind, if only implicitly, in linguistic descriptions of non-native varieties. This is exemplified by the following two quotations, which describe LEs and NEs with reference to standard, native English (see parts in bold):

[T]he findings indicate that the L2 writers, particularly Chinese learners, employ significantly fewer "downtoners" (i.e., *might, would, possible*) but more "boosters" (i.e., *sure, can, should*) **when compared with the native speakers** in academic writing. (Chen 2010: 44, emphasis added)

IndE [Indian English] is found to **differ from native usage** by associating *had + V-ed* with present perfect and preterite meanings. (Sharma 2001: 343, emphasis added)

On the other hand, the possible links between the Outer Circle and the Expanding Circle have hardly been investigated. Typically, the two types of varieties are discussed in separate journals (compare, for example, *World Englishes* and *Studies in Second Language Acquisition*) and at different conferences (like the conferences organized by the International Association of World Englishes, IAWE, or by the European Second Language Association, Eurosla). Specialists in the two camps also tend to use different terminologies, e.g. "substrate influence" versus "L1 transfer" to refer to the influence of the speaker's mother tongue, and "accepted local form", "innovation" or "emerging feature" versus "mistake" or "error" to describe a feature that is not found in native English.

In the next section, we will see that this lack of communication between research on the Outer and Expanding Circles was highlighted several decades ago, but that up until recently not much has been done by way of rapprochement between the two. The rationale for comparing the two types of varieties will then be outlined, by showing how the expectation of both similarities and differences between NEs and LEs makes the comparison between them worthwhile. The

core of the paper, Section 4, will present four case studies that use a corpus-based comparative approach to examine linguistic phenomena in NEs and LEs, covering different fields of linguistics: embedded inversion (syntax), phrasal verbs with *up* (lexis), word clusters (phraseology) and discourse markers (pragmatics). Section 5 will briefly explain how the descriptive analysis of the differences and (especially) similarities between NEs and LEs can lead to a more explanatory account, by considering the role played by general cognitive processes of language acquisition. Finally, Section 6 will present some challenges facing researchers who wish to pursue this line of research and will offer concluding remarks.

2. NEs and LEs: An old (and persistent) paradigm gap

NEs and LEs do not just represent different types of varieties of English. They are also embedded within two different frameworks, or paradigms, namely contact linguistics and second language acquisition (SLA) research. The Outer Circle includes countries where minimally two languages are commonly used, namely an indigenous language and English. Most of the time, an even larger number of languages (or language varieties) co-exist in these multilingual communities. Contact linguistics is concerned with the interaction between these languages and language varieties, which may result in “complex patterns of contact linguistics, including lexical transfer, code switching and code mixing, and discoursal and syntactic change and accommodation” (Bolton 2006: 261). Pidgins and creoles, as prime examples of contact languages, have attracted a great deal of attention among specialists in contact linguistics, but NEs, being varieties “used in a multilingual environment and function[ing] as a lingua franca for daily interactions among speakers of different languages” (Siegel 2010: 831), are also a natural object of study for them (see Mesthrie 2006). SLA research, on the other hand, despite its label, is not traditionally concerned with institutionalized second-language varieties of English. It investigates “the acquisition of any language after the acquisition of the mother tongue” (Ellis 2008: 6), and although this could in principle cover NEs (see Sridhar and Sridhar 1986: 3), its scope is normally limited to the acquisition of foreign languages, as well as what are (rather confusingly) termed second languages, i.e. languages that are acquired by learners (usually immigrants) in a country where the target language is the majority first language (e.g. English in the USA).

The distance and lack of dialogue between contact linguists working on NEs and SLA researchers working on LEs was brought to light as early as 1986, when Sridhar and Sridhar (1986: 3) pointed to the “paradigm gap that has prevented research on second-language acquisition theory and indigenized varieties of English from making substantive contributions to each other”. Although Sridhar and Sridhar called for a better integration of the two paradigms, their call was not acted upon until very recently. In 2011 Hundt and Mukherjee, referring to this initial call, were still deploring that “these two objects of inquiry have continued to be treated as fundamentally different and unrelated areas of research ever since” (Hundt and Mukherjee 2011: 1).

Sridhar and Sridhar (1986) put forward two reasons for this paradigm gap. The first one has to do with theory-internal limitations, and more particularly a failure of SLA theories to incorporate (descriptive) findings on NEs. The second reason is a lack of empirical foundation in NEs studies. In their paper, they plead for “detailed studies of the acquisition of IVEs [indigenized varieties of English] using systematic data collection procedures, rigorous analytical methods, and explicit reporting conventions” (Sridhar and Sridhar 1986: 4). While such studies may have been difficult to carry out in the 1980s, thanks to projects such as the International Corpus of English (ICE) this has now become feasible. ICE includes subcorpora representing several native varieties of English and NEs, compiled according to the same principles, which facilitates the comparison of the varieties they represent.¹ In addition, a number of individual corpora representing certain specific NEs have become available, e.g. the Kolhapur Corpus of Indian English² the Xhosa English Corpus or the SAVE (South Asian Varieties of English) Corpus. Interestingly, the advent of corpora has equally benefited the Expanding Circle. Instead of the small sets of data on which SLA studies used to rely, we now have access to large, balanced and carefully designed corpora of LEs. One prominent such corpus is the International Corpus of Learner English (ICLE), which is made up of several components containing essays written by EFL learners from sixteen mother tongue backgrounds (Granger et al. 2009). Other learner corpora of English have also appeared, like the Cambridge Learner Corpus, the Japanese EFL Learner Corpus and LINDSEI, the Louvain International Database of Spoken English Interlanguage (Gilquin,

¹ See, however, the articles in the *ICAME Journal* 34 (2010) on some of the challenges involved in applying the same principles to the compilation of certain ICE components.

² The Kolhapur Corpus uses the Brown sampling frame, a frame used by other corpora as well, which again allows for the direct comparison between several (geographically or chronologically distinct) varieties of English.

De Cock and Granger 2010). Hence, linguists now have the necessary resources with which to empirically study both NEs and LEs and, *de facto*, to compare the two types of varieties, applying the “rigorous analytical methods” advocated by Sridhar and Sridhar (1986: 4).

Over the last few years, contact linguistics and SLA research have come a little closer together, thanks to a few studies which have exploited the tools and techniques of corpus linguistics to compare NEs and LEs. The first such study, to my knowledge, is Nesselhauf’s (2009) analysis of co-selection phenomena in several components of ICE (ICE-India [ICE-IND], Jamaica [ICE-JAM], Kenya [ICE-KEN] and Singapore [ICE-SIN]) and ICLE (German, French, Finnish and Polish L1 groups) – see also Nesselhauf (2011). Although the phenomena investigated are very specific and the frequencies, therefore, are quite low, Nesselhauf is able to observe similar preferences for certain collocations (e.g. *take into consideration* rather than *take account of*) and certain patterns of use (e.g. *come into contact* and *come in contact*), as well as similar non-standard prepositional verbs (e.g. *comprise of*, *discuss about* or *emphasize on*) in NEs and LEs, thus demonstrating that the two types of varieties are not as dissimilar to each other as the theoretical distance between them might have suggested. Nesselhauf’s study has been followed by a (modest) number of similar studies: Gilquin (2011); Götz and Schilk (2011); Hundt and Vogel (2011); Szmrecsanyi and Kortmann (2011); Davydova (2012); Laporte (2012); and Meriläinen and Paulasto (fc.).³ Most of these studies are limited in one way or another, like Nesselhauf’s study was. Some are restricted to a single variety of NEs and LEs, e.g. Davydova (2012), where NEs are merely represented by IndE and LEs by Russian learner English. Others rely on rather small corpus samples, e.g. Hundt and Vogel (2011), where NEs are represented by samples of some 40,000 words. Finally, most of them focus on one isolated aspect of the English language, a major exception being Szmrecsanyi and Kortmann (2011), who consider a wide range of grammatical markers such as pronouns, the negator *not*, the auxiliary *do* and inflected verbs. Comprehensive studies comparing various aspects of several varieties of NEs and LEs on the basis of large corpus samples are still lacking. However, the few studies listed above have the merit of having demonstrated both the feasibility and desirability of comparisons between NEs and LEs, and thus having paved the way for further explorations.

³ In addition, a few studies have also contributed to bridging the paradigm gap between NEs and LEs by viewing EFL and ESL as situated along a continuum, although they relied on data representing either only NEs (e.g. Mukherjee and Gries 2009) or only LEs (e.g. Gilquin and Granger 2011).

3. Rationale for a comparative approach

Besides the fact that the studies mentioned in the preceding section have yielded promising results, there are a number of reasons why linguists should further pursue the line of research first evoked by Sridhar and Sridhar (1986). These mainly have to do with the kind of status that English enjoys in the Outer and Expanding Circles, which leads one to expect both similarities and differences between NEs and LEs, making the comparison between the two worthwhile. We can expect differences first of all because NEs and LEs are acquired in essentially different settings: mainly a naturalistic one for NEs (which are acquired in everyday interactions) and mainly an instructional one for LEs (which are learned in the context of the classroom).⁴ Second, NEs have both intranational and international functions, whereas LEs normally only have international functions. The cultural and historical contexts are also different for the two types of varieties: for NEs speakers, English is historically rooted, being the language of the colonizers, but for EFL learners, English is culturally foreign. Furthermore, NEs are claimed to be in the process of developing their own norms and some non-standard features are viewed as emerging features of the varieties. LEs, on the other hand, arguably still rely exclusively on the native, Inner Circle norm of English; differences from Inner Circle varieties are thus considered to be errors. In addition there is the obvious argument (which also accounts for the specificities of individual NEs and LEs varieties) that these varieties are geographically distant, being used by speakers from different parts of the world, and that they are associated with different mother tongue backgrounds, which is likely to result in different transfer phenomena.

Yet, despite these differences, we can also expect similarities. One compelling reason for this is related to the non-native nature of both NEs and LEs. Since English is acquired/learnt as an L2, we can expect universal language learning strategies to play a role (Winford 2009). Also, since the borderline between EFL and ESL is fuzzy (see Nayar 1997; Görlach 2002: 99–117), we can hypothesize that certain features will take different positions on a continuum between NEs and LEs, with some overlap in-between. Finally, the mere experience of going through the NEs literature as an SLA specialist (or the other way round) reveals characteristics that look strikingly familiar. Below are some examples (1–7) taken from the *electronic World Atlas of Varieties of English*

⁴ All this, of course, is a matter of degree, as English is taught at school in the Outer Circle too, and learners in the Expanding Circle can also have access to English in their everyday life, for example through the Internet.

(*eWAVE*; Kortmann and Lunkenheimer 2011), which represent various NEs. Anyone with experience in teaching foreign learners of English is bound to recognize certain features occurring in their students' production: the absence of marking on plural nouns (1), the use of plural marking for uncountable nouns (2), the use of a zero article where a definite article would be required (3), the invariant use of *don't* for all persons in the present tense (4), the use of a simple present instead of a present perfect (5), the lack of inversion to form questions (6) and the extension of the progressive aspect to stative verbs (7). Indeed, most of these features have been shown to characterize (certain varieties of) LEs (see e.g. Chuang and Nesi 2006; Thewissen 2013).

- (1) The three **girl** there don't want to talk to us. (Hong Kong English [HKE], emphasis added)
- (2) We gave him **advices** not to contest with his sister. (Ugandan English, emphasis added)
- (3) **Ø** Ideal candidate will be a recently qualified accountant. (Kenyan English [KenE], emphasis added)
- (4) She **don't** like that paint. (IndE, emphasis added)
- (5) I **work** for them since I was fifteen. (Malaysian English [MalE], emphasis added)
- (6) **You want** more rice? (Sri Lankan English [SLE], emphasis added)
- (7) Are you **understanding** me? (KenE, emphasis added)

All these elements point to the existence of some common ground between NEs and LEs, which underlines the relevance of an integrated approach seeking to compare the two types of varieties. Having established the legitimacy of such an approach, I can now turn to the presentation of four case studies, each of which tackles a different linguistic aspect of the varieties under investigation using naturally occurring data from NEs and LEs corpora. These case studies do not claim to offer a complete panorama of the differences and similarities between NEs and LEs. Yet they show that several aspects of language may contribute to bringing contact linguistics and SLA research closer to each other, and thus removing the "theory-internal limitations" denounced by Sridhar and Sridhar (1986: 4).

4. Four case studies

In this section, four case studies will be described, each one dealing with a different linguistic phenomenon: embedded inversion (Section 4.1), phrasal verbs with *up* (Section 4.2), word clusters (Section 4.3) and discourse markers (Section 4.4). By considering a variety of phenomena and relying on large samples of corpora representing a wide range of populations, these case studies try to overcome the limitations of previous studies which combine NEs and LEs perspectives (see Section 2). However, given space constraints, it will not be possible to provide a detailed account of each of the phenomena. I will therefore focus on some of the most interesting findings, highlighting in particular non-standard features (i.e. differences in use and/or frequency from ENL) which are shared by NEs and LEs. The case studies cover four different fields of linguistics: syntax (for embedded inversion), lexis (for phrasal verbs with *up*), phraseology (for word clusters) and pragmatics (for discourse markers). They also concern both topics that have been studied in contact linguistics and SLA research (this is the case for phrasal verbs and discourse markers), and other topics that have only or mainly been studied in one of these two paradigms (contact linguistics for embedded inversion and SLA research for word clusters). For each case study, the topic of investigation will be briefly introduced, then the data and methodology will be described, and finally the main results will be outlined. Table 1 gives an overview of the corpora that were used in the different studies. The NEs data come from ICE (the components that were publicly available – and complete – at the time the studies were carried out): ICE-Hong Kong [ICE-HK], ICE-IND, ICE-KEN, ICE-Philippines [ICE-PHI], ICE-SIN and ICE-Tanzania [ICE-TAN]. They represent varieties at different developmental stages according to Schneider's (2003, 2007) dynamic model of the evolution of postcolonial Englishes.⁵ The LEs data come from the second version of ICLE (ICLEv2) for writing and from LINDSEI for speech. They were produced by EFL learners with different mother tongues (from several language families): 16 in ICLE and 11 in LINDSEI (ten are common to the two corpora) – see the

⁵ Schneider (2007) claims that Hong Kong English has reached the third phase of nativization (133–139), as have Kenyan English (189–197), Philippine English (140–144) and Tanzanian English (197–199). India is somewhere between phase three and phase four, that of endonormative stabilization (161–173), while Singapore English (SgE) is said to have clearly reached phase four (153–161).

complete list in Table 1.⁶ In addition, I used corpora of native British English as a baseline, representing both speech and writing: ICE-GB (the British component of ICE), LOCNEC (Louvain Corpus of Native English Conversation) and the BNC (British National Corpus) Baby edition. While the first corpus is a good match for the NEs data, the second one resembles the LEs more closely, and so each one was used alternately in the first two case studies as a reference corpus. The BNC was useful when larger samples of data than those offered by ICE-GB and LOCNEC were necessary (third case study). Finally, the last case study, instead of relying on a single reference corpus, used a yardstick made up of a combination of ICE-GB and LOCNEC, which made it unnecessary to favour either NEs or LEs when looking for a comparable native corpus. All the corpora, native and non-native, were queried by means of the concordancer WordSmith Tools (WST) version 5 (Scott 2008). The statistical significance of the results was tested by means of the log-likelihood ratio.

Table 1. Overview of the NEs, LEs and ENL data used in the case studies⁷

Corpus	ICE	ICLEv2	LINDSEI	ICE-GB	LOCNEC	BNC Baby
Variety	NEs	LEs	LEs	ENL	ENL	ENL
Origin	HK, IND, KEN, PHI, SIN, TAN	BG, CH, CZ, DU, FI, FR, GE, IT, JP, NO,	BG, CH, DU, FR, GE, GR, IT, JP, PL, SP,	BR	BR	BR

⁶ The LINDSEI corpus was kept in its entirety. The ICLE data were restricted to those produced by learners whose first language at home was also their mother tongue and who lived in the country associated with this mother tongue (e.g. Sweden for the Swedish learners); for the Chinese component of ICLE, the data produced by learners living in or coming from Hong Kong were excluded.

⁷ Origin: HK=Hong Kong, IND=Indian, KEN=Kenyan, PHI=Philippine, SIN=Singaporean, TAN=Tanzanian, BG=Bulgarian, CH=Chinese, CZ=Czech, DU=Dutch, FI=Finnish, FR=French, GE=German, GR=Greek, IT=Italian, JP=Japanese, NO=Norwegian, PL=Polish, RU=Russian, SP=Spanish, SW=Swedish, TSW=Tswana, TR=Turkish, BR=British. Genre: direct conv.=direct conversations, acad. wr.=academic writing, stud. wr.=student writing (essays and examination scripts), arg./lit. essays=argumentative/literary essays. Note that not all of the (sections of the) corpora are used in every case study. Details are found in the “data and methodology” section of each study.

		PL, RU, SP, SW, TSW, TR	SW			
Genre	direct conv., acad. wr., stud. wr.	arg./lit. essays	informal interviews	direct conv., acad. wr., stud. wr.	informal interviews	acad. wr.
Total size	1,864,711 words	3,163,142 words	792,141 words	313,873 words	118,398 words	1,019,697 words

Although efforts were made to select (sub)corpora that were similar in genre across NEs, LEs and ENL, it should be noted that it was not always possible to find corpus data that were fully comparable. The case study on embedded inversion (Section 4.1), for instance, uses data from ICE, LINDSEI and LOCNEC. While LINDSEI and LOCNEC are exact replicas of each other, having been compiled according to the same design criteria, the informal interviews they are made up of may be said to differ from the spontaneous conversations contained in ICE.⁸ The two genres correspond to more or less casual speech, but their specificities may have an impact on the comparison. It must be emphasized, however, that if linguistic similarities can be brought to light despite certain differences in genre between some of the corpora, this arguably constitutes quite strong evidence that parallels can be drawn between NEs and LEs.

Not only genre may have an impact on the analysis (see e.g. Zipp and Bernaisch 2012 on NEs), but also the type of usage that is represented in the corpora. The corpora used in the four case studies mainly represent acrolectal (or upper mesolectal) NEs usage and upper intermediate to advanced stages of LEs, that is, those varieties that are expected to be closer to standard native English. Other varieties may present other features, as shown by Davydova (2012) for NEs and by Dagneaux, Denness and Granger (1998) for LEs. One must therefore bear in mind that the findings below do not cover the whole spectrum of NEs and LEs usage.

Finally, the aggregate approach that I adopt in the case studies by treating NEs, LEs and ENL as groups should not overshadow the possible variation

⁸ Interestingly, however, the difference between the two may not always necessarily be so clear as some ICE corpora occasionally include interviews rather than spontaneous conversations in the sampling (Marianne Hundt, p.c., December 2013).

(linked to different mother tongues, different proficiency levels, different stages in the evolution of the varieties, idiolectal preferences, etc.) between the varieties belonging to these groups (e.g. HKE versus IndE versus SgE), as well as between the individuals speaking these varieties (e.g. IndE speaker A versus IndE speaker B versus IndE speaker C). Such inter-varietal and intra-varietal variation does exist, as becomes evident from analyses carried out as a complement to the more global results presented here, but on the whole the general tendencies arguably remain valid for a majority of the varieties and speakers considered. This contributes to legitimizing the deliberate strategy of adopting a bird's-eye perspective so that broad similarities can emerge from the data.

4.1 Embedded inversion⁹

4.1.1 *Topic of investigation*

The first case study considers a syntactic phenomenon, namely embedded inversion (EI). EI consists of the use of inverted word order in indirect questions. It may occur in WH-questions (e.g. *could you tell me what time is it*) or in Yes/No-questions (e.g. *I'm wondering is it the normal way*). This phenomenon is well documented in the NEs literature (see Kolbe and Sand 2010), where it is described as an “angloversal” (Mair 2003: 84), a non-standard feature shared by several geographically distant varieties of English (especially NEs). In *eWAVE* (Kortmann and Lunkenheimer 2011), it is said to be attested in 61 per cent of the varieties investigated (including IndE, HKE, Male and KenE). By contrast, EI has so far hardly been dealt with in SLA research. Meriläinen (2010) is one of the few existing studies that considers this feature in LEs, revealing the presence of EI in the written production of Finnish-speaking learners of English.

4.1.2 *Data and methodology*

The results for NEs come from a study by Meriläinen and Paulasto (fc.) on spoken data from ICE-IND, ICE-KEN and ICE-SIN. In addition, the French, German and Swedish components of LINDSEI were selected to represent LEs, and the native counterpart of LINDSEI, LOCNEC, was used as a baseline (see Table 2). Spoken data were preferred to written data as EI is more likely to occur in speech (Kolbe and Sand 2010: 38). Instances of EI were retrieved by first

⁹ This section is based on a study that I conducted with Lea Meriläinen (Gilquin and Meriläinen 2012).

extracting all the occurrences of six verbs claimed by Biber et al. (1999) to be the most common verbs with *wh*-, *whether*- and *if*-clauses, viz. *ask*, *know*, *see*, *tell*, *understand* and *wonder*. The concordances were then manually examined to retrieve cases of indirect questions (with and without EI).¹⁰

Table 2. Corpora used for the study of embedded inversion

NEs			LEs			ENL		
<i>Corpus</i>	<i>Genre</i>	<i>Size</i>	<i>Corpus</i>	<i>Genre</i>	<i>Size</i>	<i>Corpus</i>	<i>Genre</i>	<i>Size</i>
ICE (IND, KEN, SIN)	Speech	c. 1,490,000	LINDSEI (FR, GE, SW)	Informal interviews	249,156	LOCNEC	Informal interviews	118,398

4.1.3 Results and discussion

As can be seen in Table 3, the proportion of EI, calculated out of all the instances of indirect questions with the six verbs under study, is the highest among the speakers of NEs, representing 18.6 per cent in WH-questions and 4.1 per cent in Yes/No-questions. Embedded inversion is less common in LEs but it does occur, at least in WH-questions, with a percentage of 2.2. The native corpus that I used for this comparison did not include any instances of embedded inversion – although the phenomenon is attested in native English as well (see Quirk et al. 1985: 1052).

Table 3. Proportion of EI in NEs, LEs and ENL

	WH-questions	Yes/No-questions
NEs	18.6% (155/833)	4.1% (16/393)
LEs	2.2% (5/226)	0.0% (0/116)
ENL	0.0% (0/111)	0.0% (0/38)

¹⁰ Note that this stage of identification was sometimes difficult as it is not always possible to distinguish between cases of direct questions and indirect questions with EI in speech. In addition, some examples of embedded inversion seem to be the result of planning problems typical of speech, rather than deliberate syntactic choices. All cases of embedded inversion, except those which clearly corresponded to a direct question, were included in the analysis, regardless of the (possible) explanation for their production.

In an attempt to identify specific patterns of use of EI, and given the very small number of instances found in the LINDSEI components, LEs data from ICLE and from a corpus of Finnish Matriculation Examination Compositions (MEC) by Finnish- and Swedish-speaking students (Meriläinen 2010) were also scrutinized. This investigation revealed that EI is governed by certain similar linguistic principles in NEs and LEs. First, collocational patterns appear to play a role in the use of EI. Thus, the chunks *what's* and *how's*, which often occur together in English, are likely to be kept together in indirect questions and thus result in embedded inversion,¹¹ as in (8) and (9). Occasionally, this leads to self-corrections, and in effect to a double marking of the verb *be*, as also pointed out by Hilbert (2011: 132) for NEs (see Examples (10) to (12)).

(8) But I don't know **what's** the prize like. (ICE-SIN:S1A-014, emphasis added)¹²

(9) she asked us **what's** our language (LINDSEI-SW:009, emphasis added)

(10) Don't understand **how's** it **is** coming up. (ICE-SIN:S1A-053, emphasis added)

(11) they're still looking and . trying to find out **what's** . **what** the: . right . courses of studies **are** for them (LINDSEI-GE:042, emphasis added)

(12) I know (mm) (eh) where the: (er) (er) company is **what's** the surroundings **are** and and everything (LINDSEI-GE:033, emphasis added)

The second principle that turns out to guide the use of EI in both types of varieties is related to the distance between the matrix verb and the embedded verb. Lengthy subjects may give rise to embedded inversion, as in Examples (13) and (14), where the subjects have been italicized. It should be noted that the influence of long subjects has been acknowledged as a characteristic of EI even in Standard English (Quirk et al. 1985: 1052). On the other hand, it seems to have gone unnoticed in the literature on EI that coordination may have the same effect, with the second part of the question being more likely to display embedded inversion, as shown in Examples (15) and (16).

¹¹ This was also noticed by Kolbe and Sand (2010: 34), which led them to exclude *what's* from their database on the grounds that “this lexical chunk does not allow variation”.

¹² All corpus examples are reproduced as is and followed by the (abbreviated) name of the corpus and the file identification.

(13) Now you might well ask **what is** *this enormous waste of resources and money* to sent three hundred observers (ICE-IND:S1B-022, emphasis added)

(14) Only a few even know **who are** *the Finnish Members of European Parliament*. (MEC-FI, emphasis added)

(15) they were asking about uh this religion uhn **and what is** the philosophy of this uh Indian culture (ICE-IND:S1A-036, emphasis added)

(16) we have to ask what it is than that brings human society forward, **and**, more important, **what is it**, that spoils every attempt to establish a really just and democratic society. (ICLE-GE:DRE-0017.1, emphasis added)

Summing up, this first case study has shown that EI, which is often dealt with in the literature as an angloversal typical of NEs, is also found in LEs, though in smaller proportions. In addition, EI can occur in similar contexts in the two types of varieties, namely with repeated chunks or when the matrix verb and the embedded verb are quite far from each other, thus adhering to principles which are equally valid for NEs and LEs.

4.2 Phrasal verbs with *up*

4.2.1 *Topic of investigation*

The second case study is more lexical in nature, focusing on phrasal verbs (PVs) with *up*. From an SLA point of view, PVs have been described as “one of the most notoriously challenging aspects of English language instruction” (Gardner and Davies 2007: 339). As a consequence, PVs often appear to be avoided by foreign learners of English, as demonstrated by Dagut and Laufer (1985) for Hebrew-speaking learners, Hulstijn and Marchena (1989) for Dutch-speaking learners, and Liao and Fukuya (2004) for Chinese-speaking learners. A similar underuse has been demonstrated by Schneider (2004) for the NEs he investigated (except for SgE). However, from the perspective of NEs, it is especially the distinctiveness of PVs that has been underlined. For SLE, for example, it has been shown by Meyler (2007, cited in Mendis and Rambukwella 2010: 189) that the particle is sometimes omitted (e.g. *Please don't throw my letter* instead of *throw away*) or unwanted (e.g. *She couldn't bear up the pain* instead of *bear the pain*); examples of PVs with new meanings are also given, e.g. *put on* in the sense of ‘gaining weight’, *pass out* in the sense of ‘graduating from a university or technical college’ and *come down* in the sense of ‘failing an examination or test’. Similarly, Aceto (2006: 215) refers to the creation of phrasal verbs in

Caribbean Englishes, with examples like *kiss up* (meaning ‘to kiss’) and *wet up* (meaning ‘to soak’).

4.2.2 Data and methodology

The data for this case study represent both writing and speech (see Table 4). One typically written genre and one typically spoken genre were selected, namely academic writing and direct conversations, which were examined in ICE-HK, ICE-IND, ICE-KEN, ICE-PHI, ICE-SIN and ICE-TAN (for NEs) and in ICE-GB (for ENL). For want of a better match, these data were compared to the argumentative and literary essays from ICLEv2 and the informal interviews from LINDSEI. For the sake of simplicity, I will describe the stylistic comparison as one between writing and speech, but it should be borne in mind that other written and spoken text types may display different features (see Zipp and Bernaisch 2012: 176–186 on written NEs).

Table 4. Corpora used for the study of phrasal verbs with *up*

NEs			Les			ENL		
<i>Corpus</i>	<i>Genre</i>	<i>Size</i>	<i>Corpus</i>	<i>Genre</i>	<i>Size</i>	<i>Corpus</i>	<i>Genre</i>	<i>Size</i>
ICE (HK, IND, KEN, PHI, SIN, TAN)	Acad. writing	544,612	ICLEv2 (all)	Arg./lit. essays	3,163,142	ICE-GB	Acad. writing	87,128
	Direct conv.	1,035,122	LINDSEI (all)	Informal interviews	792,141		Direct conv.	183,366

The particle *up* was chosen because it is claimed to be the most frequent particle used in PVs (Biber et al. 1999: 413), and also because a large proportion of the occurrences of the word *up* corresponds to particles used in PVs, which makes it feasible to carry out the analysis on the basis of a raw, untagged corpus (see Hundt and Mair 1999: 230). Nevertheless, the automatic extraction of *up* had to be followed by manual work to discard the few cases where *up* was not part of a PV.

4.2.3 Results and discussion

In terms of overall frequency, Figure 1 shows that NEs and LEs speakers are similar in that they underuse phrasal verbs with *up* when compared to native

speakers of English. The difference in frequency between NEs and ENL is statistically significant, as is the difference between LEs and ENL ($p < 0.0001$).

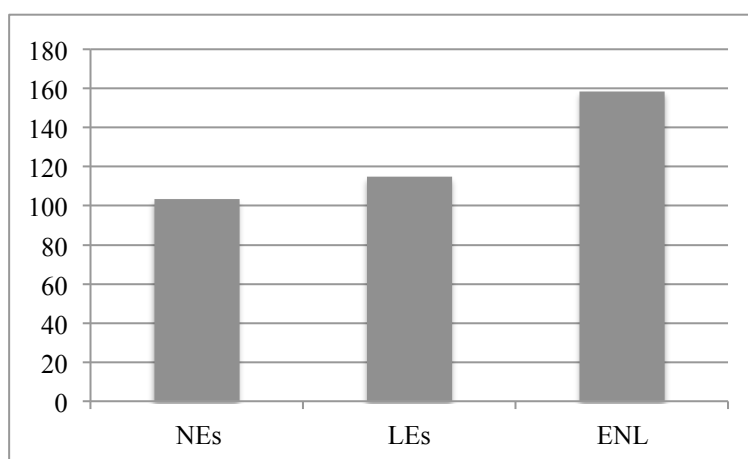


Figure 1. Relative frequency per 100,000 words of PVs with *up* in NEs, LEs and ENL

What distinguishes NEs and LEs is the relation between PVs with *up* and the type of texts (written versus spoken) in which they occur (see Figure 2). NEs speakers follow the native trend in using more PVs in speech than in writing, which corresponds to the generally recognized informal character of phrasal verbs (see Siyanova and Schmitt 2007: 121). EFL learners do the exact opposite and use fewer PVs in speech than in writing, which results in a heavy overuse of PVs with *up* in writing (as compared to ENL) and an even heavier underuse in speech ($p < 0.0001$). NEs speakers, by contrast, underuse PVs in speech ($p < 0.0001$) but neither overuse nor underuse them in writing.

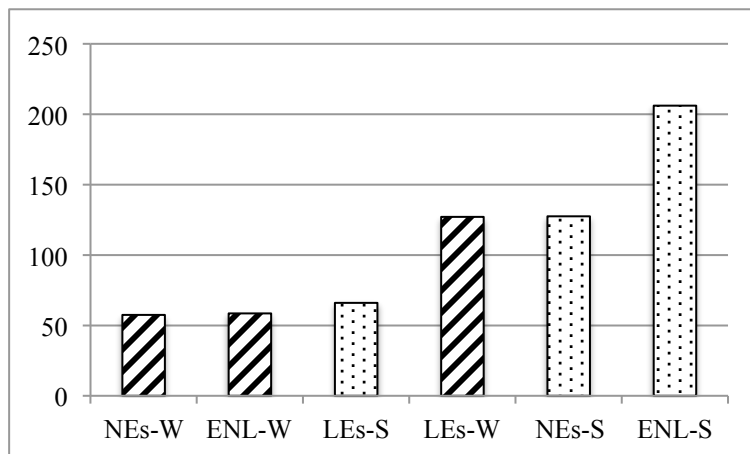


Figure 2. Relative frequency per 100,000 words of PVs with *up* in written (W) and spoken (S) NEs, LEs and ENL

Such stylistic differences may be related to the context in which English is acquired in the Outer and Expanding Circles. Since PVs are more typical of spoken/informal English (see above), they are more likely to be encountered in a naturalistic environment than in a classroom setting, which would account for NEs speakers' advantage here and the fact that they seem to have a fairly accurate view of the type of register in which these verbs are the most appropriate. Foreign learners of English, by contrast, have been shown to mix registers by using spoken-like features in writing (Altenberg and Tapper 1998; Gilquin and Paquot 2008) and written-like features in speech (Kieweg 2000; De Cock 2011), which is reflected in the above results for phrasal verbs with *up*.¹³

Qualitatively, certain similarities emerge if we examine the actual PVs that are produced by speakers from the two groups. One common phenomenon is the use of an additional particle (see also Zipp and Bernaisch 2012: 189–190), as in (17) to (19), where the particle *up* is redundant because the meaning of upward movement or telicity is already expressed by the verb itself (*surface*, *complete* or *rise*).

¹³ It should be noted that this lack of register awareness does not seem to be a developmental feature that could characterize young speakers of NEs too, at least not in the case of PVs with *up*. Using the collection of student texts from the ICE corpora as a basis for comparison with the ICLE essays, it turns out that NEs students use hardly more PVs than NEs expert writers: 65 instances per 100,000 words for the former compared with 57 for the latter (and 127 in the ICLE essays).

(17) Certain problems can **surface up** and need to be expected. (ICE-TAN:W2A-037, emphasis added)

(18) fifty percent of the class girls will be married by the time MA **completes up** (ICE-IND:S1A-062, emphasis added)

(19) Although the time passes they cannot share the things and the tension **rises up**. (ICLE-TU:CUK-0121.1, emphasis added)

Interestingly, NEs and LEs also share certain PVs which represent innovations not listed in dictionaries of phrasal verbs. A case in point is the PV *cope up with*, a deviation from *cope with*, which seems to be built by analogy with phrasal-prepositional verbs like *come up with*, *meet up with* and *put up with*. *Cope up with* is described in the contact linguistics literature as a feature of certain NEs (see e.g. Platt 1989: 399 on SgE). However, the corpus analysis shows that it not only occurs in NEs (20), but also in LEs (21).

(20) **Coping up with** stress amen (ICE-SIN:S1A-018, emphasis added)

(21) While man kept only his role of the bread-winner of the family, woman had to **cope up with** both, her work and maternal duties (ICLE-CZ:PRAG-0042.3, emphasis added)

Other examples of innovations are given in (22) and (23). These innovations are not necessarily shared between NEs and LEs, but they testify to the common creative potential of both types of varieties. Thus, *meddle up* in (22) could be a mixture between *meddle with* and *mess up*, referring to the act of ruining or spoiling something by touching it when you are not supposed to. *Fashion up* in (23) appears to describe the act of doing something up, at the same time making it more fashionable. Of course, such innovations occur in native English too, as exemplified in (24), taken from BNC Baby, where *sprayed up* seems to be built according to the same principle as *fashion up* in (23), referring to the act of doing something up by means of spray (in this case car spray paint). However, their occurrence in NEs and LEs demonstrates that linguistic creativity is not the sole preserve of (norm-providing) native speakers.

(22) I've been trying my best to keep it clean then he comes and he **meddle up** all over again (ICE-SIN:S1A-054, emphasis added)

(23) because of the instructions to **fashion** your jeans **up** by stone-washing and colouring them (ICLE-GE:AUG-0048.3, emphasis added)

(24) our cars worth over a grand if I'd **sprayed it up** and put it on the market (BNC Baby:KD3, emphasis added)

To summarize this second case study, we have seen that phrasal verbs with *up* are underused in both NEs and LEs, but foreign learners of English, unlike NEs speakers, display a lack of style sensitivity in their use of PVs, which is most probably linked to the limited range of genres to which they are exposed in the classroom, and the particularly low exposure to speech. In addition, NEs and LEs show similar (or similarly built) non-native uses of PVs, for example involving the use of a redundant particle.

4.3 Word clusters

4.3.1 *Topic of investigation*

For this third case study, we turn to phraseology, and more particularly word clusters (also called lexical bundles), i.e. repeated sequences of words, with a focus on three-word clusters in academic writing. The topic of phraseology in academic writing has been studied quite extensively (including via the analysis of word clusters) in both ENL and LEs (e.g. Howarth 1996; Chen and Baker 2010). The prefabricated nature of language, and academic writing in particular, has been underlined, as well as EFL learners' difficulty in emulating native speakers in this respect. The problems brought to light by SLA specialists include learners' reliance on a small number of phrases (to be compared with the wider variety displayed in native production) and the influence of the L1 on the choice of certain expressions, often resulting in infelicitous collocations which correspond to calques from the learner's mother tongue. By contrast, phraseology seems to have attracted much less attention in contact linguistics (see Nesselhauf 2011: 159), especially when it comes to the study of word clusters. Three-word clusters are investigated in NEs by Götz and Schilk (2011), but this is within the context of a study that seeks to compare IndE with German learner English (see Section 2).

4.3.2 *Data and methodology*

As outlined in Table 5, the present case study exploits the student writing sections (untimed student essays and student examination scripts) of ICE-HK, ICE-IND, ICE-KEN, ICE-PHI, ICE-SIN and ICE-TAN, as well as the whole of ICLEv2 (with its 16 mother tongue backgrounds). As for the native reference corpus, one must note that the two previous case studies relied on a native corpus that was built either on the model of the LEs data (LOCNEC in the case study on embedded inversion) or that of the NEs data (ICE-GB in the case study on phrasal verbs with *up*). The problem with these corpora is that they are relatively

small in comparison with the non-native data (which combine several components representing different populations). In the case of a (keyword) word cluster analysis, however, it is essential to have access to a sufficiently large reference corpus for the results to be tangible. I therefore decided to work with the academic section of BNC Baby as a baseline. This corpus contains more than one million words and it represents the kind of norm that novice writers are expected to strive for. The clusters were extracted automatically from this and the other corpora, using the WordList function of WST. Only clusters with a minimum frequency of three occurrences were included. On the basis of these lists of clusters and by means of the KeyWords function of WST, a keyword analysis was performed which compared NEs with ENL, and another one which compared LEs with ENL, using the log-likelihood as a statistical test. This made it possible to identify the clusters of the non-native varieties which can be categorized as more or less distinctive compared to the native corpus.

Table 5. Corpora used for the study of word clusters

NEs			Les			ENL		
<i>Corpus</i>	<i>Genre</i>	<i>Size</i>	<i>Corpus</i>	<i>Genre</i>	<i>Size</i>	<i>Corpus</i>	<i>Genre</i>	<i>Size</i>
ICE (HK, IND, KEN, PHI, SIN, TAN)	Student writing	284,977	ICLEv2 (all)	Arg./lit. essays	3,163,142	BNC Baby	Acad. writing	1,019,697

4.3.3 Results and discussion

Table 6 lists the three-word clusters that present the most significant differences in frequency between native English and one of the two types of non-native English varieties considered here. From the upper part of the table, which shows clusters that are underused in NEs and LEs, it appears that all the clusters that are underused in NEs, viz. *per cent of*, *likely to be*, *a number of* and *terms of the*, are also underused in LEs. Foreign learners, in addition, underuse a number of other clusters, like *in relation to* or *extent to which* (which is part of the larger cluster *the extent to which*), and many others which are not listed in Table 6. Beyond some shared underused items, which are typical of academic writing, NEs speakers thus seem to fare better than EFL learners, who underuse a larger number and wider variety of clusters.

Table 6. Under- and overused three-word clusters in NEs and LEs

Clusters	NEs	LEs
Underused	<i>per cent of</i> <i>likely to be</i> <i>a number of</i> <i>terms of the</i>	<i>per cent of</i> <i>in terms of</i> <i>terms of the</i> <i>the house of</i> <i>a number of</i> <i>the effect of</i> <i>likely to be</i> <i>the context of</i> <i>in relation to</i> <i>extent to which</i>
Overused	<i>code switching is</i> <i>the teacher should</i> <i>a lingua franca</i> <i>is very important</i> <i>invitation to treat</i> <i>that the accused</i> <i>in Hong Kong</i> <i>the English language</i> <i>in the society</i> <i>the accused had</i>	<i>a lot of</i> <i>I think that</i> <i>in my opinion</i> <i>dreaming and</i> <i>imagination</i> <i>in the world</i> <i>more and more</i> <i>they do not</i> <i>in order to</i> <i>first of all</i> <i>the real world</i>

In terms of overuse (lower part of Table 6), it turns out that almost all of the most overused clusters in NEs are topic-dependent, that is, related to the topic on which the students had to write, e.g. *code switching is*, *a lingua franca* or *in Hong Kong*. For the present purpose, such clusters can be disregarded, as they are simply due to the fact that the NEs speakers and the native speakers in the corpora were asked to write on different topics. In LEs, on the other hand, several of the overused clusters are topic-independent and underline interesting aspects of (novice) EFL academic writing, some of which have been emphasized in learner corpus research (see e.g. Gilquin, Granger and Paquot 2007a). *A lot of*, which is the most overused cluster in LEs, reflects the propensity of EFL learners to adopt a more spoken-like tone when writing (see Section 4.2.3), since *a lot of* is rather informal, in comparison with *a great deal of*, for example. *More and more* is another cluster that one is more likely to encounter in informal conversations than in academic writing, where *increasingly* would be more

common (Gilquin, Granger and Paquot 2007b: IW3). The overuse of *I think that* and *in my opinion* points to another feature of foreign learners' academic texts, namely their overly involved style (see Petch-Tyson 1998). Foreign learners usually make themselves very visible as writers when they produce academic discourse, especially when they express their opinions, whereas native (expert) writers often prefer more impersonal devices such as *it is important that*, *it seems that* or *interestingly* (Gilquin, Granger and Paquot 2007b: IW16). Admittedly, the presence in ICLE of such involvement markers may also have to do with the type of academic discourse that is found in the corpus: these markers, which are "used to express personal feelings and attitudes and to interact with readers" (Petch-Tyson 1998: 108), are more likely to occur in the argumentative (and literary) essays included in ICLE than in the student essays and examination scripts of ICE or the academic papers of BNC Baby. At the same time, this result echoes the above finding that EFL learners prefer clusters that are typical of speech, as involvement is more characteristic of speech than of writing (Gilquin and Paquot 2008: 48–50). The last type of effect that is apparent from the list of overused clusters in LEs may be described as a (possible) teaching-induced effect. *In order to* is a case in point, since it is often taught as a way of expressing purpose, when the simple form *to* would usually be sufficient and even more adequate. Another example might be the overuse of *first of all* (which is to be related to the overuse of *on the other hand* among four-word clusters). The overuse of such structuring devices may be explained by the fact that "[s]tudents are **drilled** in the categorical use of a short list of expressions – often those functioning as connectives [...] – regardless of whether they are used primarily in spoken or written language (if indeed at all), or to which text types they are appropriate" (Milton 1998: 190, emphasis added). In other words, students are taught to use cohesive devices in order to make the structure of their texts clearer. The outcome, however, can be an overuse of such expressions (see also Milton and Tsang 1993). Some of these devices (like *in order to* or *so as to*, which both perform the same function) are overused in NEs too, but to a much lesser extent than in LEs. This could be explained by the smaller proportion which formal teaching represents in NEs students' total experience with the English language.

In summary, we have seen in this case study that NEs and LEs writers underuse the same clusters (although foreign learners underuse other clusters as well), but that it is mainly LEs writers who heavily overuse certain topic-independent clusters. It has also been suggested that teaching might have an influence on the (over)use of certain cohesive devices and that this influence, if it

exists, seems to be stronger in LEs than in NEs, as one would expect from the settings in which the two types of varieties are acquired/learned.

4.4 Discourse markers

4.4.1 *Topic of investigation*

The final case study looks at the use of discourse markers (DMs) in NEs and LEs. The use of DMs in both native and non-native Englishes has been widely discussed, with the latter covering NEs (e.g. Valentine 1991; Lange 2009) and LEs (e.g. Müller 2005; Aijmer 2011). The contact linguistics literature has mostly focused on indigenous DMs (or discourse particles) and the sociolinguistic functions that these particles can fulfil in NEs. In SLA research, on the other hand, it is mainly the misuse of English DMs by EFL learners that has been the centre of attention, together with a learner's tendency to overproduce certain DMs, like *well*, to the detriment of others, like *sort of* (see Gilquin 2008). The focus here will be on two-word DMs, and more precisely the five DMs *and so*, *and then*, *I mean*, *sort of* and *you know*, which correspond to the two-word DMs that reach a minimum relative frequency of 250 occurrences per 100,000 words in at least one of the corpus components.

4.4.2 *Data and methodology*

This study relies on the data taken from the direct conversations of six components of ICE (ICE-HK, ICE-IND, ICE-KEN, ICE-PHI, ICE-SIN and ICE-TAN), as well as all the informal interviews contained in LINDSEI and its eleven components. As for the native baseline, given the possible influence of context on the use and frequency of DMs, especially between conversations and interviews (see Fuller 2003), I decided to combine the direct conversations of ICE-GB and the informal interviews of LOCNEC into a single reference corpus. This combined benchmark represents an attempt to avoid the kind of bias that could result from the sole reliance on a corpus representing either the context of the NEs data or that of the LEs data. An overview of the corpora used can be found in Table 7.

Table 7. Corpora used for the study of discourse markers

NEs			LEs			ENL		
<i>Corpus</i>	<i>Genre</i>	<i>Size</i>	<i>Corpus</i>	<i>Genre</i>	<i>Size</i>	<i>Corpus</i>	<i>Genre</i>	<i>Size</i>
ICE (HK, IND, KEN, PHI, SIN, TAN)	Direct conv.	1,035,122	LINDSEI (all)	Informal interviews	792,141	ICE-GB	Direct conv.	183,366
						LOCNEC	Informal interviews	118,398

All the occurrences of the five DMs under investigation were extracted from these corpora. Each concordance line was then carefully examined to distinguish the DM uses of the bigrams from their non-DM uses (i.e. cases where the bigram did not behave as a syntactically non-obligatory string).¹⁴ In addition, the study probes into the use of indigenous/L1 DMs by making concordances of some DMs selected on the basis of the literature (e.g. Wong 2004; Gilquin 2008).

4.4.3 Results and discussion

With respect to the overall frequency of the five DMs (Figure 3), NEs, LEs and ENL speakers appear to form three distinct groups, with the differences between them being statistically highly significant ($p < 0.0001$). Both types of non-native varieties present an underuse of DMs, but NEs, which are situated in-between LEs and ENL, contain more DMs than the foreign varieties (while still being closer to LEs than to ENL).

¹⁴ The results for LINDSEI and LOCNEC are based on a study that I carried out with Sylviane Granger (see Gilquin and Granger *fc.*).

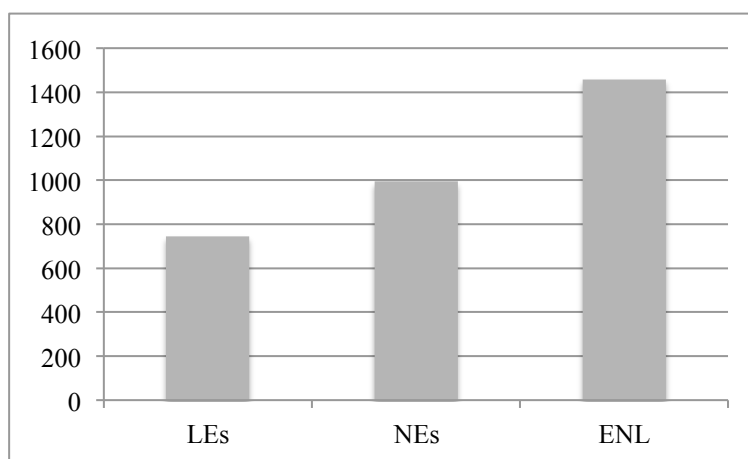


Figure 3. Overall relative frequency per 100,000 words of the five DMs in LEs, NEs and ENL

If we consider the five DMs individually (see Figure 4), we notice that *and so* is relatively uncommon in the corpus data, and that *and then* presents quite similar frequencies across LEs, NEs and ENL. Among the other DMs, *I mean* and *sort of* are underused by NEs speakers and EFL learners alike ($p < 0.0001$), *sort of* being the most underused of the two. In addition, *you know* is also underused in LEs ($p < 0.0001$), but it is overused in NEs ($p < 0.0001$). Of the two most common DMs found in ENL, viz. *you know* and *I mean*, both are more frequent in NEs than in LEs, a difference which is statistically significant ($p < 0.0001$) and which points to a more native-like use of DMs in the Outer Circle than in the Expanding Circle, at least from a quantitative point of view.

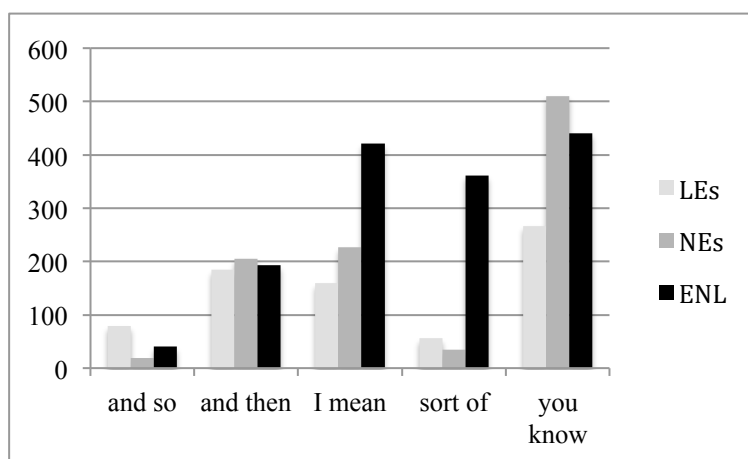


Figure 4. Individual relative frequency per 100,000 words of the five DMs in LEs, NEs and ENL

This greater ability among NEs speakers compared with EFL learners to approximate the native norm in their use of DMs can be accounted for by referring to the acquisition setting in the Outer and Expanding Circles. DMs have been described as “language which is not taught” (Hellermann and Vergun 2007). It is therefore not altogether surprising that foreign learners, who acquire English in a classroom setting, may not have been able to learn them. In addition, DMs are associated with authentic speech, which EFL learners rarely get exposed to (see Gilquin and Paquot 2008: 52). By contrast, NEs speakers, who also acquire and use English in everyday life, are more likely to be exposed to natural English and hence (re)produce common DMs when speaking English.

While the focus of this study is on a selection of five DMs found in native English, it is interesting to notice that both NEs and LEs also regularly include indigenous/L1 DMs. French-speaking EFL learners, for example, can be shown to use the French DMs *enfin* and *allez* when speaking English. *Allez* is also found among (Belgian) Dutch learners, who often use this French word when speaking Dutch but occasionally use it when speaking English too. Examples (25) to (27) illustrate these uses. Note how in (26) the French DM *allez* is followed within a few words by the English DM *you know*.

- (25) and I thought well . if I if I go to university and here the lessons **enfin**
the lectures in Dutch and . but . when you hear it it's not the same
(LINDSEI-FR:018, emphasis added)

(26) I didn't know how to: to: to note my (er) ... (erm) .. **allez** th= the the date (eh) the datas<?> of my exams . the timetable you know (LINDSEI-FR:046, emphasis added)

(27) when you go sailing you don't need anything so **allez** you need wind that's all (LINDSEI-DU:035, emphasis added)

In NEs, the DM *la(h)* is quite widespread in several varieties, including SgE and HKE, but other indigenous DMs are found to occur in NEs, including *meh*, *na* and *sha*. Examples are given in (28) to (30). As was the case in (26) above, indigenous DMs appear to be compatible with native English DMs: in (28) *lah* co-occurs with *like* and *you know*.

(28) And they are saying they are trying to like get away **lah** you know (ICE-SIN:S1A-087, emphasis added)

(29) You mean at home cannot sleep with air-con one meh (ICE-SIN:S1A-014, emphasis added)

(30) Don't look me at me in this way **la** (ICE-HK:S1A-086)

Some of the indigenous/L1 DMs can actually reach a frequency which is as high as, and sometimes higher than, that of English DMs. Figure 5 reveals that *allez* in the French component of LINDSEI is not very common in comparison with the other (native) DMs, but *enfin* is more frequent than *sort of*.

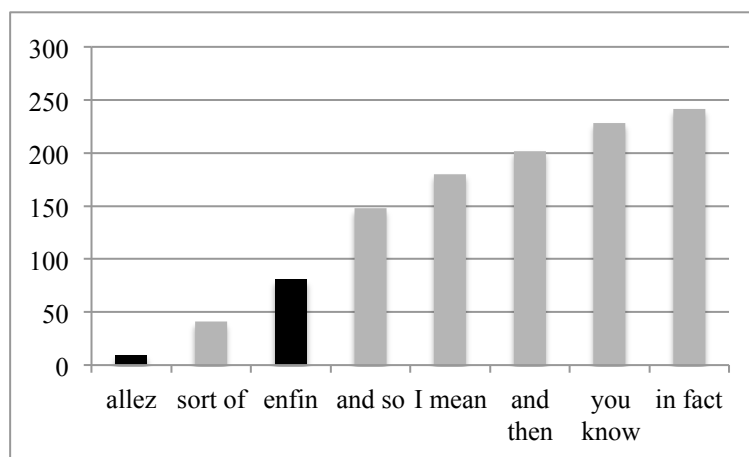


Figure 5. Relative frequency per 100,000 words of DMs in LINDSEI-French

In SgE, most of the indigenous DMs that I looked at are relatively infrequent, but the DM *lah* is so common that it exceeds the frequency of most of the English DMs, with the exception of *you know* (see Figure 6).

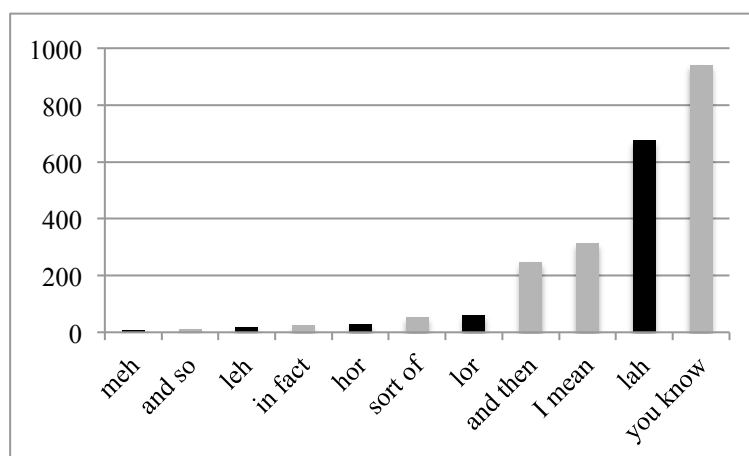


Figure 6. Relative frequency per 100,000 words of DMs in ICE-SIN

One possible explanation for the (sometimes massive) presence of indigenous/L1 DMs in NEs and LEs has to do with the pressure of online planning in speech, which is even higher for EFL learners due to their lower proficiency level in English and the relative lack of practice they tend to have in using English for spoken communication. Because of this pressure, most of the speaker's attention has to be devoted to the content of his/her message, rather than to more pragmatic aspects of it, such as the use of DMs. It seems that DMs might partly escape monitoring, with the speakers perhaps not even realising that they are using DMs from their L1. In this respect, it is probably not a coincidence that among the non-English words found in the LINDSEI corpus, 12 per cent represent DMs. An alternative explanation in the case of NEs is that indigenous DMs serve as "ethnic identity markers" (Boas and Weilbacher 2007: 33) or, as Gold and Tremblay (2006) call them, "national icons", which allow NEs speakers to maintain their local identity. This function may only become available to very advanced and proficient speakers who are able to monitor the use of these items in speech. It might also be that NEs speakers' (persistent) use of indigenous DMs is due to the input they receive, which is mostly from people for whom English is also a second language.¹⁵

¹⁵ I thank Marianne Hundt for pointing this out to me.

To sum up, what this case study has shown is that NEs are situated in-between LEs and ENL in terms of overall (English) DMs frequency, although both groups tend to significantly underuse DMs in comparison with ENL. In addition, NEs and LEs share the use of indigenous/L1 DMs, a feature that is especially striking in NEs, which could be explained by the fact that DMs enable NEs speakers to make their ethnic identity visible.

5. From description to explanation

The above case studies are mainly exploratory, and only some of the most interesting findings have been presented. Each of the four linguistic phenomena investigated would deserve to be further researched, for example through more qualitative analyses taking particular instances of the phenomena into account. Yet these studies (and those that were introduced in Section 2) demonstrate that NEs and LEs, despite their differences, share a number of features, a factor which justifies a unified approach to contact linguistics and second language acquisition research. However, such a descriptive account is only the first step towards a truly integrated framework combining the two fields, and needs to be followed by an explanatory account and, ultimately, a theoretical one. In this section, I will very briefly illustrate how the descriptive account of the above studies can lead to an explanatory account, which represents a necessary attempt at generalization and can provide a deeper understanding of the relation between NEs and LEs.

The differences between the two types of varieties can easily be explained by referring to the different contexts of acquisition, the different mother tongues, etc. To explain the similarities between them, on the other hand, it is possible to appeal to general cognitive processes of language acquisition (and possibly language change), such as the ones that are listed by Schneider (2012). Thus, “redundancy”, which Schneider (2012: 65) describes as “the unmotivated repeated (or double) marking of the same piece of information”, could explain the use of the particle *up* with a verb that already expresses upward movement or telicity, e.g. *rise up* or *complete up* (Section 4.2). “Analogy”, which corresponds to “the transfer of some principle or procedure from a source domain to a target domain, so that one sub-system is structured on the model of another” (Schneider 2012: 66), might be responsible for the general phenomenon of embedded inversion (Section 4.1), since the inverted word order of the indirect question is structured on the model of the direct question. It could also account for a case

like *cope up with*, probably built by analogy with *come up with*, *meet up with* or *put up with* (Section 4.2). Although it is not mentioned as such by Schneider, who presents it as part of the redundancy process, we could add to this list the process of “explicitation”, which refers to the “maximizing [of] transparency” (Schneider 2012: 67) and could explain the overuse of explicit structuring expressions like *first of all*, *on the other hand* or *in order to* (Section 4.3).

The other processes mentioned by Schneider (“simplicity”, “generalization”, “regularization”, etc.) and other similar processes characterizing the acquisition of a non-native language might serve to explain other features, yet to be uncovered, that are shared between NEs and LEs. The effect of these processes could be additive so that, for instance, the cognitive process of explicitation could combine with the influence of teaching, resulting in a heavier overuse of structuring expressions in LEs than in NEs (see above). Of course, one should be careful when attributing linguistic features to certain processes since, for example, surface similarities may hide different phenomena (see Davydova et al. 2011: 317). The explanatory step should therefore be executed very carefully, ideally on the basis of converging evidence, to substantiate one’s claims and avoid hasty conclusions or misinterpretation. Only then could it provide a sound basis for a theoretical rapprochement between contact linguistics and SLA research, which should be the ultimate goal of comparative approaches to NEs and LEs.

6. Challenges and some concluding remarks

By way of conclusion, I would like to point out what I see as the main challenges facing researchers who want to compare NEs and LEs. The first has to do with the information available about the subjects investigated. SLA specialists have always been aware of the possible influence of a wide range of factors on L2 production, which is why learner corpora tend to be accompanied by rich metadata about the learners whose writing or speech is included in the corpus (in LINDSEI, for example, no less than 23 variables are recorded per interview). Corpora of NEs, by contrast, often lack such metadata, which means that it is usually impossible to study the influence of sociolinguistic variables like the age or gender of the speakers, or their amount of exposure to English.¹⁶ The issue of

¹⁶ This type of information is available for some of the ICE corpora but is not necessarily published (Marianne Hundt, p.c., December 2013).

the mother tongue is also a delicate one for NEs, given the multilingual environment in which the users of these varieties live. A corpus of NEs usually provides an indication of the speakers' country of origin, but this does not necessarily say anything about their L1, which may vary from one region to another – for all we know, there might even be native speakers of English among them, as is the case in some of the ICE corpora. We also do not know whether the speakers have a single native language or whether other languages may have an influence on their use of English. NEs corpora including metadata similar to those found in learner corpora would thus be a welcome addition to the list of resources available to conduct research in contact linguistics but also, obviously, to draw comparisons between NEs and LEs.

Another challenge, already evoked at the beginning of Section 4, has to do with corpus comparability. It is practically impossible to find corpora that are fully comparable with each other across Kachru's three circles. With NEs and LEs having essentially been treated separately so far, it seems as if nobody ever really thought about or cared to compile comparable corpora for these two types of varieties. We can find similarly designed corpora of NEs and ENL (like ICE) and similarly designed corpora of LEs and ENL (like LINDSEI and LOCNEC), but comparable data representing NEs and LEs (plus possibly ENL) are very hard to find – or the samples are quite small (as is the case with the student writing sections of ICE, which are reasonably comparable with the ICLE data but limited in size). Two such corpora have, however, recently been initiated: ICNALE, the International Corpus Network of Asian Learners of English, which includes essays written by Asian speakers from the Expanding and Outer Circles, and ENODASE, the ENO Database of Student English, coordinated by Heli Paulasto at the University of Eastern Finland – Joensuu, which will include student writing produced by speakers from the three circles.

In-depth comparisons of NEs and LEs are also made difficult by the restricted range of genres that are represented in the corpora. This, naturally, is related to the limited number of functions that the language varieties can fulfil, especially the learner varieties. Learner corpora will normally have to be collected within the setting of the classroom, through tasks which the learners perform as part of their regular classroom activity. NEs data, on the other hand, can be collected in more spontaneous communicative settings, although in a number of contexts the default option would be the use of the local language rather than English.

Finally, because we generally have to rely on corpora that were collected several years ago, it is difficult to keep track of the evolution of the varieties.

NEs often have a natural tendency to evolve towards a more stabilized and endonormative form (as described in Schneider's 2003 dynamic model), while EFL learners may benefit from increased exposure to (semi-)natural English in their everyday life, for example through internet forums or through American TV series downloaded from the Web, which are becoming increasingly popular among young EFL learners. Such more recent developments may not be reflected in the corpora that are currently available.

These challenges, however, should not discourage us from digging deeper into the relation between NEs and LEs, in an attempt to bring contact linguistics and SLA research closer together and lead them to cross-fertilize each other. As suggested by the four case studies discussed in this paper, this involves taking into account both diverging forces, those that bring about differences between the two types of varieties (like the acquisition setting), and converging forces, which result in similarities between the two types of varieties (like general cognitive processes of language acquisition). Furthermore, a complete integration between the fields of contact linguistics and SLA research will only be possible if, next to empirical comparative studies like those presented here, we work on the development of a common terminological, methodological and theoretical apparatus to provide a truly integrated framework for the analysis of NEs and LEs. Such a framework, ultimately, might even be expanded to include the comparison with other language varieties like English as a Lingua Franca (see Schneider 2012), pidgins and creoles, dialects or translated language, and with other phenomena such as child language acquisition (ontogeny), diachronic evolution (phylogeny), or indeed with corresponding varieties and phenomena in other languages (typology). If certain principles could empirically be shown to be valid for any type of language change or development, this would push back many frontiers and would yield fundamental insights into the workings of the human mind.

Sources

BNC (British National Corpus) Baby.
International Corpus of English (ICE).
International Corpus of Learner English, Version 2 (ICLEv2).
Louvain Corpus of Native English Conversation (LOCNEC).
Louvain International Database of Spoken English Interlanguage (LINDSEI).

References

- Aceto, Michael. 2006. "Caribbean Englishes". In Braj B. Kachru, Yamuna Kachru, and Cecil L. Nelson, eds. *The Handbook of World Englishes*. Oxford: Blackwell Publishing, 203–222.
- Aijmer, Karin. 2011. "Well I'm not sure I think... The Use of *well* by Non-Native Speakers". *International Journal of Corpus Linguistics* 16: 231–254.
- Altenberg, Bengt, and Marie Tapper. 1998. "The Use of Adverbial Connectors in Advanced Swedish Learners' Written English". In Sylviane Granger, ed. *Learner English on Computer*. London and New York: Longman, 80–93.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow, Essex: Pearson Education Limited.
- Boas, Hans C., and Hunter Weilbacher. 2007. "How Universal Is the Pragmatic Detachability Scale? Evidence from Texas German Discourse Markers". In Frederic Hoyt, Nikki Seifert, Alexandra Teodorescu, and Jessica White, eds. *The Proceedings of the Texas Linguistic Society IX Conference: The Morphosyntax of Underrepresented Languages*. Stanford: CSLI Publications, 33–58.
- Bolton, Kingsley. 2006. "World Englishes Today". In Braj B. Kachru, Yamuna Kachru, and Cecil L. Nelson, eds. *The Handbook of World Englishes*. Oxford: Blackwell Publishing, 240–269.
- Chen, Hsin-I. 2010. "Contrastive Learner Corpus Analysis of Epistemic Modality and Interlanguage Pragmatic Competence in L2 Writing". *Arizona Working Papers in SLA & Teaching* 17: 27–51.
- , and Paul Baker. 2010. "Lexical Bundles in L1 and L2 Academic Writing". *Language Learning & Technology* 14: 30–49.
- Chuang, Fei-Yu, and Hilary Nesi. 2006. "An Analysis of Formal Errors in a Corpus of L2 English Produced by Chinese Students". *Corpora* 1: 251–271.
- Coulmas, Florian, ed. 1981. *A Festschrift for Native Speaker*. Berlin and New York: Mouton de Gruyter.
- Dagneaux, Estelle, Sharon Denness, and Sylviane Granger. 1998. "Computer-Aided Error Analysis". *System* 26: 163–174.
- Dagut, Menachem, and Batia Laufer. 1985. "Avoidance of Phrasal Verbs: A Case for Contrastive Analysis". *Studies in Second Language Acquisition* 7: 73–79.
- Davydova, Julia. 2012. "Englishes in the Outer and Expanding Circles: A Comparative Study". *World Englishes* 31: 366–385.
- , Michaela Hilbert, Lukas Pietsch, and Peter Siemund. 2011. "Comparing Varieties of English: Problems and Perspectives". In Peter Siemund, ed. *Linguistic Universals and Language Variation*. Berlin and New York: Mouton de Gruyter, 291–323.
- De Cock, Sylvie. 2011. "Contracted Forms in LINDSEI and LOCNEC". Paper presented at Learner Corpus Research 2011, Louvain-la-Neuve.
- Ellis, Rod. 2008. *The Study of Second Language Acquisition* (2nd ed.). Oxford: Oxford University Press.
- Fuller, Janet M. 2003. "The Influence of Speaker Roles on Discourse Marker Use". *Journal of Pragmatics* 35: 23–45.

- Gardner, Dee, and Mark Davies. 2007. "Pointing out Frequent Phrasal Verbs: A Corpus-Based Analysis". *TESOL Quarterly* 41: 339–359.
- Gilquin, Gaëtanelle. 2008. "Hesitation Markers among EFL Learners: Pragmatic Deficiency or Difference?". In Jesús Romero-Trillo, ed. *Pragmatics and Corpus Linguistics: A Mutualistic Entente*. Berlin and New York: Mouton de Gruyter, 119–149.
- . 2011. "Corpus Linguistics to Bridge the Gap between World Englishes and Learner Englishes". In Leonel Ruiz Miyares, and María Rosa Álvarez Silva, eds. *Comunicación Social en el siglo XXI, Vol. II*. Santiago de Cuba: Centro de Lingüística Aplicada, 638–642.
- , and Magali Paquot. 2008. "Too Chatty: Learner Academic Writing and Register Variation". *English Text Construction* 1: 41–61.
- , and Sylviane Granger. 2011. "From EFL to ESL. Evidence from the *International Corpus of Learner English*". In Joybrato Mukherjee, and Marianne Hundt, eds. *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*. Amsterdam and Philadelphia: John Benjamins, 55–78.
- , and Lea Meriläinen. 2012. Combining Learner English and World Englishes Perspectives: The Case of Embedded Inversion. Paper presented at the 22nd Annual Conference of the European Second Language Association (EUROSLA), Poznań.
- , and ———. fc. "Learner Language". In Douglas Biber, and Randi Reppen, eds. *Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press.
- , Sylviane Granger, and Magali Paquot. 2007a. "Learner Corpora: The Missing Link in EAP Pedagogy". *Journal of English for Academic Purposes* 6: 319–335.
- , ———, and ———. 2007b. "Improve your Writing Skills (Writing Sections)". In Michael Rundell, ed. *Macmillan English Dictionary for Advanced Learners* (2nd ed.). Oxford: Macmillan Education, IW1–IW29.
- , Sylvie De Cock, and Sylviane Granger. 2010. *Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM*. Louvain-la-Neuve: Presses universitaires de Louvain.
- Gold, Elaine, and Mireille Tremblay. 2006. "Eh? and hein?: Discourse Particles or National Icons?". *Canadian Journal of Linguistics/Revue canadienne de linguistique* 51: 247–263.
- Görlach, Manfred. 2002. *Still More Englishes*. Amsterdam and Philadelphia: John Benjamins.
- Götz, Sandra, and Marco Schilk. 2011. "Formulaic Sequences in Spoken ENL, ESL and EFL: Focus on British English, Indian English and Learner English of Advanced German Learners". In Joybrato Mukherjee, and Marianne Hundt, eds. *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*. Amsterdam and Philadelphia: John Benjamins, 79–100.
- Granger, Sylviane, Estelle Dagneaux, Fanny Meunier, and Magali Paquot. 2009. *The International Corpus of Learner English. Handbook and CD-ROM. Version 2*. Louvain-la-Neuve: Presses universitaires de Louvain.
- Hellermann, John, and Andrea Vergun. 2007. "Language Which Is Not Taught: The Discourse Marker Use of Beginning Adult Learners of English". *Journal of Pragmatics* 39: 157–179.

- Hilbert, Michaela. 2011. "Interrogative Inversion as a Learner Phenomenon in English Contact Varieties: A Case of Angloveralls?". In Joybrato Mukherjee, and Marianne Hundt, eds. *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*. Amsterdam and Philadelphia: John Benjamins, 125–143.
- Honna, Nobuyuki. 2006. "East Asian Englishes". In Braj B. Kachru, Yamuna Kachru, and Cecil L. Nelson, eds. *The Handbook of World Englishes*. Oxford: Blackwell Publishing, 114–129.
- Howarth, Peter Andrew. 1996. *Phraseology in English Academic Writing. Some Implications for Language Learning and Dictionary Making*. Tübingen: Max Niemeyer Verlag.
- Hulstijn, Jan H., and Elaine Marchena. 1989. "Avoidance. Grammatical or Semantic Causes?". *Studies in Second Language Acquisition* 11: 241–255.
- Hundt, Marianne, and Christian Mair. 1999. "'Agile' and 'Uptight' Genres: The Corpus-Based Approach to Language Change in Progress". *International Journal of Corpus Linguistics* 4: 221–242.
- , and Joybrato Mukherjee. 2011. "Introduction: Bridging a Paradigm Gap". In Joybrato Mukherjee, and Marianne Hundt, eds. *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*. Amsterdam and Philadelphia: John Benjamins, 1–5.
- , and Katrin Vogel. 2011. "Overuse of the Progressive in ESL and Learner Englishes – Fact or Fiction?". In Joybrato Mukherjee, and Marianne Hundt, eds. *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*. Amsterdam and Philadelphia: John Benjamins, 145–165.
- ICAME Journal. 2010. *ICE Age 2: ICE Corpora of New Englishes in the Making*. 34: 5–88.
- Kachru, Braj B. 1985. "Standards, Codification and Sociolinguistic Realism: The English Language in the Outer Circle". In Randolph Quirk, and Henry G. Widdowson, eds. *English in the World: Teaching and Learning the Language and Literatures*. Cambridge: Cambridge University Press, 11–30.
- Kieweg, Werner. 2000. "Zur Mündlichkeit im Englischunterricht". *Der Fremdsprachliche Unterricht Englisch* 34: 4–9.
- Kolbe, Daniela, and Andrea Sand. 2010. "Embedded Inversion Worldwide". *Linguaculture* 1(1): 25–42.
- Kortmann, Bernd, and Kerstin Lunkenheimer, eds. 2011. *The Electronic World Atlas of Varieties of English [eWAVE]*. Leipzig: Max Planck Institute for Evolutionary Anthropology <<http://www.ewave-atlas.org/>> (3 March 2014).
- Lange, Claudia. 2009. "'Where's the Party yaar!'. Discourse Particles in Indian English". In Thomas Hoffmann, and Lucia Siebers, eds. *World Englishes – Problems, Properties and Prospects*. Amsterdam and Philadelphia: John Benjamins, 207–226.
- Laporte, Samantha. 2012. "Mind the Gap! Bridge between World Englishes and Learner Englishes in the Making". *English Text Construction* 5: 265–292.
- Liao, Yan, and Yoshinori J. Fukuya. 2004. "Avoidance of Phrasal Verbs: The Case of Chinese Learners of English". *Language Learning* 54: 193–226.
- Mair, Christian. 2003. "Kreolismen und verbales Identitätsmanagement im geschriebenen jamaikanischen Englisch". In Elisabeth Vogel, Antonia Napp, and Wolfram Lutterer, eds. *Zwischen Ausgrenzung und Hybridisierung*. Würzburg: Ergon, 79–96.

- Mendis, Dushyanthi, and Harshana Rambukwella. 2010. "Sri Lankan Englishes". In Andy Kirkpatrick, ed. *The Routledge Handbook of World Englishes*. London and New York: Routledge, 181–196.
- Meriläinen, Lea. 2010. *Language Transfer in the Written English of Finnish Students*. Publications of the University of Eastern Finland. Dissertations in Education, Humanities, and Theology, no 9. Joensuu: University of Eastern Finland.
- , and Heli Paulasto. fc. "Embedded Inversion as an Angloversal: Evidence from Inner, Outer and Expanding Circle Englishes". In Markku Filppula, Juhani Klemola, and Devyani Sharma, eds. *The Oxford Handbook of World Englishes*. Oxford: Oxford University Press.
- Mesthrie, Rajend. 2006. "Contact Linguistics and World Englishes". In Braj B. Kachru, Yamuna Kachru, and Cecil L. Nelson, eds. *The Handbook of World Englishes*. Oxford: Blackwell Publishing, 273–288.
- Meyler, Michael. 2007. *A Dictionary of Sri Lankan English*. Colombo: Michael Meyler.
- Milton, John. 1998. "Exploiting L1 and Interlanguage Corpora in the Design of an Electronic Language Learning and Production Environment". In Sylviane Granger, ed. *Learner English on Computer*. London and New York: Longman, 186–198.
- , and Elza Shuk-ching Tsang. 1993. "A Corpus-Based Study of Logical Connectors in EFL Students' Writing: Directions for Future Research". In Richard Pemberton, and Elza Shuk-ching Tsang, eds. *Studies in Lexis*. Hong Kong: The Hong Kong University of Science and Technology, 215–246.
- Mukherjee, Joybrato, and Stefan Th. Gries. 2009. "Collostructional Nativisation in New Englishes: Verb-Construction Associations in the International Corpus of English". *English World-Wide* 30: 27–51.
- Müller, Simone. 2005. *Discourse Markers in Native and Non-Native English Discourse*. Amsterdam and Philadelphia: John Benjamins.
- Nayar, P. Bhaskaran. 1997. "ESL/EFL Dichotomy Today: Language Politics or Pragmatics?". *TESOL Quarterly* 31: 9–37.
- Nesselhauf, Nadja. 2009. "Co-Selection Phenomena across New Englishes: Parallels (and Differences) to Foreign Learner Varieties". *English World-Wide* 30: 1–25.
- . 2011. "Exploring the Phraseology of ESL and EFL Varieties". In Thomas Herbst, Susen Faulhaber, and Peter Uhrig, eds. *The Phraseological View of Language. A Tribute to John Sinclair*. Berlin and New York: Mouton de Gruyter, 159–177.
- Petch-Tyson, Stephanie. 1998. "Writer/Reader Visibility in EFL Written Discourse". In Sylviane Granger, ed. *Learner English on Computer*. London and New York: Longman, 107–118.
- Platt, John. 1989. "The Nature of Indigenized Englishes: Interference – Creativity – Universals". *Language Sciences* 11: 395–407.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London and New York: Longman.
- Schneider, Edgar W. 2003. "The Dynamics of New Englishes: From Identity Construction to Dialect Birth". *Language* 79: 233–281.
- . 2004. "How to Trace Structural Nativization: Particle Verbs in World Englishes". *World Englishes* 23: 227–249.

- . 2007. *Postcolonial English: Varieties Around the World*. Cambridge: Cambridge University Press.
- . 2012. "Exploring the Interface between World Englishes and Second Language Acquisition – and Implications for English as a Lingua Franca". *Journal of English as a Lingua Franca* 1: 57–91.
- Scott, Mike. 2008. *WordSmith Tools Version 5*. Liverpool: Lexical Analysis Software.
- Sharma, Devyani. 2001. "The Pluperfect in Native and Non-Native English: A Comparative Corpus Study". *Language Variation and Change* 13: 343–373.
- Siegel, Jeff. 2010. "Contact Languages of the Pacific". In Raymond Hickey, ed. *The Handbook of Language Contact*. Oxford: Wiley-Blackwell, 814–836.
- Siyanova, Anna, and Norbert Schmitt. 2007. "Native and Nonnative Use of Multi-Word vs. One-Word Verbs". *IRAL* 45: 119–139.
- Sridhar, Kamal K., and Shikaripur N. Sridhar. 1986. "Bridging the Paradigm Gap: Second Language Acquisition Theory and Indigenized Varieties of English". *World Englishes* 5: 3–14.
- Szmrecsanyi, Benedikt, and Bernd Kortmann. 2011. "Typological Profiling: Learner Englishes versus Indigenized L2 Varieties of English". In Joybrato Mukherjee, and Marianne Hundt, eds. *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*. Amsterdam and Philadelphia: John Benjamins, 167–187.
- Thewissen, Jennifer. 2013. "Capturing L2 Accuracy Developmental Patterns: Insights from an Error-Tagged EFL Learner Corpus". *The Modern Language Journal* 97: 77–101.
- Valentine, Tamara, M. 1991. "Getting the Message Across: Discourse Markers in Indian English". *World Englishes* 10: 325–334.
- Winford, Donald. 2009. "The Interplay of 'Universals' and Contact-Induced Change in the Emergence of New Englishes". In Markku Filppula, Juhani Klemola, and Heli Paulasto, eds. *Vernacular Universals and Language Contacts: Evidence from Varieties of English and Beyond*. London and New York: Routledge, 206–230.
- Wong, Jock. 2004. "The Particles of Singapore English: A Semantic and Cultural Interpretation". *Journal of Pragmatics* 36: 739–793.
- Zipp, Lena, and Tobias Bernaisch. 2012. "Particle Verbs across First and Second Language Varieties of English". In Marianne Hundt, and Ulrike Gut, eds. *Mapping Unity and Diversity World-Wide. Corpus-Based Studies of New Englishes*. Amsterdam and Philadelphia: John Benjamins, 167–196.

Author's address

Gaëtanelle Gilquin
Centre for English Corpus Linguistics
SSH/ILC
Université catholique de Louvain
Collège Erasme
Place Blaise Pascal 1, bte L3.03.33
1348 Louvain-la-Neuve (Belgium)
gaetanelle.gilquin@uclouvain.be