

GAËTANELLE GILQUIN
Université catholique de Louvain
Louvain-la-Neuve
Belgium
gaetanelle.gilquin@uclouvain.be

Corpus linguistics to bridge the gap between World Englishes and Learner Englishes

1. Introduction

This paper aims to bridge the paradigm gap that exists between World Englishes (new, indigenised varieties of English spoken in countries like India or Singapore) and Learner Englishes (foreign varieties of English spoken in, e.g., France or Spain), by employing the tools and methods of corpus linguistics. More precisely, it relies on data from the *International Corpus of English* and the *International Corpus of Learner English* (among others) to investigate and compare the use of phrasal verbs with *up* in the two types of varieties, both quantitatively and qualitatively, highlighting the differences between them but also (and perhaps more importantly) the features that unite them.

2. World Englishes vs Learner Englishes

Braj Kachru is well-known for his model of the three concentric circles, a model which seeks to represent “the types of spread, the patterns of acquisition and the functional domains in which English is used across cultures and languages” (Kachru 1985: 12). The so-called “inner circle” includes countries where English is used as a mother tongue (viz. United Kingdom, United States of America, Canada, Australia and New Zealand). The “outer circle” includes former British or American colonies where English is an official or semi-official language, used for example in education, administration, literature or the media (e.g. India, Nigeria, Singapore). Although people in these countries normally have another language as their mother tongue, they learn English as a second language (ESL), that is, they learn it in a naturalistic environment and with no or little formal instruction (i.e. classes specifically devoted to the learning of the target language). These non-native varieties are also called “World Englishes” (Bhatt 2001) or “indigenized varieties of English” (Sridhar & Sridhar 1986). The last circle in Kachru’s model is the “expanding circle”, which includes non-native speakers who use English as a lingua franca in areas like business, diplomacy, politics, research or tourism. Unlike the outer circle, English in the expanding circle is not an official language in the countries where it is used. It is taught as a foreign language (EFL), that is, in the classroom context and with limited exposure to the language outside the classroom. Examples of countries belonging to the expanding circle are Brazil, Japan and Spain. These varieties are also referred to as learner English, or (by analogy with “World Englishes” and in an attempt to emphasise the diversity that may exist among learners from different mother tongue backgrounds) as “Learner Englishes” (see Gilquin & Granger in press). Kachru’s model of the three concentric circles is illustrated in Figure 1.

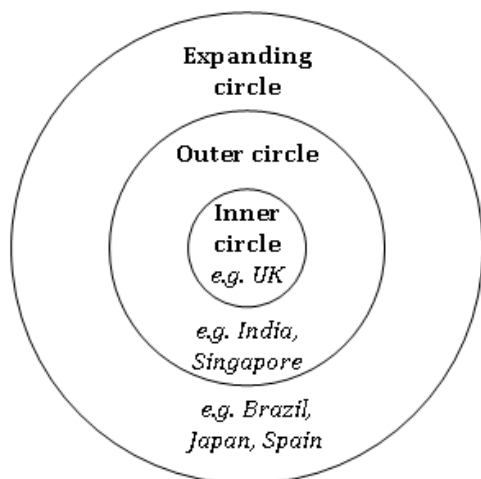


Figure 1. Kachru's (1985) model of the three concentric circles

Despite the fact that the outer circle and the expanding circle correspond to different contexts of acquisition of the language (ESL vs EFL) and despite the geographical distance that separates the countries belonging to the two respective circles, some linguists have called for a rapprochement between World Englishes and Learner Englishes, underlining the benefit that both fields could gain from it. Thus, more than twenty years ago, Sridhar & Sridhar (1986) already evoked the idea of such a rapprochement. Their proposal, however, has hardly been followed by any concrete effects. World Englishes and Learner Englishes have both led to a great deal of research, but with no real contact between them. In this paper, it is claimed that a first step towards the type of theoretical rapprochement advocated by Sridhar & Sridhar (1986) could be an empirical investigation of the

possible similarities between World Englishes and Learner Englishes. The framework that is proposed for such an analysis is the framework of corpus linguistics, as described in the next section.

3. Corpus linguistics and non-native English

Corpus linguistics can be defined as a methodology that relies on the use of a corpus, that is, “a collection of machine-readable authentic texts (including transcripts of spoken data) which is sampled to be representative of a particular language or language variety” (McEnery et al. 2006: 5). The first major corpora represented varieties from the inner circle (i.e. native English): American English for the Brown corpus (Francis & Kučera 1964) and British English for the LOB (Lancaster-Oslo-Bergen) corpus (Johansson et al. 1978). From the 1990s onwards, however, non-native varieties of English (outer and expanding circles) started to make their way into corpora too.

Sidney Greenbaum launched the *International Corpus of English* (ICE) project (see Greenbaum 1996), meant to collect data from countries where English is the first language (inner circle varieties, cf. ICE-GB, the British component of ICE), but also countries where English is a second official language (outer circle varieties, cf. ICE-Fiji or ICE-Philippines). Each ICE component consists of one million words of spoken and written English and follows a similar structure, with a number of genres and sub-genres, such as direct conversations, business transactions, spontaneous commentaries or broadcast news (Nelson 1996). Today, over twenty components are ready or in the process of being compiled (see <http://ice-corpora.net/ice/index.htm>), and numerous linguists have already exploited these data to study various features of World Englishes (see for example the special issue of the journal *World Englishes* published in 2004).

Originally part of the ICE project (cf. Granger 1996), but later growing apart from it, the *International Corpus of Learner English* (ICLE) contains argumentative essays written by EFL learners from several mother tongue backgrounds, and thus represents the last circle in Kachru's model, the expanding circle. The first version of the corpus included data produced by learners from 11 mother tongue backgrounds, for a total of some 2.5 million words (Granger et al. 2002), while the second version includes writing from learners representing 16 different mother tongue backgrounds and a total of 3.7 million words (Granger et al. 2009). Like ICE, this corpus has led to many investigations of the typical features of Learner Englishes (see, e.g., Granger 1998).

Corpora like ICE (for the outer circle) and ICLE (for the expanding circle) have greatly contributed to a better understanding of the nature of non-native English. Interestingly, their use is not limited to the study of World Englishes or Learner Englishes as separate and independent varieties of English. In fact, combining these corpora makes it possible, probably for the first time ever, to systematically compare features of World Englishes and of Learner Englishes, as demonstrated very recently by Nesselhauf (2009). The next section presents the results of a case study that uses the tools and methods of corpus linguistics to compare ESL and EFL speakers' use of phrasal verbs with *up*, thus bridging the gap between outer circle and expanding circle varieties.

4. Case study: phrasal verbs with *up*

This case study deals with phrasal verbs, which may be described as “one of the most notoriously challenging aspects of English language instruction” (Gardner & Davies 2007: 339) – due to their sheer quantity (Cornell 1985), their frequent opaqueness (Lindner 1982), their polysemy (Sjöholm 1995), and the syntactic problems they involve, most notably in terms of particle placement (Gries 2001). More particularly, we will focus on phrasal verbs with *up*, the most frequent particle in the English language according to Biber et al. (1999) and Kennedy (2003).

Table 1. List of corpora used

Type	Medium	Corpus	Size
Native English (NE)	Writing	BNC-baby_acad	1,049,331
	Speech	BNC-baby_spoken	947,799
	Writing (student)	LOCNESS	325,764
	Speech (student)	LOCNEC	118,555
Learner Englishes (LE)	Writing	ICLE	3,586,883
	Speech	LINDSEI	784,021
World Englishes (WE)	Writing	ICE-EA-KEN (acad)	82,003
		ICE-EA-TAN (acad)	83,060
		ICE-IND (acad)	93,537
		ICE-SIN (acad)	97,148
	Speech	ICE-EA-KEN (conv)	55,622
		ICE-EA-TAN (conv)	6,582
		ICE-IND (conv)	195,034
		ICE-SIN (conv)	194,251
TOTAL			7.619.581

Besides data from the *International Corpus of English* (Kenyan, Tanzanian, Indian and Singaporean components) and from the *International Corpus of Learner English* (see above on these two corpora), use was made of a spoken learner corpus, LINDSEI (*Louvain International Database of Spoken English Interlanguage*; cf. Gilquin et al. forthcoming), as well as several native corpora serving as a baseline: sections from the *British National Corpus* (BNC), the *Louvain Corpus of Native English Essays* (LOCNESS) and the *Louvain Corpus of Native English*

Conversation (LOCNEC). This wide range of corpora made it possible to compare the use of phrasal verbs with *up* in native English, World Englishes and Learner Englishes, but also in speech (spontaneous conversation) and writing (academic essays). The different corpora and their sizes are summarised in Table 1.

Figure 2 shows the frequency of phrasal verbs with *up* in native English and learner Englishes and reveals that while foreign learners tend to overuse phrasal verbs in writing (compare ICLE with BNC-acad and LOCNESS), they heavily underuse them in speech (compare LINDSEI with LOCNEC and BNC-spoken), which seems to point to a phenomenon of register confusion. A more detailed analysis of the components of ICLE also brings to light the great diversity that exists among learners from different mother tongue backgrounds (see Figure 3) – a diversity which may be explained, among others, by the influence of the mother tongue (speakers of a Germanic language, who have phrasal verbs in their mother tongue, use more phrasal verbs than speakers of a Romance language, who do not; see also Waibel 2007) and the degree of exposure to the target language (Tswana learners, who have a higher degree of exposure to English by being taught through the medium of English from the fifth grade in primary school, display the highest frequency of phrasal verbs).

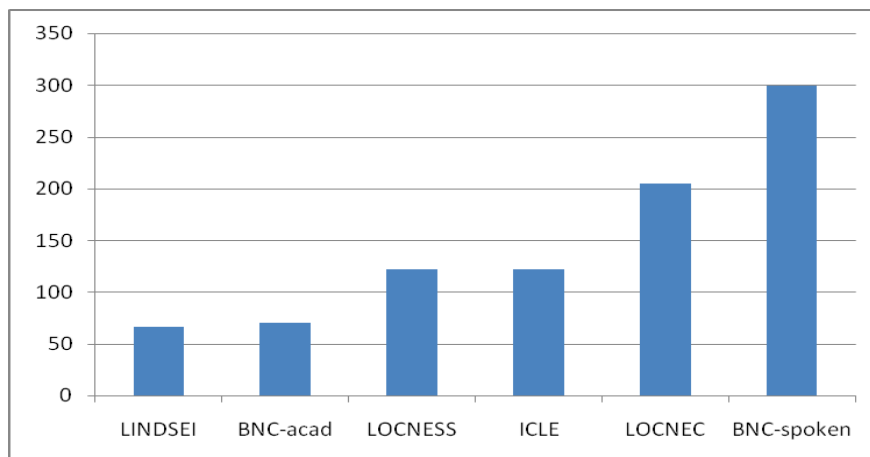


Figure 2. Frequency of phrasal verbs with *up* in NE and LE (relative frequency per 100,000 words)

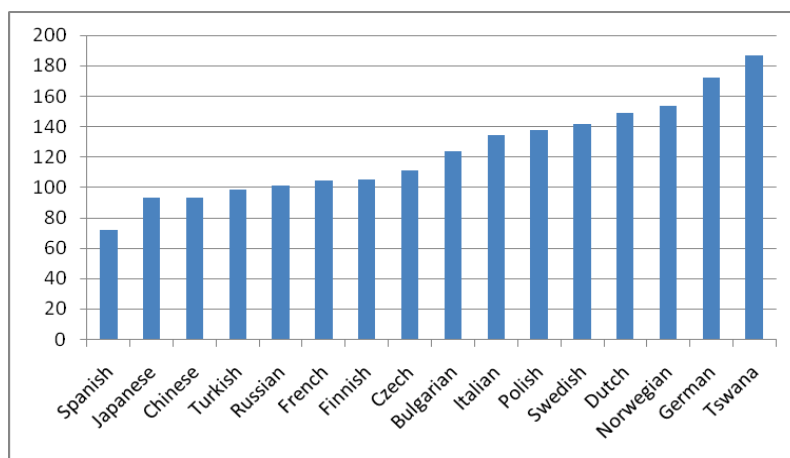


Figure 3. Frequency of phrasal verbs with *up* in ICLE (relative frequency per 100,000 words)

Turning to the use of phrasal verbs in World Englishes (see Figure 4), it appears that speakers of an indigenised variety of English underuse phrasal verbs with *up* in speech, as is the case in Learner Englishes. In writing, on the other hand, there is no notable difference between native English and World Englishes, which seems to suggest that speakers of indigenised varieties of English are perhaps slightly more sensitive to register variation than foreign learners. With the exception of the Tanzanian spoken corpus (which is very small and contains only two phrasal verbs with *up*), all the written corpora contain comparatively fewer phrasal verbs than their spoken counterparts.

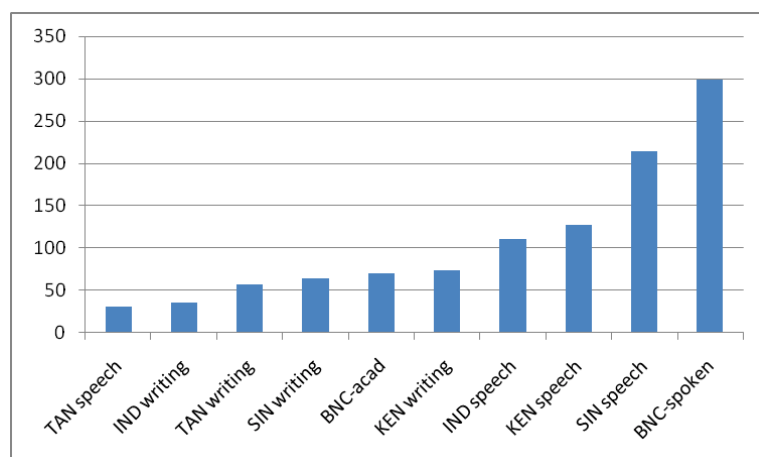


Figure 4. Frequency of phrasal verbs with *up* in NE and WE (relative frequency per 100,000 words)

From a more qualitative point of view, it turns out that World Englishes and Learner Englishes also share a number of features, although they present some differences too. Both types of varieties include forms of phrasal verbs that are not standard (i.e. would not occur in native English). Some of these forms are limited to Learner Englishes. Thus, while ICLE contains some examples of misplaced particles, as in (1), the data from ICE do not include any such examples. On the other hand, both ICE and ICLE display cases of inappropriate phrasal verbs (compare [2] and [3]) and instances of redundant particles (compare [4] and [5]).

- (1) *The government is planning to put up these on the island of Öland.* <ICLE-SW>
- (2) *There is no other way we can do we just walk up to be parents* <ICE-KEN S1A-017>
- (3) *Nowadays the impulse to build up also springs up from the need as it did when out of necessity to move quickly men made cars* <ICLE-BU>
- (4) *Fifty percent of the class girls will be married by the time MA completes up* <ICE-IND S1A-062>
- (5) *The population cannot increase up naturally.* <ICLE-CH>

Particularly interesting too are phrasal verbs that could be described as innovations, because they are not listed in common dictionaries of phrasal verbs. Some of these are simply incomprehensible, e.g. (6), and could be rejected as errors. Others, however, are perfectly understandable and testify to the creative skills of the speaker. Sentences (7) to (10) are just some of the examples, taken from ICE and ICLE, which are not included in any of the phrasal verbs dictionaries that were consulted, but still make perfect sense in context. Surprisingly, some innovations are even shared by World Englishes and Learner Englishes, as for example the phrasal verb *to cope up (with)*, which is identified by Platt (1989) as a typical feature of Singapore English, but actually occurs in other indigenised varieties of English (11) as well as in Learner Englishes (12).

- (6) *The combination of this suppositions make this contrast flaps up strongly and clearly.* <ICLE-SP>
- (7) *And you press it down so that when the pie bakes it wouldn't pluff up and break* <ICE-SIN S1A-039>
- (8) *I've been trying my best to keep it clean then he comes and he meddle up all over again* <ICE-SIN S1A-054>
- (9) *When the darkness curves the town, the lights blind us up by different colours* <ICLE-RU>
- (10) *... because of the instructions to fashion your jeans up by stone-washing and colouring them* <ICLE-GE>
- (11) *If they miss that bus it's very difficult for them to cope up and adjust* <ICE-IND S1A-069>
- (12) *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>

Such shared innovations, incidentally, raise at least two important questions: (i) how can such similarities be explained, considering that contact between all these varieties is extremely unlikely to have taken place, and (ii) how should these similarities be treated, given that the expanding circle is normally seen as norm-dependent, whereas the outer circle is recognised as norm-developing (cf. Berns 1995). Answering fundamental questions such as these is the next step of this research, and could actually lead to the kind of theoretical rapprochement that Sridhar & Sridhar (1986) probably had in mind.

5. Conclusion

Using the case of phrasal verbs with *up* as an illustration, this paper has demonstrated that, despite the geographical distance between them and the different contexts of acquisition, varieties from the outer circle and from the expanding circle may share certain features, which suggests that the theoretical rapprochement advocated by Sridhar & Sridhar (1986) over twenty years ago may not be such an impossible dream after all.

DRAFT

Further research, however, is much needed, both to investigate other aspects of language and to examine the theoretical and pedagogical implications of any similarities that may be uncovered between World Englishes and Learner Englishes.

References

- Berns, Margie (1995) "English in the European Union". *English Today* 11(3): 3-11.
- Bhatt, Rakesh M. (2001) "World Englishes". *Annual Review of Anthropology* 30: 527-550.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan (1999) *Longman Grammar of Spoken and Written English*. Harlow, Essex: Pearson Education Limited.
- Cornell, Alan (1985) "Realistic goals in teaching and learning phrasal verbs". *International Review of Applied Linguistics in Language Teaching (IRAL)* 23(4): 269-280.
- Francis, W. Nelson and Henry Kučera (1964) *Manual of Information to Accompany 'A Standard Sample of Present-Day Edited American English, for use with Digital Computers'* (revised 1979). Providence, Rhode Island: Department of Linguistics, Brown University.
- Gardner, Dee & Mark Davies (2007) "Pointing out frequent phrasal verbs: a corpus-based analysis". *TESOL Quarterly* 41(2): 339-359.
- Gilquin, Gaëtanelle and Sylviane Granger (in press) "From EFL to ESL: Evidence from the *International Corpus of Learner English*". In Marianne Hundt & Joybrato Mukherjee (eds) *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Gilquin, Gaëtanelle, Sylvie De Cock and Sylviane Granger (forthcoming) *The Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM*. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Granger, Sylviane (1996) "Learner English around the world". In Sidney Greenbaum (ed.) *Comparing English Worldwide: The International Corpus of English* (pp. 13-24). Oxford: Clarendon Press.
- Granger, Sylviane (ed.) (1998) *Learner English on Computer*. London & New York: Addison Wesley Longman.
- Granger, Sylviane, Estelle Dagneaux and Fanny Meunier (eds) (2002) *The International Corpus of Learner English. Handbook and CD-ROM*. Louvain-la-Neuve: Presses universitaires de Louvain.
- Granger, Sylviane, Estelle Dagneaux, Fanny Meunier and Magali Paquot (eds) (2009) *The International Corpus of Learner English. Handbook and CD-ROM. Version 2*. Louvain-la-Neuve: Presses universitaires de Louvain.
- Greenbaum, Sidney (ed.) (1996) *Comparing English Worldwide: The International Corpus of English*. Oxford: Clarendon Press.
- Gries, Stefan Th. (2001) "Particle placement in English: A multifactorial investigation". In Ruth M. Brend, Alan K. Melby and Arle R. Lommel (eds) *LACUS Forum XXVII: Speaking and Comprehending* (pp. 19-32). Fullerton, CA: LACUS.
- Johansson, Stig, Geoffrey Leech and Helen Goodluck (1978) *Manual of Information to Accompany the Lancaster-Oslo-Bergen Corpus of British English, for Use with Digital Computers*. Oslo: Department of English, Oslo University.
- 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* (pp. 11-30). Cambridge: Cambridge University Press.
- Kennedy, Graeme (2003) *Structure and Meaning in English. A Guide for Teachers*. Harlow: Pearson Education Limited.
- Lindner, Susan (1982) "What goes up doesn't necessarily come down: The ins and outs of opposites". In Kevin Tuite, Robinson Schneider and Robert Chametzky (eds) *Paper from the Eighteenth Regional Meeting Chicago Linguistic Society, April 15-16 1982* (pp. 305-323). Chicago: Chicago Linguistics Society.
- McEnery, Tony, Richard Xiao and Yukio Tono (2006) *Corpus-Based Language Studies. An Advanced Resource Book*. Oxon & New York: Routledge.
- Nelson, Gerald (1996) "The design of the corpus". In Sidney Greenbaum (ed.) *Comparing English Worldwide: The International Corpus of English* (pp. 27-35). Oxford: Clarendon Press.
- Nesselhauf, Nadja (2009) "Co-selection phenomena across New Englishes: Parallels (and differences) to foreign learner varieties". *English World-Wide* 30(1): 1-25.
- Platt, John (1989) "The nature of indigenized Englishes: Interference — Creativity — Universals". *Language Sciences* 11(4): 395-407.
- Sjöholm, Kaj (1995) *The Influence of Crosslinguistic, Semantic, and Input Factors on the Acquisition of English Phrasal Verbs: A Comparison between Finnish and Swedish Learners at an Intermediate and Advanced Level*. Åbo: Åbo Akademi University Press.
- Sridhar, Kamal K. and S.N. Sridhar (1986) "Bridging the paradigm gap: Second language acquisition theory and indigenized varieties of English". *World Englishes* 5(1): 3-14.
- Waibel, Birgit (2007) *Phrasal verbs in learner English: A corpus-based study of German and Italian learners*. PhD dissertation. Albert-Ludwigs-Universität Freiburg.