One norm to rule them all? Corpus-derived norms in learner corpus research and foreign language teaching

Gaëtanelle Gilquin

Université catholique de Louvain

gaetanelle.gilquin@uclouvain.be

Abstract

This paper considers the issue of the norm in the context of learner corpus research and its implications for foreign language teaching. It seeks to answer three main questions: Does learner corpus research require a native norm? What corpus-derived norms are available and how do we choose? What do we do with these norms in the classroom? The first two questions are more research-oriented, reviewing the types of reference corpora that can be used in the analysis of learner corpora, whereas the third one looks into the pedagogical use of corpus-derived norms. It is shown that, while studies in learner corpus research can dispense with a native norm, they usually rely on one, and that a wide range of native and non-native norms are available, from which choosing the most appropriate one(s) is of crucial importance. This large repertoire of corpus-derived norms is then reconsidered in view of the reality of the foreign language classroom.

1. Introduction

Let me start this paper with two quotes.[1] The first one, by Palmer (1922: vi, ix; emphasis original), states that

> [the foreign student of English] will soon learn that his own native phonetic system is of no more use in England than his own native monetary system; in both cases he must use the currency of the country. Sooner or later another fact will also become impressed on his mind; namely, that he must use the same sort of English sentences as those which are used in England. (…) authentic models the student *must* have.

The second one, by Jenkins (2000: 160), claims that

> There is really no justification for doggedly persisting in referring to an item as 'an error' if the vast majority of the world's L2 English speakers produce and understand it. Instead, it is for L1 speakers to move their own receptive goal posts and adjust their own expectations as far as international (…) uses of English are concerned.

These quotes illustrate the two ends of a continuum of normative attitudes towards foreign and second language (L2) learners, ranging from a very strict attitude in which only the native (L1) speaker is deemed a worthy model for learners to a very lenient attitude in which learners are free to use the target language as they wish (provided they can understand each other) and native speakers should adapt to learners' standards (or lack thereof).

The place of the native speaker in foreign language learning and teaching has been the topic of much discussion and debate. In this paper, I will tackle the issue from the perspective

of learner corpus research (LCR) and foreign language teaching (FLT), and will broaden the discussion beyond the native speaker to include different types of norm. The norm has been defined and categorized in different ways, but I will use the term in its meaning of point of reference, as a synonym of baseline, benchmark or indeed yardstick, which evokes the etymological origin of the word (from Latin *norma*, which refers to a carpenter's square). More precisely, I will follow Bamgbose's (1987: 105) definition of the norm as 'a standard language form or practice that serves as a reference point for other language forms or practices'. In the context of LCR, which seeks to investigate L2 varieties on the basis of learner corpora (see Gilquin 2020), the norm serves as a reference point for learner language and is derived from a reference corpus which is part of the LCR analysis.

The overarching question that I will be addressing in this paper is whether, to paraphrase the inscription found on the One Ring in J.R.R. Tolkien's *The Lord of the Rings*, we should provide for learners 'One Norm to rule them all, One Norm to find them, One Norm to bring them all and in the darkness bind them'. This question will be broken down into three sub-questions: (i) Does LCR require a native norm? (ii) What corpus-derived norms are available and how do we choose? (iii) What do we do with these norms in the classroom? The first two questions are more research-oriented, while the third one looks into how corpus-derived norms can make their way into teaching practices. My main focus will be on the English language and the context of English as a foreign language (EFL), although many of the points made about English are valid for other languages as well.

2. The status of the native norm at the emergence of LCR

The end of the 1980s and the beginning of the 1990s witnessed a 'paradigm shift' (Kachru 1991: 5), epitomized by the Quirk-Kachru controversy, which Kachru (1991) described as an opposition between 'deficit linguistics' and 'liberation linguistics'. Quirk saw standard native English as the only acceptable variety. As a proponent of 'deficit linguistics', he considered that '[i]t is neither liberal nor liberating to permit learners to settle for lower standards than the best' (Quirk 1990: 9). Kachru, on the other hand, as a representative of 'liberation linguistics', claimed that the English language is not owned by native speakers and that 'English is now the language of those who use it' (Kachru 1985: 20). In particular, he promoted institutionalized second-language varieties, those non-native varieties found in former British or American colonies where English has a (semi-)official status, and underlined their norm-developing nature. Around this period, several publications appeared that questioned the place of the native speaker: *The Native Speaker is Dead!* (Paikeday 1985a); *May I Kill the Native Speaker?* (Paikeday 1985b); *Displacing the 'Native Speaker'* (Rampton 1990); *ELT: The Native Speaker's Burden?* (Phillipson 1992); *Going Beyond the Native Speaker in Language Teaching* (Cook 1999).

It is in the context of this 'paradigm shift' that LCR emerged in the late 1980s. With its focus on non-native varieties, LCR was well anchored in the new paradigmatic context, thus departing from the general field of corpus linguistics at the time, which 'had shown great potential in investigating a wide range of native language varieties (…) but had neglected the non-native varieties' (Granger et al. 2015a: 1). In fact, the International Corpus of Learner English (ICLE), one of the first learner corpora, initially started as part of the International Corpus of English (ICE), a project that is essentially known for its contribution to the study of institutionalized second-language varieties and that is hence a prime example of 'liberation

linguistics'. Unlike the norm-developing institutionalized second-language varieties, however, the foreign-language varieties investigated in LCR are said to be norm-dependent (Kachru 1985), which necessarily brings the native norm more to the fore.

In addition to this special context, LCR, because of its focus of interest, also developed links with the field of second language acquisition (SLA) research, which at the time was influenced by two opposing forces. The first one was the (dominant) tradition of analyzing learner language relative to a native norm, as illustrated by the application of notions such as 'error analysis', which seeks to identify non-native features in learner language, or 'fossilization', which characterizes a stage in language development that fails to reach a native level. According to this tradition, summarized by Cook (1997: 38) in a literature review, '[t]he L2 learner's goal is (…) "full native speaker competence", paraphrased as "perfect command" or "language mastery". L2 learners are failures because they do not attain the same competence as native speakers'. The second force in SLA was the idea of comparative fallacy (Bley-Vroman 1983), which advocates the study of learner language as a system with its 'own internal logic' (ibid. 14). Comparisons of learner language with native language are rejected on the grounds that 'the learner's system is worthy of study in its own right, not just as a degenerate form of the target system' (ibid. 4). Interestingly, these two approaches are reflected in one of the earliest and most popular methodologies in LCR, namely contrastive interlanguage analysis (CIA; Granger 1996). This methodology consists of two types of comparison: (i) a comparison between native and non-native language, which corresponds to the traditional approach of judging learner language against a native norm, and (ii) a comparison between different types of non-native language, which does not involve any native reference, following the precept of comparative fallacy.

3. Does LCR require a native norm?

Despite a context that could have favoured both a 'liberation' approach and a 'deficit'
approach, it must be recognized that the native norm has always occupied a major role in
LCR. Flowerdew (2015: 469) points out that '[a] key facet of learner corpus research is that
the learner corpus is usually compared with a native-speaker control corpus'. Typically, the
results obtained through the analysis of learner corpus data are checked against the results
from native corpus data representing the target language, which makes it possible to identify
traces of non-nativeness, including errors, but also cases of under- or overuse, that is, the
presence of fewer or more instances of a certain feature in learner language than in native
language. To assess the role of the native norm in LCR, I have carried out a literature review
on the basis of the 57 publications that are described as representative studies in the chapters
of the *Cambridge Handbook of Learner Corpus Research* (Granger et al. 2015b). Of these
studies, 32 (56%) involve a comparison of the learner data with a native norm, while the
others do not involve any such comparison. However, of these remaining studies, 15 (26%)
still refer to and somehow rely on a native norm, for example by identifying and sometimes
annotating errors, or by using a native corpus as a resource from which word frequencies are
extracted or collocations are checked. All in all, there are only ten studies out of 57 (18%)
that do not include any kind of native reference.

The use of a native norm in learner corpus studies, however, does not mean that
learner language must be seen as 'deficient' if it differs from native language, since the norm
need not be prescriptive. Hunston (2002: 212) observes that 'information about differences
between learner and native-speaker usage is not necessarily taken as a criticism of the
learners involved'. The terms 'underuse' and 'overuse', for example, are merely descriptive

concepts referring to a statistically significant difference in frequency between native and non-native language.

In addition, the norm used in LCR has the advantage of being abstracted from corpus data and hence 'based on the performance of many individuals in various communication situations' (Mukherjee 2005: 14). As explained by Klippel & Mukherjee (2007), this results in a norm that is empirical and frequency-based, and that therefore represents an objective yardstick to measure learners' progression towards a certain target. The norm in LCR also tends to be explicit, being clearly identified as such in LCR studies. This is to be contrasted with some SLA studies that allegedly avoid the trap of comparative fallacy but actually rely on some hidden native norm, which leads Granger (2009: 18) to talk about 'comparative hypocrisy' in SLA.

The objectivity and explicitness of the native norm in LCR does not necessarily imply that such a norm is required or, indeed, desirable. The so-called 'English Native Norm Irrelevance Hypothesis' (House 2008: 355) put forward by some advocates of English as a lingua franca (ELF) suggests that native models might not be relevant (see also Jenkins' (2000) quote in Section 1). Amorim (2013: 111), writing from an ELF perspective, claims for example that '[w]hen the aim is practising communication, the native speaker "yardstick" (…) seems inappropriate', adding that 'EFL learners do not aim at becoming mother tongue imitators'. This may be true of certain populations of non-native speakers, and in particular L2 *users* rather than *learners* (to use Mauranen's (2011) distinction), but for many learners, the preferred norm is still native. Every year at the beginning of a course I teach on English varieties, I ask my students what norm they would like to approximate to. They always predominantly choose a native variety. Between 2008 and 2020, 55% (out of a total of 406 students) chose British English and almost 20% chose American English, plus some 2% for Canadian English and 2% for Australian English. About 12% chose 'English as it is used by

competent non-native speakers in international contexts'[2] as their preferred target. This tendency was even more striking when the students were asked what English norm they would teach. Some 67% chose British English and 24% chose American English, leaving less than 5% for competent non-native use. This strong preference for one of the two main native varieties is probably partly due to the fact that most of these students are English majors, but since many of them are to become English teachers, their preferences are important as they might influence the variety of English that future students will learn.

While we have seen that a (corpus-based and explicit) native norm is the rule in LCR and that it makes sense in view of (certain) learners' needs and wishes, it must be underlined that, as already suggested, LCR studies can and do dispense with a native norm. In the CIA model described earlier, the comparison between learner and native corpus data is only one of two components. The other one is a comparison of different types of learner corpus data, such as data produced by learners with different L1s (the most frequent kind of comparison for this component), data produced by learners with low vs high exposure to the target language, data produced by learners at different stages of acquisition, data produced in timed vs untimed conditions, etc. This makes it possible to study learner language independently of any native norm. Vyatkina (2013a), for example, adopts a longitudinal design to compare the written production of learners over four semesters and 14 measurement occasions. This allows her to provide a fine-grained analysis of the development of syntactic complexity among the learners, without comparing them against any native baseline; the point of reference is that of the previous measurement occasion(s). LCR can thus do without a native norm and investigate learner language in its own right, as advocated by certain SLA researchers.

4. What corpus-derived norms are available and how do we choose?

One of the lessons of corpus linguistics is that a language, especially an international language like English, is far from being a monolithic entity. Even if we limit ourselves to native English, there is a range of varieties, and hence a range of norms, to choose from. These norms vary according to factors such as geography, social class, age, medium or formality. I will deal with three factors here: regional variety, level of literacy and text type. I will then review some potential non-native norms and will propose general criteria to choose the most appropriate norm(s).

4.1. Regional variety

The same (native) language used in different places often displays regional variation. This can make things difficult when it comes to the choice of a norm (see Leech 1998: xix). Besides the many national varieties of English (British, American, Australian, etc.), dialectal varieties could also be viewed as potential norms. Importantly, choosing one regional variety or another as a norm in an LCR study may have consequences on the findings. In Chen's (2013) study of phrasal verbs in British, American and Chinese learner English writing, for example, Chinese learners significantly underuse phrasal verbs in comparison to an American norm, but in comparison to a British norm there is no significant difference in frequency.

While there is no perfect rule to choose between different regional varieties, the best advice is probably to opt for the variety that the learners represented in the corpus are most likely to have been exposed to. Using a reference corpus of British English for the analysis of corpus data produced by learners mainly exposed to American English, for example, would not make much sense (see Chen 2013: 435). In this respect, it would be desirable that more learner corpora should include in their metadata information about the language variety (or

varieties) that learners have been exposed to, as is the case for instance in the Process Corpus of English in Education (PROCEED; Gilquin 2022).

4.2. Level of literacy

Another issue to bear in mind when choosing a native reference corpus is the level of literacy, which covers the difference between expert and novice usage, that is, whether the writers or speakers have experience in producing the kind of text included in the corpus. Native students writing an academic text, for example, have the advantage of their native knowledge of the language, but they may be novice in the art of writing an academic text, which involves using a certain style and conforming to certain conventions. The level of literacy may have an impact on the findings. Thus, a comparison between a corpus of learner writing representing the production of non-native university students and a corpus of expert writing representing the production of professional writers (more precisely, academics) shows that learners overuse phrasal verbs with *up*, but a comparison of the same learner corpus with a corpus of novice writing representing the production of native university students reveals no significant difference (Gilquin 2011).

If choosing between an expert and a novice norm may have such an impact, it is important to consider the arguments for and against each of the two types of norms. Native novice data are more comparable to learner data, because in both cases the subjects are supposed to be more or less the same age and hence at approximately the same stage of general cognitive development. However, as noted by Leech (1998: xix), '[n]ative-speaking students do not necessarily provide models that everyone would want to imitate'. While expert writing undoubtedly represents a better model for learners in this respect, some researchers have objected to its use in LCR, underlining the 'unrealistic standard of "expert writer" models' (Hyland & Milton 1997: 184) and claiming that it is 'both unfair and

descriptively inadequate' to compare learner writing with professional native writing (Lorenz 1999: 14). Since both types of reference corpora present advantages and disadvantages, the choice between a novice and an expert norm should essentially depend on the goal of the comparison (see Ädel 2006: 205-208). If the goal is evaluative, a comparable corpus of novice writing should preferably be used, because it would be unfair to evaluate learner writing against the high standards of expert writing. If, on the other hand, the goal is pedagogical, expert writing is arguably a more ambitious target for learners to aim for. I will deal more extensively with the norm in pedagogical applications when turning to the third question.

4.3. Text type

When comparing a learner corpus and a reference corpus, it is crucial that the two corpora should be as comparable as possible in terms of text type. Studies that do not meet this criterion run the risk of neglecting an important factor and possibly invalidating the findings. Gries & Wulff (2013), for example, investigate the use of the genitive alternation in ICLE, consisting of argumentative essays, and in the whole British component of ICE, consisting of various written but also spoken genres. The two corpora are compared with each other, despite the 'highly register-, genre-, and dialect-dependent nature of the distribution of the two genitive forms' (ibid. 331).

The possible impact of text type can be illustrated by Granger & Tyson's (1996) study on connectors. Their comparison of ICLE data with native data taken from the various written genres included in the Lancaster-Oslo-Bergen corpus reveals no significant difference between the two corpora in the frequency of the three connectors *however*, *therefore* and *thus*. However, a comparison of ICLE with the Louvain Corpus of Native English Essays

(LOCNESS), which represents the same text type (argumentative essays), shows that learners actually underuse these three connectors.

It should be underlined that comparability of text type can also be considered at a more micro level, by taking into account the metadata that describe the different texts included in a corpus. Lin & Lin (2019), for example, not only restrict their analysis of the verb *make* to the written learner and native subcorpora of the International Corpus Network of Asian Learners of English (ICNALE), but they also select the texts to be included in their study by controlling for their length, their topic and the conditions in which they were written (timing and access to reference materials). Interestingly, Gries & Deshors (2014) go even further in the degree of comparability. Through the so-called MuPDAR statistical approach, they argue for extending comparability to the linguistic context itself, by using native corpus data to predict what native speakers would have done in a certain linguistic context and then comparing this prediction to learners' choices in the learner corpus.

Before turning to non-native norms, I would like to mention another possible native norm for LCR, namely the pedagogic corpus, which is made up of texts that the learner has encountered – or is likely to have encountered – in an educational setting: textbook materials, graded readers, transcripts of spoken texts that learners have been exposed to, etc. Pedagogic corpora were originally created for teaching purposes, but they have sometimes been used in research, for example by Vyatkina (2013b), who examines part-of-speech variety in a corpus of learner writing and a corpus of pedagogical materials. The main advantage of such reference corpora is that they provide a very good indication of what we may expect from learners, since they correspond to the kind of input that learners (especially in a foreign language situation) have received in the target language. Yet, pedagogic corpora do not necessarily represent the target that learners want to reach, as some of them will arguably want to go beyond what they have been taught in class.

4.4. Non-native norms

Although native corpora represent the most obvious choice when looking for a norm against which to compare learner corpora, other norms are available. One that has been regularly used in LCR is the expert norm, already mentioned earlier in relation to native data. In some cases, expertise may outweigh nativeness, as Swales (2004: 56) suggests with reference to academic genres, arguing that the main distinction is 'between those who know the academic ropes in their chosen specialisms and those who are learning them'. Reinhardt (2010), for example, uses an expert norm in his investigation of spoken directive language use in the context of office hour consultations. He compares data from a learner corpus with data produced by 'practicing academics with native or near-native English proficiency' (ibid. 96). Although expert corpora do not rely (exclusively) on native speaker models, they can be seen as suitable references in LCR for at least two reasons. First, experts are assumed to have acquired a knowledge of the language and of the genre conventions that is close to that of native speakers. Second, for genres such as academic writing, it is likely that most of the texts included in expert corpora have been checked at some point by a native speaker, for example an editor or copyeditor, which enhances their validity as models for learners.

Among non-native norms, institutionalized second-language varieties, also called 'New Englishes', could be considered too. Largely thanks to Kachru's intervention (see Section 2), these varieties are now viewed as 'systems unto themselves as opposed to deviant forms of traditional native varieties' (Balasubramanian 2015: 147). New Englishes come with their own, localized norms, which could arguably serve as a baseline in LCR. There have been attempts, recently, to compare New Englishes and learner English (e.g. Mukherjee & Hundt 2011; Gilquin 2015), but these comparisons have been drawn in a descriptive rather than normative perspective: the New Englishes data are seen on a par with the learner data rather than as a reference corpus. A normative approach could however make sense in certain

contexts. New Englishes norms might thus be relevant to populations of learners that are in contact, or are likely to be in contact, with speakers of New Englishes. The Tswana EFL learners whose production is included in ICLE, for example, are probably more likely to interact with non-native speakers of South African English than with native speakers of British or American English. To illustrate the potential of such an approach, consider Meriläinen & Paulasto's (2017) finding that embedded inversion is proportionally more frequent in the Tswana component of ICLE than in other ICLE components and in LOCNESS. This finding could possibly be related to the fact that embedded inversion is a feature of some African Englishes (Schmied 2008). Using such a norm might therefore lead to another interpretation of the results.

Finally, despite some references to the 'new ELF norm' (House 2008: 364) and despite a slow rapprochement between ELF research and LCR (Mauranen 2011), the ELF construct has not really made its way into LCR as a possible baseline. Some comparisons have been drawn between learner English and ELF (cf. Paulasto et al. 2011), but again, not in a normative perspective. As with New Englishes, such a norm might be relevant in certain contexts, such as that of a corpus of business students learning English, not with the aim of achieving native-like competence, but with the aim of communicating with other non-native speakers of English for professional purposes.

4.5. Choosing the most appropriate norm(s)

With so many norms to choose from, it is essential to think carefully about which one to use in an LCR study. While it is probably fair to say that there is no perfect reference corpus, I have already suggested two important criteria to take into account. First, the reference corpus should be as comparable as possible to the learner corpus. Comparing a written learner corpus with a spoken reference corpus, for example, or using as a baseline data produced

under different conditions (e.g. with regard to time constraints) could skew the results and invalidate the conclusions. In this respect, it is an interesting development that more and more learner corpora are created with a native equivalent, built according to the same design criteria and collected among similar populations, like the Louvain Corpus of Native English Conversation (LOCNEC), which is an exact replica of the Louvain International Database of Spoken English Interlanguage (LINDSEI). Such native counterparts can safely be used as legitimate reference corpora. The second criterion when choosing a reference corpus is that it should fit the research purposes. If the aim is evaluative, for instance, the reference corpus should reflect the type of language that learners are thought to have been exposed to and that can reasonably be expected from them (the pedagogic corpus is a very good norm in this respect). If, on the other hand, the aim is pedagogical, the reference corpus should reflect the type of language that learners need and/or want to approximate to (which is often native language, as we saw earlier).

While these criteria should guide the choice of a reference corpus, it must be recognized that sometimes researchers resort to what I would call a 'convenience norm', that is, a norm that may not be the most suitable one to compare the learner data to, but which is convenient because it is represented in a corpus that is easily accessible. I suspect that LOCNESS, which is predominantly made up of writing by American university students, has often been used as a convenience norm in comparisons with ICLE because the corpus is readily available. Given the choice, however, several users might have preferred a reference corpus mainly or even exclusively made up of writing by British university students, considering the still very important role of British English in the EFL classroom, especially in European countries (see Gilquin 2018).

So far, I have only considered the use of a single norm. However, it may be interesting to use multiple norms in LCR. The revised version of the CIA model, CIA[2]

(Granger 2015), recognizes this possibility by using the term 'reference language varieties'. This not only opens the door to other norms besides native language, but also indicates, through the plural form, that there could be several norms, possibly represented by different reference corpora. That it is now acceptable to rely on manifold norms can be seen as a sign of liberation linguistics and its recognition of the 'pluricentricity and multi-identities of English' (Kachru 1991: 4). Combining several reference corpora could actually help represent the range of norms that may be relevant to a certain learner population. A baseline mixing British and American English corpus data, for instance, might be said to reflect the type of language that many European learners are exposed to, with British English being taught at school and American English being heard outside school (e.g. in American TV series).

The use of several norms is also very much in line with the current popularity of converging evidence (see, e.g., Schönefeld 2011; Gilquin & McMichael 2018). Triangulating evidence from different corpora makes it possible to strengthen the conclusions of the analysis if the evidence converges. Lee & Chen (2009), for example, compare their written learner data with two reference corpora of native writing, underlining that this gives them 'added confidence' (ibid. 154) that they are focusing on the right items. Such triangulation is particularly important given that (i) results may vary between (similar) corpora and even within a single corpus (see Gablasova et al. 2017) and (ii) a corpus, even a very large one, may not be sufficiently representative of the target variety.

Finally, the use of additional reference corpora may be useful for explanatory purposes. Thus, comparing a written learner corpus with both written and spoken native corpora may reveal that some overused items in learner writing correspond to inappropriate stylistic choices, being more typical of native speech than of native writing. Comparing a written learner corpus with both novice and expert native writing may show that certain

16

learner features are found in novice writing as well, which arguably points to aspects of language that need to be acquired at the same stage of development by both native and non-native writers.

5. What do we do with these norms in the classroom?

From the early days of LCR, its potential for language pedagogy has been emphasized. While concrete pedagogical applications of LCR so far have not been as numerous as we might have hoped, its pedagogical implications are widely recognized and have created strong links between LCR and the field of FLT (see, e.g., Meunier 2016). It therefore seems relevant, for the last question, to turn to the pedagogical implications of corpus-derived norms.

Let us first ponder whether the trend of liberation linguistics initiated by Kachru (see Section 2) should lead to a kind of 'liberation pedagogy'. Not following any kind of norm does not seem like an option in the foreign language classroom, since a norm is required to decide what kind of language should be taught and to assess students' progress towards the set target. However, we may wonder whether a linguistic feature in learner language can be considered acceptable as long as it respects *some* norm. Corpora give access to many linguistic norms, some of them characterized by features that are not normally found in standard native varieties, which would potentially make many learner features acceptable. Actually, corpus studies comparing learner English and New Englishes have brought to light non-standard features shared by the two types of varieties. In such cases, it may seem unfair to condemn learners who produce features accepted in certain varieties. For example, the use of *discuss about* has been described as an 'interesting phenomenon at the lexis-grammar interface' in Indian English (Mukherjee 2007: 175) but is seen as an error in pedagogical

materials for learners (e.g. Turton & Heaton 1996). The main reason why such a difference in treatment may seem unfair, however, is that we look at it from our own perspective as linguists. Chances are that it does not seem unfair to teachers and learners, most of whom may not even be aware that *discuss about* is an innovative feature in certain New Englishes. Besides, learners who produce this phrase most likely do not produce it because they know that it is acceptable in Indian English, but because they think that it is correct in standard native English, probably by analogy with the noun *discussion* or with verbs like *talk* or *speak* (see Gilquin 2017). This underlines the importance of making a distinction between research norms, those that linguists use, and so-called pedagogical norms, those that 'guide the selection and sequencing of target language features for language teaching and learning' (Bardovi-Harlig & Gass 2002: 1). While the pluricentric, multi-norm model advocated by Kachru (1991: 4) is appealing from a research point of view, it is 'simply not a practical model for classroom teaching purposes' (Ur 2010: 89). Teaching several norms in the classroom would involve a number of practical problems, including the fact that time constraints would not allow teachers to teach and learners to learn multiple norms thoroughly, that textbooks and other pedagogical materials almost exclusively conform to standard British or American English norms, and that the repertoire of norms that teachers master is necessarily limited. It is symptomatic, in this respect, that scholars who suggest using norms such as New Englishes or ELF in the EFL classroom usually do not offer much in the way of concrete proposals on how to do this (possible exceptions are Jenkins' (2000) Lingua Franca Core, which is not without controversy though – see Ugarte Olea (2019) – as well as the Global Englishes Language Teaching (GELT) framework and some other proposals summarized in Rose et al. (2020), which however seek to adopt a New Englishes or ELF *approach* rather than teaching a New Englishes or ELF *variety* per se).

In the same way as native norms are still relevant in LCR, they arguably still have a key role to play in the foreign language classroom, not only because many learners have native language as their target, as we saw earlier, but also because, in real life, learners are likely to be judged according to native standards. Think, for example, of standardized tests like TOEFL, which relies on a native norm. Even IELTS, which aims to treat 'all test takers with the utmost fairness and respect by actively avoiding cultural bias', accepts 'all *standard* varieties of *native-speaker* English, including North American, British, Australian and New Zealand English' (emphasis added).[3] The use of non-standard/non-native language features has also been shown to have negative consequences in job applications (e.g. McMasters 2004) or publishing (e.g. Flowerdew 2008). Teachers would therefore be doing learners a disservice by denying them access to native norms. Actually, a survey by Young & Walsh (2010) shows that the teachers who were polled almost exclusively taught American or British English and that none of them taught English as a lingua franca or as an international language.

This does not mean that we can simply take a reference corpus of native language and start teaching learners the usage it reflects. Prodromou (1998: 267) asks 'how much "native-speaker reality" can the non-native speaker and the EFL classroom take?'. As I see it, there are at least two aspects to consider. The first one is that, as Widdowson (2000) emphasizes, native corpora may lack authenticity for learners, who may not relate to texts produced in a different, foreign context (although the compilation of reference corpora with data produced by subjects whose profiles closely match those of the learners may reduce this risk). The second aspect is that corpus-derived findings may be too detailed for learners. The information contained in the corpus-based *Collins COBUILD Grammar Patterns* volumes (Francis et al. 1996, 1998), for example, is extremely valuable from a research point of view, but the degree of detail is such that much of it would probably be lost on learners.

Another key issue about corpus-derived information is that it should not be acted on blindly. For one thing, native corpora, even expert ones, may include non-standard or non-politically correct features that teachers would not want their students to reproduce. For another, as noted by Granger (2009: 22), learner features uncovered by LCR 'need not necessarily lead to targeted action in the classroom'. Deviations from the native norm, in the form of misuse, underuse or overuse, should be considered in relation to criteria such as learners' needs, teaching objectives or teachability, and on that basis should be selected or disregarded for classroom action (ibid.).

It was suggested earlier with respect to research that the norm should ideally be adapted to learners' needs. This is true of teaching too. Corder (1981: 49) already said so when he recommended 'adopting more realistic norms/standards given the particular sort of student we have, or promoting alternative norms/standards for some subgroup of the student population'. However, since mixed classrooms have become the rule rather than the exception, it is almost impossible for teachers to use tailor-made norms, since what one student wants or needs is not necessarily what the other students want or need. A one-size-fits-all norm may therefore be the only practical solution. In that case, choosing among the major native varieties (mainly standard British or American English) arguably makes it possible to opt for a norm that is likely to be most useful to a majority of the learners and/or most familiar to the teacher. This norm can then be imposed more or less strictly depending on the context. For example, it may be less strictly imposed on students learning English for Occupational Purposes (e.g. business English) than on students specializing in the study of English. It may also be less strictly imposed in oral communication, where communicative competence and mutual intelligibility may prevail, than in written usage, where linguistic competence and language accuracy may be more crucial.

I would like to stress that this choice of a single native norm does not imply that other norms should disappear from the classroom altogether. If possible at all, learners should be made aware of the existence of some of these other norms and they should be taught to respect these norms, and possibly understand them, for example through the use of audio and video materials in different language varieties. This, incidentally, points to yet another relevant distinction, namely between productive and receptive norms.

Finally, it is important to underline that the use of corpus-derived norms should not deny learners their right to creativity. Kreyer (2015: 161) asks whether we should view the norm as 'binding' or allow for some flexibility, letting learners 'experiment with the L2 and, as a result, come up with something that in part is different from the target norm'. That learners are capable of creative language use, just like native speakers, can be illustrated by the following example, taken from Gilquin (2015: 107): *... because of the instructions to fashion your jeans up by stone-washing...* (ICLE-GE). The phrasal verb *fashion up*, which one is unlikely to find in a native reference corpus (and which is not listed in any dictionary of the English language), is nevertheless perfectly intelligible. This simple example shows that 'linguistic creativity is not the sole preserve of (norm-providing) native speakers' (ibid.), an essential point to bear in mind when evaluating learners' production.

6. Conclusion

I started this paper with a paraphrase of the One Ring inscription: 'One Norm to rule them all, One Norm to find them, One Norm to bring them all and in the darkness bind them'. Let me now assess the truthfulness of this statement. What should be clear by now is that LCR need not restrict itself to *one single norm*, and not even a native one at that. There is a wide

range of corpus-derived norms to choose from, and although from a pedagogical perspective it might be desirable to reduce the repertoire to just one or two native norms, in LCR many different norms can be suitable – native and non-native norms, novice and expert norms, single and multiple norms, etc. – as long as we respect certain general principles (e.g. comparing like with like) and are aware of the possible consequences of our choice of norm(s) (as illustrated by the examples I gave of divergent findings resulting from the use of different reference corpora). It would also be wrong to think that we can *bring all learners together* under a common norm. Norms should ideally vary according to the specific learners that are under consideration (e.g. future specialists in the English language vs business students, learners in Europe vs learners in South Africa) and according to one's purposes (descriptive, evaluative, pedagogical, etc.). In addition, norms should not be *binding*. We have seen that LCR has the capacity to liberate itself from the shackles of the norm, either by not relying on any comparison with native language or by considering the norm as purely indicative of differences, rather than deficiencies. In the foreign language classroom, a norm seems to be required for teaching and assessment purposes, but it should leave room for creative uses. Finally, norms are not meant to bring learners *into the darkness*, but on the contrary to enlighten them as to how they could reach the target they have set for themselves and to make them more tolerant of different norms that other speakers conform to or aim for.

For learner corpus researchers and teachers, having access to such a variety of norms in the form of reference corpora may seem like a challenge, because it involves finding our way through a myriad of norms and dealing with potential divergence among them. At the same time, however, this is a chance to explore different avenues, test the impact of certain choices or propose customized solutions – all the while benefiting from an authentic, explicit, representative yardstick that can be shared with other researchers or adapted to students.

Corpus-derived norms can be a very precious resource, provided we handle them with caution and are careful not to fall into the fires of Mount Norm.

References

Ädel, A. 2006. *The Use of Metadiscourse in Argumentative Texts by Advanced Learners and Native Speakers of English*. Amsterdam: John Benjamins.

Amorim, R.M. 2013. Transforming passive listeners into active speakers: A study with Portuguese undergraduates in 'English for the Social Sciences'. *English Language Teaching* 6(4): 110-119.

Balasubramanian, C. 2015. Corpus linguistics and New Englishes. In V. Cortes & E. Csomay (eds) *Corpus-based Research in Applied Linguistics. Studies in Honor of Doug Biber*, pp. 147-175. Amsterdam: John Benjamins.

Bamgbose, A. 1987. Language norms. In W. Bahner, J. Schildt & D. Viehweger (eds) *Proceedings of the International Congress of Linguists*, pp. 105-113. Berlin: Akademie-Verlag.

Bardovi-Harlig, K. & S.M. Gass. 2002. Introduction. In S. M. Gass, K. Bardovi-Harlig, S. M. Pierce & J. Walz (eds) *Pedagogical Norms for Second and Foreign Language Learning and Teaching. Studies in Honour of Albert Valdman*, pp. 1-12. Amsterdam: John Benjamins.

Bley-Vroman, R. 1983. The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning* 33(1): 1-17.

Chen, M. 2013. Overuse or underuse: A corpus study of English phrasal verb use by Chinese, British and American university students. *International Journal of Corpus Linguistics* 18(3): 418-442.

Cook, V. 1997. Monolingual bias in second language acquisition research. *Revista Canaria de Estudios Ingleses* 34: 35-50.

Cook, V. 1999. Going beyond the native speaker in language teaching. *TESOL Quarterly* 33(2): 185-209.

Corder, S.P. 1981. *Error Analysis and Interlanguage*. Oxford: Oxford University Press.

Flowerdew, J. 2008. Scholarly writers who use English as an Additional Language: What can Goffman's "Stigma" tell us? *Journal of English for Academic Purposes* 7(2): 77-86.

Flowerdew, L. 2015. Learner corpora and language for academic and specific purposes. In S. Granger, G. Gilquin & F. Meunier (eds) *The Cambridge Handbook of Learner Corpus Research*, pp. 465-484. Cambridge: Cambridge University Press.

Francis, G., S. Hunston & E. Manning (eds). 1996. *Collins COBUILD Grammar Patterns 1: Verbs*. HarperCollins Publishers.

Francis, G., S. Hunston & E. Manning (eds). 1998. *Collins COBUILD Grammar Patterns 2: Nouns and Adjectives*. HarperCollins Publishers.

Gablasova, D., V. Brezina & T. McEnery. 2017. Exploring learner language through corpora: Comparing and interpreting corpus frequency information. *Language Learning* 67(S1): 130-154.

Gilquin, G. 2011. Corpus linguistics to bridge the gap between World Englishes and Learner Englishes. In L. Ruiz Miyares & M. R. Álvarez Silva (eds) *Comunicación Social en el siglo XXI, Vol. II*, pp. 638-642. Santiago de Cuba: Centro de Lingüística Aplicada.

Gilquin, G. 2015. At the interface of contact linguistics and second language acquisition research: New Englishes and Learner Englishes compared. *English World-Wide* 36(1): 91-124.

Gilquin, G. 2017. Applied cognitive linguistics and second/foreign language varieties: Towards an explanatory account. In J. Evers-Vermeul & E. Tribushinina (eds) *Usage-based Approaches to Language Acquisition and Language Teaching*, pp. 47-71. Berlin: de Gruyter.

Gilquin, G. 2018. American and/or British influence on L2 Englishes – Does context tip the scale(s)? In S.C. Deshors (ed.) *Modeling World Englishes: Assessing the Interplay of Emancipation and Globalization of ESL Varieties*, pp. 187-216. Amsterdam: John Benjamins.

Gilquin, G. 2020. Learner corpora. In M. Paquot & S. Gries (eds) *A Practical Handbook of Corpus Linguistics*, pp. 283-303. Berlin: Springer.

Gilquin, G. 2022. The Process Corpus of English in Education: Going beyond the written text. *Research in Corpus Linguistics* 10(1).

Gilquin, G. & A. McMichael. 2018. Through the prototypes of *through*: A corpus-based cognitive analysis. *Yearbook of the German Cognitive Linguistics Association* 6(1): 43-69.

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

Granger, S. 2009. The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In K. Aijmer (ed.) *Corpora and Language Teaching*, pp. 13-32. Amsterdam: John Benjamins.

Granger, S. 2015. Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research* 1(1): 7-24.

Granger, S., G. Gilquin & F. Meunier. 2015a. Introduction: Learner corpus research – past, present and future. In S. Granger, G. Gilquin & F. Meunier (eds) *The Cambridge Handbook of Learner Corpus Research*, pp. 1-5. Cambridge: Cambridge University Press.

Granger, S, G. Gilquin & F. Meunier (eds). 2015b. *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press.

Granger, S. & S. Tyson. 1996. Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes* 15(1): 17-27.

Gries, S.T. & S.C. Deshors. 2014. Using regressions to explore deviations between corpus data and a standard/target: Two suggestions. *Corpora* 9(1), 109-136.

Gries, S.T. & S. Wulff. 2013. The genitive alternation in Chinese and German ESL learners: Towards a multifactorial notion of context in learner corpus research. *International Journal of Corpus Linguistics* 18(3): 327-356.

House, J. 2008. (Im)politeness in English as Lingua Franca discourse. In M. Locher & J. Strässler (eds) *Standards and Norms in the English Language*, pp. 351-366. Berlin: de Gruyter.

Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

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

Jenkins, J. 2000. *The Phonology of English as an International Language*. Oxford: Oxford University Press.

Kachru, B.B. 1985. Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H.G. Widdowson (eds) *English in the World: Teaching*

and Learning the Language and Literatures, pp. 11-30. Cambridge: Cambridge University Press.

Kachru, B.B. 1991. Liberation linguistics and the Quirk concern. *English Today* 7(1): 3-13.

Klippel, F. & J. Mukherjee. 2007. Standards and norms in language description and language teaching: An introduction. In S. Volk-Birke & J. Lippert (eds) *Anglistentag 2006 Halle, Proceedings*, pp. 303-306. Wissenschaftlicher Verlag Trier: Trier.

Kreyer, R. 2015. Review of Susan Nacey, "Metaphors in Learner English". *ICAME Journal* 39: 157-163.

Lee, D. & S. Chen. 2009. Making a bigger deal of the smaller words: Function words and other key items in research writing by Chinese learners. *Journal of Second Language Writing* 18(4): 149-165.

Leech, G. 1998. Preface. In S. Granger (ed.) *Learner English on Computer*, pp. xiv-xx. London: Longman.

Lin, C.-H. & Y.-L. Lin. 2019. Grammatical and lexical patterning of *make* in Asian learner writing: A corpus-based study of ICNALE. *3L, Language, Linguistics, Literature* 25(3): 1-15.

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

Mauranen, A. 2011. Learners and users – Who do we want data from? In F. Meunier, S. De Cock, G. Gilquin & M. Paquot (eds) *A Taste for Corpora. In Honour of Sylviane Granger*, pp. 155-171. Amsterdam: John Benjamins.

McMasters, M. 2004. How effective use of the English language can make or break your job search. *Academy of Organizational Culture, Communications and Conflict, Proceedings* 9(1): 41-46.

Meriläinen L. & H. Paulasto. 2017. Embedded inversion as an angloversal: Evidence from Inner, Outer, and Expanding Circle Englishes. In M. Filppula, J. Klemola & D. Sharma (eds) *The Oxford Handbook of World Englishes*, pp. 676-696. Oxford: Oxford University Press.

Meunier, F. 2016. Learner corpora and pedagogical applications. In F. Farr & L. Murray (eds) *The Routledge Handbook of Language Learning and Technology*, pp. 376-387. London: Routledge.

Mukherjee, J. 2005. The native speaker is alive and kicking – Linguistic and language-pedagogical perspectives. *Anglistik* 16(2): 7-23.

Mukherjee, J. 2007. Steady states in the evolution of New Englishes. Present-day Indian English as an equilibrium. *Journal of English Linguistics* 35(2): 157-187.

Mukherjee, J. & M. Hundt (eds). 2011. *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*. Amsterdam: John Benjamins.

Paikeday, T.M. 1985a. *The Native Speaker is Dead! An Informal Discussion of a Linguistic Myth with Noam Chomsky and Other Linguists, Philosophers, Psychologists, and Lexicographers*. Toronto: Paikeday Pub.

Paikeday, T.M. 1985b. May I kill the native speaker? *TESOL Quarterly* 19(2): 390-395.

Palmer, H.E. 1922. *Everyday Sentences in Spoken English*. Cambridge: W. Heffer & Sons Ltd.

Paulasto, H., L. Meriläinen & E. Ranta. 2011. Syntactic features in global Englishes: How 'global' are they? Paper presented at ISLE2, Boston, 17-21 June 2011.

Phillipson, R. 1992. ELT: The native speaker's burden? *ELT Journal* 46(1): 12-18.

Prodromou, L. 1998. Correspondence. *ELT Journal* 52(3): 266-267.

Quirk, R. 1990. Language varieties and standard language. *English Today* 6(1): 3-10.

Rampton, M.B.H. 1990. Displacing the 'native speaker': Expertise, affiliation, and inheritance. *ELT Journal* 44(2): 97-101.

Reinhardt, J. 2010. Directives in office hour consultations: A corpus-informed investigation of learner and expert usage. *English for Specific Purposes* 29(2): 94-107.

Rose, H., J. McKinley & N. Galloway. 2020. Global Englishes and language teaching: A review of pedagogical research. *Language Teaching* (First View): 1-33.

Schmied, J. 2008. East African English (Kenya, Uganda, Tanzania): Morphology and syntax. In R. Mesthrie (ed.) *Varieties of English, Vol. 4*, pp. 451-471. Berlin: de Gruyter.

Schönefeld, D. (ed.). 2011. *Converging Evidence: Methodological and Theoretical Issues for Linguistic Research*. Amsterdam: John Benjamins.

Swales, J.M. 2004. *Research Genres: Explorations and Applications*. Cambridge: Cambridge University Press.

Turton, N.D. & J.B. Heaton. 1996. *Longman Dictionary of Common Errors*. London: Longman.

Ugarte Olea, M.S.A. 2019. The Lingua Franca Core: A plausible option? *HOW* 26(2): 75-87.

Ur, P. 2010. English as a Lingua Franca: A teacher's perspective. *Cadernos de Letras (UFRJ)* 27: 85-92.

Vyatkina, N. 2013a. Specific syntactic complexity: Developmental profiling of individuals based on an annotated learner corpus. *The Modern Language Journal* 97(S1): 11-30.

Vyatkina, N. 2013b. Analyzing part-of-speech variability in a longitudinal learner corpus and a pedagogic corpus. In S. Granger, G. Gilquin & F. Meunier (eds) *Twenty Years of Learner Corpus Research: Looking back, Moving ahead*, pp. 479-491. Louvain-la-Neuve: Presses universitaires de Louvain.

Widdowson, H.G. 2000. On the limitations of linguistics applied. *Applied Linguistics* 21(1): 3-25.

Young, T.J. & S. Walsh. 2010. Which English? Whose English? An investigation of 'non-native' teachers' beliefs about target varieties. *Language, Culture and Curriculum* 23(2): 123-137.

Endnotes

---

[1] This is a revised version of a plenary talk given at the 12th Teaching and Language Corpora (TaLC) conference at Justus-Liebig-Universität Giessen. Thanks are due to Sandra Götz and Joybrato Mukherjee, chairs of the conference, for their kind invitation and for encouraging me to think about the concept of the norm. I would also like to thank the editor of *Language Teaching* as well as anonymous reviewers for their helpful comments on an earlier version of this article.

[2] The phrasing comes from Mukherjee (2005).

[3] See https://www.ielts.org/online-tutorial/overview (last accessed on 22 February 2021). It should be noted, however, that the latest version of the Common European Framework of Reference for Languages (CEFR) no longer refers to native norms.

Bionote

Gaëtanelle Gilquin is Professor of English Language and Linguistics at the University of Louvain, Belgium, and a member of the Centre for English Corpus Linguistics. Her research interests include corpus linguistics and learner corpus research, as well as varieties of English and the link between Learner Englishes and New Englishes. She is the co-director of the Louvain International Database of Spoken English Interlanguage (LINDSEI) and the director of the Process Corpus of English in Education (PROCEED), a new type of learner corpus that reproduces the writing process through keylogging and screencasting. She is one of the editors of the *Cambridge Handbook of Learner Corpus Research* (2015) and a founding member of the Learner Corpus Association.