Contrastive interlanguage analysis
A reappraisal

Sylviane Granger
Université catholique de Louvain

Since its introduction in 1996, Contrastive Interlanguage Analysis (CIA) has become a highly popular method in Learner Corpus Research. Its comparative design has made it possible to uncover a wide range of features distinctive of learner language and assess their degree of generalizability across learner populations. At the same time, however, the method has drawn criticism on several fronts. The purpose of this article is threefold: to provide a brief overview of CIA research, to discuss the main criticisms the method has faced in recent years and to present a revised model, CIA², which makes the central role played by variation in interlanguage studies more explicit and is generally more in line with the current state of foreign language theory and practice.

Keywords: Contrastive Interlanguage Analysis, native speaker, norm, comparative fallacy, interlanguage variation

1. Introduction

Learner corpora started to emerge in the late 1980s. From the early days, questions were raised about the best methodological approach to investigate this new type of data. For the first time ever researchers interested in foreign or second language (L2) learning and teaching had at their disposal very large and rich collections of L2 data in electronic format and powerful tools with which to analyse them. The main challenge was to design methods that could help researchers uncover the distinctive characteristics of learner language. A comparative framework suggested itself quite naturally as a good way of making the learner corpus speak. The basic approach in comparative methodology is to juxtapose two or more distinct subjects in order to identify similarities and differences between them, and thus “reveal aspects of the subjects that may not have been as easily seen if each was located in isolation” (Strawn 2009: 117). At the time of the emergence of Learner Corpus

ISSN 2215–1478 / E-ISSN 2215–1486 © John Benjamins Publishing Company
Research (LCR), this type of approach had already been used with great success in both monolingual and multilingual corpus linguistic studies. As observed by Johansson (1991:306), “[v]ariety is probably the area where corpus workers can make the most significant contribution”. This is readily apparent from the very wide range of studies that have compared (1) different varieties of the same language: spoken vs. written, general vs. domain-specific, current-day vs. earlier varieties, standard vs. other regional or social varieties; and (2) different languages.

In the case of learner language, two types of comparison appeared to be particularly worthwhile:

– a comparison with native language, seen as the ultimate attainment of learning a foreign/second language;
– a comparison of one sample of learner language with other samples of learner language, particularly from learners with different mother tongue backgrounds.

This approach was represented in diagram form in the 1996 article in which I introduced the term Contrastive Interlanguage Analysis (CIA) (see Figure 1).

In this first presentation of CIA, native and learner languages were abbreviated as NL and IL (interlanguage) respectively and the language used to illustrate the method was English: E1 for native English and E2 for English as a foreign language. The IL/IL branch of the method was illustrated with interlanguages representing different mother tongue backgrounds: English of French learners (E2F), German learners (E2G), Swedish learners (E2S) and Japanese learners (E2J). In subsequent presentations of the approach, other labels were used: NS for native speaker and NNS for non-native speaker (Granger 2002: 12); L1 for first/native, and L2 for foreign/second language (Granger 2012a: 18). These labels happily co-existed for a long time without raising much discussion within the LCR community.

Contrastive Interlanguage Analysis was not the first term used to refer to the approach. In my earliest paper on learner corpora (Granger 1993), I borrowed

\[
\text{CIA} \quad \begin{array}{c}
\text{NL vs IL} \\
E1 \leftrightarrow E2 \\
\text{IL vs IL} \\
E2F \leftrightarrow E2G \\
E2S \leftrightarrow E2J
\end{array}
\]

Figure 1. Contrastive Interlanguage Analysis (Granger 1996)
the term ‘comparative interlanguage research’ from Péry-Woodley (1990) to refer to a type of contrastive approach that consists in “comparing and contrasting what non-native and native speakers of a language do in a comparable situation” (1990: 144). I made clear, however, that I was using the term in a slightly broader sense, as in my own version of the approach the L2 could be compared not only with L1 but also with other L2 varieties. This two-pronged approach was considered ideal to meet the theoretical and applied objectives of LCR, namely to gain a better understanding of the mechanisms of foreign or second language acquisition and to design more efficient language teaching tools and methods.

CIA rapidly caught on and has spawned a large and highly diversified body of research. At the same time, however, it has also drawn criticism on several fronts. More than twenty years on, it is time to take stock of CIA research to date, address the criticisms it is facing and, if appropriate, revise the model. Section 2 provides a brief overview of CIA studies. Section 3 addresses the main objections levelled at CIA. In Section 4, I present a new version of CIA that brings it more in line with current language learning and teaching theory and practice. The final section contains some concluding remarks and pointers for future research.

2. Overview of CIA studies

The CIA methodology has been used in a wide range of studies. The learner corpus bibliography stored on the Learner Corpus Association website\(^1\) contains over 1,000 publications, most of which are at least partly CIA-based. It is well beyond the scope of this paper to give a comprehensive survey of CIA studies (see Hasselgård & Johansson 2011 for a recent review of the field). Instead I will provide a brief overview structured around the following issues: range of L2s, medium (spoken vs. written), linguistic phenomena investigated (lexis, grammar, discourse), proficiency level and type of CIA approach (L1 vs. L2 or L2 vs. L2). Special attention will be paid to emerging trends that provide evidence of the vitality of CIA studies.

**Target languages**

At first characterized by a near-exclusive focus on L2 English, LCR has gradually been extended to other L2s. The “Learner corpora around the world” website\(^2\) lists 128 corpus compilation projects, 60% of which focus on L2 English. The remaining 40% are distributed over twenty different L2s. English is still dominant, but the

---

diversification in learner corpus collection has led to a welcome increase in the number of CIA studies focusing on languages other than English (for example, Brecke & Zinsmeister 2012 for L2 German and Jantunen & Brunni 2013 for L2 Finnish).

Medium
Writing dominates CIA studies. This can be explained by the fact that writing, especially argumentative writing, is a key skill for the type of learners on whom CIA has mainly focused, i.e. advanced learners (see below). Nevertheless, attention for learner speech has grown in recent years. At first limited to an analysis of broad orthographic transcriptions, it is now beginning to rely on audio recordings of spoken language (see Ballier & Martin forthcoming for a recent survey). It is interesting to note that while the shift in LCR has been from writing to speech, mainstream Second Language Acquisition (SLA) research is undergoing an inverse evolution. Ortega & Byrnes (2008: 284) note that while “[t]raditionally, the field of SLA has tended to see oral language as a privileged site for L2 learning (…), the balance seems to be tipped on the side of academic and written language registers”. This evolution might well contribute to a rapprochement between the two fields.

Linguistic phenomena
One of the most notable contributions of LCR, and CIA studies in particular, is the impressive expansion of the range of linguistic phenomena investigated. The focus of SLA research has long been — and to a large extent still is — on “the nuts and bolts of language as a system — that is, mostly on the morphological and syntactic domains” (Ortega & Byrnes 2008: 283). LCR has added several new dimensions, in particular in the areas of lexis and discourse. The dominant view of lexis in CIA has been phrasal rather than single-word-based. The literature abounds in studies of collocations, colligations, lexical bundles and collocations, which have considerably enhanced our knowledge of the L2 phrasicon (for a survey see Paquot & Granger 2012). Studies of discourse features are also lexically driven to a large extent, with many studies on connectors and discourse markers, as well as routine formulae that play a major role in discourse. Comparatively fewer CIA studies focus on grammatical phenomena. The categories that have been investigated most extensively are “lexically-based” (Gilquin 2002: 186), such as the closed class of modals (Aijmer 2002) or phenomena like raising constructions that are triggered by particular verbs or adjectives (Callies 2008). But a large range of grammatical phenomena (noun phrase complexity, relative clauses, word order, etc.) cannot readily be investigated on the basis of raw corpus data. Greater use of part-of-speech tagged and, ideally, parsed learner corpora is clearly a key desideratum for
future CIA research. It is important to bear in mind, however, that it is always possible to manually tag any aspect of language. Few researchers have taken advantage of this option, most likely because of the tremendous amount of time and effort needed to design an annotation scheme and apply it to L2 data. However, works such as Springer’s (2012) study of discourse features in L2 writing give compelling evidence of the benefits it can bring.

Proficiency level
The focus on the upper stages of acquisition that characterizes CIA was intentional. It was meant to fill a gap in SLA research. SLA tended to centre more on the beginning stages of acquisition, which resulted in a lack of knowledge of advanced interlanguage and a corresponding lack of pedagogical materials addressing advanced learners’ needs. As a result of the many CIA studies carried out over the years, we now have a much clearer picture of the complex interplay of lexico-grammatical and discourse features that characterizes advanced interlanguage. Although this focus on the upper limits of acquisition is a positive development, it is also a limitation and recent initiatives concentrating on the lower stages are a welcome new direction for CIA studies (see Giagkou et al. 2013).

L1 vs. L2
The more popular branch of CIA has involved a comparison of learner data with native data. Although this type of comparison is controversial (see Section 3), it is understandable that researchers should have embraced it with such enthusiasm in view of the dominant focus on the upper stages of the acquisition continuum: it makes much more sense to compare advanced learners with native speakers than to do so for beginners. The gap is narrower and L1-L2 comparisons are very powerful heuristic tools to circumscribe it. Patterns of over- and underuse of linguistic features can readily be identified with the appropriate software tools and methods and provide impetus for further analysis. Misuse is much more difficult to spot with this type of comparison and is usually investigated with Computer-aided Error Analysis rather than CIA. However, L1-L2 comparisons can help reveal some errors, as overuse is often an indicator of error. For example, the overuse of on the contrary by French learners of English results from a faulty one-to-one equivalence with the French connector au contraire.

The native corpora used are of two types: large representative corpora like the British National Corpus3 (BNC) or the Corpus of Contemporary American English4

(COCA); and smaller corpora like the Louvain Corpus of Native English Essays\(^5\) (LOCNESS), a corpus of novice native writing that has served as the counterpart to student non-native writing in many CIA studies. Recent years have seen growing use of comparisons of learner data with student writing assessed as proficient, regardless of native speaker status,\(^6\) a trend that reflects researchers’ growing questioning of the native vs. non-native dichotomy (see Section 3.2 for further discussion of this issue).

**L2 vs. L2**

It is very difficult to draw reliable conclusions from the analysis of one single L2 variety. This is something that was very quickly realized by the designers of the International Corpus of Learner English (ICLE; Granger et al. 2009): “In order to be able to distinguish those features of L2 English that were L1-dependent, i.e. the result of transfer from the mother tongue, from those that were common to all learners, irrespective of mother tongue, i.e. the cross-linguistic invariants, it was essential to enlarge the corpus and include learners from different language backgrounds” (Granger 1993: 60). Distinguishing general L2 features from features that are specific to one particular learner population is of crucial importance for both L2 theory and practice. CIA has made it possible to revisit the notion of transfer on a more solid empirical basis and to gain a much better understanding of the workings of transfer, including new notions such as “transfer of primings” (Paquot 2010: 192). At first, as is apparent from the above quote, L2-L2 comparisons were interpreted solely in terms of different language backgrounds. But there are many other variables that influence learner output and generate varieties of interlanguage that also deserve to be compared. Future research would benefit from more studies focused on variables other than L1 transfer, such as the effect of foreign vs. second language setting, task effects or differences in proficiency level. Studies such as Ådel (2008), which investigates the effect of timing on L2 performance, are examples to emulate.

---


\(^6\) The most widely used are the Michigan Corpus of Upper-Level Student Papers (MICUSP) (http://legacyweb.lsa.umich.edu/eli/eli1/micusp/Index.htm, accessed 14 September 2014) and the British Academic Written English (BAWE) corpus (http://www.coventry.ac.uk/research/research-directory/art-design/british-academic-written-english-corpus-bawe/, accessed 14 September 2014).
3. Criticisms of CIA

Over the years, CIA has been subjected to a range of criticism, most targeted at the L1/L2 branch. Two main issues have been raised: the first relates to the notion of ‘comparative fallacy’, while the other concerns the thorny issue of the norm.

3.1 Comparative fallacy

The notion of ‘comparative fallacy’ was coined by Bley-Vroman (1983), who argued that “work on the linguistic description of learners’ languages can be seriously hindered or sidetracked by a concern with the target language” (1983: 2). In a recent volume entitled “Interlanguage — Forty years later”, Selinker (2014: 230) reaffirms “the principle that interlanguages must be described in their own terms” and Larsen-Freeman (2014: 217) observes that “[b]y continuing to equate identity with idealized native speaker production as a definition of success, it is difficult to avoid seeing the learner’s IL as anything but deficient”. Bley-Vroman’s warning has been widely heeded by the SLA community, although the debate is far from closed (cf. Lardiere 2003).

Since CIA, at least in one of its branches, involves a comparison between learner language and the target language, it is understandable that it should have been accused of falling prey to the comparative fallacy. The following comment by Barbieri (2010: 149–150) is a good illustration of this critical stance towards CIA: “Learner corpus research [is] a discipline that, although growing steadily, appears to have remained trapped in the comparative fallacy — an approach to learner language that relies almost exclusively on notions of transfer and underuse or overuse and on contrastive models such as CIA and ICM”.

Similarly, Hunston (2002: 211) criticizes CIA on the grounds that “it assumes that learners have native-speaker norms as a target”.

Several arguments can be advanced to counter these criticisms. Tenfjord et al. (2006) insist that theory and method should not be confused. Although their main argument relates to Computer-aided Error Analysis, it is also relevant for CIA: “the TL [target language] description imposed upon the text by the analyst provides an objective method for pointing out phenomena that any adequate theory of SLA must be able to account for on a theoretical level” (2006: 101–102). Another powerful argument is that “any SLA study implicitly has a built-in notion of interlanguage with the target language lurking in the background” (Sung Park 2004: 3). As I pointed out in my own rebuttal of the comparative fallacy in LCR (Granger 2009), all the studies that compare learners of different proficiency levels

---

are in fact based on an underlying L1 norm, as proficiency is usually assessed with an L1 target in mind. The same can be said of SLA studies reporting the results of grammaticality judgement tests, which presuppose target-like answers. Examples of this type have led Cook (2012: 4173) to conclude that “[t]he native speaker has been the omnipresent ghost in the machine”. It should be added that the comparative fallacy does not enjoy unanimous support within the SLA community. The view held by Lardiere (2003: 140), for example, seems to be very close to the rationale behind CIA: “[I]t is true, as Bley-Vroman observes, that we researchers cannot guess in advance ‘the very large number of imaginable subsystems’ learners make use of in constructing the target grammar (…). But we have to start somewhere. It is a legitimate goal to determine where the likeliest jumping-off points are, that is, where and to what extent divergence from the target language is occurring”. White (2003: 27) adds that “[I]t is not the case that one should never compare L2 speakers to native speakers of the L2 (…). What is problematic is when certain conclusions are drawn based on failure to perform exactly like native speakers’. Ellis & Barkhuizen (2005: 360) provide another reason for questioning the validity of the comparative fallacy. They observe that “learners are typically targeted on native-speaker norms and as such themselves perform ‘cognitive comparisons’ in the process of learning an L2” and conclude that “an analysis based on external norms can be seen to be psycholinguistically valid”.

None of the arguments levelled against CIA are convincing enough to lead learner corpus researchers to abandon L1-L2 comparisons. However, the whole debate surrounding the comparative fallacy is a healthy reminder that interlanguages can — and indeed should — also be studied in their own right. There has been a strong tendency to set L2 against L1 rather than study it independently of L1. The second branch of CIA makes it possible to focus on L2, either cross-sectionally or longitudinally. In fact, the two branches of CIA can be combined: L1-L2 comparisons can be used to bring to light features of learner language that, once uncovered, can be analysed from a strictly L2 perspective. There is one additional important consideration that should be borne in mind: LCR — and within it, CIA — does not merely pursue a theoretical aim. It has always had a strong commitment to L2 teaching. From a pedagogical point of view, the benefit of L1-L2 comparisons is even more obvious, as they provide language teaching professionals with precious information on what learners do right or wrong or partly wrong in a particular skill or task, which can be used to inform a wide range of pedagogical applications.
3.2 The issue of the norm

The constant reference to native speaker language in CIA studies has been interpreted as recognition of the existence of one single monolithic norm in L2 studies. In the current language landscape, and more particularly in the case of English given its global status, nativeness has become a highly elusive and controversial notion. In their article entitled “Transcending the nativeness paradigm”, Brutt-Griffler & Samimi (2001) criticize the native vs. non-native dichotomy and conclude that “[t]he more English becomes an international language, the more the division of its speakers into ‘native’ and ‘non-native’ becomes inconsistent” (2001: 105). Criticism has emanated mainly from two fronts: World Englishes (WEs) and English as a Lingua Franca (ELF).

Taking as an example “the slowly emerging variety of Thai English”, Tan (2005) insists on better recognition of learners’ local environment. She deplores the fact that in current LCR, “[t]here is little consideration given to how the social and cultural practices of learners’ local contexts could influence their use of English” (2005: 127). In her view, comparisons with native language usage and, in particular, labels such as overuse, underuse and misuse reflect “imperialistic assumptions about the ownership of English” (ibid.: 128). This criticism of CIA is unfair. The idea that there are several possible standards for CIA has been underlined from the early days of LCR, as is evident from the following quote: “NL/IL comparisons require a control corpus of native English and (...) there is no lack of them. A corpus such as the International Corpus of English (ICE) even provides a choice of standards: British, American, Australian, Canadian, etc.” (Granger 1998: 13). ICE corpora are currently available for thirteen varieties (Canada, East Africa, Great Britain, Hong Kong, India, Ireland, Jamaica, New Zealand, Nigeria, The Philippines, Singapore, Sri Lanka, USA), all of which can be used as reference corpora for CIA studies.

The emergence of English as a Lingua Franca has also accounted for a number of critical remarks. ELF refers to communication among non-native users of English with different first languages rather than with native speakers. Using wording reminiscent of Selinker’s position on IL, Seidlhofer (2001) insisted that ELF should be viewed as language use in its own right, not as a deficit variety. In a similar vein, Cogo & Dewey (2012: 37) claim that “the native/non-native dichotomy is neither appropriate nor useful for an exploration of ELF interactions”. Aston (2011: 10–11) deplores the tendency of LCR to focus on negative features, i.e. “perceived inadequacies in learner behaviour”. He promotes the idea of the successful L2 user and suggests that “[R]ather than comparing their performances

---

with those of native speakers, it might well be more appropriate to compare them with performances by successful users of English as a Lingua Franca” (2011: 11). It is beyond the scope of this article to open up a debate around ELF. Two observations are in order though: first, as pointed out by Ferguson (2009: 127), ELF comes in a range of proficiency levels and not enough is known about this proficiency cline “to define more precisely the language of the expert ELF user or, indeed, to provide a means of distinguishing between ELF acceptable features and others not so acceptable”. Second, ELF is far from being incompatible with CIA. Cogo & Dewey (2012: 39) explicitly mention comparison with English as a native language (ENL) as a method used by ELF researchers to “identify which lexical items and grammatical structures regularly appear in ELF but which are not normally found in ENL”. The method is similar to CIA, the only difference being the general outlook on the features uncovered. As observed by Seidlhofer (2001) herself, the same phenomena can be viewed as deviations from a native speaker norm in an EFL perspective and as features of successful communication in ELF.

While neither WEs nor ELF pose a particular threat to CIA, it is important to pay heed to the ideas that underlie them. Both point to a plurilithic view of language (Hall et al. 2013) and lead us to reconsider the native vs. non-native dichotomy, which has been largely presented as uncontroversial in CIA. Progressively, however, the evolution of the L2 learning and teaching field has led to a growing feeling of unease with the dichotomy. It is symptomatic in this respect that my recent definitions of CIA (Granger 2012a, 2012b) include the notion of expert variety in addition to native variety. The conclusion is not to abandon the terms native and non-native altogether but to avoid using them as de facto generic terms in CIA.

4. CIA reappraised

4.1 Reference and learner varieties

The many changes that have affected the language landscape, and L2 studies in particular, in the past thirty years challenge the validity of some of the tenets of CIA. It is time to design a new version of the methodology, one that takes stock of the evolution both within and outside LCR. The new version, abbreviated as CIA², is presented in Figure 2.

9. Granger (2012a: 18): comparison of “either two learner varieties (L2 vs. L2) or one learner variety and one native (or expert) variety (L2 vs. L1)”. Granger (2012b: 3238): “Target language comparisons (…) set learner corpus data against a corpus of expert production, usually a native speaker corpus”.
In keeping with the strong variationist trend that characterizes current language studies, the new version of CIA promotes the notion of ‘varieties’: reference language varieties on the one hand, and interlanguage varieties on the other. The term ‘Reference Language Varieties’ (RLV) makes it clear that there are a large number of different reference points against which learner data can be set. Alongside the traditional inner circle varieties, which will continue to be used by many researchers, the model also incorporates the possibility of using outer circle varieties as well as corpora of competent L2 user data, as suggested for example by proponents of English as a Lingua Franca. Besides these dialectal variables, the diagram also makes explicit reference to diatypic variables, which are essential to ensure text-type comparability. As suggested by the dotted arrows, RLVs can be used concurrently and compared. This option has been used with success in several CIA studies. Lee & Chen (2009), for example, have used as RLVs a corpus of expert academic writing consisting of journal articles written by experts, who may or may not be native speakers, and a corpus of undergraduate dissertations written by novice native writers and observe that “[h]aving two corpora to compare our learner data against gives us added confidence that we are focusing on the right words” (2009: 154). The word ‘reference’ makes it clear that the corpus does not necessarily need to represent a norm (see Ådel 2006, Appendix 2, for a discussion of reference corpora in LCR). A good example to illustrate this is the LOCNESS. The reason it has been used so extensively is that it contains argumentative essays and is therefore arguably a more reliable basis for quantitative comparisons with L2 corpora like the ICLE than more general corpora. Connector density is a case in point. As shown by Granger & Tyson (1996), a corpus of argumentative essays like the LOCNESS contains a much higher number of connectors than a general
corpus of native writing like the *Lancaster-Oslo-Bergen Corpus* (LOB)\(^\text{10}\) and is therefore a more adequate basis for quantitative comparison with ICLE data. The rationale behind the use of a corpus of novice native-speaker writing has led to some misunderstandings, as evidenced by the following question raised by Aston (2008: 350): “[S]hould we really assume that native-speaker student essays, such as those used for comparison in the ICLE project, constitute a model to imitate?” The answer to this question is a clear and resounding no. The LOCNESS is a reference corpus, not a norm for EFL learners.

Parallel to this, CIA\(^2\) introduces the term ‘Interlanguage Varieties’ (ILV) to refer to learner language. The addition of the term ‘varieties’ stems from a wish to highlight the highly variable nature of interlanguage. It acts as a prompt for researchers to take variability in learner language even more into account in learner corpus studies. So far, the variable that has been most thoroughly investigated is undeniably the learners’ mother tongue. The explicit mention of task and learner variables in the diagram draws attention to the large number of variables whose effect should be investigated. It is noteworthy that many of these variables have been duly recorded as metadata in many learner corpus projects, but their potential effect on learner production has been relatively neglected in the analyses. There is clearly much to be gained from reading the abundant SLA literature on interlanguage variation (Dörnyei & Skehan 2003; Romaine 2003). This will enable us not only to gain increased knowledge of this key issue in acquisition studies, but also to add new dimensions to those traditionally recorded in the form of learner corpus metadata (see, for example, Möller 2013 for the inclusion of motivation and language aptitude scores).

\subsection*{4.2 Over- and underuse}

Criticism of CIA has also regularly involved a condemnation of the terms ‘overuse’ and ‘underuse’. Aston (2011: 11) speaks of “the employment of the damning terms ‘overuse’ and ‘underuse’”. According to the same author (2008: 343), these terms should be banned, as they imply that “the learner should at all times attempt to conform to native-speaker norms”. This criticism derives from a faulty interpretation of the two terms. The learner corpus literature makes it clear that the terms are neutral. Gilquin & Paquot (2008: 38), for example, note that “the terms ‘overuse’ and ‘underuse’ are descriptive, not prescriptive, terms: they merely refer to the fact that a linguistic form is found significantly more or less in the learner corpus than in the reference corpus”. Likewise, Lee & Chen (2009: 153) point out that in their study, “the term *overused* is a purely statistical term — there are no

\textsuperscript{10}. [http://clu.uni.no/icame/manuals/LOB/INDEX.HTM (accessed 14 September 2014).]
prior assumptions that all such overused items necessarily represent bad writing practices. They merely point to differences which merit further, qualitative investigation”. In my critical review of LCR (Granger 2009), I insisted on the need to apply a range of filters (learner needs, teaching objectives and teachability) to decide whether the frequency differences revealed by CIA needed to be acted upon or not.

The fact remains that dictionary definitions of the two terms indicate that they have negative connotations. To overuse, for example, is defined as ‘to use too much or too often’ or ‘to use so much as to make less appealing’. With the benefit of hindsight, I think it might have been better to use the term ‘overrepresentation’, which is less negative. The Oxford Online Dictionary\textsuperscript{11} defines the verb overrepresent as ‘to include a disproportionately large number of (a particular category or type of person), as in a statistical study’. The same reasoning applies to the term ‘underuse’. However, the two terms are well established now and, in my view, they can continue to be used in their technical meanings of ‘containing more or less than’. Aston’s criticism is not merely terminological, however. He certainly has a point when he states that “many analyses of learner corpora have treated all quantitative differences from reference group behaviour as undesirable — as things that learners should learn not to do, rather than as things that may help them communicate and learn more effectively” (2011: 11). Along the same lines, Shea (2009: 1) insists that “although research into the characteristics of learner text is valid in its own right, research which aims to inform pedagogy should look to move beyond designations of the relative frequencies of lexical items in texts and begin to address the question of whether these differences have a substantive effect on a learner’s writing”. In other words, using something too much may be a perfectly understandable, even desirable, feature of learner language. To take one example, using a large number of high-frequency verbs may be a very wise strategy in the beginning stages of the acquisition process. In more advanced stages, however, this overuse is probably a sign that the learners’ lexical repertoire needs to be expanded.

5. Conclusion

Since its inception in the late 1980s, CIA has become a highly popular methodological approach in LCR. More than twenty years on, it seemed worthwhile to revisit the method and propose a revised version that incorporates the many changes that have been taking place in language theory and practice in recent years, without sacrificing any of its strengths. The new model underscores the status of CIA as a

\footnote{http://www.oxforddictionaries.com (accessed 14 September 2014).}
method: it is not a theory of SLA and it has no direct link with teaching. It clarifies the ambiguity around the terms native/non-native and, most importantly, it makes the central role played by variation in interlanguage studies even more explicit.

The criticism levelled at CIA over the years should serve as an incentive to extend our model beyond interlanguage varieties. As shown in Figure 3, interlanguage has close links with other crosslingual varieties, a generic term I use to refer to language varieties that have specific characteristics due to the interplay of two or more languages. For English, the term includes Learner Englishes (LEs), World Englishes (WEs), Lingua Franca Engishes (LFEs) and Translated Englishes (TEs), each of which has its own privileged domains of reference. Some researchers have begun to cross disciplinary boundaries to compare crosslingual varieties. Gilquin (forthcoming), for example, shows that comparisons of WEs and LEs, though challenging, help us delve deeper into the phenomenon of language variation, highlighting 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)”. Mauranen (2011: 169), on the other

Figure 3. Crosslingual varieties of English
hand, remarks that although LEs and ELFs display fundamental differences, “they share certain features which make them yield results which are of great mutual interest”. Paquot (2014) highlights yet another area of fruitful cross-fertilization, i.e. the interface between LEs and TEs.

All scientific approaches can benefit from a regular questioning of their basic tenets. As aptly put by Johansson (1991:313), “[w]ise linguists, like experienced craftsmen, sharpen their tools and recognize their appropriate uses”. A first sharpening of CIA has been proposed here. I have no doubt that, as our knowledge of L2 acquisition continues to evolve, there will be scope for further sharpening in years to come.

References


DOI: 10.1111/j.1473-4192.2009.00225.x


DOI: 10.1075/lllt.6.04gra


Larsen-Freeman, D. 2014. “Another step to be taken – Rethinking the end point of the interlanguage continuum”. In Z. Han & E. Tarone (Eds.), Interlanguage. Forty years later. Amsterdam and Philadelphia: John Benjamins, 203–220. DOI: 10.1075/lllt.39.11ch9

Lee, D. & Chen, S. 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, 149–165. DOI: 10.1016/j.jslw.2009.05.004


Paquot, M. 2014. “Cross-linguistic influence and formulaic language. Recurrent word sequences in French learner writing”. In L. Roberts, I. Vedder & J. Hulstijn (Eds.), EUROSLA Yearbook


Author’s address

Sylviane Granger
Centre for English Corpus Linguistics
Institut Langage et Communication
Université catholique de Louvain
Place Blaise Pascal 1 bte L3.03.33
1348 Louvain-la-Neuve
Belgium
sylviane.granger@uclouvain.be