

# Learner corpora and learning context

Joybrato Mukherjee and Sandra Götz

## 1 Introduction

It has been noted by various scholars that second language acquisition (SLA) theory needs to take into account learning context as a determining factor in the acquisition/learning process (e.g. Norris and Ortega 2001). That said, SLA theoreticians face a number of challenges in systematising the infinity of individual learning processes and the multitude of specific learning contexts that affect and involve learners. These challenges are caused by the fact that ‘learning context’, defined by Ellis (1994: 197) as ‘the different settings in which second/foreign language (L2) learning can take place’, is by no means a variable that applied linguistic theory and description can easily come to grips with, as these settings are shaped by many different factors, ranging from the learner’s mother tongue (L1) and the language learning situation in his/her country of origin to specific learning materials used in the classroom context. Firstly, any specific learning context includes a variety of context-determining factors, e.g. the teacher’s or interlocutor’s quality and quantity of input in terms, for example, of the underlying linguistic standards, the acceptability of deviations from norms and the error sensitivity of the teacher or interlocutor. Secondly, and more importantly, one always needs to find methodological shortcuts, as it were, in order to reduce the sheer complexity of the variable ‘context’ (and its effects on the learner) by abstracting away a limited number of context types and, hence, by setting up certain typologies of learning contexts for theory and description. Housen et al. (2011), for example, distinguish between three levels of learning context: the curricular learning context as it is shaped by educational curricula and as it manifests itself in the language classroom; the extra-curricular context as transcending all situations, including out-of-school settings, in which language learning takes place; and the individual learning context of a specific learner inside and outside the classroom. In their study they focus on

the effect of one particular context-determining factor, namely the role of language prominence, in the extra-curricular and curricular context. In general, they are perfectly right in emphasising that ‘*context* is ill-defined in the literature and the term is interpreted and operationalised in many different ways’ (Housen et al. 2011: 106, emphasis in original). This is undeniably also a major challenge for learner corpus research, as the learning contexts in which learners have reached the interlanguage stage represented by a learner corpus have to be taken into consideration when the learners’ output is analysed.

## 2 Core issues

### 2.1 Learning contexts and abstract learning situation types

To begin with, it is necessary to carefully distinguish between the ongoing process of language acquisition/learning (which, in essence, is a sequence of many concrete learning contexts that in turn exert their influence on the learner), on the one hand, and any given interlanguage stage in the process that manifests itself in a snapshot of the learner’s output as a product (e.g. in a learner corpus), on the other (e.g. Ellis and Schmidt 1997). It is clear that a complete and detailed log of the entirety of a language learner’s exposure to the second/foreign language at hand can only be produced, if at all, in a longitudinal study for very few individual learners. Thus, if we want to abstract from individual learners to the general picture of the output of a greater number of typical learners (and to general issues of typical acquisition/learning processes), we need to take into account comparable data from many learners that have been exposed to a defined set of typical and representative learning contexts. This is one of the central rationales in compiling learner corpora as repositories of learners’ natural language use (Ellis 1994; Granger 2002). It goes without saying that learner corpora also enable us to zoom in on individual learners and, thus, on potential correlations between individual sets of learning contexts and their effects on a specific learner’s output. In the light of the great variability of learners’ performances even in fairly homogeneous cohorts as included in learner corpora, it has been repeatedly noted that learner corpus research should also pay due attention to the individual learner’s output (e.g. Mukherjee and Rohrbach 2006; Granger et al. 2009).

Against this background, a conceptual and terminological distinction should be drawn between individual learning contexts and learning situation types. Individual learning contexts are concrete in that they are specifiable in terms of the learner(s) and the context-determining factors. An individual learning context is, for example, a specific language learning lesson to an identifiable learner in a specific classroom context. On the other hand, learning situation types are those generalised kinds of

language-contact scenarios that can be considered as typical of any given acquisition/learning process. Such scenarios can be arranged from more specific types, e.g. the classroom context for learning English as a Foreign Language as a German learner in a typical German secondary school, to more general types, e.g. the study-abroad context in all its versatility. What these situation types, largely reminiscent of Housen et al.'s (2011) 'broad contexts', have in common is the fact that they are abstract and transcend individual learners and contexts. A very useful way of capturing a wide range of both individual learning contexts and abstract learning situation types is the compilation of so-called 'language contact profiles' for all learners in a given data set (see also Chapter 2, this volume). The method of language-contact profiling goes back to Seliger (1977) and has gone through a series of reincarnations, the most ambitious version perhaps being the one suggested by Freed et al. (2004a) for learners of Spanish as a foreign language in the United States. This version includes, for example, detailed information on learners' domain-specific use of Spanish as well as the kind and extent of interactions with interlocutors in Spanish inside and outside school. For learner corpora such as the *International Corpus of Learner English* (ICLE; Granger et al. 2009) and the *Louvain International Database of Spoken English Interlanguage* (LINDSEI; Gilquin et al. 2010), profiles of the abstract learning situation types experienced by the learners included in the corpus, e.g. the number of years of L2 learning in class and the time spent abroad in the target language community, are available as meta-information on the corpus.

Among the typical learning situation types that SLA theory has traditionally focused on are three situation types in particular: the formal language classroom ('at home', AH) context, the 'domestic immersion' (IM) context and the 'study-abroad' (SA) context (Collentine and Freed 2004a). These contexts refer to vastly different sets of specific situations in which individual learners of a foreign language typically find themselves. In the AH type of situations, learning a foreign language takes place in the instructional setting of a classroom, with a clear focus on form, on accuracy and on reaching certain curricular and language-pedagogical goals, in short: on truly 'learning' the language. The IM type manifests itself most typically in a summer programme in the L2 context in which learning the L2 is still in focus but in which genuine communicative functions of the L2 also come into play (when, for example, interacting in one's spare time with other summer-programme participants using English as a lingua franca, as it were). SA contexts in particular have spawned a vast literature in SLA research because of the (assumed) natural use of the L2 for a wide range of communicative functions in the target culture over a longer period of time. It has been shown in a number of empirical studies that – contrary to widely held beliefs in language pedagogy – SA contexts, though beneficial in many regards for learners, do not always and automatically facilitate or accelerate the acquisition/learning process

(e.g. Freed 1995; Barron 2002). The *LeaP* corpus (*Learning Prosody in a Foreign Language*; Gut 2006) and the error-tagged *Falko* corpus (Reznicek et al. 2010) are two examples of (German) corpora which contain information on the SA contexts that the language learners have been exposed to.

## 2.2 Extending the scope of learner corpus research: focus on English as a world language

The terms 'acquisition' and 'learning' are often used side by side in SLA theory, in applied linguistics and in language pedagogy. With regard to English as a global language with many more non-native users than native speakers, the conceptual background of these two terms is rather complex and, thus, warrants special attention and clarification.

There is unanimous agreement that the English language, as a consequence of British colonisation from the sixteenth to the twentieth century and American dominance in the twentieth and the beginning of the twenty-first century, has emerged as the most significant world language, and that it may well continue to be so in the light of the widespread use of English as a Second Language in newly industrialising economies like India (Graddol 1997, 2006). However, the great variety of Englishes spoken and learned around the world are not monolithic – quite the opposite: Kachru's (1985) well-known three-circles model offers a powerful (yet oversimplistic) distinction between three major types of Englishes: English as a Native Language (ENL; as used in the original English-speaking countries such as England), English as a Second Language (ESL; as used mostly by non-native, yet competent speakers – and officially recognised – in many former colonial territories, e.g. Nigeria) and English as a Foreign Language (EFL; as learned by an increasing number of people in non-English-speaking countries, e.g. China). Kachru (1985) assumes that while ENL varieties are 'norm-producing' (also in the sense of linguistic standards for *foreign* language classrooms around the world), ESL varieties are 'norm-developing' in that they have developed – or are in the process of developing – their own linguistic standards, resulting in their general acceptance, use and codification, that is, in 'endonormative stabilisation'. In contradistinction, EFL learners (and their individual variants) are entirely dependent on 'exonormative' standards, and their deviations from these norms are usually considered as learner errors, not as exponents of variety formation. In this general context of World Englishes, it continues to make sense, following Gass's (1997) distinction, to relate the terms and concepts of 'second language' and 'language acquisition' to ESL varieties, while 'foreign language' and 'language learning' seem to be more fitting for the EFL setting.

Above and beyond such seemingly neat, categorical distinctions, there is a growing awareness in applied linguistics and SLA research, corpus linguistics and learner corpus research, as well as in the World Englishes

community, that in the realities and histories, in the contexts of language use of billions of speakers and learners of English around the world, we deal with continua rather than categories. From a diachronic perspective, this is shown by Schneider's (2003, 2007) 'dynamic model' of the evolution of New Englishes and its application to a wide range of test cases. For example, there are gradual shifts (which can be described in terms of five overlapping phases) in the interactions between the settlers and the indigenous population, in language-contact situations (including the dominant situation types for English language learning available to the indigenous people) and in the overarching political and societal setting, that foster the development of English as a Foreign Language in a new colonial territory towards becoming an accepted ESL variety. Once such an ESL variety has become an accepted part of the local linguistic repertoire, it is also possible for a language shift from an ESL variety to an ENL variety to set in, as can be currently observed amongst young Singaporeans. In essence, Schneider's (2003, 2007) dynamic model assumes a historical EFL-ESL-ENL continuum as underlying the formation of new varieties of English. From a synchronic perspective, Gilquin and Granger (2011), Housen et al. (2011) and others, too, view the distinction between ESL and EFL as a cline with many in-between categories. On the basis of analysing and comparing various national components of *ICLE*, Gilquin and Granger (2011), for example, argue the case for establishing a cline between more EFL-like learner variants (e.g. English as used by Spanish L1 learners in Spain) and more ESL-like ones (e.g. English as used by Tswana L1 learners in South Africa). They also assume that factors related to differences in the dominant learning situation types are largely responsible for this cline, namely the different degree of focus on form in the classroom and the different exposure to English outside the classroom.

In the light of the clines between the various types of English in general and the shared features of postcolonial second-language varieties and learner Englishes in particular (e.g. the largely non-native status of the speakers/learners, the additional-language status of English, transfer effects from speakers'/learners' L1 and the importance of the classroom-based learning situation type), it comes as no surprise that the past few years have seen intense discussions as to how to bridge the long-established 'paradigm gap' (Sridhar and Sridhar 1986) between SLA theory and learner language analysis (including learner corpus research), on the one hand, and research into postcolonial second-language varieties in the World Englishes community, on the other. The studies included in Mukherjee and Hundt (2011) as well as Nesselhauf (2009), for example, show that it is worth developing an integrated view of EFL and ESL and, thus, analysing and comparing corpora that capture the entire range from foreign language learners and competent speakers of postcolonial second-language varieties to native speakers of English as a first language. For the various types of English and their users, the major

abstract learning situation types and the individual learning contexts are not categorically different but provide a complex picture of gradual shifts (e.g. the growing importance of formal instruction from ENL to ESL to EFL), partially shared features (e.g. the high level of exposure to acrolectal newspaper English in both ENL and ESL varieties) and clear differences (e.g. the absence of English from most contexts outside the classroom in EFL settings).

### **2.3 A note on (learner) corpus research and cognitive linguistics**

From the very beginning of computer-based corpus linguistics in the late 1950s, there has been a great divide between the corpus-linguistic community, traditionally interested in the description of natural language use in general and recurrent patterns in linguistic performance in particular, and the generativist camp with its clear focus on the native speaker's competence. This great divide kept the corpus-linguistic research paradigm and the cognitive-linguistic/psycholinguistic community apart for a number of decades. Over the past few years, however, there seems to have been general agreement that these two paradigms in their modern versions and with their state-of-the art empirical methodologies are not at all mutually exclusive but can be combined very effectively to shed light on the 'performance' and 'competence' of language users in a holistic way (e.g. Bybee 2002; Mukherjee 2004; Gries 2010a). Thanks to this welcome development of convergence, links are growing between learner corpus research and the dominant cognitive strand in SLA theory (e.g. Granger 2009a). More specifically, various areas of potential cross-fertilisation between learner corpus research, cognitive linguistics and SLA theory have been identified and explored in a number of recent corpus-based interlanguage studies (e.g. Deshors 2015; Littré 2015).

With regard to learning context as a decisive factor in the language-learning process, there is a potential link between a corpus-based analysis of the individual contexts in which learners come across certain words and structures, on the one hand, and their relevance to a model of the underlying cognitive state as a result of the learning processes, on the other hand (e.g. in terms of a mental grammar), when considering Hoey's (2005: 11) well-known suggestion that 'the mind has a mental concordance of every word it has encountered, a concordance that has been richly glossed for social, physical, discursal, generic and interpersonal context'. That said, it remains clear that a learner corpus does not offer an immediate window onto learners' minds. In a discussion of the implications and applications of findings obtained from learner corpora, a general caveat applies: in order to provide converging evidence for plausible cognitively oriented models and explanations, corpus data need to be combined with experimental data (e.g. elicitation tests, acceptability tasks) or other appropriate data (e.g. Gilquin and

Gries 2009). In the following sample studies, the focus will be on how learner context information and learning situation type information in (learner) corpora can be utilised for the description and analysis of learner language use.

### 3 Representative studies

There is a substantial body of empirical studies that have looked at different learning contexts and/or learning situations. In the following, we will present four representative studies from different areas of research and with different disciplinary backgrounds as illustrations of the wide range of research perspectives on learning context.

The first sample study (see Section 3.1) represents the wealth of research in the SLA paradigm. Most of the research has focused on longitudinal studies that have investigated foreign language learners' gains in proficiency in different areas of language use before and after a stay abroad in the target language community or an intensive programme compared to learners in stay-home control groups (see also Chapter 18, this volume). One thoroughly investigated and representative study of this kind was carried out by Freed et al. (2004b) on learners of French. Specifically, they investigate fluency gains in the three aforementioned groups.

In Section 3.2, we will report on one of the few studies at this stage that have utilised the learning-context information included in learner corpora. We assume that this will most probably change in the future, because learner corpora such as *ICLE* or *LINDSEI* (see above) include meta-information with learner profiles. Callies and Szczesniak (2008) put forward a methodologically very interesting suggestion of how learner profiles can be used to compile subcorpora that represent learners' proficiency according to the time they have spent abroad in addition to relying on their 'institutional status' alone.

More recently, learning context has also attracted the interest of linguists who try to bridge the gap between the SLA paradigm, focusing on EFL, and research into postcolonial ESL varieties, for example by describing and comparing learner languages and World Englishes ranging from more 'ESL-like settings' to more 'EFL-like settings'. In Section 3.3, we will report on the findings of a sample study (Mukherjee and Gries 2009) that has looked at lexico-grammatical differences between three postcolonial Englishes in different evolutionary stages, representing different points on the EFL-ESL-ENL cline.

Finally, in Section 3.4, we will present De Knop's (2015) study of French L1 learners' acquisition and conceptualisation of the metaphorical meanings of the German posture verb *sitzen*, a study which represents a similarly recent trend in learner corpus research: the combination of learner corpora, cognitive-linguistic models and data, and SLA theory.

**3.1 Freed, B. F., Segalowitz, N. and Dewey, D. P. 2004b.** 'Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs', *Studies in Second Language Acquisition* 26(2): 275-301.

Freed et al.'s (2004b) study was one of the first studies to compare quantitatively the acquisition of various dimensions of fluency. It focused on twenty-eight students of French who studied in three different learning contexts: (1) formal language classrooms in an institution at home (AH); (2) intensive summer-immersion programmes (IM); (3) a study-abroad setting (SA). Specifically, Freed et al. (2004b) analysed and compared gain scores as a function of the learning context and as a function of the time reported using French outside the classroom. The study aims at systematically and quantitatively comparing differences in fluency acquisition of American adult learners (i.e. college students) of French in an AH vs IM vs SA context. In particular, it addresses the following research questions (Freed et al. 2004b: 280):

1. Are there salient differences in the acquisition of oral fluency by students who have studied abroad, compared to students whose learning takes place in IM programmes or the regular AH language classroom?
2. Do time-on-task factors (e.g. instructional time, out-of-class-time spent interacting orally with native speakers or using the language within the literate domain) vary in each of these contexts?
3. To what extent are the measured differences in oral fluency associated with these time-on-task features?

The study investigated twenty-eight American college students of French as a foreign language (eighteen female, ten male) at an average age of 21.29 years (range = 17-48). All of them were native English speakers and none of them had spent time abroad in the target language community before this study. Six of the students were majoring in French, the rest were majoring in the social sciences, humanities and science. The AH group consisted of eight students (three female, five male), the IM group had twelve students (eight female, four male), and the SA group comprised eight students (seven female, one male).

The different learning contexts are characterised by Freed et al. (2004b) as follows: the students of the AH group had two to four years of prior instruction of French, had a weekly average of three to four hours of French instruction taking place mostly in French, but they were not extensively exposed to the French language outside the classroom. The students who were enrolled in the IM programme also had two to four years of prior instruction and were enrolled in one to three language courses, but had a weekly average of 17.5 hours of formal classroom study that took place exclusively in French. There were a number of activities the students were supposed to take outside the classroom; they were also supposed to live together with other French school students. The students of the SA

group, who spent the summer in Paris, also had an average of two to four years of prior instruction and a weekly average of 16.4 French class hours per week. Additionally, these students could also take courses in the social sciences in French, were exposed to a variety of social activities with other French students, and lived either with French host families, in student residences or by themselves, and they were pledged to speak only French.

Freed et al. (2004b: 285ff.) investigated six quantifiable fluency measurements for their study, which they defined in the following way: (1) speech rate in words per minute, (2) hesitation-free speech runs, i.e. runs without dysfluencies consisting of silent pauses of 400 ms or more, (3) filler-free speech runs, i.e. mean number of words spoken without filled dysfluent pauses (regardless of length), (4) fluent runs, i.e. the number of words in the longest run of speech without any pauses or dysfluencies, (5) repetition-free speech runs, i.e. the mean number of words in a run of speech that did not contain any kinds of repeats or false starts, (6) grammatical-repair-free speech runs, i.e. number of words in the longest run without self-correction of a grammatical mistake. Additionally, they included three other measures of the learners' overall performance, viz. (7) total words spoken, (8) duration of speaking time and (9) longest turn in words.

For the data collection, all the students participated in 15–30-minute oral interviews at the beginning and at the end of the semester. These interviews were transcribed, and two one-minute segments spoken by each learner (one at the seventh, one at the eleventh minute) were extracted from both interviews. Altogether, four interview segments were thus analysed for oral fluency features per learner. In addition to these interview segments, the learners had to submit two language-contact profiles (LCP, Freed et al. 2004a) before and after the semester, which specified their participation in the extra-curricular activities and other means of exposure to the French language outside the classroom in more detail. These data were analysed by calculating analyses of variance (ANOVAs); correlations were tested by calculating one-tailed or two-tailed *t*-tests. Multiple regression analyses were used to test possible correlations of gains in fluency as the dependent variable and hours spent using French outside the classroom as the predictor variable.

Concerning the fluency measurements before and after the semester, the study revealed that the IM and the SA groups showed significant gains in fluency after the semester, with the IM group showing the greatest increase, whereas the AH group did not show any gains at all. The findings can be summarised in more detail as follows:

1. The AH group showed no overall trends for significant gains in any of the investigated variables over time; in six of the investigated variables, only half of the students had an equal performance or showed an increase from the pre- to the post-test, and in three of the variables, the majority of students even performed worse in the post-test.

2. The IM group made significant gains in oral performance in terms of the total number of spoken words, length of the longest turn, rate of speech and speech fluency based on a composite of fluency measures.
3. The SA group showed improvements in their performances in four of the variables, although not all of them to a significant extent. Significant gains could only be documented in terms of speech fluency compared to the AH group, but the SA group showed fewer gains than the IM group.

The second step of the study correlated out-of-class language contact to the French language with the learners' oral fluency performance using the LCP data in the analysis. Here, the IM group reported significantly more contact hours with the French language than any other group in speaking and writing, as well as in global contact hours. The SA group, on the other hand, reported significantly more contact situations with the English language than the IM class, except for listening. The SA group also had more out-of-class activities in English than in French (the difference not being significant, however). In contrast, the IM group showed significantly more out-of-class contact in French than in English.

Gains in fluency of the learners' foreign language ability and their overall performance were correlated with the sum of the number of out-of-class hours spent using (i.e. speaking, reading, writing and listening to) French. The multiple regression analyses revealed that, surprisingly, the reported hours per week spent writing in French outside the classroom were associated with gains in oral fluency.

The study offers an innovative way of operationalising the concept of speech fluency by analysing a variety of quantifiable variables and suggests a highly appealing methodological procedure of describing how different variables of learning context are connected with gains in fluency over a semester. The study has very important language-pedagogical implications and proves that different learning contexts result in significant differences in the acquisition of fluency features, yet (and contrary to intuitive expectations one may have had) studying abroad does not necessarily show the greatest positive effects. In fact, it was the participants of the IM group that showed the greatest overall gains in most of the investigated variables. Since it was also the IM group that reported the greatest exposure to the target language outside the classroom, the study corroborates that exposure to the target language is strongly associated with the learners' increase in fluency. The study has also shown that this cannot be taken for granted, even in an SA context, as the questionnaires showed that the learners in the SA group spent more hours a week using English than French. These results stress the paramount importance that authentic exposure to the target language outside the classroom has for achieving gains in language proficiency. Finally, Freed et al. (2004b) argue that these effects cannot be described by looking at the learning context

alone. Rather, one has to take into account additional meta-information gathered from LCPs.

**3.2 Callies, M. and Szczesniak, K. 2008.** 'Argument realization, information status and syntactic weight: A learner-corpus study of the dative alternation', in Walter, M. and Grommes, P. (eds.), *Fortgeschrittene Lernervarietäten: Korpuslinguistik und Zweitspracherwerbsforschung*. Tübingen: Niemeyer, pp. 165–78.

Callies and Szczesniak (2008) investigate advanced German and Polish learners' use of dative verbs in English as their target language with regard to aspects of argument realisation, information status and syntactic weight. Their study shows how learner profiles in learner corpora can be used to determine the proficiency level of learners by taking into consideration the learning context (i.e. the amount of time spent abroad in the target language community) in order to compile homogeneous sub-corpora of learners with specific proficiency levels.

This study investigates the two alternative constituent orderings of the dative alternation in advanced German and Polish learners' written English in order to test if major principles of information-structure influence the learners' choice of the possible variants.

The data investigated in this study were written data of intermediate to advanced German and Polish learners of English as compared to the writing of native speakers of English. The data were taken from the argumentative essays of learners in *ICLE* (Granger et al. 2002) and compared to the native-speaker reference corpus, the *Louvain Corpus of Native English Essays* (*LOCNESS*, divided into a British and an American subcorpus). All the learners were undergraduate university students of English in their third or fourth year of studies; all of them were in their twenties. Additionally, this study makes use of learning context variables (here: number of months the learners had spent abroad in the target language community) as an indicator for determining each individual learner's proficiency level more accurately. Thus, Callies and Szczesniak only included learners who had spent up to twelve months abroad in an English-speaking country in order to balance their data. Accordingly, they excluded more proficient learners from their study, i.e. those who had spent an extended time abroad (i.e. more than twelve months). Although the correlation between learning context and proficiency level is a complex one and has to be taken with considerable caution, the researchers nonetheless defined a plausible methodological shortcut, as it were.

The study focused on fifteen highly frequent dative verbs which were taken from the research literature on the dative alternation (i.e. Levin 1993; Wasow 1997; Biber et al. 1999) and checked against frequencies in the *British National Corpus* (*BNC*).<sup>1</sup>

<sup>1</sup> [www.natcorp.ox.ac.uk/](http://www.natcorp.ox.ac.uk/) (last accessed on 13 April 2015).

The fifteen most frequent dative verbs were extracted from the four subcorpora (i.e. US *LOCNESS*, BRIT *LOCNESS*, Polish *ICLE* (only including learners who had spent up to twelve months abroad), German *ICLE* (only including learners who had spent up to twelve months abroad)) and were then manually categorised according to dative verbs that occurred in a double-object construction (DOC, e.g. *John gave Mary the book*) and those which occurred in a prepositional construction (PC, e.g. *John gave the book to Mary*). In order to prevent fixed expressions that only allow one of the variants from skewing an investigation of dative alternations, fixed-goal expressions<sup>2</sup> which discourage/block the dative alternation (e.g. *bring somebody to life/power* vs *\*bring somebody life/power*) were excluded from the analysis, while fixed-theme expressions<sup>3</sup> which can occur in alternations (e.g. *give a break*) were included. All cases were manually analysed according to the corpus and the variant used. In a qualitative, content-related analysis, it was determined whether the choice of variant can be predicted by information status and/or syntactic weight.

Callies and Szczesniak (2008) show that the investigated verbs have a preferred tendency to occur in one construction across all corpora, and no deviations from the native target norm can be attested for the Polish or the German learners. For instance, the verb *tell* is used in the DOC in the great majority of cases in all three corpora. The findings also suggest that the lexical constraints of dative verbs do not cause problems for advanced learners who had spent up to twelve months abroad. In a similar vein to the native speakers, the learners' use of either the PC or DOC construction is also determined by information status and syntactic weight. Although the stay-abroad variable does not have a significant influence on the learners' performance in this study, it opens up interesting methodological perspectives on incorporating learning context variables into learner corpus research.

Methodologically, this study shows how learner corpora along with the accompanying learning profiles, as provided in *ICLE*, for example, can be used to create subcorpora that represent homogeneous learner groups according to different learning context variables, e.g. the time spent abroad in the target language community. Although this approach is a methodological shortcut, as it were, it provides one way of compiling *quasi*-longitudinal corpora and, thus, of dispensing with the troublesome and time-consuming procedures of conducting pre- and post-tests before and after the semester abroad by compiling subcorpora of learner groups who, for example, exclusively studied in an AH context (e.g. learners who

<sup>2</sup> Fixed-goal expressions involve some sort of caused motion and a spatial goal and do not include a recipient. They can only occur in the *to*-variant (e.g. *send somebody to the devil* vs *\*send the devil somebody*).

<sup>3</sup> Fixed-theme expressions involve some sort of caused possession meaning and, logically, a recipient, rather than a purely spatial goal. They can either occur in idiomatic constructions where no variation is possible (e.g. *show somebody the ropes* vs *\*show the ropes to somebody*) or can allow for alternations (e.g. *to give somebody the cold shoulder* vs *give the cold shoulder to somebody*).

spent zero months abroad) and learner groups who studied English in an SA context (e.g. who spent six to twelve months abroad).

**3.3 Mukherjee, J. and Gries, S. Th. 2009.** ‘Collostructional nativisation in New Englishes: Verb-construction associations in the International Corpus of English’, *English World-Wide* 30(1): 27–51.

Mukherjee and Gries’s (2009) study looks at lexico-grammatical tendencies in three postcolonial Asian Englishes: Hong Kong English, Indian English and Singapore English. In all the three postcolonial contexts, new second-language varieties of English have emerged. But according to Schneider’s (2003, 2007) evolutionary model of variety formation, the three varieties have to be plotted onto different developmental stages, with Hong Kong English being the least advanced (and the most EFL-like) variety and Singapore English being the most advanced one: the extra-curricular and curricular learning contexts are, therefore, vastly different. The study compares the use of collostructions, i.e. verb-construction associations as defined by Stefanowitsch and Gries (2003), between the three postcolonial Englishes and the shared historical input variety, namely British English.

Mukherjee and Gries (2009) are interested in identifying correlations between the evolutionary stage of a postcolonial variety of English with its specific learning context, on the one hand, and the similarity/dissimilarity of the use of the collostructions with the historical input variety, on the other. The hypothesis is that the more advanced (and the least EFL-like) a postcolonial variety is, the more dissimilar it is in its collostructional profile from British English.

In Schneider’s (2003, 2007) evolutionary model of variety formation, which has been applied to a wide range of postcolonial Englishes worldwide, it is assumed that the emergence of a New English variety follows an essentially uniform process of five phases: (1) foundation, (2) exonormative stabilisation, (3) nativisation, (4) endonormative stabilisation, (5) differentiation. Essentially, this process implies that for the indigenous people in a colonial territory, the English language develops from a foreign language into a ‘nativised’ second language, which is retained in the postcolonial era as a (co-)official language fulfilling a wide range of communicative functions in administration and politics, in the press, in the economy and in the educational system. Mukherjee and Gries (2009) argue that in the light of Schneider’s (2003, 2007) criteria and parameters, Hong Kong English is still in an early stage (phase 2/3) with many characteristic features of a truly foreign language (EFL), e.g. with hardly any widespread acceptance of local norms and standards: ‘There is thus no societal basis for any nativised variety of “Hong Kong English”’ (Li 1999: 95). Indian English, on the other hand, is a much more advanced variety of English. Local norms and standards have emerged at virtually all linguistic levels, which have also been codified in a number of

dictionaries and usage guides. These Indian features of English are also widely accepted as admissible forms by many local speakers. According to Mukherjee (2007) and others, Indian English as a typical example of an ESL variety represents an evolutionary stage somewhere between phase 3 and phase 4 in Schneider's (2003, 2007) model. Singapore English is the most advanced variety in the fully endonormatively stabilised phase 4. In fact, there is a growing tendency among young speakers to acquire English as an L1.

In this study, the three postcolonial groups of speakers thus represent different positions on the EFL–ESL–ENL cline. The different postcolonial contexts are characterised by fundamentally different learning contexts for the acquisition of English (in the sense of abstract learning situation types, see Section 2.1), with a focus on classroom-internal use only in Hong Kong, the widespread use as a national lingua franca and an interethnic link language in India, and as a national language symbolising the integration of ethnicities and cultures into one nation in Singapore.

In this study, the Hong Kong, Indian, Singapore and British components of the *International Corpus of English* (*ICE-HK*, *ICE-IND*, *ICE-SIN* and *ICE-GB*, respectively) are used. The *ICE* components are designed to be comparable corpora; they have the same size (1 million words each) and the same design (40% written, 60% spoken, same genres, same number of text samples per genre). *ICE-GB* served as the baseline variety: on this basis the relevant collocations were identified for the comparative study.

Four different types of verbs were looked at: fifteen verbs attracted to the ditransitive construction in *ICE-GB*, fourteen verbs attracted to the intransitive construction, fifteen verbs attracted to the monotransitive construction and fifteen verbs with no particular collocational preferences ('neutral verbs'). A proportional sample of instances of these fifty-nine high-frequency verbs in the three constructions was taken. In total, 11,487 instances of the fifty-nine verbs in the four *ICE* components were handcoded for the three constructions. On this basis, a 'multiple distinctive collexeme analysis' was conducted, resulting in a matrix of (binary) constructional preferences versus dispreferences for each of the fifty-nine verbs across the four corpora. For example, while the most prototypical ditransitive verb *give* was found to prefer the ditransitive construction (e.g. *she gave them a lecture*) and disprefer both the intransitive and the monotransitive construction in *ICE-GB*, *ICE-HK* and *ICE-SIN*, the same verb turned out to prefer not only the ditransitive construction but also the monotransitive construction in *ICE-IND* (e.g. *she gave a lecture*).

The study shows that while the collocational preferences for verbs that attract the monotransitive construction in *ICE-GB* are more or less the same across the three Asian *ICE* components, the collocational tendencies for verbs that are attracted to the ditransitive construction show intervarietal variation: various verbs that prefer the ditransitive construction in *ICE-GB* do not prefer the ditransitive construction in one of the

Asian Englishes (e.g. *cost* in *ICE-IND*, *lend* in *ICE-SIN*); in all of these cases, the verbs at hand prefer the monotransitive construction in the Asian English variety. The most significant finding of the study is a clear overall correlation between the evolutionary stage of the postcolonial variety, on the one hand, and its degree of dissimilarity from present-day British English, on the other, thus confirming the initial hypothesis. More specifically, from *ICE-HK* to *ICE-IND* and *ICE-SIN*, the number of collocational preferences/dispreferences shared with *ICE-GB* decreases. That is, the less EFL-like the learning context becomes, the more the speaker of a postcolonial variety diverges from the patternings of the historical input variety.

The implications of this study for research into learning contexts across postcolonial Englishes are twofold. Firstly, different learning contexts in different postcolonial settings may result in different patterns of usage in learners'/speakers' output, depending on the extent to which a full-fledged variety has already emerged. Secondly, future research ought to analyse and compare in a much more systematic way language use in different postcolonial settings as abstract learning situation types in general and more specific learning contexts in individual postcolonial settings in particular (e.g. language use in Singaporean families in interactions between L2-English-speaking parents and L1-English-speaking children).

**3.4 De Knop, S. 2015.** 'Conceptual tools for the description and acquisition of the German posture verb *sitzen*', *Corpus Linguistics and Linguistic Theory* 11(1): 127–60.

De Knop (2015) looks at the range of meanings of the German posture verb *sitzen* ('sit') in general and the underuse of its metaphorical meanings by Belgian-French L1 learners of German in particular. Relevant metaphorical uses include, for example, *er sitzt in der Falle* ('he is trapped', lit. 'he sits in the trap') and *das Kleid sitzt nicht* ('the dress does not fit well', lit. 'the dress does not sit'). This verb has been widely neglected in research into German posture verbs, which so far has concentrated on *stehen* ('stand') and *liegen* ('lie'). This study aims to fill this gap by (1) providing a cognitive-linguistic description of the postural, locational and metaphorical meanings of *sitzen* and (2) offering SLA-related and language-pedagogical applications that highlight the importance of the actual input offered in the learning context. In her study, De Knop (2015) combines corpus data and experimental data.

De Knop (2015) argues that the acquisition of the metaphorical meanings of *sitzen* can be enhanced if the context of learning is enriched with cognitively accessible concepts as well as with more data in the learning context representing these meanings. Her main research questions are twofold: (1) how can the various meanings of *sitzen* be described and categorised? and (2) how can the cognitive-linguistic descriptive framework

be utilised with regard to the acquisition of *sitzen* by French L1 learners of German?

The study includes two tests. In the first cloze test, the test group consisted of thirty French-speaking Bachelor students of German in translation and interpretation studies in Brussels. Their proficiency in German is at the B1/B2 level according to the Common European Framework of Reference for Languages (Council of Europe 2001). The control group consisted of thirty native speakers of German of a similar age. De Knop (2015) had looked at the major teaching materials and grammar books used in Belgium in order to find out if and to what extent metaphorical meanings of *sitzen* occur in the typical German foreign language classroom in Belgium (with classroom lessons serving as the major abstract learning situation type for the learners at hand). Against this background, the cloze test is intended to establish to what extent the learners and the native speakers have the metaphorical use of *sitzen* at their disposal. In the second cloze test, eighteen Belgian-French learners of German in the same study programme (with a B2/C1 proficiency level) took part. This second test consisted of a pre-test (with the learners taking part with the same assumed low level of exposure to metaphorical uses of *sitzen* in their previous classroom experience as the first test group), a teacher-centred lesson on German posture verbs in general and the verb *sitzen* and the semantic network of its meanings in particular, and a post-test (with the learners now taking part with the additional experience of the teacher-centred lesson).

In order to establish which meanings of *sitzen* are used by foreign language learners of German, De Knop (2015) consulted the *Falko* corpus. For the general picture of native-speaker usage, she looked at the core corpus of the digital dictionary of the German language (*Digitales Wörterbuch der Deutschen Sprache, DWDS*).<sup>4</sup> In the first cloze test, thirty sentences were presented to the test group and the control group; twelve sentences were taken from the *DWDS* corpus with slots to be filled with a form of *sitzen* (either locational or metaphorical meaning). The other eighteen sentences were control sentences. The significance of the differences between the two groups was shown with the help of three separate hierarchical logistic regression models (using the *lme4* R package). In the second cloze test, a similar methodology was used: both the pre-test and the post-test included sixteen sentences each, but only in eleven sentences would it have been natural for the participants to choose a form of *sitzen*; again, an appropriate hierarchical logistic regression model was fitted to the data. For the conceptualisation of the meanings of *sitzen* – which also served as the basis for the teacher-centred lesson between the pre-test and the post-test – a semantic network was established linking the metaphorical meanings of *sitzen* to the basic postural meaning of the verb (e.g. *eine Frau*

<sup>4</sup> [www.dwds.de](http://www.dwds.de) (last accessed on 13 April 2015).

*sitzt am Tisch*, lit. ‘a woman sits at the table’) and its locational meanings (e.g. *die Brille sitzt auf der Nase*, lit. ‘the glasses sit on the nose’), using conceptual metaphors (and their visualisation) as tools to explain the meaning extension (e.g. CATEGORIES ARE CONTACT as in *das Kleid sitzt nicht*, lit. ‘the dress does not sit’).

The study shows that while 5 per cent of all instances of *sitzen* in German native usage include metaphorical verb meanings, Belgian-French learners of German restrict themselves entirely to postural and locational meanings and use other strategies than metaphorical *sitzen* in cloze tests (e.g. the general verb *sein*, ‘to be’). One main reason for the underuse of metaphorical *sitzen* is its underrepresentation in the typical learning contexts in the German language classroom in Belgium, as teaching materials and grammar books indicate. De Knop (2015) shows in her study that an enriched learning context, offering learners a cognitively accessible network of meanings of *sitzen* (including the metaphorical ones) with concrete examples, can benefit the learners in that they use metaphorical *sitzen* significantly more frequently: from the pre-test to the post-test in the second cloze test, the percentage of answers including *sitzen* went up from 1% to 56%. Although the experimental set-up has to be taken with a measure of caution (because, for example, the teacher-centred lesson on German posture verbs immediately preceded the post-test), the study confirms, firstly, that the learning context is a decisive factor in the acquisition process and, secondly, that cognitive linguistics, learner corpus research and language pedagogy are natural allies.

#### 4 Critical assessment and future directions

As the representative studies in Section 3 indicate, all the factors related to learning context in all its versatility exert a great influence on the language-learning process and learners’ L2 output. There are great differences in proficiency gains depending on the learning context and other socio-biographic variables. However, empirical research into interrelations between learning contexts and language learners’ performance so far has mainly focused on a small number of learners and/or a very restricted set of context-related variables. This is due to the different levels of granularity at which learning contexts can be described (from abstract learning situation types to specific learning contexts), the multitude of factors that are relevant when analysing learning contexts and the lack of standardised (learner) corpora with context-related meta-information on the learners and their language-learning experiences.

Future research will no doubt profit immensely from the advent of relevant large-scale (learner) corpora. For English, a number of resources are already available. The most important learner corpora for written and spoken learner data are *ICLE* and *LINDSEI*; for the analysis

of interrelations between different postcolonial settings and L2 speakers' output, *ICE* provides an unprecedented database with a variety of meta-information for some (but unfortunately not all) of the national *ICE* components. Also, the recently compiled *International Corpus Network of Asian Learners of English (ICNALE)*; (Ishikawa 2011, 2014) represents another massive database of c. 1.5 million words of speech and writing produced by learners from ten Asian countries. Due to its conscientiously collected meta-information on each learner of the corpus, *ICNALE* allows for exhaustive analyses of different aspects of learning context, for 'contrastive (interlanguage) analyses' (Granger 1996), as well as for analyses of ESL varieties and EFL variants in Asia. For other target languages, an increasing number of learner corpora, some of which have taken new steps in corpus annotation, have become available recently, too, e.g. for German as a foreign language, the *Falko* corpus and the *LeaP* corpus include innovative annotation schemes and unprecedented meta-information with regard to error tagging (see Chapter 7, this volume) and the phonology of learner language (see Chapter 6, this volume), respectively.

Given the increasing interest in correlations between learning context (both at the level of individual learning contexts and abstract learning situation types) and the process of acquiring and using a foreign language, it is to be hoped that more corpora become available that provide as detailed information on the learners included and the contexts of language learning they have experienced as provided by e.g. *LINDSEI*.

From a methodological perspective, it should be noted that monofactorial analyses of the kind that have been characteristic of much research into learning contexts so far can never shed light, due to their very nature, on the complexity of the wide range of context-related factors and their interrelation in shaping a given language-learning situation. Therefore, future research into correlations between learning contexts, on the one hand, and language learners' output, on the other, should be multifactorial in nature (e.g. Gries and Deshors 2014; Chapter 8, this volume). Given that any language acquisition process is an essentially individual learning process, it is necessary in future research to explore in detail the correlations between individual learners' learning contexts and their individual second language acquisition processes. This picks up on the suggestion made by Coniam (2004) and Barlow (2013) that – in spite of the corpus-linguistic interest in more or less large groups of speakers representing language varieties and languages in their entirety – individual language-using profiles and individual usage-based grammars, respectively, be taken into account, too. As language acquisition is a process in time, future research should also look more closely at the development of learning contexts across time and the implications of these changes for the language acquisition process from a longitudinal perspective (as, for example, suggested by De Knop (2015) with regard to the second cloze test in her own study, see Section 3.4).

Finally, and from a theoretical viewpoint, it is necessary to define 'context' more clearly in terms of the descriptive level of granularity and the related factors that need to be addressed. In the present chapter we have argued that a distinction be made between abstract learning situation types and individual learning contexts. These two conceptions should be viewed as endpoints on a cline covering the entire gradient from macroscopic approaches to context (e.g. in terms of the general English language-learning setting in Hong Kong) to minute microscopic analyses (e.g. by looking at correlations in a specific classroom context between individual teachers' input and the output of language learners in specific cohorts).

## Key readings

Collentine, J. and Freed, B. F. (eds.) 2004b. *Learning Context and its Effects on Second Language Acquisition*. Special issue of *Studies in Second Language Acquisition* 26(2).

Howard, M. (ed.) 2011. *Role of Learning Context in Second Language (L2) Acquisition*. Special issue of *International Review of Applied Linguistics in Language Teaching* 49(2).

Both special issues include studies of specific linguistic structures (e.g. at the level of phonology) and how learners' output of the structures at hand is affected by learning-context-related factors, ranging from instructed classroom settings to study-abroad contexts. They also address theoretical and methodological issues, e.g. the question of how to define and conceptualise learning context.

DuFon, M. A. and Churchill, E. (eds.) 2006. *Language Learners in Study Abroad Contexts*. Clevedon: Multilingual Matters.

This volume provides a wide-ranging collection of empirical studies of learners in various study-abroad learning contexts. Each individual study offers both a specific perspective on learner experiences abroad and a discussion of theoretical issues related to study-abroad contexts and their impact on language learning.

Meunier, F. and De Knop, S. (eds.) 2015 *Learner Corpus Research, Cognitive Linguistics and Second Language Acquisition*. Special issue of *Corpus Linguistics and Linguistic Theory* 11(1).

This special issue includes various studies on different target languages in which cognitive-linguistic models and methods are combined with learner corpus research in order to better describe and model aspects of second language acquisition. By drawing on specific

learner corpora and, thus, learners with defined learning experiences, all the studies provide interesting insights into potential correlations between (aspects of) learning context and the learning process (and progress).

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

This book includes a number of studies that are intended to bridge the paradigm gap between SLA-based investigations of foreign language learners of English and the field of World Englishes, focusing on postcolonial second-language varieties of English. This collection reveals that it is useful to analyse and compare learning contexts in all the predominantly non-native settings in which English is used.

Kirkpatrick, A. (ed.) 2010. *The Routledge Handbook of World Englishes*. London: Routledge.

This handbook offers a comprehensive overview of contexts of usage and learning contexts of English, including regional varieties of English, and various contexts in which English is used as an international language (e.g. 'online Englishes'). Several chapters also address language-pedagogical implications arising out of the multitude of (learning) contexts in which English is used.