

Research in corpora in language teaching and learning

Elaine Vaughan

Michael McCarthy

1 Introduction

A corpus (plural corpora) is simply a large collection of texts, typically nowadays stored on a computer. The idea of understanding individual texts or any part of a text by assembling large quantities of texts that have something in common long predates modern computerized corpora (for a brief history, see McCarthy and O’Keeffe 2010: 3-5). The use of the word *corpus* to refer to a body of texts (e.g. all the works of one author) goes back to at least the early 18th century, and for many centuries before that, biblical scholars treated the Bible as a corpus, seeking out all instances of particular words and bringing them together in manually-produced concordances. Equally, scholars of literature have long sought to understand an author’s mindset or ‘fingerprint’ by examining their work as a body. And that is what we still do when we use corpora in language teaching: we seek evidence as to how the language is used and how it is learnt by looking at lots of texts assembled together. This way of approaching language is considered by corpus linguists to be a more reliable yardstick than looking at any single text or expression in a text, or else putting aside the textual evidence and looking inwardly, the process of introspection, which some linguists believe to be equally valid.

Discussion of the relationship between corpora, corpus research and second language pedagogy is in some ways well-trodden ground; in a 2011 overview, McEneaney and Xiao identify at least twenty-five authored or edited volumes with the pedagogical applications of language corpora more or less loosely as their theme (McEneaney & Xiao 2011). In 1996, Jan Svartvik, a key figure in the establishment of corpus linguistics as a disciplinary paradigm with a distinctive voice and flavor, claimed that corpora were becoming ‘mainstream’, no longer the province of the ‘boffin’ (as they had been perceived hitherto). The relatively recent appearance of handbooks specific to the domain of corpus linguistics reflects its growing maturity as

a diverse but cohesive domain of research (O’Keeffe & McCarthy 2010; Biber & Reppen 2015).

Rundell and Stock (1992) used the memorable phrase, the “corpus revolution” to refer to the impact that corpora had had – and still have – on lexicography and lexicology. In contrast, the impact of corpus linguistics on second language teaching and learning has been more gradual. In what follows, we trace the impact that access to corpora has had on key areas such as reference works and classroom materials. We look at ways in which corpus research has challenged and supported the edifices of second language teaching and learning, and we give an overall flavor of some of the ways in which research into various types of corpora has changed the way we look at language.

2 Pioneers: Corpora and reference books for language learners

2.1 Learners’ dictionaries

The first major effect computerized corpora had on language teaching was the publication of corpus-informed dictionaries for learners. Traditionally, dictionaries had been written based on time-consuming manual annotation of how words were used in the works of great writers and with what meanings. Using a computer had four major effects on this process: (1) the amount of data to be examined could be increased to a scale previously unimagined and could include thousands of books, newspapers, magazines, official documents, letters, transcripts of speech, and, more recently, the vast textual resources of the world wide web, (2) the massive task of producing a comprehensive dictionary of a language could be sped up to an extent never achieved before, (3) dictionaries could be better targeted towards groups of users with specific needs, and (4) the evidence of usage presented by the computer often challenged conventional knowledge and ideas about language and the way linguists understood the process of meaning-making.

Point (4) above is particularly important. For instance, a computer storing millions of words of English texts is able very quickly to examine its data and produce a frequency list of all the words in the corpus ranging from the most frequent to the rare and obscure, something impossible to achieve purely through introspection. Such information is invaluable in the design of dictionaries for learners and teaching

Vaughan, E. & M. McCarthy, (forthcoming 2017). Research in Corpora in Language Teaching and Learning. In: Hinkel, E. (ed.) *Handbook of Research in Second Language Teaching and Learning. Volume III*. New York: Routledge, pp. 173-185.

materials. Knowing what the most frequent words of a language are and what their most frequent meanings are enables lexicographers to produce dictionaries targeted at learners at different levels of proficiency. For an elementary learner of English who wants to progress to a higher level, for example, a dictionary the size of the massive *Oxford English Dictionary* might be a less than practical purchase, whereas one that covers the 5-10,000 or so most frequent words could be very useful indeed, for example the *Collins COBUILD Essential English Dictionary* (COBUILD 2010), aimed at learners approximately up to the B1 (lower intermediate) level of the CEFR. COBUILD dictionaries were pioneers in the field of corpus-based dictionaries, thanks to the work of John Sinclair and his team at the University of Birmingham, UK, but other major publishers soon followed suit and there now exists a popular range of corpus-based dictionaries for learners of English at different levels.

Crucially, and also in relation to point 4, above, pioneers in the field of corpus-based dictionaries such as Sinclair and his associates were struck by the way computer analysis revealed very heavy, regular patterning in the use of English (Sinclair 1991). For example, a concordance of the word form *see* in the five-million-word CANCODE corpus of everyday spoken English (see McCarthy 1998 for details of the corpus) yields some 10,800 lines of examples, far too many for one human individual manually to find and examine. However, corpus software quickly reveals a visual and statistically-supported pattern: around 4,500 of the concordance lines (45%) are accounted for by either *You see* or *I see*, which occur as frequent discourse markers in conversation. *You see* and *I see*, along with thousands of other strings of two or more words in both speaking and writing, have become frozen or fossilized into routine formulae, also referred to as chunks, lexical bundles, clusters or phrasal/fixed expressions (McCarthy & Carter 2002; Biber et al. 2004; O’Keeffe et al. 2007; Martinez & Schmitt 2012). Many of these chunks proved to be as frequent as or even more frequent than many common, everyday single words (for a discussion, see Shin and Nation 2008). In this respect, the by-products of corpus analysis for dictionary-making were to have a profound impact on thinking about what and how to teach.

However, it was not just chunks of the pragmatic/discoursal kind that impacted on reference works; soon specialized dictionaries of phrasal verbs and idioms based on

corpus evidence were made available for learners (e.g. the *Cambridge International Dictionary of Phrasal Verbs* of 1997 and its partner *Cambridge International Dictionary of Idioms* of 1998). Equally, dictionary research had revealed the ubiquity of collocations (combinations of words of varying degrees of fossilization). For example the adverb *strictly*, although common in the pragmatic chunk *strictly speaking*, also shows a very strong tendency in the written segment of the British National Corpus (BNC)¹ to collocate with adjectives such as *controlled*, *limited*, *necessary*, *confidential* and *forbidden*, suggesting that it commonly functions as an intensifier with a preference for contexts of restriction and authority. Knowing what the typical collocations are for thousands of words is not necessarily something that native speakers or non-native expert users of a language can access through intuition and introspection, and corpus-based dictionaries of collocations for learners have come to the rescue (e.g. the *Oxford Collocations Dictionary for Students of English* of 2002 and the *Macmillan Collocations Dictionary* of 2010). More recent corpus research into collocation has confirmed its centrality in meaning-making (notably Cheng et al. 2009).

Dictionaries for learners have come of age and it would now seem odd for any major publisher to trumpet a learner's dictionary that was *not* corpus-based, such is the expectation that has built over the last 30 years.

2.2 Grammar reference works

As with dictionaries, the evidence of corpora forced a rethink of what aspects of grammar should be taught and how. Once again, the COBUILD team occupied the vanguard of publishing corpus-based grammar reference for learners (COBUILD 1990). Sinclair, in his introduction to the first COBUILD grammar, was clear in his commitment to the use of natural examples to illustrate grammar in functional contexts, saying: "It is sad that many teachers seem doomed to work with invented material", and, in a clear rejection of introspection, "The mind plays tricks, and, specially, is unreliable when one is thinking about very short utterances, without a clear context to support them" (Sinclair 1990: xi). For Sinclair, the only proper source for understanding grammar was the corpus.

¹ Data cited herein has been extracted from the British National Corpus Online service, managed by Oxford University Computing Services on behalf of the BNC Consortium. All rights in the texts cited are reserved.

Vaughan, E. & M. McCarthy, (forthcoming 2017). Research in Corpora in Language Teaching and Learning. In: Hinkel, E. (ed.) *Handbook of Research in Second Language Teaching and Learning. Volume III*. New York: Routledge, pp. 173-185.

The extent to which research into corpora brought into relief distinctions between spoken and written grammars of various types (so that we can now talk about grammars in the plural, and not necessarily mean ‘grammar reference books’) has been one of the major contributions of corpus research. Traditional, sentence-based grammar (mostly created in an era before easy voice recording was possible) relied heavily on written evidence, but spoken corpora revealed a number of grammatical features that distinguished speaking from writing in important ways. In their major reference grammar, Biber et al. (1999) observed significant differences between the grammar of, on the one hand, written registers (fiction writing, news writing and academic writing) and on the other, conversation. Earlier, Carter and McCarthy (Carter and McCarthy 1995) had argued for greater attention to spoken grammar in language pedagogy. Spoken corpora showed that (a) there were often major differences in the frequency of grammatical items and structures between writing and speaking (especially everyday conversation), (b) a good deal of spoken ‘grammar’ could be subsumed within fossilized chunks, and (c) some common features of conversational grammar in corpora had gone unnoticed and had never found their way into materials or the classroom.

The steady transformation of grammar reference books has mirrored that of dictionaries, with the major language publishers putting out titles that boast corpus-informed content, for example Biber et al. (2002), Swan (2005), Carter and McCarthy (2006), Parrott (2010), Carter et al. (2011). In his introduction to the *Collins-COBUILD English Grammar*, Sinclair optimistically concluded: “...in a few years’ time teachers will look back and wonder how they coped with the lifeless examples they used to work with” (Sinclair 1990). The popularity and widespread acceptance of corpus-based grammar reference books suggest he was right.

3 Teaching and learning materials

3.1 Vocabulary and grammar learning

An important spin-off from the lexicographical research of the 1980s and 1990s was the move towards bringing corpus information into learners’ materials. Lexical research into corpora had thrown up information revealing the phrasal nature of large swathes of the vocabulary of English, including fossilized chunks, phrasal verbs and

idioms, and collocations. The 1990s and 2000s were to see corpus-informed materials covering all these aspects. The special nature of some of these materials was that they also included information based on research into learner corpora. The evidence of large corpora such as the Cambridge Learner Corpus (<http://languageresearch.cambridge.org/index.php/cambridge-english-corpus>, hereafter CLC) has been used in a variety of ways that materials writers have benefitted and continue to benefit from.

Learner corpora can provide both ‘negative’ and ‘positive’ evidence of how learners use the target language. If the corpus is error-coded (i.e. what raters perceive to be errors of usage are coded in some way in the data to make them searchable), recurring problems can be spotted and addressed in the materials. Moreover, if the data is also linked to a database containing information about the learners (e.g. age, gender, first language, CEFR level, etc.) then more specifically-targeted content concerning errors, such as errors common to a particular first-language user group or language family, or common at a particular proficiency level can be incorporated into the material. For example, McCarthy and O’Dell (2012: 172-3), based on their research into the error-coded CLC, select for teaching and practice a range of uncountable nouns (e.g. *information, advice, transport, knowledge*) which seem to remain stubbornly problematic even at B2 and advanced levels of the CEFR.

While error warnings might seem a negative way of using learner corpora information (i.e. what a learner should *not* do), much positive information has been exploited too. The English Profile project (see <http://www.englishprofile.org/>), a multi-disciplinary research program into how learners actually use the target language at the different levels of proficiency described in the CEFR, has led to the creation of the *English Vocabulary Profile*, a free online resource based on corpus research, where teachers and learners can see what vocabulary typical, internationally-profiled learners have command of at any particular CEFR level (for further information, see <http://www.englishprofile.org/wordlists>). This database has become a major source of information for materials writers, with more and more published materials based on it becoming available. There now exist a large number of learner corpora, not just for English, which will undoubtedly influence and contribute to the creation of corpus-informed vocabulary materials over the coming years. At the time of writing, the

Université catholique de Louvain's Centre for English Corpus Linguistics lists more than 150 learner corpora of various types on its *Learner Corpora around the World* webpage (<http://www.uclouvain.be/en-cecl-lcworld.html>).

The COBUILD project, as before, led the way with various books exploiting general and specialized vocabulary from corpora (e.g. Goodale 1995; Maskull 1995, 1996, 1997), and continues to provide vocabulary materials for all levels of the CEFR (COBUILD 2013a). Materials using corpus data include teaching and practice books on idioms, phrasal verbs, collocations, and specialized areas such as business vocabulary and academic vocabulary (Maskull 2010; McCarthy and O'Dell 2016). The COBUILD *Key Words* series (COBUILD 2013b) gives corpus-based examples and practice in vocabulary on a range of special subjects such as accounting, hospitality, chemical engineering, etc.

One notable feature of corpus-informed vocabulary materials is the growing focus on chunks as a central aspect of vocabulary learning. Traditionally, vocabulary learning was dominated by single words, often in lists and often in the form of words in the L2 with a translation into the learner's L1; only phrasal verbs and colorful idioms provided variation. However, in recent years, more and more attention has been paid in materials to the forms and functions of common chunks extracted from corpora.

In the case of grammar teaching and learning, at the time of writing, less is available in terms of corpus-informed material than for vocabulary, though this is changing rapidly. Grammar materials reflect the corpus insights incorporated into grammar reference works, principally the distinction between spoken and written grammar, along with a move towards blurring the distinction between grammar and lexis. Hughes and McCarthy (1998) argued for greater attention to spoken contexts for the understanding of grammatical meanings and functions in teaching (e.g. the uses of the past perfect in conversation for providing background information), while McCarthy and Carter (2001) offer 10 criteria for establishing a spoken grammar in pedagogical contexts, based on corpus research.

Spoken grammar materials include features such as situational ellipsis, that is to say, the non-occurrence of items often taught as obligatory with regard to writing. So the golden rules of English finite clauses always requiring subject pronouns and

requiring auxiliary or modal verbs in questions are routinely absent in banal, everyday utterances such as “Finished yet?”, “Sorry, don’t know” and “You still got that book I lent you?”. Such natural and fully grammatical utterances present a two-pronged dilemma for teachers. On the one hand, stakeholders in the language teaching enterprise (teachers and students, curriculum bosses, publishers, parents, politicians) often perceive such usage as impoverished or downright wrong (see Hughes 2010). On the other hand, features like situational ellipsis are typically context-embedded and not amenable to out-of-context, sentence-based grammar teaching, requiring considerable skill on the part of material writers and/or teachers in creating appropriate situations for their non-forced use in the classroom or self-study exercises. Despite this, practical guidelines have been offered (e.g. Hilliard 2014) and spoken grammar materials have met with a positive response from teachers and learners (Carter and McCarthy, 1997; Carter et al. 2000; Carter et al. 2011b; Paterson et al. 2012).

With regard to the grammar-lexis distinction and the way corpus evidence challenges the division, considerable attention has been paid to the implications for language teaching of pattern grammar (Hunston & Francis 2000; Willis 2003) and the lexical approach (Willis 1990; Lewis 1997). In these paradigms, the emphasis is on the patterned relationship between grammatical configurations and the lexical items they regularly co-occur with. For example, in the North American COCA corpus, the English subjunctive form (where all persons utilize the base form of the verb without inflexion, as in “I insist that she *be* allowed to attend”) is not a free-floating form but seems strongly wedded to contexts where it is governed by lexical verbs such as *request, demand, require, suggest, urge, direct, insist*, i.e. verbs related to directive or persuasive speech acts. The suggestion for pedagogical grammar is that it would be more helpful to learners, when teaching structures, to list at the same time, the lexical verbs most frequently to be found in their environments. Thornbury (2005) is an extension of this principle, working around the lexico-grammar of the most frequent words in English. These approaches thus blur the distinction between the ‘grammar’ lesson and the ‘reading and vocabulary’ lesson which was often the hallmark of syllabuses and class timetables. The potential of evidence from corpora to challenge intuited or cherished notions of ‘grammar’ or ‘correct’ and ‘incorrect’

usage can be said to have been both instructive and disruptive where second language teaching and learning is concerned.

3.2 General course books and skills teaching

An interesting direction in the pedagogical research has been to critique the nature of second language learning materials, such as textbooks, and how representative they are of the sort of language learners will encounter outside of the classroom, based on corpus evidence. Mindt (1996) suggested that, based on his analysis of English language textbooks designed for the German market, there was “a kind of School English which does not seem to exist outside the language classroom” (p. 232). Learners exposed to such material were (in his view) ill-equipped to handle authentic interactions with ‘native’ [sic.] speakers (ibid.). Meanwhile, Römer (2004, 2005), examines the disparity between input in textbooks and the samples of naturally occurring use in corpora, such as the use of modal verbs and progressives, Conrad (2004) examines the authentic use of the linking adverbial *though*, and Rühlemann (2009) focuses on an area of language – the reporting of direct speech – where the incongruity between everyday use and pedagogical grammar are missing an essential element. Rühlemann (2009) notes the concentration in pedagogical materials on indirect reporting of speech typical primarily of spoken to written discourse, involving backshift, rather than the more overwhelmingly common, narratized versions typical of naturally occurring spoken language (ibid: 429; see also McCarthy 1998: ch.8). This disparity has been addressed in some more recent textbooks, and is discussed in depth by McCarten (2010).

Perhaps the biggest challenge of the corpus revolution has been the exploitation of corpus information beyond the more formal aspects of language encoded in the grammar and vocabulary. Already in the 1990s and early 2000s, applied linguists were advocating more discourse-and pragmatics-based approaches to language teaching, including, in addition to contextual and discorsal approaches to grammar, discourse-level teaching of speaking and writing skills (McCarthy 1991, 1998; Celce-Murcia 1991; Hinkel 1997a and b, 2002; Celce-Murcia & Olshtain 2000). However, considerable obstacles faced those interested in exploiting corpora and in mediating corpus information to create practical, usable and appropriate teaching materials for the four skills, especially speaking. The earliest corpus-informed general course

Vaughan, E. & M. McCarthy, (forthcoming 2017). Research in Corpora in Language Teaching and Learning. In: Hinkel, E. (ed.) *Handbook of Research in Second Language Teaching and Learning. Volume III*. New York: Routledge, pp. 173-185.

books once again came out of the COBUILD project (Willis & Willis 1988, 1989). The COBUILD course was ahead of its time in several ways. Not only did it adopt a task-based approach (new enough in its day) and an emphasis on oral performance, it also incorporated the lexical approach and graded its language content according to the frequency of the core vocabulary of English as measured by the COBUILD corpus (e.g. Part 1 focused on the first 700 most frequent words, Part 2 extended coverage to the next 850 words, and so on). This corpus-led approach resulted in what was perhaps an unfamiliar set of tools for many language teachers steeped in more conventional materials – in many parts of the world, the notional-functional and communicative approaches were still struggling to make an impact – and the course was not without its critics for the demands it placed on its users (e.g. Fillips McCreary 1991). It was to be more than a decade later that authors and publishers took up again the gauntlet of corpus-informed general courses. McCarthy et al. (2005-6 and 2012-14) was an attempt to put into a more familiar package for teachers and learners (in terms of design, activity types etc.) corpus-derived content which included practice with pragmatic markers, lexical chunks, rhetorical strategies in writing, interpersonal speaking strategies, along with information about corpus frequency of items in graphic displays. The whole syllabus, not just word lists, was corpus-informed, based on corpora compiled with the end users in mind rather than simply exploiting existing corpora gathered for purposes such as dictionary writing.

4 Corpora and L2 pragmatics

Spoken corpora are an invaluable resource for studying pragmatic phenomena in interaction (Vaughan & Clancy 2013), given the recent focus of pragmatic research on ‘authentic’ or naturally occurring spoken language. Many of the linguistic resources whose patterns spoken corpus research has uncovered are characterised by their contribution to the interpersonal functioning and structural flow of conversation (see Clancy & McCarthy 2015). Where second language teaching and learning is concerned at least, there exists a twinned postulation of what have been termed *pragmalinguistic* and *sociopragmatic* knowledge (e.g. Thomas 1983). These refer to, in the first place, knowledge of the conventional linguistic formulations by which an act may be accomplished (e.g. a greeting, or a compliment) and awareness that a number of different options for its accomplishment may exist (pragmalinguistic

knowledge). This implies another type of competence for learners: control of appropriate pragmalinguistic choices, or sociopragmatic knowledge. From a research perspective, there are many different ways to get at what learners know in this regard, and what they need to know. Pragmatic competence as a critical aspect of second language use has been a strong focus of research for quite some time. A combination of attention to what has been called *interlanguage pragmatics*, or how second language speakers use and acquire L2 pragmatics (see Blum-Kulka & Kasper 1993) and exploitation of language corpora of various types has been advanced as a way of complementing or augmenting conventional modes of investigation. For example, specific pragmatic phenomena such as speech acts have often been studied using discourse completion tasks (DCTs). This work has been usefully supplemented by comparative studies of speech act realizations, such as Golato's (2003) research on compliments, and Schauer and Adolphs' (2006) focus on expressions of gratitude, where the evidence of DCTs is compared with evidence from corpora of naturally occurring language.

In a more general sense, access to language corpora has meant that varieties of language can be described in systematic ways. One of the ways in which different spoken varieties of pluricentric languages like English can tend to make their uniqueness visible is in 'small' items, such as pragmatic markers – items such as *well*, *you know*, *sort of* and *kind of*, amongst other things. It has been noted that items such as these are perceived as connected to fluency (see Hasselgren 2002; McCarthy 2010). Patterns of fossilization of pragmatic markers in learner language have been discussed by Romero-Trillo (2002), as have distinctive patterns of use in the speech of second language speakers of English (e.g. Fung & Carter 2007; Gilquin 2008) and developmental aspects of their acquisition (e.g. Polat 2011). Other pragmatic features of "conversational grammar" (Rühlemann 2007), such as vague language forms, and politeness and hedging strategies, have suggested new directions for pedagogy, especially in the teaching of speaking. McCarthy (2010) highlights the relationship between perceptions of fluency and how fluency is interactively constructed via aspects of naturally occurring language such as turn-taking, using corpus evidence. Thus corpora would seem to have much to contribute to pragmatics and the study of learner pragmatic competence. However, much

remains to be achieved in the general field of learner interlanguage and what contributions corpus linguistics might make to it.

5 Learner corpora and Second Language Acquisition (SLA)

Research in mainstream SLA has traditionally relied on learner language data elicited in experimental conditions and less so on natural language data, which may explain what might appear on the surface to be a rather surprising lack of direct interaction between SLA and corpus linguistics. Corpus research has given birth to hybrid disciplines such as corpus-based stylistics, semantics, pragmatics, and sociolinguistics (cf. McEnery & Wilson 1996 and the chapters in O’Keeffe & McCarthy 2010) and, as we have shown, has generated a critical mass of applications to language pedagogy. Nonetheless, Granger has noted a tendency for researchers in the fields of SLA and foreign language teaching to maintain “a certain distance” (2009:13). While there had long been an interest in how second and foreign languages are and should be learned (see some early examples cited in Carter and McCarthy 2015), the first empirical studies in SLA date from the 1960s (Ellis and Shintani 2014) – around the time, in fact, when the first computer corpora were being compiled. But though both enterprises started to crystallize around the same time, practical and theoretical interactions between the two are not as well developed.

However, there were some early voices advocating the application of corpus-linguistic approaches in SLA, for example, the idea of applying corpus-based L1 acquisition techniques to L2 acquisition (Rutherford & Thomas 2001), while Granger has consistently suggested learner corpus research may provide a “bridge between the two disciplines” (2009: 13). Myles (2015) also highlights the importance of a bidirectional relationship between the two, advocating wider use of learner corpora in SLA research, as well as learner corpora to be designed and analyzed with SLA theory in mind. One small irony is that SLA researchers often collect large amounts of elicited data from student activities such as paired-tasks, role-plays or discourse completion tasks (DCTs), data which are, in effect, learner corpora, albeit small and limited in scope (Myles 2005) but which are typically the object of analysis using statistical software packages (sometimes accompanied by qualitative analysis) rather than corpus-linguistic tools. Golato (2003) offers an interesting discussion of

Vaughan, E. & M. McCarthy, (forthcoming 2017). Research in Corpora in Language Teaching and Learning. In: Hinkel, E. (ed.) *Handbook of Research in Second Language Teaching and Learning. Volume III*. New York: Routledge, pp. 173-185.

how different data elicitation techniques, including naturalistic data, offer distinct resources for SLA researchers, but it is still clear that corpus-analytical approaches live somewhat in the shadow of statistics-based empirical studies.

Many of the research questions central to SLA are amenable to corpus-based research, and there are potentially a number of *post facto* connections that could be made between studies and findings that emerged while corpora and SLA research were developing along parallel lines. For example, Ellis & Shintani (2014) connect Sinclair's (1991: 110) corpus-derived *idiom principle*, or the observation that language users have available to them "...a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments" with Pawley & Syder's (1983) claim that second language learners' linguistic competence could be characterized as consisting of "a store of familiar collocations" (p. 192) or "memorized chunks" (Ellis & Shintani 2014: 56; see also McCarthy 2010).

A considerable boost to cross-fertilization between corpus linguistics and SLA has come from the English Profile project, mentioned in section 3.1 above. The English Profile sets out to create profiles of typical learners at each level of the CEFR, based on learner corpora gathered worldwide. Needless to say, noting every individual characteristic of every one of the thousands of learners represented in the learner corpora available to the project would be both interminable and impractical, which has prompted researchers instead to develop the notion of *critical features*. These are core features of acquisition which emerge and are controlled by most or all learners at any given level, offering a sort of identikit picture of the interlanguage of the typical learner at each level. Hawkins and Buttery (2010) give the examples of mastery of ditransitive verbs (e.g. *She gave him a present*), which are critical for learners at CEFR level B1 and upwards, while object + complement structures (e.g. *He made everyone angry*) are critical for B2 and upwards – 'upwards' meaning that once the feature is controlled at any given level, it remains actively present in all the levels above that level. Murakami and Alexopoulou (2015), also researchers on the English Profile project, have used learner corpora to great effect to challenge the long-standing SLA doctrine of a universal order of acquisition of grammatical morphemes, suggesting instead, on the basis of their corpus, that L1 influence is a

decisive factor in morpheme acquisition and accuracy. The SLA-oriented research within the English Profile both shows how corpora can be used to address classic SLA issues and also how corpora might support or challenge SLA findings based on other types of data and analysis.

6 Corpora, varieties and variation

There are a number of studies that give an overview of different types of corpora and exemplars of freely or commercially available corpora (for different types, see Hunston 2002: 14-16; for types and exemplars, see Vaughan and O’Keeffe, 2015). One other issue relevant for modelling second language pedagogy is the nature of the language targeted during corpus compilation. This is particularly so in the case of English, with its plurality of varieties in common currency. This goes beyond our discussion of specialized corpora in section 5: specialized corpora have been compiled to flesh out descriptions of particular varieties, but of equal significance is broadening the scope of what are considered to be legitimate *models* of English, a challenge greatly relevant to the practice of second language pedagogy. We refer to the concept of ‘World Englishes’, and Kachru’s (1985) model of describing varieties of English in terms of Inner, Outer and Expanding concentric circles. The Inner Circle refers to countries where English is the primary native language, such as the UK or Australia; the Outer Circle describes former colonies of these Inner Circle countries, where English is used as an official or second language; the Expanding Circle refers to countries where English is not used as a first or second language but is used as a lingua franca, for business, for example. There is much to problematize within these definitions; for example, the multiplicity of native varieties would justify referring to native Englishes (cf. Gilquin 2015), and there are obviously complex issues of contact history, status and prestige for some of what Kachru classes as Inner Circle varieties (Irish English is a good example here). There is also the question of whether English as a lingua franca is properly a variety or simply a set of functional strategies adopted for immediate communicative purposes and as varied as the number of situations in which lingua franca communications occur (for a discussion, see Seidlhofer 2001).

In order to capture the variety of native Englishes for comparative purposes, projects such as the International Corpus of English (ICE) developed. Starting in 1990, the

Vaughan, E. & M. McCarthy, (forthcoming 2017). Research in Corpora in Language Teaching and Learning. In: Hinkel, E. (ed.) *Handbook of Research in Second Language Teaching and Learning. Volume III*. New York: Routledge, pp. 173-185.

ICE corpora have been built according to a common design, making them amenable for direct comparison. There are currently fourteen varieties of English represented in ICE (<http://ice-corpora.net/ICE/INDEX.HTM>). With regard to non-native user corpora (as distinct from learner corpora), at the time of writing, corpora such as the VOICE corpus (see <https://www.univie.ac.at/voice/page/index.php>) and the ELFA corpus (a corpus of spoken academic English in lingua franca situations: see <http://www.helsinki.fi/englanti/elfa/elfacorpus.html>) are providing important additions to the potential for modelling English in international pedagogical contexts.

Variety and variation have been considerably enhanced by the trend in recent years towards smaller, more carefully-targeted corpora. Examples include classroom corpora of the type exploited by Seedhouse (2004) and Walsh (2011), where the manageable size of the corpus allows for close scrutiny using conversation-analytical techniques. Equally important have been business English corpora (Handford 2010), which have led to a new generation of corpus-based business English materials (McCarthy et al. 2009; Handford et al. 2012). What is clear is that there is plenty of room in the arena of corpus-based teaching and learning for all sorts of corpora to play a role, from large-scale native-user corpora, through targeted, specialized corpora, to learner corpora and non-native user corpora, and the field can only be enriched by such variety.

7 Conclusion

There can be no doubt that the corpus revolution in language teaching and learning has had a major impact on thinking about what we teach, how we approach it, and on teaching and learning resources. In the case of English, not only have corpora changed the face of reference works and other materials, they have also increasingly enhanced our understanding of the differences between speaking and writing and generated fresh interest in aspects of SLA such as the development of interlanguage as evidenced in learner corpora, along with a better understanding of social and pragmatic aspects of competence. Alongside these developments there have been healthy debates on the models of English which corpora can offer and a move from the early days of mega-corpora compiled to assist the writing of dictionaries and reference grammars to more varied, smaller and specialized corpora that have underpinned academic, professional and vocational English teaching and given a

Vaughan, E. & M. McCarthy, (forthcoming 2017). Research in Corpora in Language Teaching and Learning. In: Hinkel, E. (ed.) *Handbook of Research in Second Language Teaching and Learning. Volume III*. New York: Routledge, pp. 173-185.

new impetus to the study of the great variety of English used in the contemporary world.

The future will see more and more multi-modal corpora (Knight 2011), where audio, video and transcript can be viewed simultaneously on-screen, offering greater potential for the understanding of language in its contexts of use and consequent benefits for language pedagogy. Learner corpora will also undoubtedly prove more and more influential in language testing and assessment (see the discussion in Barker 2010). We can also expect huge spin-offs from the data accumulated in blended and online learning programs, which will give us massive corpora of learner production that can be exploited both in the general senses discussed in this chapter and also in technology-led developments such as adaptive learning (San Pedro & Baker 2016) and automated assessment, as exemplified in the work of Buttery and her associates (see <http://apc38.user.srcf.net/research/#alta>). We can safely conclude that in research in language teaching and learning, corpora are here to stay.

References

- Barker, F. (2010) How can corpora be used in language testing? In O’Keeffe, A. & McCarthy, M. J. (eds.) *The Routledge Handbook of Corpus Linguistics*. Abingdon, Oxon: Routledge, 633-645.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999) *The Longman Grammar of Spoken and Written English*. London: Longman.
- Biber, D., Conrad, S. & Cortes, V. (2004) *If you look at ...: Lexical Bundles in University Teaching and Textbooks*. *Applied Linguistics* 25 (3): 371-405.
- Biber, D., Conrad, S. & Leech, G. (2002) *Longman Student Grammar of Spoken and Written English*. Harlow: Pearson Education Limited.
- Biber, D. & Reppen, R. (2015) (eds.) *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press.
- Vaughan, E. & M. McCarthy, (forthcoming 2017). Research in Corpora in Language Teaching and Learning. In: Hinkel, E. (ed.) *Handbook of Research in Second Language Teaching and Learning. Volume III*. New York: Routledge, pp. 173-185.

- Blum-Kulka, S. & Kasper, G. (1993) *Interlanguage Pragmatics*. New York: Oxford University Press.
- Carter, R. A., Hughes, R. & McCarthy, M. J. (2000) *Exploring Grammar in Context*. Cambridge: Cambridge University Press.
- Carter, R. A. & McCarthy, M. J. (1995) Grammar and the spoken language. *Applied Linguistics* 16 (2): 141-58.
- Carter, R. A. & McCarthy, M.J. (1997) *Exploring Spoken English*. Cambridge: Cambridge University Press.
- Carter, R. A. & McCarthy, M.J. (2002) From conversation to corpus: A dual analysis of a broadcast political interview. In Sánchez-Macarro, A. (ed.) *Windows on the world: Media discourse in English*. Valencia: University of Valencia Press, 15-39.
- Carter, R. A. & McCarthy, M. J. (2006) *Cambridge Grammar of English*. Cambridge: Cambridge University Press.
- Carter, R. A. & McCarthy, M. J. (2015) Spoken grammar: where are we and where are we going? *Applied Linguistics*. First published online January 30, 2015. doi:10.1093/applin/amu080.
- Carter, R. A., McCarthy, M. J., Mark, G. & O’Keeffe, A. (2011a) *English Grammar Today*. Cambridge: Cambridge University Press.
- Carter, R. A., McCarthy, M. J., Mark, G. & O’Keeffe, A. (2011b) *English Grammar Today. Workbook*. Cambridge: Cambridge University Press.
- Celce-Murcia, M. (1991) Grammar pedagogy in second and foreign language teaching. *TESOL Quarterly* 25 (3): 459-480.
- Celce-Murcia, M. & Olshtain, E. (2001) *Discourse and context in language teaching. A guide for language teachers*. Cambridge: Cambridge University Press.
- Cheng, W., Greaves, C., Sinclair, J. McH. & Warren, M. (2009) Uncovering the extent of the phraseological tendency: towards a systematic analysis of Concgrams. *Applied Linguistics* 30 (2): 236-252.
- Vaughan, E. & M. McCarthy, (forthcoming 2017). Research in Corpora in Language Teaching and Learning. In: Hinkel, E. (ed.) *Handbook of Research in Second Language Teaching and Learning. Volume III*. New York: Routledge, pp. 173-185.

Clancy, B. & McCarthy, M.J. (2015) Co-constructed turn-taking. In Aijmer, K. & Rühlemann, C. (eds.) *Corpus Pragmatics*. Cambridge: Cambridge University Press, 430-453.

COBUILD (1990) *Collins COBUILD English Grammar*. London: HarperCollins Publishers.

COBUILD (2010) *Collins COBUILD Essential English Dictionary*. Glasgow: HarperCollins Publishers.

COBUILD (2013a) *Work on Your Vocabulary*. [five volumes covering A1 to C1 CEFR levels] London: HarperCollins Publishers.

COBUILD (2013b) *Collins COBUILD Key Words*. [series of volumes on special types of English vocabulary. See http://www.collins.co.uk/category/English+Language+Teaching/English+for+Specific+Purposes/Collins+COBUILD+Key+Words#esp_pg=1.

Conrad, S. (2004) Corpus linguistics, language variation, and language teaching. In J. Sinclair (ed.) *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins, 67-85.

Ellis, R. & Shintani, N. (2014) *Exploring Language Pedagogy through Second Language Acquisition Research*. London: Routledge.

Fillips McCreary, C. (1991) Review of Collins COBUILD English Course, Level III. *The Modern Language Journal* 75 (2): 248-249

Fung, L. & Carter, R. (2007) Discourse markers and spoken English: Native and learner use in pedagogic settings. *Applied Linguistics* 28 (3): 410-439.

Gilquin, G. (2008) Hesitation markers among EFL learners: Pragmatic deficiency or difference? In Romero Trillo, J. (ed.) *Corpus and Pragmatics: A Mutualistic Détente*. Berlin: Mouton de Gruyter, 119-149.

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

Vaughan, E. & M. McCarthy, (forthcoming 2017). Research in Corpora in Language Teaching and Learning. In: Hinkel, E. (ed.) *Handbook of Research in Second Language Teaching and Learning. Volume III*. New York: Routledge, pp. 173-185.

- Golato, A. (2003) Studying compliment responses: a comparison of DCTs and recordings of naturally occurring talk. *Applied Linguistics* 24 (1): 90-121.
- Goodale, M. (1995) *Collins COBUILD Concordance Samplers: Book 2: Phrasal Verbs*. London: HarperCollins.
- Granger, S. (2009) The contribution of learner corpora to second language acquisition and foreign language teaching: a critical evaluation. In Aijmer, A. (ed.) *Corpora and Language Teaching*. Amsterdam: John Benjamins, 13-32.
- Handford, M. (2010) *The Language of Business Meetings*. Cambridge: Cambridge University Press.
- Handford, M., Lisboa, M., Koester, A. & Pitt, A. (2012) *Business Advantage*. Cambridge: Cambridge University Press.
- Hasselgren, A. (2002) Learner corpora and language testing: Smallwords as markers of learner fluency. In Granger, S., Hung, J. & Petch-Tyson, S. (eds.) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins, 143-174.
- Hawkins, J. & Buttery, P. (2010) Criterial Features in Learner Corpora: Theory and Illustrations. *English Profile Journal*, 1(1): e5. doi:10.1017/S204153621000010.
- Hilliard, A. (2014) Spoken grammar and its role in the English language classroom. *English Teaching Forum* 4: 2-13.
- Hinkel, E. (1997a) Indirectness in L1 and L2 academic writing. *Journal of Pragmatics* 27 (3): 360-386.
- Hinkel, E. (1997b) The past tense and temporal verb meanings in a contextual frame. *TESOL Quarterly* 31 (2): 289-314.
- Hinkel, E. (2002) Teaching grammar in writing classes: Tenses and cohesion. In E. Hinkel and S. Fotos (eds.) *New Perspectives on Grammar Teaching in Second Language Classrooms*. Mahwah, NJ: Lawrence Erlbaum Associates, 233-260.
- Hughes, R. & McCarthy, M. J. (1998) From sentence to discourse: Discourse grammar and English language teaching. *TESOL Quarterly* 32 (2): 263–287.
- Vaughan, E. & M. McCarthy, (forthcoming 2017). Research in Corpora in Language Teaching and Learning. In: Hinkel, E. (ed.) *Handbook of Research in Second Language Teaching and Learning. Volume III*. New York: Routledge, pp. 173-185.

- Hughes, R. (2010) What a corpus tells us about grammar teaching materials. In O’Keeffe, A. & McCarthy, M.J. (eds.) *The Routledge Handbook of Corpus Linguistics*. Abingdon, Oxon: Routledge, 401-413.
- Hunston, S. (2002) *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Hunston, S. & Francis, G. (2000) *Pattern Grammar. A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.
- Kachru, B.B. (1985) Standards, codification and sociolinguistic realism: The English language in the Outer Circle. In Quirk, R. & Widdowson, H.G. (eds.) *English in the world: Teaching and learning the language and literatures*. Cambridge: Cambridge University Press, 11-30.
- Knight, D. (2011) *Language, Corpus and Gesture: A Multimodal Corpus Approach*. London: Continuum.
- Lewis, M. (1997) *Implementing the Lexical Approach: Putting Theory Into Practice*. Hove: Language Teaching Publications.
- Martinez, R. & Schmitt, N. (2012) A phrasal expressions list. *Applied Linguistics* 33 (3): 299-320.
- Maskull, B. (1995) *Collins COBUILD Key Words in the Media*. London: Collins Educational.
- Maskull, B. (1996) *Collins COBUILD Key Words in Business*. London: Collins Educational.
- Maskull, B. (1997) *Collins COBUILD Key Words in Science & Technology*. London: Collins Educational.
- Maskull, B. (2010) *Business Vocabulary in Use*. Second Edition. Cambridge: Cambridge University Press.
- McCarten, J. (2010) Corpus-informed course book design. In O’Keeffe, A. & McCarthy, M.J. (eds.) *The Routledge Handbook of Corpus Linguistics*. Abingdon, Oxon: Routledge, 413-427.
- Vaughan, E. & M. McCarthy, (forthcoming 2017). Research in Corpora in Language Teaching and Learning. In: Hinkel, E. (ed.) *Handbook of Research in Second Language Teaching and Learning. Volume III*. New York: Routledge, pp. 173-185.

- McCarthy, M. J. (1991) *Discourse Analysis for Language Teachers*. Cambridge: Cambridge University Press.
- McCarthy, M. J. (1998) *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press.
- McCarthy, M. J. (2008) Accessing and interpreting corpus information in the teacher education context. *Language Teaching* 41 (4): 563-574.
- McCarthy, M.J. (2010) Spoken fluency revisited. *English Profile Journal* 1 e4 doi:10.1017/S2041536210000012
- McCarthy, M. J. & Carter, R. (1995) Spoken Grammar: what is it and how can we teach it? *English Language Teaching Journal* 49 (3): 207-218.
- McCarthy, M. J. & Carter, R. A. (2001) Ten criteria for a spoken grammar. In E. Hinkel and S. Fotos (eds.) *New Perspectives on Grammar Teaching in Second Language Classrooms*. Mahwah, NJ: Lawrence Erlbaum Associates, 51-75.
- McCarthy, M. J. & Carter, R. A. (2002) 'This, that and the other': Multi-word clusters in spoken English as visible patterns of interaction. *Teanga (Yearbook of the Irish Association for Applied Linguistics)* 21: 30-52.
- McCarthy, M.J., McCarten, J. & Sandiford, H., (2005-2006). *Touchstone Student's Books 1, 2, 3 and 4*. Cambridge: Cambridge University Press.
- McCarthy, M. J., McCarten, J. & Sandiford, H. (2012-14). *Viewpoint Student's Books 1 and 2*. Cambridge: Cambridge University Press.
- McCarthy, M. J., McCarten, J., Clark, D. & Clark, R. (2009) *Grammar for Business*. Cambridge: Cambridge University Press.
- McCarthy, M. J. & O'Dell, F. (2012) *English Vocabulary in Use Upper Intermediate*. Third Edition. Cambridge: Cambridge University Press.
- McCarthy, M. J. & O'Dell, F. (2016) *Academic Vocabulary in Use*. Second Edition. Cambridge: Cambridge University Press.
- Vaughan, E. & M. McCarthy, (forthcoming 2017). Research in Corpora in Language Teaching and Learning. In: Hinkel, E. (ed.) *Handbook of Research in Second Language Teaching and Learning. Volume III*. New York: Routledge, pp. 173-185.

- McCarthy, M. J. & O’Keeffe, A. (2010) Historical perspective: what are corpora and how have they evolved? In O’Keeffe, A. & McCarthy, M. J. (eds.) *The Routledge Handbook of Corpus Linguistics*. Abingdon, Oxon: Routledge, 3-13.
- McEnery, T. & Wilson, A. (1996) *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- McEnery, T. & Xiao, R. (2011) What can corpora offer in language teaching and learning. In Hinkel, E. (ed.) *Handbook of Research in Second Language Teaching and Learning, Volume II*. London: Routledge, 364-380.
- Mindt, D. (1996) English corpus linguistics and the foreign language teaching syllabus. In Thomas, J. & Short, M. (eds.) *Using Corpora for Language Research. Studies in Honour of Geoffrey Leech*. London: Longman, 232-247.
- Murakami A. & Alexopoulou, T. (2015) L1 influence on the acquisition order of English grammatical morphemes. A learner corpus study. *Studies in Second Language Acquisition* 0: 1-37.
- Myles, F. (2005) Interlanguage corpora and second language acquisition research. *Second Language Research* 21 (4): 373–391.
- Myles, F. (fc. 2015) Second language acquisition theory and learner corpus research. In Granger, S., Gilquin, G. & Meunier, F. (eds.) *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press.
- O’Keeffe, A. & McCarthy, M.J. (2010) (eds.) *The Routledge Handbook of Corpus Linguistics*. Abingdon, Oxon: Routledge.
- O’Keeffe, A., McCarthy, M.J. & Carter, R. (2007) *From Corpus to Classroom. Language Use and Language Teaching*. Cambridge: Cambridge University Press.
- Parrott, M. (2010) *Grammar for English Language Teachers*. Second Edition. Cambridge: Cambridge University Press.
- Paterson, K., Caygill, C. & Sewell, R. (2012) *A Handbook of Spoken Grammar*. Peaslake, Surrey: Delta Publishing.
- Vaughan, E. & M. McCarthy, (forthcoming 2017). Research in Corpora in Language Teaching and Learning. In: Hinkel, E. (ed.) *Handbook of Research in Second Language Teaching and Learning, Volume III*. New York: Routledge, pp. 173-185.

- Pawley, A. & Syder, F. (1983) Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In Richards, J. & Schmidt, R. (eds.) *Language and Communication*. London: Longman, 191-226.
- Polat, B. (2011) Investigating acquisition of discourse markers through a developmental learner corpus. *Journal of Pragmatics* 43 (15): 3745-3756.
- Römer, U. (2004) A corpus-driven approach to modal auxiliaries and their didactics. In Sinclair, J. McH. (ed.) *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins, 185-199.
- Römer, U. (2005) *Progressives, Patterns, Pedagogy. A Corpus-Driven Approach to English Progressive Forms*. Amsterdam: John Benjamins.
- Romero Trillo, J. (2002) The pragmatic fossilization of discourse markers in non-native speakers of English. *Journal of Pragmatics* 34 (6): 769-784.
- Rühlemann, C. (2007) *Conversation in Context: A Corpus-Driven Approach*. London: Continuum.
- Rühlemann, C. (2009) Discourse presentation in EFL textbooks: A BNC-based study. In Renouf, A. & Kehoe, A. (eds.) *Corpus Linguistics: Refinements and Reassessments*. Amsterdam: Rodopi, 415-435.
- Rundell, M. & Stock, P. (1992) The corpus revolution. *English Today* 8 (3): 21-32.
- Rutherford, W. and Thomas, M. (2001) The child language data exchange system in research on second language acquisition. *Second Language Research* 17: 195–212.
- San Pedro, M. O. Z. & Baker, R. S. (2016) Adaptive Learning. In McCarthy, M. J. (ed.) *The Cambridge Guide to Blended Learning for Language Teaching*. Cambridge: Cambridge University Press, 234-247.
- Schauer, G. A. & Adolphs, S. (2006) Expressions of gratitude in corpus and DCT data: Vocabulary, formulaic sequences and pedagogy. *System* 34 (1): 119-134.
- Seedhouse, P. (2004) *The Interactional Architecture of the Language Classroom: A Conversation Analysis Perspective*. Oxford: Blackwell.
- Vaughan, E. & M. McCarthy, (forthcoming 2017). Research in Corpora in Language Teaching and Learning. In: Hinkel, E. (ed.) *Handbook of Research in Second Language Teaching and Learning. Volume III*. New York: Routledge, pp. 173-185.

Seidlhofer, B. (2001) Closing a conceptual gap: the case for a description of English as a Lingua Franca. *International Journal of Applied Linguistics* 11: 133-158.

Shin, D. & Nation, P. (2008) Beyond single words: the most frequent collocations in spoken English. *ELT Journal* 62 (4): 339-348.

Sinclair, J. (1990) Introduction. In *Collins COBUILD English Grammar*. London: HarperCollins Publishers, v-xi.

Sinclair, J. (1991) *Corpus, Concordance and Collocation*. Oxford: Oxford University Press.

Svartvik, J. (1996) Corpora are becoming mainstream. In Thomas, J. & Short, M. (eds.) *Using Corpora for Language Research*. London: Longman, 3-13.

Swan, M. (2005) *Practical English Usage*. Third Edition. Oxford: Oxford University Press.

Thornbury, S. (2005) *Natural Grammar*. Oxford: Oxford University Press.

Thomas, J. (1983) Cross-cultural pragmatic failure. *Applied Linguistics* 4 (2): 91-112.

Vaughan, E. and O'Keeffe, A. (2015) Corpus analysis. In Tracy, K. (ed.) *The International Encyclopedia of Language and Social Interaction*. Hoboken, NJ: John Wiley & Sons Inc., 1-17.

Vaughan, E. and Clancy, B. (2013) Small corpora and pragmatics. In Roméro-Trillo, J. (ed.) *The Yearbook of Corpus Linguistics and Pragmatics: New Domains and Methodologies*, 1: 53-73.

Walsh S. (2011) *Exploring Classroom Discourse: Language in Action*. London: Routledge.

Willis, J & Willis, D. (1988/1989) *Collins COBUILD English Course*. Parts 1, 2 and 3. London: HarperCollins Publishers.

Willis, D. (1990) *The Lexical Approach*. London: Collins ELT.

Vaughan, E. & M. McCarthy, (forthcoming 2017). Research in Corpora in Language Teaching and Learning. In: Hinkel, E. (ed.) *Handbook of Research in Second Language Teaching and Learning. Volume III*. New York: Routledge, pp. 173-185.

Willis, D. (2003) *Rules, Patterns and Words: Grammar and Lexis in English language Teaching*. Cambridge: Cambridge University Press.

Vaughan, E. & M. McCarthy, (forthcoming 2017). Research in Corpora in Language Teaching and Learning. In: Hinkel, E. (ed.) *Handbook of Research in Second Language Teaching and Learning. Volume III*. New York: Routledge, pp. 173-185.