

6 *Applying Corpus Research Indirectly to Language Teaching Materials and Assessment Development*

Niall Curry, Geraldine Mark, Hyoshin Lee, Tony McEnery, Graham Burton, Tony Clark, and Dongkwang Shin

6.1 Introduction

As a field of research, applied linguistics is decidedly broad, as the chapters of this volume exemplify.* Yet, upon its initial conception in the 1940s, applied linguistics held a much narrower focus and was largely synonymous with research on language teaching and learning (Berns & Matsuda, 2006). In this way, studies in applied linguistics saw a concentration of research dedicated to the identification of effective means to teach language, to apply knowledge of form and meaning in the language classroom, and to use knowledge of linguistic structures to support language acquisition (Voegelin & Harris, 1947) – issues that remain pertinent in language education circles today. Notably, as the field evolved, it shaped and influenced, and was shaped and influenced by, other subfields of education and linguistics, with one such subfield being corpus linguistics.

Amid the corpus revolution of the early 1990s (Rundell & Stock, 1992), corpus linguists and language teachers saw promise in the potential for corpora and corpus linguistics methodologies to solve extant challenges in language education by offering data-driven means for informing and enhancing language teaching and learning. Following technological advances and the development of then state-of-the-art corpus analysis software, researchers sought to inform language classroom practices as well as the resources used by teachers and learners in and out of the classroom. In Leech's (1997) conception, the former constituted direct applications whereby learners engage with corpora in the language classroom and participate in so-called data-driven learning (Johns, 1991; for more on data-driven learning, see

* The research presented herein is part-funded by a number of national and international funded research projects. These include funding from the ESRC, part of UK Research and Innovation, under grant ES/W010615/1; the King Sejong Institute Foundation; and the Cambridge University Press and Assessment Funded Research Programme.

also Apavaloae & Farr, Chapter 4, Crosthwaite & Gazmuri Sanhueza, Chapter 2, and Rees, Chapter 3 in this volume). The latter pertained to indirect applications which involved the use of corpora to develop learner dictionaries, pedagogical grammars, and classroom resources based on examples of attested language use. The general premise supporting such research practices was the view that exposing learners to examples of language in use, as opposed to the likes of idealised coursebook language (Hale et al., 2018), could help them to learn useful and relevant language. This focus remains central to the ongoing corpus revolution (Curry & McEnery, 2025; Chambers, 2019) which has broadened its scope to considerations of intelligent technologies for direct applications as well as corpus use for the indirect development of language coursebooks and language assessments, for example.

Adopting Leech's view of corpus applications, this chapter reflects on indirect applications of corpus linguistics to language teaching materials and assessment development, focusing specifically on research that has engaged with stakeholders and practitioners in order to influence practices. To support this reflection, Section 6.2 discusses relevant research on indirect applications of corpus linguistics, with a closing focus on materials and assessment development. This is followed, in Section 6.3, by a reflective account of corpus applications to language education, drawing on studies by Burton (2012), Curry and Clark (2020); Curry et al. (2022), Curry and Mark (2024; forthcoming), and Lee et al. (2023). To close this chapter, Section 6.4 offers critical reflections on applications and impact based on these studies in order to issue guidance for those wishing to use their corpus linguistics research to inform the practices of various stakeholders involved in language teaching, including teachers, coursebook publishers, and assessment developers.

6.2 Indirect Corpus Applications to Language Teaching and learning

Indirect applications of corpus linguistics to language education gained prominence in the 1990s. At the time, the apparent affordances of corpora for language education led researchers and practitioners to develop practices for informing the resources used by teachers to teach a language (Rundell & Stock, 1992). The corpus revolution in language education came to shape lexicography for language learning, reference materials development, teaching materials development, assessment development, and teacher education. While the revolution

has developed apace, it has not had an equal impact on each of these different yet connected facets of language education (Chambers, 2019; Curry & McEnery, 2025). While areas such as corpus-informed assessment development remain in their infancy, arguably lexicography and learner dictionary development have seen the greatest influence at the hands of corpus linguistics.

The widespread uptake of corpus linguistics for informing lexicography and learner dictionaries is explicable. Prior to the use of corpora, lexicographers largely drew on intuition to divine definitions of meaning in context (Hanks, 2009). However, at the cusp of the corpus revolution, Sinclair's early work on the COBUILD dictionary (1987) set in motion a research agenda that would change learner dictionary development forever. This work led to lexicographical practices becoming informed by everyday corpus techniques – for example, the use of collocation analysis to develop collocation-based dictionaries (e.g., Kjellmer, 1994). The result of these developments was a new approach to dictionary making where information gleaned from corpora was used to add complexity and context to learner dictionary definitions (e.g., Siepmann, 2015).

The second-most well espoused indirect application of corpus linguistics to language education pertains to reference materials. Reference grammars, for example, use corpus data to contextualise language, offer examples of attested language use, and draw distinctions between the language used in different registers and modes. While what followed from the corpus revolution arguably made the greatest contribution to the development of corpus-informed grammars, it is worth noting that the notion of using corpora or corpus-like data predates the field as we know it. In fact, in as early as the late nineteenth century, there is evidence of grammarians using personal corpora to inform their decisions when producing grammars (see, e.g., Burton, 2023 for more on the history of grammar) – a process later made more systematic by Fries (1952). More recent grammars, such as those produced by Biber et al. (1999; 2021) and Carter and McCarthy (2006), exemplify contemporary practices in grammar development and, critically, their emergence afforded those involved in language education the necessary resources to question prescriptive views of language that were embedded in language teaching curricula. Traditionally, for example, many features of spoken language, such as those identified in Carter and McCarthy's spoken grammar, would have been seen as errors or non-standard forms owing simply to their lack of compliance with the norms of (published) written language (Timmis, 2013). Nowadays, the notion of spoken grammar as being distinct from written grammar is widely accepted within the academy.

However, barriers to wider uptake of this notion, such as a focus on so-called standard varieties, remain apparent.

In the context of teacher education, corpus linguistics was initially promoted as a tool for enhancing teachers' language awareness, offering them a resource with which to challenge and potentially validate their assumptions and intuitions (Apavaloae & Farr, Chapter 4 in this volume; Tsui, 2006; O'Keeffe & Farr, 2012; Farr & Leńko-Szymańska, 2024). Over time, corpus linguistics has become central to teacher education, with task-based approaches being used to develop teachers' linguistic knowledge and digital literacies (Frankenberg-Garcia, 2012). However, the perceived benefits of corpus linguistics are often overshadowed by the challenges it poses, such as issues of access, time, interest, and adequate support, leading many educators to question its value (Poole, 2022). This sentiment is reflected across various teaching contexts (e.g., Crosthwaite & Schweinberger, 2021), underscoring the need for innovative strategies to better engage practitioners in corpus research and bridge the gap between research and practice.

Corpus linguistics has also had an indirect influence on the development of teaching materials. McCarthy (2008) differentiates between corpus-based and corpus-informed coursebooks, noting that corpus-based coursebooks rely entirely on corpus data, whereas corpus-informed coursebooks integrate corpus data with a broader range of data sources, such as market data, pedagogical research, and consumer feedback. The majority of coursebooks utilising corpora are corpus-informed (O'Keeffe *et al.*, 2007), with publishers typically leveraging corpus research to highlight useful and frequently occurring features of a language (McCarten, 2012). Despite these advancements, the adoption of corpus linguistics in materials development has faced challenges. Although materials developers have acknowledged some use and understanding of corpus linguistics (Burton, 2012), its application remains limited, as corpora and corpus linguistics are often perceived as inaccessible to practitioners (Ur, 2017). Recent research efforts have involved key stakeholders – including publishers – in the research process to enhance their engagement with relevant corpus linguistic research (Curry *et al.*, 2022; Curry & Mark, 2024, *forthcoming*). However, such studies are relatively rare in the literature. Therefore, while indirect applications of corpus linguistics have significantly influenced materials development, there is still much to be done to better support the vast number of teachers and learners worldwide who rely on such materials to teach and learn languages.

In the field of language assessment, the application of corpus linguistics is arguably limited. Nevertheless, the growing integration of

corpus linguistics in various aspects of language teaching and learning has prompted scholars to advocate for a more cohesive approach that includes assessment (O’Keeffe et al., 2007). Language assessment, which is concerned with ensuring the validity and reliability of tests, must navigate the delicate balance between descriptive and prescriptive views of language. For language produced by learners to be evaluated and classified, it must be measurable (Barker et al., 2015), which necessitates the development of specific criteria for assessment. To address this need, learner corpora are increasingly employed to draw conclusions about how learners typically use language at different stages of their learning journey (Chapelle & Plakans, 2013). This trend has introduced an empirical, data-driven approach to language assessment (Callies & Götz, 2015; Curry & Clark, 2020). Learner corpora, such as the Cambridge Learner Corpus (see Barker et al., 2015) and the Trinity Lancaster Corpus (Gablasova et al., 2019), are instrumental in informing assessment preparation materials (Qin et al., 2016; Gablasova & Brezina, 2018), in shaping localised test preparation (Curry & Clark, 2020), and in developing test items that are appropriately levelled (Barker et al., 2015). However, the convergence of assessment and corpus linguistics has progressed more slowly compared to other indirect applications, likely due to fundamental epistemological differences between the two fields (Schissel, 2023).

It is to these final two areas that the research applications discussed in this chapter pertain. The following sections build on this literature and demonstrate how, through a collection of work, the authors of this chapter have sought to advance the application of corpus linguistics indirectly. The goal of these sections is to illustrate the potential affordances of corpora for language education and the need to work with stakeholders to affect change.

6.3 Case Studies

This section focuses on the authors’ experiences in using corpora indirectly to inform language assessment and teaching materials development. Section 6.3.1 reflects on the contributions of learner corpora to language assessment, beginning with a reflection on a study by Curry and Clark (2020) which utilised the Cambridge Learner Corpus to examine learner language in order to support those undertaking language assessment. This is followed by a detailed discussion of a study by Lee et al. (2023), in which corpus and standard-setting approaches were combined to support the development of a Korean language test for learners of Korean. In Section 6.3.2, the discussion shifts to the role of corpus linguistics in the development of language

teaching materials. The section first reflects on Burton's (2012) research with materials developers, before addressing Curry et al.'s (2022) and Curry and Mark's (2024, forthcoming) work with teachers, publishers, and assessment developers.

6.3.1 Indirect Corpus Applications to Language Assessment

In Curry and Clark (2020), corpus approaches are used to study spelling errors made by Arabic-speaking learners of English who undertook the Cambridge Preliminary English B1 examination. This research, set within the broader field of corpus linguistics and language assessment and conducted in conjunction with assessment developers at Cambridge University Press and Assessment, examines the error-annotated component of the Cambridge Learner Corpus. The study is motivated by well-evidenced challenges facing Arabic-speaking learners of English regarding difficulties with spelling in English (e.g., Allaith & Joshi, 2011; Al-Zuoud & Kabilan, 2013). The primary goal was to identify and analyse spelling errors across different subcorpora of Arabic-speaking learners of English, comparing them with non-Arabic speakers to determine patterns and implications for language assessment in Middle Eastern and North African (MENA) contexts. The study sought to provide insights into how the first language of these learners influences their English spelling, with the ultimate aim of localising assessment for learners in MENA contexts.

The study focuses on the Cambridge Learner Corpus' 30-million-word error-annotated component. This corpus includes examination scripts from students who have taken Cambridge English exams, with a subset of the data focusing on the B1 Preliminary exam. The analysis compares five subcorpora of learners who took the B1 Preliminary examination: learners from Libya, the United Arab Emirates, other Arabic-speaking regions, a combined corpus of all Arabic speakers, and a reference corpus of non-Arabic speakers composed of B1 Preliminary examinations taken by learners of every other first-language background represented in the corpus, which accounts for more than 130 language backgrounds.

In analysing the subcorpora, the study employed both quantitative and qualitative corpus techniques. First, the most frequent errors in each subcorpus were identified through the use of the error annotations (see Nicholls, 2003). The frequencies and types of spelling error were identified and computed for each subcorpus, categorising errors by their part of speech. To offer a more in-depth and qualitative perspective on errors produced by learners in this Preliminary B1 examination, two specific case studies were presented: one focusing

on commonly misspelled verbs and another on words beginning with 'wh'/'w'. Dispersion tests (range and Juilland's D) were applied to ensure that the spelling errors were not skewed by uneven dispersion across the texts, and confidence intervals were calculated to compare the significance of these errors across the different subcorpora. The dispersion tests allowed for the identification of outliers skewing the data of which there were none.

The findings demonstrate that spelling errors are significant among Arabic-speaking learners, particularly those from Libya and the United Arab Emirates. These learners, especially those aged between twenty-one and thirty, showed a higher frequency of spelling errors compared to their non-Arabic-speaking counterparts. The analysis identified a range of commonly misspelled words, including nouns, adjectives, adverbs, and verbs, accessed through the analysis of parts of speech and recurring word-form errors. The study also highlighted specific challenges related to phonological differences between English and Arabic, such as issues with orthographic representation of diphthongs, schwa sounds, and silent letters, which contribute to the high incidence of spelling errors for a range of word types. The two case studies provided further qualitative insights, revealing that errors related to the spelling of verbs and words beginning with 'wh'/'w' were particularly prevalent.

Overall, the study's findings had important implications for language assessment, particularly in the context of developing reflexive assessment support for Arabic-speaking learners. While learners are not penalised for spelling errors in the Preliminary B1 examination, issues of intelligibility may impact their capacity to perform well in their studies and assessments and communicate effectively through their writing. The identification of common spelling errors and their phonological roots suggests a need to reconsider how spelling is addressed in examination preparation for learners from this region. The research highlights the potential for localised assessment practices that account for first-language interference, thereby offering more personalised preparation materials designed to limit the likelihood of spelling errors that may be unintelligible. Specifically, Curry and Clark note the potential value of using corpus linguistics insights to support such a practice and they encourage the provision of targeted pedagogical support to help learners from this region avoid these common pitfalls.

Drawing on this experience of working with assessment bodies on corpus linguistics research, Lee et al. (2023) present a study of Korean learner language, using examination scripts from the Sejong Korean Language Assessment (SKA). The study was conducted in conjunction

with colleagues at the King Sejong Institute – a foundation established by the South Korean government that encourages the learning of the Korean language around the world. The study primarily focused on analysing the language proficiency of Korean learners to align the language proficiency levels demonstrated in the SKA test with the Common European Framework of Reference for Languages (CEFR) with a specific focus on the written component. Overall, the project aimed to develop a comprehensive grading system, establish language proficiency standards that are internationally recognised, and enhance the usability of SKA in educational and professional contexts globally.

Given the wide-reaching aims of this project, Lee *et al.*'s (2023) research employed a variety of methods. Initially, the SKA grading system and test items were compared to those aligned to the CEFR to assess the compatibility of the CEFR Framework for SKA and to identify necessary adjustments for alignment. In the second phase of the study, Sketch Engine (Kilgarriff *et al.*, 2014) was used to analyse a corpus of written responses to the SKA by Korean learners. In terms of metadata, the corpus was structured according to test items and for each item, anonymised student IDs, scores, and a range of other demographic information were captured. The corpus was divided into several subcorpora based on different test items and score bands, allowing for a detailed analysis of the language used in various contexts and by learners at different proficiency levels. The corpus analysis combined two primary approaches: keyword analysis and sentential grammar analysis.

The keyword analysis, computed with Sketch Engine (Kilgarriff *et al.*, 2014) and its simple maths approach, identified key words and phrases in each subcorpus, revealing the vocabulary commonly used by learners at different levels. Keyword analysis involved comparing each subcorpus with the rest of the corpus to identify words that were unusually frequent in certain items and proficiency bands. The sentential grammar analysis focused on the structure of sentences, examining patterns such as the use of nouns, verbs, adjectives, and other grammatical elements. The analysis focused on sentence structures that occurred frequently, allowing the researchers to identify common grammatical patterns associated with different levels of proficiency. This dual approach provided insights into both the lexical and structural aspects of the learners' language use.

The findings revealed a clear progression in language complexity and proficiency as learners moved from lower to higher score bands. In the lower score bands (0–9 and 10–19), learners predominantly used basic vocabulary and simple sentence structures, reflecting an A1 level of proficiency according to the CEFR. As the score bands increased,

there was a noticeable expansion in vocabulary, with learners beginning to use more complex words and phrases typical of higher proficiency levels. By the time learners reached the 60–79 score bands, their language use was consistent with a B1 level, characterised by a broader range of vocabulary and more sophisticated sentence structures. The highest score bands (80–100) demonstrated an advanced vocabulary and complex grammatical constructions, aligning with the C1 level of the CEFR. Notably, the study was unable to align the SKA to the C2 level of the CEFR due to the lower proficiency level of the test-takers. In fact, no Korean language learner reached a C2 proficiency level according to the findings. The study also identified specific variation associated with different test items. For instance, vocabulary related to technology, food, and daily activities was prevalent in test items that engage with those themes, while narrative test items elicited language pertaining to emotions, actions, and temporal aspects. This allowed the analysis to account for the presence of language that is skewed by the test item and identify language more generally used across the test items. The progression in grammatical complexity was similarly evident in test items, with later items showcasing more intricate sentence structures and a greater variety of grammatical elements. Again, this insight was used to inform and contextualise the findings.

While the alignment to the CEFR and corpus analyses offered granular detail on the language of the tests and the learners, expert panels were also used to review and discuss grading standards, sample answers, and the cut-offs of grading bands, which help in determining the proficiency levels that correspond to CEFR levels. These different approaches were brought together to create a manual of standard-setting questions for the SKA writing section which afforded a clear alignment of scores and proficiency bands to the CEFR.

This study was designed in pursuit of a larger goal than that of Curry and Clark (2020) and, as such, utilised a more complex methodology of which corpus approaches were only one part. By combining the outputs of the different approaches employed, the findings of this project clarified the alignment of the SKA to the CEFR. Lee et al. (2023) argue that by establishing this linkage, SKA can gain broader acceptance and reliability as a global standard for assessing Korean language proficiency. As an initial phase of research, the development of a comprehensive rating system and the alignment of SKA with CEFR would appear to enhance the usability of SKA in various contexts; however, the study emphasises the need for continuous improvement and international cooperation to ensure the SKA remains relevant and reliable as a language assessment tool that develops apace with international standards.

The study's corpus analysis, which combines keyword and sentential grammar analysis, demonstrates the potential role of corpus approaches in research on language assessment. This approach allows for a nuanced understanding of learner language, capturing both the lexical richness and grammatical sophistication that characterise different proficiency levels. Building on Curry and Clark (2020), the study also highlights the importance of localised assessment practices that consider the specific linguistic features of Korean learners, contributing to the development of more equitable and effective language assessments.

6.3.2 Indirect Corpus Applications to Language Teaching Materials Development

In the context of materials development, Burton (2012) offers an early critical evaluation of the limited impact of corpus research on materials development in English language teaching contexts. The paper investigates the reasons behind the limited uptake of corpora in this facet of language education by examining the attitudes and practices of coursebook authors regarding the use of corpus data. Burton's study highlights the potential benefits of integrating corpus linguistics into coursebook production but also addresses the challenges faced by authors in doing so.

The study is based on data collected through a survey of thirteen English language teaching coursebook authors. The survey sought to gauge the extent to which these authors had used corpus data in their coursebook writing, the reasons for or against its use, and their general attitudes towards the incorporation of corpus linguistics into English language teaching materials development. It also explored the types of corpora that authors had accessed, the purposes for which they used them, and the challenges they faced in integrating corpus data into their work. In terms of design, the survey afforded a mixed-methods analytical approach, using multiple-choice, Likert-scale, and open-ended questions. Burton (2012) analysed the quantitative data to identify patterns in the authors' use of corpus data, while qualitative responses provided insights into the specific challenges and attitudes that influence their engagement with corpus linguistics research for coursebook production.

The findings revealed that while some coursebook authors had made use of corpus data, this practice was far from widespread. Of the thirteen authors surveyed, eight reported having used a corpus in their writing, with the British National Corpus 1994 (BNC 1994)

being the most commonly referenced corpus. However, the use of corpus data was often limited and not necessarily central to the authors' work. The reasons cited for not using corpora included a lack of expertise, limited access to relevant corpora, and tight writing schedules that did not allow time for corpus consultation. The study also highlighted a significant divide in attitudes towards corpus linguistics. While some authors recognised the potential benefits of using corpora to inform grammatical content and vocabulary selection, others were unconvinced of its relevance, particularly in the context of grammar. Moreover, there was a general consensus that publishers are not actively encouraging the use of corpus data, primarily because of a perceived lack of demand from the market. Publishers, from the coursebook authors' perspectives, are more focused on producing materials that they are confident will sell, which often leads to a conservative approach that favours traditional methods over innovative ones.

Burton's study underscores the complex dynamics that influence the adoption of corpus linguistics in English language teaching coursebook production. One of the key outcomes of the study is its illumination of the barriers to integrating corpus data into coursebooks. These barriers are not merely technical but are deeply rooted in the commercial realities of publishing and the practical constraints faced by coursebook authors. Burton argues that these obstacles should be borne in mind when linguists make claims for the relevance of corpus linguistics research for materials design, as he notes that one should not assume that corpus findings published in academic journals will automatically make their way into published materials. Burton's study also points to the need for more accessible and user-friendly corpus tools that can be easily integrated into the coursebook writing process as well as training for authors who may not currently have the skills or knowledge to effectively use corpus data. Ultimately, Burton's findings suggested that for corpus-based or corpus-informed coursebooks to become more prevalent, there must be a shift in how both educators and publishers perceive the value of corpus linguistics in language teaching. Building on this work by Burton, a number of projects designed to access different stakeholders' perceptions of the affordances of corpus linguistics for coursebook writing were developed by Curry et al. (2022) and Curry and Mark (2024, forthcoming).

Curry et al. (2022) explore how corpus linguistics research, particularly regarding the use of adverbs in spoken British English, was utilised by English language teaching publishers to develop coursebooks. The study analysed differences in adverb usage in the spoken components of BNC 1994 and 2014 to understand how adverb usage

has evolved. The primary goal was to assess whether and how English language teaching publishers, specifically Cambridge University Press, would incorporate these corpus findings into their editorial practices and materials development.

As a multiphase, longitudinal study, the initial corpus analysis offered insight into potential changes in adverb usage over two decades in spoken British English. By focusing on five detailed case studies on specific adverbs ('like', 'so', 'just', 'well', and 'literally'), the study tracked changes in their frequency, syntactic positioning, and functional roles in spoken British conversation. In the second and third phases, the study combined the outputs of the corpus analysis with qualitative interviews and a review of published ELT materials. These interviews aimed to assess the publishers' understanding of corpus linguistics, their previous use of corpus data when making editorial decisions, and their reactions to the specific findings presented. Third, a qualitative review of new coursebooks, including *Talent* and *Evolve*, was conducted to identify whether and how the corpus findings had influenced the production of teaching materials.

The findings revealed that there have been notable shifts in adverb usage in spoken British English over the two decades studied. For example, the adverb 'like' significantly increased in frequency and expanded its functional roles, particularly as a discourse marker and in reported speech. Similarly, 'so' and 'just' also saw increases in usage, with 'so' being increasingly used as a linking adverb. In contrast, 'well' showed a decrease in frequency, although its use as a linking adverb (specifically in the phrase 'as well') increased. The adverb 'literally' exhibited the most dramatic increase, reflecting its rising prominence in casual speech, often used in a non-literal sense for emphasis.

The interviews with publishers revealed a mixed response to the application of these corpus findings. While the publishers acknowledged the value of corpus research for bringing examples of attested language use into coursebooks, they also highlighted practical challenges, such as the need for materials to be accessible to learners and the constraints of aligning corpus-based insights with existing syllabus structures. The review of the published materials showed some integration of corpus findings, particularly in the representation of contemporary spoken English. However, the extent of this integration varied, and the application of corpus insights was often selective rather than systematic.

This study had significant implications for both corpus linguistics and English language teaching materials development. It underscored the potential of corpus research to inform the creation of more up-to-date teaching materials. A direct impact of this study was the use of the

research to inform a range of educational materials. However, the findings also highlight the ongoing challenges in bridging the gap between corpus research and practical materials development. The study calls attention to the need for better integration of corpus findings in English language teaching materials development, suggesting that publishers and materials developers need to collaborate more closely with corpus linguists to fully harness the benefits of corpus research.

By working with publishers, as opposed to independent materials writers, Curry et al. (2022) identified a different set of practices and challenges to those presented in Burton (2012), evidencing that different stakeholders bring with them a range of varying – and at times contradicting – perspectives. To add to this complex landscape, Curry and Mark (2024) sought to gain further insight from those who use coursebooks, that is, teachers.

Despite the advances in corpus applications to language teaching materials development discussed in Burton (2012) and Curry et al. (2022), and more broadly in Section 6.2 of this chapter, corpus research remains under-utilised in mainstream coursebook development, particularly in representing spoken language. In Curry and Mark (2024), teachers were sought as a means to gain further insight into this limited application. The study involved data collected from two workshops with English language teachers, who were asked to engage with corpus-based findings and reflect on their implications for materials development. The workshops included eight participants, four from the English-as-a-second-language context in the UK and four from the English-as-a-foreign-language context in Malta. The teachers represented a range of different teaching backgrounds.

The format of the workshop was designed to elicit open, collective, and individual responses from teachers through a mixed-methods approach. As part of the workshop, teachers were asked to design a lesson plan based on a standard B1-level coursebook speaking task. Following this, they received a presentation on corpus findings related to spoken English, highlighting features of conversational language that are often under-represented in coursebooks – the corpus findings are presented in detail in Curry and Mark (forthcoming). The teachers then revised their lesson plans, incorporating, if they so wished, insights from the corpus research. This process was followed by group discussions and individual surveys to capture the teachers' reflections on the value of corpus linguistics for materials development and teacher education. The data gathered consisted of lesson plans designed by the teachers before and after being exposed to corpus research, recordings of workshop discussions, and responses to a post-workshop survey. Qualitative

approaches were employed to thematically analyse the data. This involved using a grounded theory approach to uncover the teachers' attitudes towards corpus linguistics and its potential applications in English language teaching.

The findings revealed that teachers recognised the value of corpus linguistics in enhancing the representation of spoken language in English language materials. After being exposed to corpus research, the teachers made significant changes to their lesson plans, focusing more on the nuances of spoken language, such as the use of frequent phrases, conversational strategies like interrupting, and the importance of 'small words' that facilitate natural conversation. The study also highlighted a shift in the teachers' perspectives, with many acknowledging the limitations of their previous approaches to language selection, which were often based on intuition rather than empirical evidence. Furthermore, the study found that teachers believed corpus-based insights could play a crucial role in both materials development and teacher education. They expressed a desire for more corpus-informed content in coursebooks, for example, frequency information, particularly in the area of spoken language, and saw corpus research as a valuable tool for professional development. However, the study also identified challenges, such as the perceived complexity of integrating corpus findings into teaching materials and the need for more training to help teachers effectively use corpus tools.

Overall, this study demonstrated once again that in working with different stakeholders, there is potential for both shared and differing perspectives, as materials writers (Burton, 2012), publishers (Curry *et al.*, 2022), and teachers (Curry & Mark, 2024) appear to place a different value on corpus linguistics. For example, frequency information was seen as valuable by teachers and less relevant by publishers. In unpacking teachers' perspectives, Curry and Mark (2024) provide compelling evidence for the inclusion of corpus linguistics in the development of English language teaching materials. By demonstrating that teachers can significantly improve their lesson planning with corpus-based insights, the study supports the argument for a more systematic integration of corpus research into coursebooks and other teaching resources. Nevertheless, this small-scale study is limited in its capacity to impact language teaching more generally. Recognising that teachers can be powerful advocates for the use of corpus linguistics in English language teaching, Curry and Mark (forthcoming) sought to leverage these teachers' voices as a means to engage with publishers and assessment developers in order to advance corpus applications to language teaching and learning.

In Curry and Mark (forthcoming), focus groups and surveys were used to access publishers' and assessment developers' perspectives on the affordances of corpus linguistics for materials development and to compare these perspectives with those of teachers. The focus groups revealed that publishers and assessment developers see research as a fundamental part of their work. Publishers saw the importance of understanding customer needs through market research and of ensuring the effectiveness of pedagogical approaches in the design of tasks. For assessment developers, research spanned a broad continuum, drawing on knowledge and expertise derived from literature on assessment, linguistics, and education. While a broadly shared concept, the notion of research for teachers, publishers, and assessment developers appears to differ in its meaning. Notably, all three of these stakeholder groups saw corpus linguistics research as an important source of knowledge for their practices.

Publishers and assessment developers acknowledged the value of corpus linguistics, particularly for informing lexical and grammatical choices in the development of English language teaching materials. However, there was a notable discussion about the challenges of representation, especially in relation to the idealisation of native speaker models in teaching materials. Publishers emphasised the need to present authentic language models that include a variety of speakers, not just native speakers. In their view, this reflects an effort to be more inclusive in their materials design and, consequently, corpora that are designed to represent standard varieties, for example, British or American English, appeared somewhat limited for them. This is because the many varieties of language they wished to represent were not widely captured in available corpora. For assessment developers, there were more specific challenges, one of which pertains to determining the standards against which language quality can be measured. They discussed the importance of aligning corpus research with the objectives of language courses and assessments, particularly in contexts where English is used as a global language. This underscores the need to balance traditional native speaker norms with the realities of English as an international lingua franca. Yet, the assessment developers acknowledged the practical limitations of this integration in the context of language assessment and, as a proxy, the development of assessment preparation materials.

In presenting the teachers' insights and reflections to publishers and materials developers, the perspectives of all three stakeholders saw a strong degree of alignment regarding the value of corpus linguistics. However, differences emerged in terms of priorities. Teachers were more focused on the practical application of corpus findings to

enhance classroom teaching, particularly in making frequent language more central in English language teaching materials. In contrast, publishers and assessment developers were more concerned with the broader implications of research for consistency, fairness, and accessibility. In view of these contrasting perspectives, the focus groups highlighted some of the barriers to applying corpus linguistics in materials development. These included issues of representation, the complexity of integrating corpus findings into existing curricula, and the cost and effort involved in developing pedagogical corpora that directly inform teaching resources. The findings suggest that a more integrated approach, whereby corpus research is more systematically applied across all stages of materials development in consultation with a range of stakeholders, could bridge the gap between research and practice in English language teaching specifically and language teaching generally.

6.4 Critical Reflections on Application and Impact

The exploration of the indirect applications of corpus linguistics in language education, as detailed in the case studies reviewed in Section 6.3, reveals a pattern that is both promising and problematic. The potential applications of corpus linguistics are a longstanding facet of the corpus revolution (Rundell & Stock, 1992) – a revolution that is still ongoing. By working with various stakeholders, there is scope for informing language assessments and enhancing teaching materials. However, despite these potential applications, the actual impact on mainstream practices in language teaching and assessment remains limited. This dichotomy wherein the potential of corpus linguistics is recognised but not fully realised speaks to a broader issue within applied linguistics: the challenge of translating research into widespread practice.

The case studies presented in Section 6.3 demonstrate varied degrees of impact, depending on the stakeholder, the context, and the goal of the study. For instance, Curry and Clark's (2020) work on spelling errors among Arabic-speaking learners offered insight into this variety of learner language and the challenge learners face when spelling words in English. In their study, they propose a more localised approach to assessment preparation whereby assessment developers for those regions highlight specific challenges faced by learners there. By working directly with the assessment developer, it was possible to share this research internally at Cambridge University Press and Assessment and offer sources for reflection and consideration in the global language assessment domain. In a more ambitious study, Lee

et al. (2023) sought to align national and international assessments developed by a South Korean government institute to the CEFR. Working as part of a wider team of education experts, linguists, and assessment researchers, this study positioned corpora as one facet of a complex research landscape. By taking an interdisciplinary approach and working with the government body dedicated to developing the SKA assessment, it was possible to feed directly into its alignment to the CEFR and its future development. The role of corpora appears more peripheral in assessment development, and this is, in part, owing to the range of considerations involved (e.g., validity, fairness, etc.). Balancing representation and expectation of language as it is used in the real world with idealised (but easily assessed) language sits at the centre of this challenge, and this is an area of applied linguistic research that will require much more investigation.

In the context of materials development, Burton (2012) and Curry et al. (2022) highlight the limited integration of corpus research into English language teaching coursebooks. Insights gleaned from the different stakeholders involved in these studies suggest that while corpus linguistics has the potential to revolutionise language teaching, its impact is often diluted by external factors, such as limited training, market demands, and institutional inertia. By working directly with publishers and assessment developers, it was possible to share insights from corpus research that have been validated by teachers (Curry & Mark, 2024, forthcoming). The accumulation of multiple relevant voices added legitimacy to the corpus research for the key stakeholders who shape materials production (i.e., publishers and assessment developers). It is worth noting that a question emerges as to whether publishers would be equally amenable to discussions and input on widely understood phenomena, such as the role of present perfect in academic writing, or whether there is something about the fuzziness of spoken language in particular that facilitates a more open perspective. Looking forward, investigating a wider array of linguistic phenomena and employing participatory approaches to research design may prove a beneficial means of advancing indirect] applications of corpus linguistics to language education.

The affordances of applying corpus linguistics to language education are clear. Corpus research provides an empirical basis for developing teaching and assessment materials, ensuring that these resources reflect examples of attested language use. By incorporating findings from large-scale corpora, educators, publishers, and assessment developers can create resources that are not only more representative of how language is used in the real world, but also more relevant to learners' needs. As corpus linguistics offers a way to address the evolving nature of

language, particularly in how it adapts to changes across time and contexts, there is an evident value in corpus linguistics for keeping the language being taught relevant and up-to-date – reflections shared by teachers and publishers alike in Curry et al. (2022) and Curry and Mark (2024, forthcoming). For assessment developers, corpus research appears to provide means to develop more valid and reliable tests by grounding assessments and preparation materials in real language use and language from a variety of contexts. This is particularly crucial in a global context, where language assessments must cater to a diverse range of learners and contexts.

Despite the capacity for corpus linguistics research to inform practices in language education, several barriers emerged in the studies discussed in Section 6.3 that appear to hinder the full integration of corpus linguistics into language education. One of the most significant barriers is the accessibility of corpus tools and data. Many educators and materials developers lack the technical skills required to use corpus tools effectively, leading to a reliance on intuition rather than empirical evidence in materials development. Another barrier is the commercial realities of the publishing industry. Publishers often prioritise materials that are guaranteed to sell, which can lead to a preference for traditional, tried-and-tested approaches over innovative, research-based methods. This is compounded by tight production schedules that leave little room for the incorporation of corpus research, which can be time-consuming. Additionally, there is the issue of alignment with existing curricula and standards. While corpus linguistics can offer new insights into language use, integrating these insights into established educational frameworks can be challenging. This is because the descriptive approaches inherent in corpus linguistics remain at odds with more prescriptive perspectives that govern many curricular decisions across the world. A change in thinking, such as that pioneered by research on English as an international language and English as a lingua franca, will be necessary to encourage those involved in designing national language curricula to embrace the oft-described messiness of corpus data.

Working with a range of stakeholders can offer valuable nuance to applied linguistics research. By co-designing research, it is possible to develop corpus projects that directly respond to practitioner needs, making pathways to application and impact clear. By offering feedback on research and data design, guidance on the foci in language research that can be of use to their practice, and legitimacy to the value of research in a wider ecosystem of potential stakeholders, the value of such impact-driven research becomes self-evident. Yet, conducting such research is not always straightforward. At times, it requires

a departure from existing epistemologies and modification in our ontological approaches to research. Moreover, building and sustaining networks of stakeholders with whom such research can be developed is challenging. Reflecting on the case studies discussed in this chapter, we propose some tentative guidelines for working with the various stakeholders involved in materials and assessment development, namely: teachers, publishers, and assessment developers.

When working with teachers, it is critical to ensure a participatory approach in research design. This will help to limit the potential for the observer's paradox by rendering the teachers co-researchers, as opposed to participants. This would also offer recourse for research engagement to act as a form of professional development. Designing research as workshops and training sessions that demystify corpus tools and demonstrate their practical applications in the classroom can be an effective way to facilitate this goal. Additionally, providing teachers with ready-made corpus-informed resources can help bridge the gap between research and practice. However, when working with teachers, we must be conscious of the potential for research fatigue and the limited time available to teachers to participate in research. One core motivator for teachers engaging in the workshops on which Curry and Mark's (2024) study was based was that their perspectives would be shared with publishers. In knowing this from the outset, the teachers saw a value in giving their views as they knew that these would be leading back to those who produce the materials they use in their classrooms every day. It should be noted that teachers' perspectives are also invaluable for enhancing the efficacy of educational research as their input can ensure that the research design addresses contemporary classroom contexts.

For publishers, the key is to balance market demands with the potential benefits of corpus-informed materials. When developing materials at a global scale, this is especially important. This could involve a phased approach where corpus findings are gradually integrated into existing products, allowing publishers to test the use of such research findings without overhauling their entire product lines. Collaboration between corpus linguists and publishers early in the development process can also ensure that research findings are translated into materials that are both pedagogically sound and commercially viable. It is unlikely that those involved in publishing will have the time or resource to read the extensive canon of research dedicated to using corpus linguistics to inform language education. Therefore, by taking a synergetic approach that involves adapting as opposed to abandoning current approaches to materials design (Burton, 2022), we could issue a return to those longstanding collaborative practices

that shaped early research in the field, such as Fries' work on the *Early Modern English Dictionary* in collaboration with Oxford University Press (see Baily, 1980).

For assessment developers, the focus should be on aligning corpus research with the goals of language assessments and taking time to engage with competing research demands that emerge from studies in language assessment. This involves not only using corpus data to inform assessment development but also ensuring that the resulting assessments are fair and reflective of real-world language use. Engaging assessment developers and education specialists in the design and analysis phases can help ensure that the research outputs are relevant and applicable to their specific needs. In this way, using corpora to inform assessment typically requires collaboration with assessment developers and/or the integration of corpus approaches within a wider interdisciplinary research project such as that presented in Lee et al. (2023).

While engaging with different stakeholders will require different pathways to application, there are some overarching principles we can adopt to put our research in the hands of practitioners. First, as researchers who use corpus linguistics, we are often concerned with the granularity of language. Yet, to move this granular insight beyond the academy we must find ways of engaging practitioners without overwhelming them with detail. One way to do this is to be selective. In Curry et al. (2022), for example, there were many documented changes in adverb use. Yet, only five case studies were selected for a detailed presentation to the publishers. The rationale for this was to have a spread of adverbs whose frequency was significantly higher and lower in the 2010s when compared to the 1990s, as well as a mix of -ly adverbs (e.g., literally) and non-ly adverbs (e.g., well). The specific five adverbs selected were also chosen owing to their prevalence in educational materials. In this way, a small number of insights was shared with the stakeholders and those insights were chosen as they were more likely to garner engagement. It would have been possible to select more than five adverbs. However, previous experience of working with publishers guided the selection, as these practitioners, who are concerned with including a range of topics, grammar points, vocabulary, and so on in the resources they develop, would not have been able to include everything. When working with stakeholders and practitioners, a smaller selection of findings can feel more accessible and applicable. With this in mind, it can be beneficial to be selective, as well as to embrace the notion that affecting *some* change is better than affecting no change at all.

The second overarching principle that we can adopt to shape our impact plans in materials and assessment development involves the inclusion of a multitude of voices and perspectives. As researchers speaking to practitioners, we are often assured of the value of our research for enhancing practice. Yet, getting practitioners to engage with and use our research is often challenging. In Lee et al. (2023) and Curry and Mark (2024, forthcoming), both projects involved bringing together different disciplinary perspectives and practitioner voices to highlight shared and diverging concerns. Bringing these voices together produced a synthesis of praxis, whereby both theory and practice from a range of professional contexts converged to offer a complex perspective on a shared issue. Ultimately, this praxis allowed for a more robust presentation of research, legitimated by researchers and practitioners alike. Looking forward, we encourage further interdisciplinary research to engage with a range of stakeholders with a view to advancing the corpus revolution and operationalising the wealth of existing knowledge of language for widespread use in language teaching, materials development, and assessment.

Ultimately, as this chapter and the many chapters in this volume demonstrate, working with stakeholders to design and produce applied and impactful research is an important and worthy facet of the fields of applied linguistics and corpus linguistics alike. Engagement from teachers, publishers, and assessment developers demonstrates the value such stakeholders see in corpus linguistics research. Yet, the process of applying corpus research places it squarely in the real world, a world of competing demands, changing fashions, and inequity. Institutional demands may limit capacity for stakeholder engagement with research, commercial demands may trump what we linguists perceive as a better practice, interdisciplinarity may curtail the potential for corpus linguistics to revolutionise teaching and assessment materials, and issues of access, training, and resource may mean that less affluent education contexts struggle to operationalise corpus research. As a field, we will arguably have much less agency to address and overcome these challenges. Yet, there may be ways in which we can influence global practices in education, nonetheless. Taking inclusive, interdisciplinary, and multi-stakeholder approaches can bolster corpus linguistics research by situating it within a wider ecosystem of education. Doing so will be challenging, but through such work we can support teachers and learners across the world to teach and learn relevant and useful language, improve the representation of a range of language varieties in educational materials and high-stakes examinations, and support the development of an approach to language assessment that embraces descriptive as opposed to prescriptive views of language.

References

- Allaith, Z. A. & Joshi, R. M. (2011). Spelling performance of English consonants among students whose first language is Arabic. *Reading and Writing*, 24(9), 1089–1110. <https://doi.org/10.1007/s11145-010-9294-3>.
- Al-Zuoud, K. M. & Kabilan, M. K. (2013). Investigating Jordanian EFL students' spelling errors at tertiary level. *International Journal of Linguistics*, 5(3), 164. <http://dx.doi.org/10.5296/ijl.v5i3.3932>.
- Bailey, R. W. (1980). Progress toward a dictionary of early modern English, 1475–1700. In Pijnenburg, W. & de Tollenaere, F. (eds.), *Proceedings of the Second International Round-Table Conference on Historical Lexicography*. Foris Publications (pp. 199–226).
- Barker, F., Salamoura, A., & Saville, N. (2015). Learner corpora and language testing. In Granger, S., Gilquin, G., & Meunier, F. (eds.), *The Cambridge Handbook of Learner Corpus Research*. Cambridge Handbooks in Language and Linguistics. Cambridge University Press (pp. 511–534). <https://doi.org/10.1017/CBO9781139649414.023>.
- Berns, M. & Matsuda, K. (2006). Applied linguistics: Overview and history. In Brown, K. (ed.), *Elsevier Encyclopedia of Language and Linguistics*. Elsevier (pp. 395–405).
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman Grammar of Spoken and Written English*. Longman.
- Biber, D., Johansson, S., Leech, G. N., Conrad, S., & Finegan, E. (2021). *Grammar of Spoken and Written English*. John Benjamins. <https://doi.org/10.1075/z.232>.
- Burton, G. (2012). Corpora and coursebooks: Destined to be strangers forever? *Corpora*, 7(1), 91–108. <https://doi.org/10.3366/cor.2012.0019>.
- Burton, G. (2022). Rehabilitating the ELT conditional system. *ELT Journal*, 76(3), 338–347. <https://doi.org/10.1093/elt/ccab064>.
- Burton, G. (2023). *Grammar in ELT and ELT materials: Evaluating its history and current practice*, Vol. 164. Channel View Publications.
- Callies, M. & Götz, S. (2015). Learner corpora in language testing and assessment: Prospects and challenges. In Callies, M. & Götz, S. (eds.), *Learner Corpora in Language Testing and Assessment*. John Benjamins (pp. 1–12). <https://doi.org/10.1075/scl.70>.
- Carter, R. & McCarthy, M. (2006). *Cambridge Grammar of English: A Comprehensive Guide; Spoken and Written English Grammar and Usage*. Cambridge University Press.
- Chambers, A. (2019). Towards the corpus revolution? Bridging the research–practice gap. *Language Teaching*, 52(4), 460–475. <https://doi.org/10.1017/S0261444819000089>.
- Chapelle, C. A. & Plakans, L. (2013). Assessment and testing: Overview. In Chapelle, C. A. (ed.), *The Encyclopedia of Applied Linguistics*. Wiley-Blackwell (pp. 241–244). <https://doi.org/10.1002/9781405198431.wbeal0603>.
- Crosthwaite, P. & Schweinberger, M. (2021). Voices from the periphery: Perceptions of Indonesian primary vs secondary pre-service teacher trainees

- about corpora and data-driven learning in the L2 English classroom. *Applied Corpus Linguistics*, 1(1). <https://doi.org/10.1016/j.acorp.2021.100003>.
- Curry, N. & Clark, T. (2020). Corpus-informed evaluation of examination criteria for MENA. In McCallum, L., & Coombe, C. (eds.), *The Assessment of L2 Written English across the MENA Region: A Synthesis of Practice*. Palgrave Macmillan (pp. 359–392). https://doi.org/10.1007/978-3-030-53254-3_15.
- Curry, N., Love, R., & Goodman, O. (2022). Adverbs on the move: Investigating publisher application of corpus research on recent language change to ELT coursebook development. *Corpora*, 17(1), 1–38. <https://doi.org/10.3366/cor.2022.0233>.
- Curry, N. & Mark, G. (2024). Using corpus linguistics in materials development and teacher education. *Second Language Teacher Education*, 2(2), 187–208. <https://doi.org/10.1558/slte.25727>.
- Curry, N. and Mark, G. (forthcoming). ‘Applications of corpus linguistics in language education: teacher, editor, and assessment developer perspectives’. *Research Notes*. In Press.
- Curry, N. & McEnery, T. (2025). *Corpus Linguistics for Language Teaching and Learning: A Research Agenda*. Cambridge University Press.
- Curry, N. & Riordan, E. (2021). Intelligent CALL systems for writing development: Investigating the use of Write and Improve for developing written language and writing skill. In Kelch, K., Byun, P., Cervantes, S., & Safavi, S. (eds.), *CALL Theory Applications for Online TESOL Education*. IGI Global (pp. 252–273). <https://doi.org/10.4018/978-1-7998-6609-1.ch011>.
- Farr, F. & Leńko-Szymańska, A. (2024). Corpus linguistics in second language teacher education. *Second Language Teacher Education*, 2(2), 117–132. <https://doi.org/10.1558/slte.28536>.
- Frankenberg-Garcia, A. (2012). Raising teachers’ awareness of corpora. *Language Teaching*, 45(4), 475–489. <https://doi.org/10.1017/S0261444810000480>.
- Fries, C. (1952). *The Structure of English*. Harcourt, Brace & World.
- Gablasova, D. & Brezina, V. (2018). Disagreement in L2 spoken English: From learner corpus research to corpus-based teaching materials. In Brezina, V. & Flowerdew, L. (eds.), *Learner Corpus Research: New Perspectives and Applications*. Bloomsbury (pp. 69–89).
- Gablasova, D., Brezina, V., & McEnery, T. (2019). The Trinity Lancaster Corpus: Development, description and application. *International Journal of Learner Corpus Research*, 5(2), 126–158. <https://doi.org/10.1075/ijlcr.19001.gab>.
- Hale, C. C., Nanni, A., & Hooper, D. (2018). Conversation analysis in language teacher education: An approach for reflection through action research. *Hacettepe University Journal of Education*, 33, 54–71. <https://doi.org/10.16986/HUJE.2018038796>.
- Hanks, P. (2009). The impact of corpora on dictionaries. In Baker, P. (ed.), *Contemporary Corpus Linguistics*. Continuum (pp. 214–236).
- Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. In Johns, T. & King, P. (eds.), *Classroom Concordancing*. University of Birmingham (pp. 1–16).

- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1, 7–36. <https://doi.org/10.1007/s40607-014-0009-9>.
- Kjellmer, G. (1994). *A Dictionary of English Collocations: Based on the Brown Corpus*. Clarendon Press.
- Lee, H., Shin, D.-G., Lee, Y.-S., Won, M.-J., McEnery, T., & Curry, N. (2023). *Sejong Korean Language Assessment (SKA) International Language Proficiency Standards Research*. King Sejong Institute Foundation.
- Leech, G. N. (1997). Teaching and language corpora: A convergence. In Wichmann, A., Fligelstone, S., McEnery, T., & Knowles, G. (eds.), *Teaching and Language Corpora*. Longman (pp. 1–23).
- McCarten, J. (2012). Corpus-informed course book design. In O’Keeffe, A. & McCarthy, M. (eds.), *The Routledge Handbook of Corpus Linguistics*. Routledge (pp. 413–427). <https://doi.org/10.4324/9780367076399>.
- McCarthy, M. (2008). Accessing and interpreting corpus information in the teacher education context. *Language Teaching*, 41(4), 563–574. <https://doi.org/10.1017/S0261444808005247>.
- Naismith, B. (2017). Integrating corpus tools on intensive CELTA courses. *ELT Journal*, 71(3), 273–283. <https://doi.org/10.1093/elt/ccw076>.
- Nicholls, D. (2003). The Cambridge Learner Corpus: Error coding and analysis for lexicography and ELT. In Archer, D., Rayson, P., Wilson, A., & McEnery, T. (eds.), *Proceedings of the Corpus Linguistics 2003 Conference*. Corpus Linguistics 16. UCREL, University of Lancaster (pp. 572–581).
- O’Keeffe, A. & Farr, F. (2012). Using language corpora in initial teacher education: Pedagogic issues and practical applications. In Biber, D. & Reppen, R. (eds.), *Corpus linguistics*, Vol. 4. Sage (pp. 335–362).
- O’Keeffe, A., McCarthy, M., & Carter, R. (2007). *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511497650>.
- Poole, R. (2022). ‘Corpus can be tricky’: Revisiting teacher attitudes towards corpus-aided language learning and teaching. *Computer Assisted Language Learning*, 35(7), 1620–1641. <https://doi.org/10.1080/09588221.2020.1825095>.
- Qin, M., Du, X., Tao, J., & Qiu, X. (2016). A study on the optimal English speech level for Chinese listeners in classrooms. *Applied Acoustics*, 104, 5056. <https://doi.org/10.1016/j.apacoust.2015.10.017>.
- Rundell, M. & Stock, P. (1992). The corpus revolution. *English Today*, 8(3), 21–32. <https://doi.org/10.1017/S0266078400006520>.
- Schissel, J. L. (2023). Bias, discrimination, and the social consequences of unproblematized assessments in TESOL. *Tesol Quarterly*, 57(2), 716–721. <https://doi.org/10.1002/tesq.3219>.
- Sinclair, J. (1987). *Collins COBUILD English Language Dictionary*. Collins.
- Siepmann, D. (2015). Dictionaries and spoken language: A corpus-based review of French dictionaries. *International Journal of Lexicography*, 28(2), 139–168. <https://doi.org/10.1093/ijl/ecv006>.

- Timmis, I. (2013). Spoken language research: The applied linguistic challenge. In Tomlinson, B. (eds.), *Applied Linguistics and Materials Development*. Bloomsbury (pp. 79–94).
- Tsui, A. B. M. (2006). What teachers have always wanted to know – and how corpora can help. In Sinclair, J. M. (ed.), *How to Use Corpora in Language Teaching*. John Benjamins (pp. 39–61).
- Ur, P. (2017). Applications of research to materials design. In Hinkel, E. (ed.), *Handbook of Research in Second Language Teaching and Learning*. Routledge (pp. 132–143). <https://doi.org/10.4324/9781315716893-10>.
- Voegelin, C. F. & Harris, Z. S. (1947). The scope of linguistics. *American Anthropologist*, 49(4), 588–600. www.jstor.org/stable/662897.