



THINKING ALLOWED

# Corpus linguistics for language teaching and learning: A research agenda

Niall Curry<sup>1</sup>  and Tony McEnery<sup>2,3</sup> 

<sup>1</sup>Manchester Metropolitan University, Manchester, UK; <sup>2</sup>Lancaster University, Lancaster, UK and <sup>3</sup>Shanghai International Studies University, Shanghai, China

**Corresponding author:** Niall Curry; Email: [n.curry@mmu.ac.uk](mailto:n.curry@mmu.ac.uk)

(Received 11 October 2024; accepted 14 October 2024)

## Abstract

This agenda identifies future research trajectories for the corpus revolution, proposing five specific research tasks designed to explore and advance the application of corpus linguistics in language education. These tasks focus on: (1) contrastive data-driven learning, (2) the development of corpus research for informing national language curricula, (3) the use of artificial intelligence for corpus informed language teaching and learning, (4) the reconsideration of the design and development of pedagogical corpora, and (5) the need for stakeholder engagement with corpus research design. Addressing these tasks requires a unified and collaborative effort as they sit at a number of key intersections in corpus applications to language pedagogy. We ask scholars executing them to engage in broad and cooperative research to meet the evolving needs of learners globally, to examine the potential of corpus linguistics for addressing new challenges in language education, and to influence and shape future directions in applied linguistics.

## 1. Introduction

Over the last 40 years, the development of methods in corpus linguistics and theories based on corpus data has brought about a series of paradigm shifts in many subfields of applied linguistics and beyond (Hunston, 2022). In the context of language teaching and learning, this development was seen to be a part of the so-called ‘corpus revolution’ (Rundell & Stock, 1992) that endeavoured to reshape language teaching and learning by transitioning from traditional, intuition-based approaches to language study and education towards empirical, data-driven methodologies grounded in the analysis of principled, representative collections of language in use, known as corpora (singular: corpus). While the corpus revolution, as we shall see, has indeed brought about major changes and shifts in thinking in the wider field of applied linguistics, arguably the corpus revolution of the 1990s is still ongoing (Chambers, 2019). As such, there remain aspects of language teaching and learning in which corpus linguistics has played a limited role, despite evident affordances for corpus linguistics to enhance and inform pedagogical practices therein (e.g., corpus linguistics and language assessment, (Gablasova et al., 2019)).

As both the fields of corpus linguistics and language teaching and learning continue to evolve and respond to wider societal developments, forging a path for the further integration of corpus linguistics within contemporary language education will require consideration of the synergies between these two fields as well as their points of divergence. This article addresses this need by laying out a research agenda for corpus linguistics and language teaching and learning that offers a path to the advancement and mutual shaping of these long-connected fields of study. Focusing on the direct and indirect applications of corpus linguistics to language teaching and learning, this article explores the multifaceted impact of corpus linguistics on applied linguistics. To this end, Section 2 presents an historical account

and state-of-the-art review of the role of corpus linguistics in data-driven learning, lexicography, materials development, teacher education, and assessment practices.

Building on the state-of-the-art, in the research agenda presented in Section 3, we propose five research activities designed to guide the expansion of the fields of corpus linguistics and language teaching and learning. This agenda focuses on a range of emerging areas of interest, such as plurilingual contexts, the implementation of corpus approaches to teaching and learning in non-tertiary contexts, the intersection of corpus linguistics and language teaching with digital pedagogies, and the integration of artificial intelligence (AI) in corpus analysis and application. Consideration is also given to indirect applications of corpus linguistics to language teaching and learning, identifying avenues for advancing research on lexicography, reference materials, teaching materials, language assessment, and teacher education. Subsequently, future directions in research are highlighted, identifying the potential for participatory approaches to research design and interrogating the universality and global relevance of research on corpus linguistics and language teaching to-date. Finally, a brief conclusion is offered in Section 4.

## 2. The corpus revolution in applied linguistics

The corpus revolution, marked by a movement in the 1990s that significantly transformed language teaching and learning, was encapsulated by the perceived potential of corpus linguistics to help develop materials and resources for language teaching and learning, based on attested examples of language in use (see Wichmann et al., 1997, for a collection of papers that reflects the research in these areas at this time). The initial wave was enabled by advances in computing power and availability, as well as a growing volume of electronic texts in general and corpora of various sorts in particular. This, in turn, spurred the development of methods and applications of corpus linguistics and had a notable impact on shifting language analysis and education from intuition-based methods to data-driven practices. This movement was made possible owing to the emergence of easy to use third-generation concordancers, such as WordSmith Tools (Scott, 2024), which became available as computers moved out of specialised labs and into the office, classroom, and home. More recently, research in this area has seen a renewal of this revolution (Chambers, 2019), propelled by technological strides in corpus and computational approaches to language analysis, technological and pedagogical developments in digital pedagogies, and new avenues for corpus application. Adopting Leech's (1997) long-standing categorisation of corpus applications to teaching and learning, Section 2.1 focuses on the *DIRECT APPLICATIONS* of corpus linguistics where corpora are accessed directly by language learners, allowing hands-on exploration of language patterns. Subsequently, Section 2.2 explores *INDIRECT APPLICATIONS*, showing how corpus insights can inform lexicography, reference materials, materials and assessment development, and teacher education.

### 2.1 Direct applications of corpus linguistics to language teaching and learning

As the primary research area within direct applications of corpus linguistics to language teaching and learning, data-driven learning (DDL) represents a pedagogical approach that integrates the analysis of language corpora directly into the teaching and learning process. The concept of DDL dates back to the early 1990s, emerging from the work of Tim Johns (1991). This approach is grounded in the belief that learners can acquire language more effectively by engaging directly with attested examples of language in use, typically extracted from large, electronically stored corpora. Initially, DDL was applied with a view to facilitating language acquisition through discovery, with teachers aiming to guide learners and to develop their learner autonomy. Through this practice, DDL encouraged learners to investigate features of language (such as collocations) and discover linguistic patterns for themselves. The approach exposed learners to contextualised uses of language in a corpus through the exploration of concordance lines (Pérez-Paredes, 2022). Learning in this way positioned the learner as a researcher (Bernardini, 2004) and afforded them the opportunity to engage critically with texts and 'notice' (Schmidt, 1990) patterns of usage in attested language samples.

DDL began with what we might call a traditional approach, with learners being guided to build corpora and to use corpus analysis software, such as WordSmith Tools (Scott, 2024) or AntConc (Anthony, 2024). In Lee and Swales (2006), for example, learners were taught to build their own specialised corpora of academic research articles. The learners were then encouraged to explore the corpus data with a view to understanding how writers in their fields and disciplinary areas write. Elsewhere, in Braun (2007), the use of ready-made frequency lists, selected concordance lines, and a bespoke concordancer typify the kinds of activities that shaped early DDL research. Often, a goal behind such studies was to evaluate the use of corpora to develop materials, and to directly and indirectly facilitate language acquisition and the development of learning skills (for more on direct and indirect DDL, see also Lusta et al., 2023).

Over the last 20 years, the cannon of research on DDL has grown exponentially and a number of key meta-analyses and systematic literature reviews by Boulton and Cobb (2017), Chen and Flowerdew (2018), and Pérez-Paredes (2022) have sought to document this development and signal directions for advancing the field. Pérez-Paredes (2022) identified concordancing and collocation analysis as being among the most effective DDL approaches reported in research, while Chen and Flowerdew (2018) localise DDL to English for academic purposes (EAP) contexts, offering guidance for both research and practice therein. As the field continues to expand, areas for development abound, including the need for further theorisation of DDL's pedagogical underpinning as well as the need for additional guidance to support teacher and learner uptake (Pérez-Paredes, 2022). Importantly, the need to enhance research practices through the likes of delayed post-testing (Boulton & Cobb, 2017) is another key outcome of these meta-analyses that should inform future work in this area.

These reviews have offered valuable insight into the methodological approaches employed across the spectrum of DDL research. For example, to access the affordances and shortcomings of DDL, many studies in this area employ mixed-methods approaches, combining observations, surveys, reflections, and other data collection and analytical tools that offer a comprehensive understanding of the operationalisation of DDL. While many studies make use of such approaches to underscore the long-standing affordances of DDL for language teaching and learning, such a complex array of methodologies also serves to identify the challenges facing the field. These include the need for specialised training for the use of corpus software – training that most teachers did not and do not have – and the limited application of DDL beyond university settings and beyond university departments that house experts in corpus linguistics.

Responding to the emergent challenges in the field, contemporary DDL has evolved in conjunction with parallel fields, such as computer-assisted language learning and digital pedagogies, to include the use of a range of corpus-based technologies. In many such cases, DDL technologies have been designed to facilitate learners' engagement with corpora without them needing to master the use of corpus consultation software. The likes of ColloCaid (Frankenberg-Garcia et al., 2019) has been developed to enhance learners' use of collocations in academic texts by offering a feedforward tool, based on several different academic corpora, including the Oxford Corpus of Academic English. In using ColloCaid, learners can engage in DDL without searching for specific language items or lexicogrammatical features, as the tool presents learners with potential collocations based on their written texts. Other tools, like Write & Improve (Write & Improve, 2024) demonstrate how corpora can be accessed by learners through receiving corpus-based corrective feedback on their writing (Wali & Huijser, 2018). This tool can be highly motivating and can facilitate noticing (Schmidt, 1990) by helping learners to analyse their own texts and identify potential errors, based on error-tagged corpus data.

As well as technological advances, contemporary DDL is characterised by a desire to implement DDL practices beyond university classrooms and English language teaching contexts. In languages for specific purposes contexts, DDL for professional purposes has centred on the teaching of language for dentistry (Crosthwaite & Cheung, 2019) and military services (Noguera-Díaz & Pérez-Paredes, 2020), *inter alia*. Working with professionals reveals further pedagogical challenges in the implementation of DDL, as they can show greater resistance to overcoming the challenges associated with the approach, *i.e.*, challenges in software acquisition, resistance to learning to use the software, and

resistance to undertaking training to use the approach, without having a clear understanding of the advantages of doing so. Yet, where challenges are overcome, results can be positive. For example, when working with primary and secondary school teachers, Crosthwaite and Schweinberger (2021) found that the latter saw advantages in using corpora for teaching language, despite the evident and sustained challenges they encountered when attempting to use corpus software. On the other hand, Crosthwaite and Schweinberger (2021) also found that primary school teachers showed markedly less interest in adopting DDL. Driven by perceptions of relevance, value, and difficulty, the uptake of DDL in professional contexts is, thus, mixed. A further recent notable development in DDL research has been its use for teaching languages other than English. This includes foci on Spanish (Yao, 2019), Mandarin Chinese (Smith, 2011), and Arabic (Harmain, 2010), for example. In each case, studies draw on the body work developed on the English language and carry with them many of the same challenges that are intertwined in the fabric of DDL research and practice. Such work is further complicated by the relatively limited availability of diverse corpora in languages other than English, the capacity for a comparable array of language research technology to accurately process languages other than English, and the quality, precision, and nuance of language taggers in languages other than English.

Overall, we would argue that the field of DDL is at a pivotal point, with its potential to significantly impact language education widely established through its emphasis on exploratory learning enabled by technological advancements; yet, its wide uptake is still to occur. Recent publications, such as Crosthwaite (2024), have sought to bridge this research–practice gap in DDL by developing a comprehensive overview of the current state of the art of DDL and its relevance in different teaching contexts. Likewise, Viana (2022) endeavoured to ground DDL by producing an edited collection with 70 chapters offering classroom-tested examples of applications of corpora to language teaching and learning. The future vision of DDL in such works is an approach that promises to offer more personalised, effective, and engaging language learning experiences. Yet, set against this perception of the approach are a number of lacunae in DDL research that have the capacity to hamper its adoption. For example, while research on DDL in languages other than English has advanced in recent years, it is still in its infancy. Therefore, there is a need to develop research not only to enhance the multilingual profile of DDL research, but also to respond to calls for plurilingual approaches to language teaching and learning (Curry, 2022) and to tease out typological challenges that the approach may encounter. Moreover, as the field develops, there is a need to learn from advances in parallel fields, such as digital pedagogies, which can support DDL’s need to develop digital literacies, engage with advances in artificial intelligence, and support language teaching in low-resource contexts (Lin, 2023; Schwartz et al., 2019). Furthermore, as DDL endeavours to expand into non-higher education contexts, there will be a need to identify the most effective means to integrate DDL within larger curricula (Anthony, 2016) and address its limited use with young learners.

Johns (1993, p. 8) noted that DDL reflected the ‘zeitgeist’ of the 1990s. However, much has changed over the last 30 years. As language teaching becomes more complex, centring around twenty-first century skills and communication practices guided by global movements in education, we must critically reflect on the future of DDL in that sphere and its capacity to support the teaching and learning of the ‘big themes’ of language teaching, such as “tenses” or “articles” (Hunston, 2002, p. 184). By addressing current challenges and exploring new directions, DDL can meet its potential as a transformative approach to language education, enriching learners’ understanding and use of language in myriad contexts.

## 2.2 Indirect applications of corpus linguistics to language teaching and learning

Indirect applications of corpus linguistics to language teaching and learning have evolved significantly since the inception of the corpus revolution in the 1990s. From an early stage, they bridged the research–practice gap and had a profound impact on language education in areas including lexicography, reference materials, teaching materials, assessment development, and teacher education. This

influence of indirect applications of corpus linguistics on language education, while notable in itself, can be seen as but one part of a broader shift towards evidence-based practices in education.

Arguably, of all of these potential areas of indirect application, lexicography has been most influenced by corpus linguistics. This is because, through the use of corpora, lexicographers need not rely solely on their intuition about language when determining meaning in context (Hanks, 2009) nor on the limited evidence that methods like quotation slips can provide.<sup>1</sup> Through the development of the COBUILD dictionary – the first corpus-based and pedagogically motivated dictionary to document language use – Sinclair (Collins COBUILD, 1987) paved the way for integrating corpus linguistics and lexicography. Building on this work, Kjellmer (1994) developed a collocation-based dictionary, and following advances in corpus linguistics and translation studies, bilingual lexicography began to draw on corpus linguistics to inform dictionary compilation (e.g., Herberg et al., 1997). Ultimately, this shift led to more accurate and comprehensive representations of words, with dictionaries moving to include information on collocations, semantic prosody, frequency, different word senses, and the use of words in specific genres and registers (Siepmann, 2015). Moreover, the advent of learner dictionaries (e.g., *Louvain English for academic purposes dictionary* (Granger & Paquot, 2015)) and specialised dictionaries (e.g., medical dictionaries) allowed lexicographers to make use of complex metadata structures and dedicated software to build corpora designed to meet learners' needs. In fact, Sketch Engine (Kilgarriff et al., 2014), a widely used corpus analysis software, has an in-built functionality for dictionary development, reflecting the centrality of corpus linguistics in lexicography, and vice versa. While corpus linguistics has allowed for seemingly exponential development in lexicography, there remain challenges in using corpora to inform dictionaries – challenges that have been amplified in the age of digitisation and artificial intelligence (Nesi, 2024). These challenges are largely linked to the quality and relevance of data, how end-users engage with dictionaries, and the capacity of corpora to ethically reflect and inform language use (Kováříková et al., 2020; Nesi, 2024).

Though paling by comparison to its influence on lexicography, in the development of reference materials (e.g., grammars, Biber et al., 1999, 2021; Fries, 1952), the impact of corpus linguistics is nonetheless well established. For the likes of grammar reference materials, Conrad (2000) notes the potential for corpus linguistics to revolutionise grammar teaching, drawing distinctions between Quirk et al. (1985) and Biber et al. (1999, 2021) in terms of their engagement with corpus data. She notes that for the latter, the corpus data afford better contextualisation, exemplification, and empirically attested identification of grammar in terms of register and mode. This work, and subsequent grammars (e.g., Carter & McCarthy, 2006) have advanced our understanding of language in context, allowing us to move past prescriptive views of appropriateness and accuracy. For example, many of the findings that have informed Carter and McCarthy's spoken grammar would once have been relegated to a category of error when judged by written language norms (Timmis, 2013). However, through corpus analysis, what may have once been viewed as error was instead presented as a distinct element of spoken language grammar. Yet, despite these advances, the propagation of such views beyond the academy remains a challenge, with the likes of native speakerism and prescriptivism sustaining socialised judgements of language and language use (Bouchard, 2020; Cushing & Snell, 2023).

Corpus linguistics has also indirectly impacted on the development of teaching materials. In the context of curriculum design, Flowerdew (1993) presents an innovative study that demonstrates how corpora could be used to select and level language for syllabus development. Further enriching this area, a pioneer in language coursebook development, McCarthy (2008), distinguishes between corpus-based and corpus-informed coursebooks. The former, he notes, pertain to coursebooks based on corpus data, while the latter see corpus data incorporated with a wider selection of other data (e.g., market data) and consumer feedback. Most coursebooks that make use of corpora are corpus-informed (O'Keeffe et al., 2007), with publishers drawing on corpus research to signal useful and frequent language (McCarten, 2012). Textbooks such as *Touchstone* (e.g., McCarthy et al., 2004) and *Evolve* (e.g., Goldstein & Jones, 2019) typify such practices, with evidence of corpus-informed useful language boxes and frequency information presented throughout. In the context of EAP, Swales



and Feak (2012) have demonstrated how corpora can be used to support the development of materials designed to help learners improve their academic writing and Flowerdew (2017) shows how research in this domain has grown, in part by engaging with increasingly complex corpora that respond to the complexity of teaching contexts – for example, the development of corpora of expert and professional language for teaching language for specific and professional purposes.

Such developments notwithstanding, the uptake of corpus linguistics for materials development has proved challenging, as, while materials developers have acknowledged some use and understanding of corpus linguistics (Burton, 2012), this use is typically marginal, as corpora and corpus linguistics are often found to be inaccessible to practitioners (Ur, 2017). This limited engagement with corpora for materials development is arguably most noticed in terms of the representation of spoken language in education materials (Norton & Buchanan, 2022). Spoken language can often be content-heavy in coursebooks when compared with spoken language corpora (Sun & Dang, 2020), with long turns and many nouns, verbs, adverbs, and adjectives making complex clauses – the likes of which are rarely heard in spoken language. Conversely, there is typically little evidence of the interactional features of spoken language, such as repair, in dialogue texts (Hale et al., 2018). Some recent research in this area has brought key stakeholders (i.e., publishers, teachers) into the research process in order to facilitate their engagement with relevant corpus linguistic research (Curry et al., 2022; Curry & Mark, 2024). However, such studies are notably infrequent in the literature. Thus, while indirect applications have advanced greatly in their impact on materials development, there remains much to be done to support the large number of teachers and learners using such materials, globally.

In the context of language assessment, the canon of research on indirect applications of corpus linguistics remains in its infancy. Yet, owing to the increased presence of corpus linguistics in other facets of language teaching and learning, there have been calls from the literature to constructively align this process through engagement with assessment (O’Keeffe et al., 2007). However, concerned with the development of valid and reliable testing, language assessment is finely balanced between descriptive and prescriptive perspectives on language. Ultimately, language produced by learners must be measurable in some way in order to be graded and classified (Barker et al., 2015), which means assessment developers need criteria against which language can be evaluated. Responding to this challenge, learner corpora are increasingly being used to make inferences about learner language use at different stages of learning (Chapelle & Plakans, 2013) and this movement has ushered in a focus on empirical, data-driven perspectives on language use for assessment (Callies & Götz, 2015). Learner corpora based on assessment, for example, the Cambridge Learner Corpus (see Barker et al., 2015), can be used to inform assessment preparation material (Qin et al., 2016), the designing of assessment criteria (Xi, 2017), and the development of appropriately levelled test items (Barker et al., 2015). Generally, however, the intersection of assessment and corpus linguistics has developed more slowly than other indirect applications. This is likely owing to evident epistemological differences in how language is addressed in both assessment and corpus linguistics research (Schissel, 2023). Attempting to bridge this gap in a large-scale study of the Trinity Lancaster Corpus, McEnery et al. (2025) use novel corpus analysis methods to explore conversational structure in spoken language assessment, providing insights that both explain existing assessment outcomes and point towards ways in which evaluation in spoken language assessment could be reconceived at a level above the sentence or utterance.

The last area of indirect applications of corpus linguistics discussed here pertains to teacher education. As the corpus revolution evolved to include the array of direct and indirect applications of corpus linguistics to teaching and learning, there has always been a parallel focus on teacher education. While corpus linguistics applications to teacher education would also feel at home among the direct applications discussed in Section 2.1, we position them indirectly here with the view that teachers who engage with corpora for preparation and reflection can influence their classrooms without necessarily bringing corpora into the space. In its initial conception in teacher education contexts, corpus linguistics was seen as a means to raise teachers’ language awareness, by offering a resource they could use to test their assumptions and intuitions (Farr & Leńko-Szymańska, 2024; O’Keeffe & Farr, 2012; Tsui, 2006). Indeed, in recent years, corpus linguistics has become a mainstay of short-term language

teacher education programmes across the world (e.g., CELTA, Naismith, 2017). Typically, task-based approaches are employed in such education programmes with a view to developing both linguistic knowledge and digital literacies (Frankenberg-Garcia, 2012). This may involve inducting teachers in the affordances of DDL (e.g., Chen et al., 2019; Leńko-Szymańska, 2014), developing corpus literacy in teachers (e.g., Heather & Helt, 2012), and facilitating critical reflections on the impact and effectiveness of corpus-based teacher education interventions (e.g., Curry & Mark, 2024; Leńko-Szymańska, 2014).

The challenges that come with this approach need to be balanced against its perceived benefits for teachers – a similar perspective that emerged in our previous discussion of DDL. Issues of access, time, interest, and support, lead to many teachers seeing limited value in corpus linguistics research (Poole, 2022). This view is echoed in a range of teaching contexts (e.g., Crosthwaite & Schweinberger, 2021), demonstrating a need to develop new ways of engaging practitioners in corpus research to bridge this research—practice gap. Le Foll (2021) presents an innovative solution for engaging trainee teachers in corpus linguistics by working with them to develop an open education resource dedicated to the application of corpus-informed materials in the language classroom. Adopting this approach may be a fruitful means to respond to teacher ambivalence towards corpus use in their language classrooms, especially if broadened in scope to work with experienced as well as trainee teachers (Curry & Mark, 2024).

Overall, the indirect applications of corpus linguistics have significantly enriched language teaching and learning, offering a more empirical, data-driven foundation for lexicography, reference and classroom materials development, assessment development, and teacher education. Continued innovation and research in these areas, particularly addressing issues of the representativeness of corpora, converging and diverging epistemologies within applied linguistics, and the widening contexts of application, will allow the corpus revolution to continue to critically enhance language teaching and learning. However, addressing such an agenda will require sustained collaborative efforts among linguists, teachers, technology developers, and a range of other stakeholders. The following research agenda endeavours to offer an initial pathway for supporting such innovations and advances.

### 3. Research agenda and research tasks

The necessarily brief literature review presented in Section 2 has exposed a number of key themes that will shape the future of corpus linguistics for language teaching and learning. In this section, we propose five research tasks that can offer recourse for addressing these emergent issues.

#### Research task 1. Develop a plurilingual approach to DDL by drawing on research on corpus linguistics, contrastive linguistics, and language pedagogy

As mentioned in Section 2, most research on corpus applications to language teaching and learning centres on the English language. As such, this research often inadvertently reinforces existing perspectives on the roles of different languages in the language classroom, as learners study and engage with their target language only. This focus on the target language can reinforce negative perspectives of the use of the first language in the classroom, for example, viewing it as a contaminant (Creese & Blackledge, 2010). Such negative views continue to emerge surrounding the use of the first language in the language classroom, despite the origins of this demonisation of the first language having limited empirical grounding – the primary critique being that overuse of first language can delay acquisition (Hanif, 2020). However, as research on translanguaging demonstrates, there is ample evidence to support the use of both target and first or other languages in the language classroom (Garcia & Wei, 2014).

For a plurilingual classroom that embraces first and other languages, what becomes important is how, not whether, other languages and linguistic competencies are utilised as teaching and learning resources. Coupled with advances in motivation research and an increased need for specialised language education, as well global movements towards culturally enriched education, leading international bodies have called for the development of plurilingual competencies in contemporary

language education. This call is evidenced in the *Council of Europe's common European framework of reference* (CEFR) companion volume, which explicitly urges teachers to engage with learners' first languages and develop plurilingual competence in the classroom (Council of Europe, 2018). In concert with this recent development, mediation strategies have also been given greater emphasis within the CEFR, with both multilingual and intralingual mediation skills necessary for contemporary learners to navigate their many modes and contexts of language use. Seeing that corpus linguistics endeavours to respond to learners' needs – through the use of personalised, bespoke resources in DDL, for example – the question emerges as to how it can do so while developing much needed plurilingual competencies and mediation strategies.

To address this concern, we argue that there is a need to revisit the potential for contrastive linguistics to inform language education. While contrastive linguistics fell from favour in the 1960s, following its limited success in error prediction (Klein, 1986), the landscape of language teaching in the twenty-first century is substantially different. Contemporary language teaching is concerned with the notion of *PAROLE*, in a Saussurean sense (e.g., Gordon, 2004). As such, issues of accuracy, error, and correction are now less central, as learners have become increasingly concerned with using language for specific purposes to communicate with speakers from across the globe. This focus on specialised and culturally diverse language sits at the core of contemporary contrastive linguistics (Curry, 2023).

Based on Curry (2022), we propose that DDL could benefit from greater engagement with the growing canon of corpus-based contrastive linguistics and seek to support teachers in the use and/or development of multilingual corpora that can act as a reference, point of comparison, and trans-linguaging resource for learners in the classroom and beyond. This would allow teachers to draw on learners' entire linguistic repertoires in the language classroom. Moreover, by using such corpora to facilitate DDL activities in the potentially multilingual classroom, teachers can encourage learners to share reflections on similarities and differences across languages and cultures, thus facilitating cultural exchange (Curry, 2022). We propose, therefore, that developing this line of research is a valuable task that will advance the application of corpus linguistics in language teaching and learning. It can do so by offering a necessary pedagogical underpinning to DDL (Pérez-Paredes, 2022) and by positioning contrastive DDL at the centre of current debates in multilingualism, plurilingual competencies, and translinguaging.

Scholars interested in developing contrastive DDL could conduct a study that tests the proposition that contrastive DDL supports the development of specialised language use, as well as plurilingual and cross-cultural competencies (for more on this, see Curry (2022)). Working with a class of multilingual learners in university or private language school contexts, teachers could support learners in building small and specialised multilingual corpora or using existing resources, provided their design meets the learners' needs. In composing or selecting corpora, it would make sense to select texts that act as a good model for the language their learners are learning. These corpora could be composed of academic texts, marketing texts, the language of advertisements, and so forth. The specificity of the text type is important. However, this specificity should be determined by the learners' needs and learning goals. Ultimately, these small and specialised multilingual corpora should be composed of the learners' target language, that is, the language they want to learn, and a first or other language in which they have expertise. The choice of language is likely to vary from learner to learner but this variation should not stop the lesson design from including group and pair work, as this form of interaction and interthinking is critical for language learning and can further facilitate cultural exchange. In this study, the corpora used could be comparable corpora, composed of comparable texts in each language or parallel corpora, consisting of source texts and their translations.

Upon developing or selecting the corpora, learners should be guided to conduct a range of typical DDL tasks (such as those discussed in Section 2), which should involve analysing lexico-grammatical elements of both their target and first or other language. Learners, for example, could focus on cohesion markers, investigating how they compare across the languages analysed. More advanced facilitators could raise the stakes of the challenge and investigate the affordances of DDL for teaching the so-called big themes of language, such as tenses. These practices should be conducted over the course



of a number of weeks; many studies investigate learner use of DDL from one to as many as 16 weeks (Pérez-Paredes, 2022), for example.

During this period, there are a number of different methodological approaches that can be adopted to support data collection and analysis. In an effort to triangulate this research, researchers could engage in classroom observation and reflection (e.g., Chen & Flowerdew, 2018) or collect learners' perspectives through the use of learning diaries and post-study focus groups (e.g., Jones & Oakey, 2024). These approaches to data collection should be guided by three primary aims: to determine (1) whether contrastive DDL helps learners to learn specialised language effectively; (2) whether contrastive DDL helps learners to better understand both their target and first or other language; and (3) whether contrastive DDL helps learners to develop cross-cultural competencies. To further triangulate this study, contrast and control groups should be analysed (e.g., Muftah, 2023), composed of learners using monolingual DDL in the former, and not using DDL in the latter.

Conducting such a study would offer welcome insight into teachers' and learners' perceptions of the affordances of contrastive DDL for language teaching and learning. It would be important, as part of this investigation, to reflect on the challenges that teachers face in the development of corpora, the advantages and disadvantages of using bespoke corpora, the kinds of textual data needed to facilitate cultural interrogation, reflection, and exchange, and the classroom management strategies and institutional support needed to run such interventions. Crucially, such a study would set DDL further along the path of multilingual research and raise additional areas of concern and development in an understudied facet of applied linguistics.

### **Research task 2. Develop contextually and culturally situated approaches to embedding corpus linguistics in language education: The case of South Korea**

An emerging challenge facing contemporary applications of corpus linguistics to language teaching and learning is closing the gap between research and practice and, more specifically, the implementation of corpus linguistics in non-university contexts. This often involves reconciling corpus linguistics approaches with teaching and learning at a national level. In the Korean context, for example, the implementation of a new national assessment in 2018 (The Korean College Scholastic Ability Test) was met with criticism, owing to its perceived inability to support curriculum development, effective assessment development, and materials development, as well as facilitate meaningful language acquisition (Lee, 2021). The implementation of this new assessment and the redevelopment of the curriculum surrounding it was well-intentioned, responding to growing mental health concerns amid the rampant 'English fever' arising from high stakes examinations (Park, 2009). Yet, for teachers, it appears to be negatively impacting their learners' proficiency, as the assessment requires limited use of communication skills (Lee, 2021). If a goal of corpus linguistics research is to develop best practices for language teaching and learning, one may wonder how corpus linguistics could support language teaching and learning in such a context.

As evidenced in Section 2, alongside developments in DDL, teaching practices, reference materials, teaching materials, and language assessments are increasingly informed by corpus linguistics. While one may think that language is the only contribution of corpora to such aspects of language teaching and learning, Yoon and Jo (2014) have demonstrated that corpora and corpus linguistics can also be used to facilitate the acquisition of metacognitive strategies (e.g., self-evaluation), cognitive strategies (e.g., processing materials), and affective strategies (e.g., strategies for lowering anxiety) – strategies that could directly respond to the aforementioned challenging circumstances surrounding English language education in South Korea. Nevertheless, despite the challenging nature of South Korea's teaching context and the focus on affective learning therein, as well as the evident affordances of corpus linguistics for informing teaching and learning, and developing affective strategies, the South Korean language teaching context has, to date, been relatively untouched by the corpus revolution. This has begun to change, as organisations, such as the Korean Association of Teachers of English, have shared research on corpus linguistics (e.g., Lee (2015) in *English Teaching*), and new journals

have emerged, dedicated to corpus linguistics in South Korea (e.g., *Corpus Linguistics Research*). Amid this change, we can see research demonstrating that learners in South Korea exhibit increased language development through the use of DDL (e.g., Hwang and Cho 2022). However, such studies are limited in number.

Recognising the challenging context in which teaching occurs in South Korea and the growing interest in corpus linguistics therein, we call for research to look closely at the interface between corpus linguistics for language teaching and learning and issues arising from the uniqueness of educational contexts worldwide, and in South Korea in particular. South Korea provides a good context for such a focus as any research agenda that responds to this context must also respond to national needs; in particular, such research must engage with the Korean National Curriculum for English language and address foci on affect and mindfulness therein (Choi, 2021). As the process of language learning itself has been driving significant mental health issues in South Korea, any interventions proposed by corpus linguistic-driven research agendas must be mindful of the demands of South Korea's complex educational history. In addressing such a task, we argue that the field must develop effective means to navigate external, top-down curricula with a view to supporting language learners effectively with corpus-based approaches, both directly and indirectly.

To begin to address this task, a study could centre on the movement from the National English Ability Test to The Korean College Scholastic Ability Test. Taking a corpus approach, the language of the reading and listening sections of both assessments could be studied to identify the primary differences in the language used in the assessments in terms of language complexity (e.g., using type-token ratio (Larsson, 2016)), level (e.g., in terms of the CEFR (McCarthy, 2016)), and register (e.g., by comparing the data with other corpora such as learner corpora, and spoken and written corpora (McEnery et al., 2025)). In so doing, the study could reveal how the language used in reading and listening sections in both assessments correspond to one another. Such an insight could be used to identify potential gaps in language input that could be addressed to solve practitioners' concerns for their learners' language acquisition (Lee, 2021). For this task, context is key and, by investigating national assessments through corpus approaches as a means to feedback into national curricula, such a study could create a roadmap for developing impactful research in corpus linguistics for language teaching and learning. Looking forward, it would be equally valuable to investigate similarly under-served contexts in this way, beyond the South Korean context. Comparative studies of such contexts could also further extend our current understanding of the potential of corpus linguistics applications in primary and secondary contexts, globally. Likewise, collaborative international projects could offer a rich, rigorous, and expedient means of advancing research in this domain while avoiding the inherent challenges in the one teacher/researcher projects that constitute much of the work in this area.

### Research task 3. Critically assess the affordances, both technological and pedagogical, of AI for informing corpus applications to language education through data-driven learning

A key, recurring critique of DDL that emerged in Section 2 is the lack of an underpinning pedagogy to frame its application in classroom contexts (Pérez-Paredes, 2022). While research continues to address this challenge (e.g., Farr & Karlsen, 2022; O'Keeffe, 2021), DDL research is facing a new evolution in the wake of developments in AI (e.g., Crosthwaite & Baisa, 2023; Flowerdew, 2024). The potential of AI for enhancing DDL is centred on its user-friendliness and attractiveness for teachers, for whom technology is often a barrier. AI-based DDL, for example, may involve using generative AI as a concordancer to investigate language generated by the technology (Lim & Wang, 2023). Yet, while generative AI may offer a panacea to many of the practical problems associated with DDL, it is not necessarily a replacement for DDL (Lin, 2023). Notably, AI does not bring with it a pedagogically robust approach nor access to attested examples of language in use (beyond language used by a generative AI technology). Moreover, as research on AI in applied linguistics attests (e.g., Putland et al., 2023), there is potential for AI to produce unreliable analyses and, based on how it has been trained,

(re)produce biases (see Yuan et al., 2024), for a learner centred discussion of issues in using of AI in the ELT classroom and Choi, [2022], for an evaluation of the capacity of AI-powered chatbots in South Korea to reinforce native speakerism). Such issues raise ethical concerns for DDL research as well as wider applications of corpus linguistics to language teaching. The question that emerges therefore is whether corpus linguists and DDL researchers can respond to this imminent proliferation of generative AI in a way that will further advance the field and not hamper it.

To respond to this question, we may look in parallel fields such as digital pedagogy. Research therein has demonstrated the pedagogical affordances and shortcomings of such technology. For example, technology has been found to increase learner engagement (Croxtton, 2014), develop learner autonomy (Godwin-Jones, 2019), motivate learners (Abdelhafez & Abdallah, 2015), and personalise learning (Kerr, 2016). Yet, despite these affordances, there are many challenges involved in the use of technology spanning issues of access (Hockly & Dudeney, 2018) as well as ethics (Sharkey, 2016). In the case of the latter, for example, learners engaging with chatbots – akin to ChatGPT – were found to be building emotional connections with fake online avatars, which, Sharkey notes, could have a detrimental impact on the development of learners' emotional intelligence. Such possibilities have given rise to growing concerns for learners' digital literacy skills (Drigas et al., 2023), with teachers and other education stakeholders seeking means to develop learners' capacity to critically engage with online and digital information.

As a form of digitally enhanced language education, DDL can support the development of digital literacies and embrace the affordances of technology for language acquisition. Already, many of the technologies discussed in Section 2 have been successful in incorporating knowledge of digital pedagogies into their design (e.g., ColloCaid, Write & Improve). With the proliferation of AI, we face a new challenge and as we move forward, it is imperative that pedagogy guide our applications (O'Keeffe, 2021). Bearing this in mind, for our third task, we propose that researchers will need to critically engage with the affordances of generative AI for DDL by ensuring its use is pedagogically and ethically grounded.

To break ground in this area, we propose a study of teachers' and learners' engagement with AI that could be used in the development of a DDL pedagogy that shapes engagement with AI. As the recent proliferation of AI sees teachers and learners using tools like ChatGPT without fully understanding their composition, it would prove invaluable to work with teachers and learners to understand how they perceive AI and interrogate how that perception influences their classroom practices. As teachers and learners are to use corpora and AI in the classroom for this study, we propose the use of AntConc (Anthony, 2024), as this tool has integrated AI functionality. To assess the engagement of teachers and learners with AI and corpus linguistics, we suggest the use of interviews through which teachers and learners could be guided to reflect on their use and understanding of AI and corpora and how these resources help them to teach and learn language. A specific focus could be placed on the affordances of AI and corpora for supporting the development of metacognitive skills (e.g., Mizumoto, 2023). These interviews could be analysed using corpus approaches combined with critical grounded theory and top-down thematic coding (e.g., Curry & Pérez-Paredes, 2023) to develop a layered understanding of teachers' and learners' emerging understandings of AI, its perceived technological affordances for language teaching and learning, and its pedagogical mediation in concert with corpora and DDL.

To underpin this understanding, scholars should draw on wider research in digital pedagogies to critically evaluate whether the underpinning pedagogy supporting the use of AI and corpus linguistics is: (a) grounded in evidence, (b) innovative in its application (e.g., Tsui & Tavares, 2021), and (c) allowing participants to truly learn language and develop language skills. Using this insight, researchers could propose guidelines for practitioners who wish to develop their learners' digital literacies through AI-use embedded in DDL activities in the language classroom. By advancing research on AI and DDL in this way, scholars will open pathways for further pedagogically situated developments in corpus approaches to language teaching and learning. This could include an investigation of the relevance and suitability of the language produced by generative AI tools to act as input for language learners.

#### Research task 4. Investigate user needs to inform the development of pedagogical corpora

Central to any corpus application to language teaching and learning is a corpus. In corpus-based DDL, teachers and learners often make bespoke, small corpora (e.g., Lee & Swales, 2006). However, for wider, indirect applications, for example, materials development and assessment development, large corpora are typically used to make reliable and empirical inferences about language (e.g., Curry et al., 2022; Gablasova et al., 2019). In many cases, such corpora are not specifically designed for pedagogical application, but are large corpora used by researchers in linguistics across the world to inform their research in a range of areas. Therefore, such corpora require pedagogical mediation to be used effectively (Widdowson, 2003). In an effort to carry out this process, Curry et al. (2022) presented a number of insights surrounding spoken language change to publishers and editors of language coursebooks, based on an analysis of the Spoken BNC 2014. The relevance of findings based on national varieties was discussed with the stakeholders, as they attempted to reconcile the research with the needs of their global markets. While the stakeholders found the information gleaned from corpus analyses useful, they noted that, in their practice, they wish to move towards materials based on English as an international language (e.g., Callies et al., 2021). As such, insights into language use in specific countries (e.g., British English) only addressed part of their language research needs. This response from a key stakeholder in global materials development echoes existing critiques of representativeness and representation in a range of indirect applications of corpus linguistics to language education, discussed in Section 2. As such, an important question emerges for the future of corpus linguistics as used by researchers in language teaching and learning. That is, how can we approach the development of pedagogical corpora to ensure their relevance for key language education stakeholders?

The notion of a pedagogical corpus is a somewhat fuzzy concept, though it largely pertains to corpora designed and constructed for pedagogical application. Thus, pedagogical corpora may be topic-driven, built around the kinds of content learners encounter in the wider curriculum (e.g., BACKBONE, Kohn, 2012). Elsewhere, pedagogical corpora are understood as those that contain texts that are used in the classroom, such as coursebooks (e.g., Meunier & Gouverneur, 2009), texts produced by learners, such as assessments (McEnery et al., 2025), or texts that have been mediated for pedagogical purposes (Braun, 2005). All of these perspectives offer valuable guidance for building corpora suitable for informing language education, with pedagogical corpora drawing on learner production and target production to differing degrees. Making clear the remit of a pedagogical corpus and its intended representation will not only serve to highlight the potential applications of pedagogical corpora, but also the potential for comparative studies of learner and target production.

Despite the evident affordances of learner corpora, there has been little critical engagement by practitioners with the centrality of the native speaker in corpora that are then used to inform teaching and learning indirectly. Therefore, as language education has moved away from focusing solely on language varieties in Kachru's inner circle (1990), towards a focus on international usage (Callies et al., 2021; Flowerdew, 2012), there is a need to revisit the notion of pedagogical corpora with a view to critiquing who they represent. This is a matter of social justice in education, as the exclusion of speakers from data informing educational materials risks reinforcing negative perspectives on so-called non-standard varieties in language education (Cushing & Snell, 2023) and the sustained exclusion of groups of learners from the materials they use in their classrooms.

As a key task facing future researchers in corpus linguistics and language teaching and learning, we propose that the concept of a pedagogical corpus be redefined in light of the globalised contexts in which language teaching and resource production take place. Drawing on existing knowledge of corpus construction and pedagogical corpora as well as wider research on English as an international language and languages other than English, future research should specifically reflect on the operationalisation of representativeness and representation in pedagogical corpora to support the development of inclusive and contextually reflexive corpora. Crucially, addressing this research task will benefit all others mentioned here.

A potential approach to undertaking such a task would be to work with stakeholders, such as materials writers (e.g., Burton, 2012), publishers (e.g., Curry et al., 2022), and teachers (e.g., Leńko-Szymańska, 2014) to identify the kinds of language varieties that they would like to represent in the resources developed for language learners. Working with contemporary concepts of community (e.g., superdiverse communities, Li et al., 2021), scholars could initiate a reconsideration of traditional approaches to sampling frame development for representative pedagogical corpora. Iteratively building this sampling frame with stakeholder engagement to capture not only diverse communities, but also a wide range of texts produced by such communities, could serve to meet the needs of stakeholders developing resources for diverse groups of learners. This is a tall order and will require the development of both very large, balanced corpora, and very specialised corpora. It is likely that no one project could address all the needs of the field. However, by beginning to work towards this goal, we can co-construct, as a field, the various resources needed to enhance global representation in education. Revisiting the notion of a pedagogical corpus and redeveloping it for contemporary language teaching and learning contexts would help to advance corpus applications in a wide range of contexts, while also serving to decolonise materials and theoretically enhance foundational concepts in corpus linguistics, such as representativeness.

### Research task 5. Expand stakeholder engagement for research on corpus linguistics and language teaching

The final task we propose relates to research design. In language education, there is a growing concern with the employment of democratic, participatory, and inclusive approaches to research design that do not see teachers and learners as subjects, but as co-researchers helping to shape a project (e.g., Vaughan & Jacquez, 2020). In the space of DDL, working with teachers in this way has proven fruitful (e.g. Crosthwaite & Schweinberger, 2021; Farr & Karlsen, 2022), yet, in more indirect applications, participatory research has made fewer inroads. Curry and Mark (2024) worked with teachers to evaluate the affordances of corpus linguistics research for informing materials development. This approach centred on the teachers, their experiences, and their perspectives, and we worked together in workshops to critique education materials with a view to developing guidelines for enhancing publishing practices. The teachers noted in particular the value they placed on having their perspectives shared with other key stakeholders, such as publishers. In Le Foll (2021), trainee teachers were guided to develop classroom materials that were published as part of an open educational resource. In this way, the trainees became knowledge producers and Le Foll's approach represents an effective implementation of participatory research design. Elsewhere, in Curry et al. (2022), work with publishers has demonstrated that key stakeholders do not always align with regard to their views of the affordances of corpus linguistics. Gray (2016) has addressed the multiplicities at work in the education 'industry' and draws attention to the many differing values and goals that shape globalised approaches to education. Offering a complementary perspective, Jordan and Long (2022) present a critical view of such stakeholders and their engagement with research, reflecting on the neoliberalisation of education and the role of capitalistic interests in guiding decision-making in education.

McCarten (2012) includes mention of publishers and assessment developers in her discussion of corpus applications in teaching and learning, and Burton (2012) notes the potential ambivalence of publishers towards corpus linguistics. Broadly, these stakeholders are noted for their power in influencing education (Thornton, 2004). Yet, they largely appear backgrounded in corpus linguistics research. In those few studies that have engaged with large educational and governmental bodies, what emerges is both a willingness to engage with research, but also a differing frame of reference that shapes the nature of that engagement. Recognising the affordances of working with publishers and assessment developers, there have been calls for further research with these stakeholders (e.g., Rodríguez-Fuentes & Swatek, 2022; Szudarski, 2023) to help advance indirect applications of corpus linguistics to materials and assessment development. Arguably, the complexity of materials production is underscored by its interdisciplinarity and globality. This complexity renders engagement with stakeholders a rich future direction in language education research, generally. Ultimately, if we are to



engage a wider array of stakeholders, we must create a space in which the perspectives of teachers, learners, publishers, and assessment developers co-exist, interact, and reconcile. Therefore, for our final proposed research task, we call for future research to enhance participatory research with key stakeholders in language education beyond teachers and learners. We encourage future researchers to bring these many voices together to support a joint-effort and engagement with language research.

One potential study could investigate the tensions between the use of language research, educational research, and market research in conjunction with research on digital pedagogies and user experience in the development of online language teaching resources. Working with publishers, scholars could investigate how corpus research fits on this wider continuum of research that informs online materials production. This could be achieved through interviews with key stakeholders designed to investigate the kinds of information that inform the decisions they make. Interviews could be analysed using thematic coding via critical grounded theory (Curry & Pérez-Paredes, 2023) and the results should indicate how language research, and specifically corpus research, co-exists in the ecosystem of knowledge and research that underpins materials production in global, print, and online materials publishing. Such a project would not only offer insight into a complex and often obscured facet of language teaching and learning, but could also shed light on the most effective means to engage stakeholders with corpus linguistics research.

The initial phase of this research task is best viewed as the planning stage of a participatory action research approach (Brydon-Miller, 1993; Maguire, 1987). Moving from the engagement with stakeholders and the grounded theory approach to understanding their concerns, the next stage would see action undertaken based on those findings. That action should involve all key stakeholders and they should play a key role in the research design. As part of the participatory action research cycle, the outcomes of the action would then feed into an evaluation of the impact of the actions by all stakeholders that, in turn, would then begin a further cycle of participatory action research, beginning with further planning. This cyclical, incremental approach to a complex problem is, in our view, likely to yield results that better meet the needs of all stakeholders and achieve the goals of this research task by emphasising ‘self-determination, the development of critical consciousness, and positive social change’ (Brydon-Miller & Maguire, 2008, p. 80). A challenge one may face in undertaking such research is that stakeholder engagement may be challenging to achieve. Working with more directly accessible stakeholders, such as teachers, could be a valuable point of departure. While stakeholders like publishers may be interested in research, they are more likely to engage with and apply research that has been legitimated by their core market, that is, teachers. Bringing teachers’ voices to publishers and evidencing the perceived value of such research for them can be a first step in demonstrating to such stakeholders that you understand their industry.

#### 4. Conclusion

This agenda has set out to contextualise the field of corpus linguistics and language teaching and learning with a view to signalling future avenues of research. Drawing on the literature review presented in Section 2, in this research agenda, we propose five research tasks that address both direct and indirect applications of corpus linguistics to language teaching and learning. With this in mind, we offer two closing reflections that we argue should be considered for each task presented herein, both for the betterment of the specific research projects as well as of the wider field.

First, while each task can be addressed individually, we have devised them for this article to support the advancement of the field in several important and complementary directions. The tasks are concerned with the direct and indirect use of corpora, and each one seeks to advance corpus applications, the development of pedagogical corpora, and the design of research projects. Moreover, each one is rooted in application with specific aims to inform and enhance practices in language education. Efforts to effect such change through DDL research have already been documented in Boulton and Vyatkins’s (2021). In their review, they analysed the conclusion sections of DDL papers to explore the recommendations that the authors made. They then investigated whether the recommendations

made were put into action and found that little happens to transform such recommendations into practice. Accordingly, much contemporary research on DDL remains part of academia's 'bounty of research [that has] such little impact' (Hattie, 2008, p. 3). With this cautionary tale in mind, we recognise that advancing corpus approaches to language education in the many ways discussed in this article is an ambitious agenda. Nonetheless, we argue that as we move forward to further develop our field, scholars must work broadly and collaboratively to address the future needs of learners across the world and interrogate the potential of corpus linguistics to respond to emergent challenges in language Education. Crucially, the bounty of research must be used.

Second, one may be surprised to note the absence of a specific task designed to develop an underpinning pedagogy for corpus applications to language education and, specifically, DDL. The lack of research on the underlying theoretical principles of DDL is an evident lacuna in the field (Boulton & Vyatkins, 2021; Chambers, 2019; O'Keeffe, 2021; Pérez-Paredes, 2022). As many corpus applications to language education were borne of the practices of teachers, materials developers, and assessment developers, we have reached a point in research on direct and indirect applications where the accompanying pedagogies merit greater interrogation and development. Research from second language acquisition, digital pedagogies, metacognition and self-regulation, and language teaching methods and approaches may offer pathways for enhancing the educational value of corpora. No one task can address this issue. Rather, we argue, a pedagogical underpinning should form a facet of any study investigating corpus applications to language education. In this vein, we propose that those interested in testing the affordances of corpus linguistics for language education must continue to interrogate and falsify taken-for-granted views that working with corpus data promotes language awareness and learner autonomy, for example.

Ultimately, attending to the tasks presented by us here will require a coherent and collective effort to build a body of work that can shape the future of applied linguistics. Looking forward, future scholars can benefit from the guidance of these research tasks, while reflecting on the interrelatedness of all areas of corpus application to language pedagogy and, specifically, the centrality of both application and pedagogy in future research.

**Acknowledgements.** We are grateful to the reviewers and editors for their helpful comments that have improved the quality of this article.

**Funding.** This work is supported by the ESRC, part of UK Research and Innovation, under grant ES/W010615/1.

## Note

<sup>1</sup> See Winchester (2003: 97–105, 113–114) for an account of the use of quotation slips in the development of the *Oxford English dictionary*.

## References

- Abdelhafez, H. A., & Abdallah, M. M. S. (2015). Making it 'authentic': Egyptian EFL student teachers' awareness and use of authentic language materials and their learning motivation. *Journal of Research in Curriculum, Instruction and Educational Technology*, 1(1), 1–12. doi:10.21608/jrciet.2015.24564
- Anthony, L. (2016). Introducing corpora and corpus tools into the technical writing classroom through data-driven learning (DDL). In J. Flowerdew & T. Costley (Eds.), *Discipline-specific writing* (pp. 176–194). Routledge. doi:10.4324/9781315519012-18
- Anthony, L. (2024). *Antconc (4.3.1) [Computer software]*. Waseda University. <https://www.laurenceanthony.net/software>
- Barker, F., Salamoura, A., & Saville, N. (2015). Learner corpora and language testing. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 511–534). Cambridge University Press. doi:10.1017/CBO9781139649414.023
- Bernardini, S. (2004). Corpora in the classroom. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 15–36). John Benjamins. doi:10.1075/scl.12.05ber
- 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., Conrad, S., & Finegan, E. (2021). *Grammar of spoken and written English*. John Benjamins. doi:10.1075/z.232

- Bouchard, J. (2020). The resilience of native-speakerism: A realist perspective. In S. A. Houghton & J. Bouchard (Eds.), *Native-speakerism: Its resilience and undoing* (pp. 17–45). Springer. doi:10.1007/978-981-15-5671-5\_2
- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348–393. doi:10.1111/lang.12224
- Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology*, 25(3), 66–89. <http://hdl.handle.net/10125/73450>
- Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents. *ReCALL*, 17(1), 47–64. doi:10.1017/S0958344005000510
- Braun, S. (2007). Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL*, 19(3), 307–328. doi:10.1017/S0958344007000535
- Brydon-Miller, M. (1993). Breaking down barriers: Accessibility self-advocacy in the disabled community. In P. Park, M. Brydon-Miller, B. Hall, & T. Jackson (Eds.), *Voices of change: Participatory research in the United States and Canada* (pp. 125–143). Bergin & Garvey.
- Brydon-Miller, M., & Maguire, P. (2008). Participatory action research: Contributions to the development of practitioner enquiry in education. *Educational Action Research*, 17(1), 79–93. doi:10.1080/09650790802667469
- Burton, G. (2012). Corpora and coursebooks: Destined to be strangers forever? *Corpora*, 7(1), 91–108. doi:10.3366/cor.2012.0019
- Callies, M., & Götz, S. (2015). Learner corpora in language testing and assessment: Prospects and challenges. In M. Callies & S. Götz (Eds.), *Learner corpora in language testing and assessment* (pp. 1–12). John Benjamins. doi:10.1075/scl.70
- Callies, M., Hehner, S., Meer, P., & Westphal, M. (2021). *Glocalising teaching English as an international language*. Routledge. doi:10.4324/9781003090106
- 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. doi:10.1017/S0261444819000089
- Chapelle, C. A., & Plakans, L. (2013). Assessment and testing: Overview. In C. A. Chapelle (Ed.), *The encyclopaedia of applied linguistics* (pp. 241–244). Wiley-Blackwell. doi:10.1002/9781405198431.wbeal0603
- Chen, M., & Flowerdew, J. (2018). A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom. *International Journal of Corpus Linguistics*, 23(3), 335–369. doi:10.1075/ijcl.16130.che
- Chen, M., Flowerdew, J., & Anthony, L. (2019). Introducing in-service English language teachers to data-driven learning for academic writing. *System*, 87, 102148. doi:10.1016/j.system.2019.102148
- Choi, L. J. (2022). Interrogating structural bias in language technology: Focusing on the case of voice chatbots in South Korea. *Sustainability*, 14(20), 13117. doi:10.3390/su142013177
- Choi, T. H. (2021). English fever: Educational policies in globalised Korea, 1981–2018. *History of Education*, 52(4), 670–686. doi:10.1080/0046760X.2020.1858192
- Collins COBUILD. (1987). *English language dictionary*. Editor in Chief: John Sinclair.
- Conrad, S. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century? *TESOL Quarterly*, 34(3), 548–560. doi:10.2307/3587743
- Council of Europe (2018). *The CEFR companion volume with new descriptors*. Language Policy Programme, Education Policy Division, Education Department. Retrieved March 15, 2021, from <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>
- Creese, A., & Blackledge, A. (2010). Translanguaging in the bilingual classroom: A pedagogy for learning and teaching? *Modern Language Journal*, 94(1), 103–115. doi:10.1111/j.1540-4781.2009.00986.x
- Crosthwaite, P. (Ed.) (2024). *Corpora for language learning: Bridging the research-practice divide*. Routledge. doi:10.4324/9781003413301
- Crosthwaite, P., & Baisa, V. (2023). Generative AI and the end of corpus-assisted data-driven learning? Not so fast!. *Applied Corpus Linguistics*, 3(3), 1–4. doi: 10.1016/j.acorp.2023.100066.
- Crosthwaite, P., & Cheung, L. (2019). *Learning the language of dentistry: Disciplinary corpora in the teaching of English for specific academic purposes*. (Vol. 93). John Benjamins Publishing Company. doi:10.1075/scl.93
- 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), 1–13. doi: 10.1016/j.acorp.2021.100003
- Croxton, R. A. (2014). The role of interactivity in student satisfaction and persistence in online learning. *MERLOT Journal of Online Learning and Teaching*, 10(2), 314–325.
- Curry, N. (2022). On contrastive analysis and language pedagogy: Reimagining applications for contemporary English language teaching. In L. McCallum (Ed.), *English language teaching in the European Union: Theory and practice across the region* (pp. 239–256). Springer. doi:10.1007/978-981-19-2152-0\_14
- Curry, N. (2023). Question illocutionary force indicating devices in academic writing: A corpus-pragmatic and contrastive approach to identifying and analysing direct and indirect questions in English, French, and Spanish. *International Journal of Corpus Linguistics*, 28(1), 91–119. doi:10.1075/ijcl.20065.cur

- Curry, N., & Mark, G. (2024). Using corpus linguistics in materials development and teacher education. *Second Language Teacher Education*, 2(2), 187–208. doi:10.1558/slte.25727
- Curry, N., & Pérez-Paredes, P. (2023). Using corpus linguistics and grounded theory to explore EMI stakeholders' discourse. In S. Curle & J. K. H. Pun (Eds.), *Qualitative research methods in English medium instruction for emerging researchers: Theory and case studies of contemporary research* (pp. 45–61). Routledge. doi:10.4324/9781003375531-5.
- 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. doi:10.3366/cor.2022.0233
- Cushing, I., & Snell, J. (2023). Prescriptivism in education: From language ideologies to listening practices. In J. C. Beal, M. Lukač, & R. Straaijer (Eds.), *The Routledge handbook of linguistic prescriptivism* (pp. 194–212). Routledge. doi:10.4324/9781003095125-14
- Drigas, A., Papanastasiou, G., & Skianis, C. (2023). The school of the future: The role of digital technologies, metacognition and emotional intelligence. *International Journal of Emerging Technologies in Learning (Online)*, 18(9), 65. doi:10.3991/ijet.v18i09.38133
- Farr, F., & Karlsen, P. H. (2022). DDL pedagogy, participants, and perspectives. In R. R. Jablonkai, & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (pp. 329–343). Routledge. doi:10.4324/9781003002901-27
- Farr, F., & Leńko-Szymańska, A. (2024). Corpus linguistics in second language teacher education. *Second Language Teacher Education*, 2(2), 117–132. doi:10.1558/slte.28536
- Flowerdew, J. (1993). Concordancing as a tool in course design. *System*, 21(2), 231–244. doi:10.1016/0346-251X(93)90044-H
- Flowerdew, J. (2012). Corpora in language teaching from the perspective of English as an international language. In L. Alsagoff, S. L. McKay, G. Hu, & W. R. Renandya (Eds.), *Principles and practices for teaching English as an international language* (pp. 226–243). Routledge. doi:10.4324/9780203819159-16
- Flowerdew, J. (2017). Corpus-based approaches to language description for specialized academic writing. *Language Teaching*, 50(1), 90–106. doi:10.1017/S0261444814000378
- Flowerdew, J. (2024). Data-driven learning: From Collins Cobuild dictionary to ChatGPT. *Language Teaching*, doi:10.1017/S0261444824000144
- Frankenberg-Garcia, A. (2012). Raising teachers' awareness of corpora. *Language Teaching*, 45(4), 475–489. doi:10.1017/S0261444810000480
- Frankenberg-Garcia, A., Rees, G., Lew, R., Roberts, J., Sharma, N., & Butcher, P. (2019). Collocaid: A tool to help academic English writers find the words they need. In F. Meunier, J. Van de Vyver, L. Bradley, & S. Thouësny (Eds.), *CALL and complexity—short papers from EUROCALL 2019* (pp. 144–150). Research-Publishing.
- Fries, C. (1952). *The structure of English*. Harcourt, Brace & World.
- 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. doi:10.1075/ijlcr.19001.gab
- Garcia, O., & Wei, L. (2014). *Translanguaging*. Palgrave Macmillan. doi:10.1057/9781137385765
- Godwin-Jones, R. (2019). Riding the digital wilds: Learner autonomy and informal language learning. *Language Learning & Technology*, 23(1), 8–25. <https://doi.org/10.125/44667>
- Goldstein, B., & Jones, C. (2019). *Evolve level 6*. Cambridge University Press.
- Gordon, W. T. (2004). *Langue and parole*. In C. Sanders (Ed.), *The Cambridge companion to Saussure* (pp. 76–87). Cambridge University Press. doi:10.1017/CCOL052180051X.006
- Granger, S., & Paquot, M. (2015). Electronic lexicography goes local: Design and structures of a needs-driven online academic writing aid. *Lexicographica - International Annual for Lexicography*, 31(1), 118–141. doi: 10.1515/lexi-2015-0007
- Gray, J. (2016). ELT materials: Claims, critiques and controversies. In G. Hall (Ed.), *The Routledge handbook of English language teaching* (pp. 95–108). Routledge. doi:10.4324/9781315676203-10
- 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. doi:10.16986/HUJE.2018038796
- Hanif, H. (2020). The role of L1 in an EFL classroom. *The Language Scholar*, 8(2), 54–62.
- Hanks, P. (2009). The impact of corpora on dictionaries. In P. Baker (Ed.), *Contemporary corpus linguistics* (pp. 214–236). Continuum.
- Harmain, H. (2010). ALKHALEEL: A corpus-based learning tool for Arabic. In *EDULEARN10 proceedings* (pp. 288–293). IATED.
- Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge. doi:10.4324/9780203887332
- Heather, J., & Helt, M. (2012). Evaluating corpus literacy training for pre-service language teachers: Six case studies. *Journal of Technology and Teacher Education*, 20(4), 415–440.
- Herberg, D., Steffens, D., & Tellenbach, E. (1997). *Schlüsselwörter der Wendezeit. Wörter-Buch zum öffentlichen Sprachgebrauch 1989/90*. Walter de Gruyter.
- Hockly, N., & Dudeney, G. (2018). Current and future digital trends in ELT. *RELC Journal*, 49(2), 164–178. doi:10.1177/00336882187773

- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press. doi:10.1017/CBO9781139524773
- Hunston, S. (2022). *Corpora in applied linguistics* (2nd ed.). Cambridge University Press. doi:10.1017/9781108616218
- Hwang, I., & Cho, M. (2022). Increasing lexical awareness through data-driven learning: Polysemy in EFL pedagogy. *Korean Journal of English Language and Linguistics*, 22, 1116–1132.
- Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. In T. Johns & P. King (Eds.), *Classroom concordancing* (pp. 1–16). University of Birmingham.
- Johns, T. (1993). Data-driven learning: An update. *TELL&CALL*, 2, 4–10.
- Jones, C., & Oakey, D. (2024). Learners' perceived development of spoken grammar awareness after corpus-informed instruction: An exploration of learner diaries. *TESOL Quarterly*, 58(3), 1138–1165. doi:10.1002/tesq.3305
- Jordan, G., & Long, M. (2022). *English language teaching now and how it could be*. Cambridge Scholars Publishing.
- Kachru, B. B. (1990). World Englishes and applied linguistics. *World Englishes*, 9(1), 3–20. doi:10.1111/j.1467-971X.1990.tb00683.x
- Kerr, P. (2016). *Personalization of language learning through adaptive technology. Part of the Cambridge papers in ELT series*. Cambridge University Press.
- Kilgariff, 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. doi:10.1007/s40607-014-0009-9
- Kjellmer, G. (1994). *A dictionary of English collocations. Based on the Brown corpus*. Clarendon Press.
- Klein, W. (1986). *Second language acquisition*. Cambridge University Press. doi:10.1017/CBO9780511815058
- Kohn, K. (2012). Pedagogic corpora for content and language integrated learning. Insights from the BACKBONE project. *The Eurocall Review*, 20(2), 3–22. doi:10.4995/eurocall.2012.11374
- Kováříková, D., Škrabal, M., Cvrček, V., Lukešová, L., & Milička, J. (2020). Lexicographer's lacunas or how to deal with missing representative dictionary forms on the example of Czech. *International Journal of Lexicography*, 33(1), 90–103. doi:10.1093/ijl/ecz027
- Larsson, T. (2016). The introductory it pattern: Variability explored in learner and expert writing. *Journal of English for Academic Purposes*, 22, 64–79. doi:10.1016/j.jeap.2016.01.007
- Lee, E.-J. (2015). Analysis of corpus-based research published in English education: Focusing on learner corpus and applied corpus linguistics research. *English Teaching*, 70(5), 193–214.
- Lee, Y. W. (2021). What on earth have we done to KCSAT English with 'jeoldaepyongga?'. *KATE Forum*, 2, 34–49.
- Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25(1), 56–75. doi:10.1016/j.esp.2005.02.010
- Leech, G. N. (1997). Teaching and language corpora: A convergence. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 1–23). Longman.
- Le Foll, E. (Ed.). (2021). *Creating corpus-informed materials for the English as a foreign language classroom. A step-by-step guide for (trainee) teachers using online resources* (3rd ed.). Open Educational Resource. <https://pressbooks.pub/elenlefol>. doi:10.5281/zenodo.4992504.
- Leńko-Szymańska, A. (2014). Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education. *ReCALL*, 26(2), 260–278. doi:10.1017/S095834401400010X
- Li, G., Anderson, J., Hare, J., & McTavish, M. (Eds.). (2021). *Superdiversity and teacher education: Supporting teachers in working with culturally, linguistically, and racially diverse students, families, and communities*. Routledge. doi:10.4324/9781003038887
- Lim, L., & Wang, V. X. (2023). The potential of using corpora and concordance tools for language learning: A case study of 'interested in (doing)' and 'interested to (do)'. In V. Bhateja, F. Carroll, J. M. R. S. Tavares, S. S. Sengar, & P. Peer (Eds.), *Intelligent data engineering and analytics. FICTA 2023. Smart innovation, systems and technologies* (Vol. 371, pp. 165–175). Springer. doi:10.1007/978-981-99-6706-3\_15
- Lin, P. (2023). ChatGPT: Friend or foe (to corpus linguists)? *Applied Corpus Linguistics*, 3(3), 100065. doi:10.1016/j.acorp.2023.100065
- Lusta, A., Demirel, Ö., & Mohammadzadeh, B. (2023). *Language corpus and data driven learning (DDL) in language classrooms: A systematic review*. Heliyon. doi:10.1016/j.heliyon.2023.e22731
- Maguire, P. (1987). *Doing participatory research: A feminist approach*. Center for International Education, University of Massachusetts.
- McCarten, J. (2012). Corpus-informed course book design. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 413–427). Routledge. doi:10.4324/9780367076399
- McCarthy, M. (2008). Accessing and interpreting corpus information in the teacher education context. *Language Teaching*, 41(4), 563–574. doi:10.1017/S0261444808005247
- McCarthy, M. (2016). Putting the CEFR to good use: Designing grammars based on learner-corpus evidence. *Language Teaching*, 49(1), 99–115. doi:10.1017/S0261444813000189
- McCarthy, M., McCarten, J., & Sandiford, H. (2004). *Touchstone series*. Cambridge University Press.
- McCarthy, M., McCarten, J., & Sandiford, H. (2004–6). *Touchstone series*. Cambridge University Press.
- McEnery, T., Clarke, I., & Brookes, G. (2025). *Learner language, discourse and interaction*. Cambridge University Press.



- Meunier, F., & Gouverneur, C. (2009). New types of corpora for new educational challenges. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 179–201). John Benjamins. <http://digital.casalini.it/9789027289988>
- Mizumoto, A. (2023). Data-driven learning meets generative AI: Introducing the framework of metacognitive resource use. *Applied Corpus Linguistics*, 3(3), 100074. doi:10.1016/j.acorp.2023.100074
- Muftah, M. (2023). Data-driven learning (DDL) activities: Do they truly promote EFL students' writing skills development? *Education and Information Technologies*, 28(10), 13179–13205. doi:10.1007/s10639-023-11620-z
- Naismith, B. (2017). Integrating corpus tools on intensive CELTA courses. *ELT Journal*, 71(3), 273–283. doi:10.1093/elt/ccw076
- Nesi, H. (2024). Are we witnessing the death of dictionaries?. *Ibérica*, 47, 7–14. doi:10.17398/2340-2784.47.7
- Noguera-Díaz, Y., & Pérez-Paredes, P. (2020). Teaching acronyms to the military: A paper-based DDL approach. *Research in Corpus Linguistics*, 8(2), 1–27. doi:10.32714/ricl.08.02.01
- Norton, J., & Buchanan, H. (2022). Why do we need coursebooks? In J. Norton & H. Buchanan (Eds.), *The Routledge handbook of materials development for language teaching* (pp. 49–64). Routledge. doi:10.4324/b22783-6/
- O'Keeffe, A. (2021). Data-driven learning, theories of learning and second language acquisition. In P. Pérez-Paredes & G. Mark (Eds.), *Beyond concordance lines: Corpora in language education* (pp. 35–55). John Benjamins.
- O'Keeffe, A., & Farr, F. (2012). Using language corpora in initial teacher education: Pedagogic issues and practical applications. In D. Biber & R. Reppen (Eds.), *Corpus linguistics*. (Vol. 4, pp. 335–362). Sage.
- O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press. doi:10.1017/CBO9780511497650
- Park, J.-K. (2009). 'English fever' in South Korea: Its history and symptoms. *English Today*, 25(1), 50–57. doi:10.1017/S026607840900008X
- Pérez-Paredes, P. (2022). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning*, 35(1–2), 36–61. doi:10.1080/09588221.2019.1667832
- 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. doi:10.1080/09588221.2020.1825095
- Putland, E., Chikodzore-Paterson, C., & Brookes, G. (2023). Artificial intelligence and visual discourse: A multimodal critical discourse analysis of AI-generated images of "Dementia". *Social Semiotics*, 1–26. doi:10.1080/10350330.2023.2290555
- 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. doi:10.1016/j.apacoust.2015.10.017
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English Language*. Pearson Longman.
- Rodríguez-Fuentes, R. A., & Swatek, A. M. (2022). Exploring the effect of corpus-informed and conventional homework materials on fostering EFL students' grammatical construction learning. *System*, 104, 102676. doi:10.1016/j.system.2021.102676
- Rundell, M., & Stock, P. (1992). The corpus revolution. *English Today*, 8(3), 21–32. doi:10.1017/S0266078400006520
- Schissel, J. L. (2023). Bias, discrimination, and the social consequences of unproblematic assessments in TESOL. *TESOL Quarterly*, 57(2), 716–721. doi:10.1002/tesq.3219
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129–158. doi:10.1093/applin/11.2.129
- Schwartz, K., Cappella, E., & Aber, J. L. (2019). Teachers' lives in context: A framework for understanding barriers to high-quality teaching within resource deprived settings. *Journal of Research on Educational Effectiveness*, 12(1), 160–190. doi:10.1080/19345747.2018.1502385
- Scott, M. (2024). *Wordsmith Tools version 9 (64 bit version)*. Lexical Analysis Software.
- Sharkey, A. J. (2016). Should we welcome robot teachers? *Ethics and Information Technology*, 18(4), 283–297. doi:10.1007/s10676-016-9387-z
- Siepmann, D. (2015). Dictionaries and spoken language: A corpus-based review of French dictionaries. *International Journal of Lexicography*, 28(2), 139–168. doi:10.1093/ijl/ecv006
- Smith, S. (2011). Corpus-based tasks for learning Chinese: A data-driven approach. In *Conference on technology in the classroom: official conference proceedings* (Vol. 48, p. 59). The International Academic Forum.
- Sun, Y., & Dang, T. N. Y. (2020). Vocabulary in high-school EFL textbooks: Texts and learner knowledge. *System*, 93, 1–13. doi: 10.1016/j.system.2020.102279.
- Swales, J. M., & Feak, C. B. (2012). *Academic writing for graduate students* (3rd ed.). University of Michigan Press. doi:10.3998/mpub.2173936
- Szudarski, P. (2023). *Collocations, corpora and language learning*. Cambridge University Press. doi:10.1017/9781108992602
- Thornton, P. H. (2004). *Markets from culture: Institutional logics and organizational decisions in higher education publishing*. Stanford University Press.
- Timmis, I. (2013). Spoken language research: The applied linguistic challenge. In B. Tomlinson (Ed.), *Applied linguistics and materials development* (pp. 79–94). Bloomsbury.
- Tsui, A. B. M. (2006). What teachers have always wanted to know—and how corpora can help. In J. M. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 39–61). John Benjamins.

- Tsui, A. B. M., & Tavares, N. J. (2021). The technology cart and the pedagogy horse in online teaching. *English Teaching & Learning*, 45(1), 109–118. doi:10.1007/s42321-020-00073-z
- Ur, P. (2017). Applications of research to materials design. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 132–143). Routledge. doi:10.4324/9781315716893-10
- Vaughan, L. M., & Jacquez, F. (2020). Participatory research methods—choice points in the research process. *Journal of Participatory Research Methods*, 1(1), 1–14. doi: 10.35844/001c.13244.
- Viana, V. (Ed.). (2022). *Teaching English with corpora: A resource book*. Taylor & Francis.
- Wali, F., & Huijser, H. (2018). Write to improve: Exploring the impact of an automated feedback tool on Bahraini learners of English. *Learning and Teaching in Higher Education: Gulf Perspectives*, 15(1), 14–34. doi:10.18538/lthe.v15.n1.293
- Wichmann, A., Fligelstone, S., McEnery, T., & Knowles, G. (Eds.) (1997). *Teaching and language corpora*. Longman. doi:10.4324/9781315842677-1
- Widdowson, H. G. (2003). *Defining issues in English language teaching*. Oxford University Press.
- Winchester, S. (2003). *The meaning of everything. The story of the Oxford English dictionary*. Oxford University Press.
- Write & Improve (2024). *Write & improve*. Retrieved May 1, 2024, from <https://writeandimprove.com/>
- Xi, X. (2017). What does corpus linguistics have to offer to language assessment? *Language Testing*, 34(4), 565–577. doi:10.1177/0265532217720956
- Yao, G. (2019). Vocabulary learning through data-driven learning in the context of Spanish as a foreign language. *Research in Corpus Linguistics*, 7, 18–46. doi:10.32714/ricl.07.02
- Yoon, H., & Jo, J.-W. (2014). Direct and indirect access to corpora: An exploratory case study of comparing students' error corrections and learning strategy use. *Language Learning Technology*, 18(1), 96–11. <http://lt.msu.edu/issues/february2014/yoonyjo.pdf>
- Yuan, Y., Li, H., & Sawaengdist, A. (2024). The impact of ChatGPT on learners in English academic writing: Opportunities and challenges in education. *Language Learning in Higher Education*, 14(1), 41–56. doi:10.1515/cercles-2023-0006

**Niall Curry** is Senior Lecturer in Applied Linguistics at Manchester Metropolitan University. He is Series Co-Editor of the Routledge Applied Corpus Linguistics and Routledge Corpus Linguistics Guides book series, Section Editor of Elsevier Encyclopedia of Language and Linguistics, and a Fellow of The Royal Society of Arts and Associate Fellow of the Global China Academy. For more information about Niall, his publications, and his ongoing projects, visit: <https://linktr.ee/niallcurry>. **Tony McEnery** is distinguished professor of Linguistics and English language in the Department of English Language and Linguistics at Lancaster University and Changjiang Chair at Xi'an Jiaotong University in China. He is also consulting professor at Shanghai International Studies University and a visiting professor at Zhejiang University of Media and Communications. He is a Fellow of the Academy of Social Sciences, the Global China Academy, and the Royal Society of Arts in the United Kingdom.