CORPUS LINGUISTICS FOR LANGUAGE TEACHING AND LEARNING: A RESEARCH AGENDA

Muminova Dilafruz Umurzoq qizi

University of Tashkent for applied sciences
1 st year master's degree student

Annotation: *This work explores the intersection of corpus linguistics (CL) and language pedagogy, proposing a research agenda to bridge theory and classroom practice. It highlights how corpus-based approaches—such as data-driven learning (DDL), concordance analysis, and frequency studies—can enhance language teaching by providing authentic linguistic examples, revealing patterns of usage, and fostering learner autonomy.*

Keywords: *Agenda, CL, lexico-grammatical elements, DDL, CEFR, metacognitive, plurilingual, generative AI.*

Research agenda and research tasks

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

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, Reference Creese and Blackledge2010). 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, Reference Hanif2020). 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, Reference Garcia and Wei2014).

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, Reference Klein1986), 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, Reference Gordon and Sanders2004). 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, Reference Curry2023).

Based on Curry (Reference Curry and McCallum2022), 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 translanguaging 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, Reference Curry and McCallum2022). 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, Reference Pérez-Paredes2022) and by positioning contrastive DDL at the centre of current debates in multilingualism, plurilingual competencies, and translanguaging.

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 (Reference Curry and McCallum2022)). 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, 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, Reference Pérez-Paredes2022), 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, Reference Chen and Flowerdew2018) or collect learners' perspectives through the use of learning diaries and post-study focus groups (e.g., Jones & Oakey, Reference Jones and Oakey2024). 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, Reference Muftah2023), 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, Reference Lee2021). 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, Reference Park2009). Yet, for teachers, it appears to be negatively impacting their learners' proficiency, as the assessment requires limited use of communication skills (Lee, Reference Lee2021). 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 (Reference Yoon and Jo2014) 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 (Reference Lee2015) 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 Reference Hwang and Cho2022). 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, Reference Choi2021). 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, Reference Larsson2016)), level (e.g., in terms of the CEFR (McCarthy, Reference McCarthy2016)), and register (e.g., by comparing the data with other corpora such as learner corpora, and spoken and written corpora (McEnery et al., Reference McEnery, Clarke and Brookes2025)). 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, Reference Lee2021). 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, Reference Pérez-Paredes 2022). While research continues to address this challenge (e.g., Farr & Karlsen, Reference Farr, Karlsen, Jablonkai and Csomay 2022; O'Keeffe, Reference O'Keeffe, Pérez-Paredes and Mark 2021), DDL research is facing a new evolution in the wake of developments in AI (e.g., Crosthwaite & Baisa, Reference Crosthwaite and Baisa 2023; Flowerdew, Reference 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, Reference Lim, Wang, Bhateja, Carroll, Tavares, Sengar and Peer 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, Reference 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., Reference Putland, Chikodzore-Paterson and Brookes 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., Reference Yuan, Li and Sawaengdist 2024), for a learner centred discussion of issues in using of AI in the ELT classroom and Choi, [Reference 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 (Croxton, Reference Croxton 2014), develop learner autonomy (Godwin-Jones, Reference Godwin-Jones 2019), motivate learners (Abdelhafez & Abdallah, Reference Abdelhafez and Abdallah 2015), and personalise learning (Kerr, Reference Kerr 2016). Yet, despite these affordances, there are many challenges involved in the use of technology spanning issues of access (Hockly & Dudeney, Reference Hockly and Dudeney 2018) as well as ethics (Sharkey, Reference
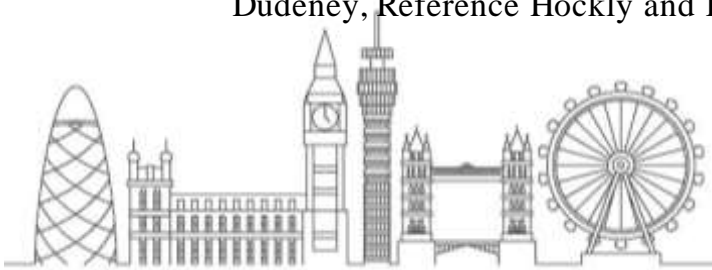
Sharkey2016). In the case of the latter, for example, learners engaging with chatbots – akin to ChatGPT – were found to be building emotional connections with fake online avatars, which, Sharkey notes, could have a detrimental impact on the development of learners' emotional intelligence. Such possibilities have given rise to growing concerns for learners' digital literacy skills (Drigas et al., Reference Drigas, Papanastasiou and Skianis2023), 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, Reference O'Keeffe, Pérez-Paredes and Mark2021). 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, Reference Anthony2024), 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, Reference Mizumoto2023). These interviews could be analysed using corpus approaches combined with critical grounded theory and top-down thematic coding (e.g., Curry & Pérez-Paredes, Reference Curry, Pérez-Paredes, Curle and Pun2023) 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, Reference Tsui and 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, Reference Lee and 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., Reference Curry, Love and Goodman 2022; Gablasova et al., Reference Gablasova, Brezina and McEnery 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, Reference Widdowson 2003). In an effort to carry out this process, Curry et al. (Reference Curry, Love and Goodman 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., Reference Callies, Hehner, Meer and Westphal 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, Reference Kohn2012). Elsewhere, pedagogical corpora are understood as those that contain texts that are used in the classroom, such as coursebooks (e.g., Meunier & Gouverneur, Reference Meunier, Gouverneur and Aijmer2009), texts produced by learners, such as assessments (McEnery et al., Reference McEnery, Clarke and Brookes2025), or texts that have been mediated for pedagogical purposes (Braun, Reference Braun2005). 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 (Reference Kachru1990), towards a focus on international usage (Callies et al., Reference Callies, Hehner, Meer and Westphal2021; Flowerdew, Reference Flowerdew, Alsagoff, McKay, Hu and Renandya2012), 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, Reference Cushing, Snell, Beal, Lukač and Straaijer2023) 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, Reference Burton2012), publishers (e.g., Curry et al., Reference Curry, Love and Goodman2022), and teachers (e.g., Leńko-Szymańska, Reference Leńko-Szymańska2014) 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., Reference Li, Anderson, Hare and McTavish2021),
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, Reference Vaughan
and Jacquez2020). In the space of DDL, working with teachers in this way has proven
fruitful (e.g. Crosthwaite & Schweinberger, Reference Crosthwaite and
Schweinberger2021; Farr & Karlsen, Reference Farr, Karlsen, Jablonkai and
Csomay2022), yet, in more indirect applications, participatory research has made fewer
inroads. Curry and Mark (Reference Curry and Mark2024) 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 (Reference Le Foll2021), 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. (Reference Curry, Love and Goodman2022), work with
publishers has demonstrated that key stakeholders do not always align with regard to
their views of the affordances of corpus linguistics. Gray (Reference Gray and
Hall2016) 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 (Reference Jordan

and Long2022) 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 (Reference McCarten, O'Keeffe and McCarthy2012) includes mention of publishers and assessment developers in her discussion of corpus applications in teaching and learning, and Burton (Reference Burton2012) notes the potential ambivalence of publishers towards corpus linguistics. Broadly, these stakeholders are noted for their power in influencing education (Thornton, Reference Thornton2004). 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, Reference Rodríguez-Fuentes and Swatek2022; Szudarski, Reference Szudarski2023) 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, Reference Curry, Pérez-Paredes, Curle and Pun2023) 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, Reference Brydon-Miller, Park, Brydon-Miller, Hall and Jackson1993; Maguire, Reference Maguire1987). 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, Reference Brydon-Miller and Maguire2008, 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.

REFERENCES:

1. Sinclair, J. (2004). How to use corpora in language teaching.

2. John Benjamins.   Seminal work on integrating corpora into pedagogy, emphasizing data-driven learning (DDL) and authentic language exposure.

3. McEnery, T,, & Hardie, A. (2011). Corpus linguistics: Method, theory and practice. Cambridge University Press. Covers corpus methodologies with implications for language teaching research.

4. Granger, S. (2015). "The contribution of learner corpora to reference and instructional materials design." In S. Granger et al. (Eds.), The Cambridge Handbook of Learner Corpus Research (pp. 485–510). Cambridge University Press.

5. Boulton, A., & Cobb, T. (2017). "Corpus use in language learning: A meta-analysis." Language Learning, 67(2), 348–393. Quantitative synthesis of DDL's effectiveness across contexts.

6. Johns, T. (1991). "Should you be persuaded: Two examples of data-driven learning." ELR Journal, 4, 1–16

7. O'Keeffe, A., & McCarthy, M. (Eds.). (2022). The Routledge Handbook of Corpus Linguistics (2nd ed.).