Generative AI for corpus approaches to discourse studies: A critical evaluation of ChatGPT

Niall Curry, Paul Baker and Gavin Brookes

Abstract

This paper explores the potential of generative artificial intelligence technology, specifically ChatGPT, for advancing corpus approaches to discourse studies. The contribution of artificial intelligence technologies to linguistics research has been transformational, both in the contexts of corpus linguistics and discourse analysis. However, shortcomings in the efficacy of such technologies for conducting automated qualitative analysis have limited their utility for corpus approaches to discourse studies. Acknowledging that new technologies in data analysis can replace and supplement existing approaches, and in view of the potential affordances of ChatGPT for automated qualitative analysis, this paper presents three replication case studies designed to investigate the applicability of ChatGPT for supporting automated qualitative analysis within studies using corpus approaches to discourse analysis.

The findings indicate that, generally, ChatGPT performs reasonably well when semantically categorising keywords; however, as the categorisation is based on decontextualised keywords, the categories can appear quite generic, limiting the value of such an approach for analysing corpora representing specialised genres and/or contexts. For concordance analysis, ChatGPT performs poorly, as the results include false inferences about the concordance lines and, at times, modifications of the input data. Finally, for function-to-form analysis, ChatGPT also performs poorly, as it fails to identify and analyse direct and indirect questions. Overall, the results raise questions about the affordances of ChatGPT for supporting automated qualitative analysis within corpus approaches to discourse studies, signalling issues of repeatability and replicability, ethical challenges surrounding data integrity, and the challenges associated with using non-deterministic technology for empirical linguistic research.

Keywords: ChatGPT; Corpus Linguistics; Discourse Analysis; Generative AI; Qualitative Analysis

Highlights

- ChatGPT is used to conduct automated qualitative analysis on three replication case studies.
- When semantically categorising keywords, ChatGPT performs reasonably well; however, the categories generated can be generic.
- For concordance analysis, ChatGPT performs poorly, making false inferences and modifying the input data.
- For function-to-form analysis, ChatGPT performs poorly as, at times, it fabricates data and fails to perform analytical procedures accurately.
- ChatGPT appears to have limited utility for corpus approaches to discourse studies, owing to issues of repeatability and replicability, ethics, and its non-deterministic nature.

1. Artificial intelligence in corpus approaches to discourse studies

Understood as a form of intelligence that is both similar to and distinct from human intelligence (Korteling et al., 2021), artificial intelligence (AI) and its subfields of machine learning and natural language processing (Shoenbill et al., 2023) play an important role in contemporary linguistics research. AI is argued to emulate human cognition (Shneiderman, 2020), a capacity which has given rise to predictive systems that can learn from text examples and models to solve problems (Jackson, 2019). Within AI, natural language processing is concerned with the means through which computers engage with and operationalise human language (Shoenbill et al., 2023), supporting text analysis and translation (Deng & Liu, 2018), for example, while machine learning serves to automate these processes by training machines to learn and self-develop (Fogel & Kvedar, 2018). From automatic transcription (Blanchard et al. 2015) to data visualisation (Hoorens et al., 2017), AI has undoubtedly improved both the quality and efficacy of linguistics research, opening new pathways for development and critique in its many of subfields, including corpus linguistics and discourse analysis. As a methodological approach, corpus linguistics is governed by its adoption of an empirical epistemology, where evidence-based approaches and critical data literacies continue to develop and (re)shape the field (Lin & Adolphs, 2023). Much of this development has occurred in synchrony with technological developments that have ushered in new approaches, tools, and expectations for corpus research (Lin & Adolphs, 2023), including considerations of issues of repeatability, replicability, and ethics (Brookes & McEnery, 2020), use of coding languages for corpus analysis (Anthony, 2021), and the incorporation of increasingly complex statistical models for corpus analysis (Larsson et al., 2020).

Subsuming both natural language processing and machine learning, AI has been used to develop software for conducting corpus analysis, such as AntConc (Antony, 2023), to enhance approaches to multimodal corpus development (Zhou & Gao, 2023), and to inform tools that make use of corpus linguistics for applied purposes (e.g., corpus-informed tools for language teaching and learning; Curry & Riordan, 2021). AI developments have also enabled automated tagging of linguistic data (Rayson, 2008), where automated part-of-speech and semantic tagging have become normalised processes in the development and analysis of corpora. Overall, AI has influenced both the processes for conducting corpus linguistics research as well as its application to real-world contexts. Arguably, any study published under the banner of corpus linguistics over the past 30 years is indebted, to some degree, to advances in AI.

A parallel analytical approach, discourse analysis, pertains to the study of language in use, focusing on how language use is influenced by context and social dynamics. As with corpus linguistics, discourse analysis necessitates the collection and analysis of texts; however, and particularly when adopting a post-structuralist perspective, the analysis typically focuses on the identification and interrogation of social, political, and cultural values embedded within texts (Baker, 2023). Critically, discourse analysis approaches respond to and reflect the ontological positioning of the analyst and serve to reveal the underlying ideological assumptions and power relations constructed and communicated through texts (Dunn & Neumann, 2016). Research in this domain branches into a range of subfields, including critical discourse analysis, positive discourse analysis, and metadiscourse analysis, and, in application, research in discourse analysis has been used to inform professional communication in a variety of contexts, e.g., healthcare communication (Baker & Brookes, 2022).

Approaches to discourse analysis have also benefited from developments in AI, albeit to a lesser degree. Automated approaches to data coding, using tools such as ATLAS.ti (Paulus & Lester, 2016), and methodological approaches including topic modelling (Brookes & McEnery, 2019) and sentiment analysis (Taufek et al., 2021), have had varying degrees of success in supporting discourse analysis. Criticisms of the affordances of AI for discourse analysis largely centre on the incapacity of AI technologies to conduct critical and contextualised analyses to the calibre of the human analyst (Brookes & McEnery, 2019). These issues notwithstanding, AI continues to inform discourse research, typically through the combination of corpus linguistics and discourse analysis, with corpus analysis software allowing researchers to study large amounts of textual data through a discourse analysis lens (Baker, 2023), including in approaches such as corpus-assisted discourse studies (or CADS; Partington et al 2013).

The recent increase in the availability of generative AI technologies, such as text-based, image-based, and code-based generators, has added another layer of potential for corpus approaches to discourse studies (Zappavigna, 2023). ChatGPT, for example, is an AI chatbot, developed by OpenAI (2023), that draws on natural language processing and deep learning to attempt to understand human-produced text and generate human-like text. Globally, as of November 2023, OpenAI reports having over 180 million registered users, with nearly 1.5 billion visits to the ChatGPT site per month and, at the time of writing, there are multiple versions of ChatGPT with both ChatGPT 3.5 and ChatGPT 4 in popular use. When asking ChatGPT to perform tasks, prompt guides can support users who wish to finetune their engagement with the chatbot. For example, users can modify temperature settings (Buruk, 2023), which can amplify or limit the creativity of the tool and render it more or less deterministic. Likewise, nucleus sampling and text length modification can allow users to determine how ChatGPT should engage with the input and deliver the output, through use of a cumulative probability threshold that limits the choice of tokens and aids in the production of text that is both coherent and non-deterministic (Agarwal, 2023). Ultimately, as a text-based generative AI, ChatGPT is worthy of consideration in the context of academic research, given claims of ChatGPT's affordances for supporting ideation (Qureshi, 202) and automating qualitative analysis (Rahman et al., 2023), its wide, international uptake (Bin-Nashwan et al., 2023), and current debates surrounding the role of ChatGPT in academic practices, including authorship and data analysis (Misra & Chandwar, 2023).

In linguistics research, ChatGPT has been argued to offer value by helping human analysts to develop search strings in regular expression for analysing social media texts in digital discourse analysis (Zappavigna, 2023). In the context of language teaching, the increased proliferation of generative AI tools, including ChatGPT, has afforded opportunities for reflection that may broaden approaches in and perspectives on data-driven learning (Crosthwaite & Baisa, 2023). Elsewhere, ChatGPT has been considered in terms of its capacity to replicate processes that typically require the use of corpus analysis software; however, initial testing indicates that the generative AI fails to effectively identify keywords and collocates in texts (Lin, 2023). In a large study of ChatGPT that included 25 tasks and over 48,000 prompts, the tool was found to perform reasonably well when conducting quantitative analysis, with a somewhat poorer performance with tasks that require more reasoning (Kocoń et al., 2023). Elsewhere, however, ChatGPT has been recognised for its potential affordances in qualitative research, with some arguing that it can assist with thematic coding (Siiman et al., 2023) while others have gone so far as to claim that it can automate qualitative analysis. On this latter point, for example, it has been argued that ChatGPT can provide adequate qualitative analysis when given qualitative data

with specific prompts to guide the analysis (Rahman et al., 2023) and that it can deliver automated inductive coding (Hämäläinen et al., 2023). Incidentally, these affordances appear to respond to recognised limitations in corpus approaches to discourse studies; namely, those pertaining to the lack of automated approaches to qualitative analysis (Gillings et al., 2023) since, to-date, automated approaches remain largely form-driven and quantitative.

Given the demonstrable capacity of AI for enhancing and informing corpus approaches to discourse studies and the understanding that the, "emergence of any technique of data […] analysis poses important questions about the extent to which that technique might supplement or even replace existing techniques in a given field" (Brookes & McEnery, 2019, p.4), the recent proliferation of generative AI technologies, such as ChatGPT, raises questions about their potential value for developing approaches in the field, further. Therefore, noting that ChatGPT has been acknowledged for its potential for conducting automated qualitative analysis (Rahman et al., 2023) and recognising the limitations of existing technologies for conducting such analysis within a broader corpus linguistic methodology (Brookes & McEnery, 2019; Gillings et al., 2023), the following section aims to investigate the affordances of ChatGPT for conducting automated qualitative analyses within research taking corpus approaches to discourse studies. To do so, three replication case studies are presented.

2. Case studies

This paper compares ChatGPT 4's approach to automated qualitative analysis with three previously published studies. The case studies centre on the use of ChatGPT for (1) semantically grouping key words (compared with Hunt & Brookes, 2020), (2) conducting concordance analysis (compared with Baker et al., 2013), and (3) conducting function-to-form analysis (compared with Curry, 2021). These case studies were chosen to investigate a range of applications of ChatGPT to corpus approaches to discourse studies, moving from limited to greater context for the analysis. For case study 1, the semantic categorisation of keywords was chosen owing to the documented potential of ChatGPT for working with decontextualised data (Rahman et al., 2023). For case study 2, the analysis of concordance lines was chosen, owing to the centrality of concordance analysis within corpus approaches to discourse studies (Baker, 2023) and, thus, the potential for its widespread use by discourse analysts. Finally, case study 3 was chosen as a response to calls from the literature to identify automated approaches to function-to-form analysis of discourse in a corpus (Gillings et al., 2023). Overall, the goal of each study was to test the capacity for ChatGPT to perform automated qualitative analysis and offer a critical reflection on its affordances for corpus approaches to discourse studies.

It is worth noting that ChatGPT was chosen for this analysis owing to it being widely used, and in response to the afore-discussed claims of its potential for supporting automated qualitative analysis. ChatGPT 4 was used also as OpenAI (2023) argues that it offers a more powerful, accurate, and reasoned generative AI that ChatGPT 3.5. In each analysis, an inductive approach was employed owing to the differing scope and focus of analysis in the three case studies. In terms of ChatGPT's parameters, it was decided that the default settings of ChatGPT would be used, as: (1) the goal of the study is to use ChatGPT as the average user would, following Brookes and McEnery (2019); and (2) in changing the temperature and nucleus sampling settings, for example, there is a risk that the analysis would be open to the charge of analytical 'cherry-picking' (Widdowson, 2004). To support the analysis and effective use of ChatGPT, the prompts developed for each case study were designed following models, examples, and advice from the wider literature e.g., Kocoń et al. (2023) and Lin (2023).

2.1 Case study 1: Semantic categorisation of keywords in online support group interactions about diabulimia (Hunt & Brookes, 2020)

The first analysis focuses on the categorisation of keywords, replicating an existing analysis by Hunt and Brookes (2020, pp. 144-145). These authors examined the discourse around three mental health conditions in the context of online support groups. The portion of the analysis replicated here was based on a corpus of forum posts about diabulimia. Diabulimia is an eating disorder in which a person living with (typically Type 1) diabetes reduces or stops taking insulin in order to lose weight (Brookes, 2018). The authors generated keywords by comparing this dataset against the Spoken BNC 2014 (see pp. 77 for details regarding procedural and statistical decisions). This comparison gave 72 keywords, which the authors manually analysed within the context of the forum posts. On this basis, the authors assigned each keyword to one of a series of broad thematic and lexical categories which they developed inductively based on the majority of the uses of each keyword (see Table 1).

Table 1: Key thematic and lexical categories and associated keywords from the diabulimia corpus, ordered alpgabetically (adpated from Hunt & Brookes 2020, pp. 144-145)[1].

Thematic/lexical category	Associated keywords
Body and weight	body, exercise, fat, gain, lose, loss, lost, weight
Diabetes	blood, complications, control, diabetes, diabetic, diagnosed, dka, high, life, low, sugar, sugars, type, years
Diabulimia and disordered behaviours	diabulimia, disorder, less, months, problem, problems, since
Feelings and emotional responses	difficult, feel, hard, needs
Food and eating	carb, carbs, diet, eat, eating, food
Forum-related	anyone, best, hi, hope, keep, luck, others, post, thanks, understand, wish
Healthcare and health professionals	care, doctor
Insulin	insulin, pump, taking
Pronouns	im, my, myself, yourself
Recovery	health, healthy, help, support
Other	also, am, etc, found, its, may, will, without, young

Using ChatGPT, the following prompt was given: 'Here's a set of keywords from online support group interactions about diabulimia. Group them into thematic and lexical categories' [followed by an alphabetised list of the 72 keywords shown in Table 1, with each keyword separated by a semicolon]. The resulting categories are shown in Table 2.

Table 2: Key thematic and lexical categories generated by ChatGPT, ordered alphabetically.

Thematic/lexical category	Associated keywords
Medical and health conditions	complications, diabetes, diabetic, diabulimia, diagnosed, disorder, dka, type
Bodily measures and states	blood, body, fat, gain, high, lose, loss, lost, low, sugar, sugars, weight
Treatment and management	carb, carbs, care, control, diet, doctor, exercise, insulin, pump, taking
Emotional and psychological aspects	difficult, feel, hard, needs, problem, problems
Timeframe	life, months, since, years
Dietary factors	eat, eating, food
Personal pronouns	Im, its, my, myself, yourself
Interpersonal and supportive language	anyone, health, healthy, help, others, support, understand

---

[1] Note that, to avoid a preponderance of categories, the authors also created an 'Other' category for words that could not be easily assigned to any of the other categories.

Sentiments and wishes	best, hope, keep, luck, thanks, wish
Misc. social interaction markers	also, am, etc, found, hi, may, post, will, without, young

ChatGPT produced a similar number of categories (n=10) to the original analysis (n=11). There are some similarities between the sets of categories, especially in terms of the following pairs: 'Food and eating'/'Dietary factors'; 'Feelings and emotional responses'/'Emotional and psychological aspects'; 'Pronouns'/'Personal pronouns'; 'Body weight'/'Bodily measures and states'; and 'Other'/'Misc. social interaction markers'. These pairs of categories share many keywords, but there are also some notable differences. Such differences seem to occur in particular where ChatGPT categorised words according to their surface meanings, whereas Hunt and Brookes (2020) categorised words based on their contextual meanings. For example, while ChatGPT assigned the keywords *problem* and *problems* to the, 'Emotional and psychological aspects' category, Brookes and Hunt (2020) found that these terms tended to be used euphemistically to refer to diabulimia, so accordingly assigned them to the category, 'Diabulimia and disordered behaviours'.

The categories themselves also exhibit some important differences between Tables 1 and 2. Sometimes, these differences result due to the reliance on surface versus contextual meaning. For example, the temporal words categorised as, 'Timeframe' by ChatGPT were instead assigned by Hunt and Brookes (2020) to categories reflecting the precise health conditions with which each term tended to be used. Specifically, *less* and *months* were assigned to the category, 'Diabulimia and disordered behaviours', as these words were consistently used to quantify the amount of time contributors had been experiencing diabulimia, while *life* and *years* were categorised under, 'Diabetes' for the same reason. There were also differences in the levels of thematic granularity employed in the original analysis and by ChatGPT. For example, where ChatGPT used the rather broad category 'Medical and health conditions', most of the keywords in this category were assigned by Hunt and Brookes to either of the more specific categories, 'Diabetes' or 'Diabulimia and disordered behaviours'. Similarly, most of the words assigned by ChatGPT to the category, 'Treatment and management' were categorised by Hunt and Brookes either as, 'Healthcare and health professionals' or 'Insulin'. Likewise, most of the keywords assigned by ChatGPT to the categories, 'Interpersonal and supportive language' and 'Sentiments and wishes' were assigned by Hunt and Brookes to categories that were more granular with respect to the topic (i.e., 'Recovery') and medium (i.e., 'Forum-related') of the texts in the corpus. In the latter case, ChatGPT assigned many of these genre-related keywords to the more generic, 'Misc. social interaction markers' category.

Overall, ChatGPT produced a manageable number of categories, with the vast majority of keywords being allocated to categories that reflect their surface meanings. The tool could also draw on the information about the context of the data that was supplied through the prompt in order to feasibly categorise technical vocabulary. For example, it accurately decoded the initialism *dka* (providing the gloss, 'diabetic ketoacidosis' in parentheses). However, there are also areas where ChatGPT's categorisation requires further explanation. For example, it was not clear why keywords such as *will*, *without* and *young* were assigned to the category, 'Misc. social interaction markers', and on what basis terms such as *health* and *healthy* were categorised as, 'Interpersonal and supportive language' (and indeed, on what basis many of the keywords assigned to this category were not assigned to the similar 'Sentiments and wishes' (and vice versa)). These could represent instances of mis-categorisation. Notably, in lowering the temperature settings of ChatGPT, the results of its analysis reflected similar differences when compared to the analysis in Brookes and Hunt (2020).

Given the topic specificity of the data (as indicated in the prompt), one might question how helpful a category as broad as 'Medical and health conditions' is for a study of a specialised corpus of health-related communication. To test an iterative approach to generating more granular categories, ChatGPT was given the same prompt again, but this time just with the keywords belonging to the 'Medical and health conditions' category. It categorised these words into the following, more granular categories: 'Medical conditions' (*diabetes*; *diabulimia*; *dka*); 'Medical classification/terminology' (*type*; *disorder*); 'Medical processes' (*diagnosed*; *complications*); and 'Patient identification' (*diabetic*). These categories are indeed more granular. However, in order to generate them, ChatGPT needed to make some assumptions about how these terms are used. For example, classifying *diabetic* as, 'Patient identification' implies that this word is used to discuss patients and while this is certainly true, its use in the data is also more diverse. Therefore, in the semantic categorisation of keywords, it is reasonable to conclude that ChatGPT performs reasonably well; however, the value of such categorisation for specialised discourse is at times questionable, as the categories generated can be quite generic.

2.2 Case study 2: Concordance analysis of the word *homosexuals* in newspaper texts about Islam (Baker et al., 2013)

The second analysis attempted to replicate an existing analysis, reported in Baker et al. (2013: 111-112), which involved an examination of a 143 million-word corpus of UK national newspaper articles about Islam and Muslims. An analysis of 106 concordance lines of *homosexuals* taken from a single year (2000) aimed to identify how often and in what ways homosexuality was linked to Islam in that period. The original manual analysis identified 8 cases of homosexuality being linked to Islam. The manual analysis then identified two types of concordance lines – those which constructed Islam as homophobic (three cases) and those which equated homosexuality with Islam as being similarly oppressed groups (5 cases). Examples of each type are shown in Table 3.

Table 3: Cases of *homosexuals* being linked to Islam, adapted from Baker et al. (2013: 111-112).

#	Examples
1	in a country where television, films and music are banned. Homosexuals have been buried alive under walls, petty thieves have had their
2	's eccentric outbursts against a variety of supposed enemies, ranging from homosexuals and Freemasons to communists and Muslims. Yesterday Elio Toaff, the

In example 1, although the concordance line does not mention Islam, the analyst had flagged this as a relevant case due to the reference to homosexuals being buried alive under walls, using contextual knowledge that this was a practice associated with the Taliban, which also refers to itself as the Islamic Emirate of Afghanistan. The second example, however, did not require contextual knowledge as *homosexuals* and *Muslims* were mentioned in the concordance line, and clearly linked together as someone's *supposed enemies*.

Using ChatGPT the following prompt was given: 'Here's a set of concordance lines from newspapers about Islam. How many of them directly relate Islam to homosexuality in some way', along with the 106 concordance lines. ChatGPT identified 7 cases, all of which were different from the 8 identified by the human analyst. However, none of ChatGPT's examples (see cases 3-9) appeared to indicate a relevant relationship between homosexuality and Islam.

Table 4: Cases of *homosexuals* being linked to Islam, as identified by ChatGPT.

#	Examples
3	Cote du Rhone and wondered why it should be the case that homosexuals have unusually long index fingers; three pregnant wives and one post-
4	Shah and Islam, it was made clear that homosexuals do form a social group for the purpose of claims for asylum
5	last year, offers legal recognition for all cohabiting couples, including homosexuals. Hailed by supporters as France's first truly modern piece of"
6	made in favour of repealing the law that it encourages hatred of homosexuals
7	Hindus and Sikhs." In 1994 the age of consent for homosexuals was lowered from 21 to 18. Downing Street said equality was
8	www.morocco-travel.com). Legislation to reduce the age of consent for homosexuals in England, Wales and Scotland to 16 reached the statute book last night after
9	stupendously incorrect politically, referring to women as 'bitches' and homosexuals as 'benders'. A disproportionate percentage of his streetwise patois

Then ChatGPT was given the prompt, 'In what ways do these 7 instances refer to the relationship between homosexuality and Islam? Can similar ways be grouped together'. This resulted in 5 categories for the 7 cases – something which perhaps should raise a note of warning from the outset.

Two of its categories only contained one concordance line, which is arguably not a convincing example of a category. One of the categories produced was similar to that identified by the human analyst, with ChatGPT calling it, 'Hate and Discrimination', and assigning a single concordance line to it (line 6). However, ChatGPT did not identify the category of viewing homosexuality and Islam as having a similar status as Baker et al. (2013) did, and the way it grouped the concordance lines into categories was not convincing. For example, with the, 'Hate and Discrimination' case, ChatGPT said that line 6 'refers to the concept of encouraging hatred of homosexuals in relation to Islam. This could be related to Islamophobic stereotypes of perceptions'. It is noted that ChatGPT hedged the possible connection using the word *could*. However, the concordance line makes no mention of *Islam*, and in fact the article from which the concordance line was extracted is focussed on a discussion of a British law called Section 28, with *Islam* only mentioned briefly, quite a long way from the concordance line. One possible explanation for this decision by ChatGPT is that it may be working across concordance lines, drawing context and content from other concordance lines and not seeing them as extracts from separate texts. Moreover, when ChatGPT quoted concordance line 6 in its response, it actually rephrased the wording, replacing the word "made" with "move".

Subsequently, with lowered temperature settings, ChatGPT was then given a further prompt, asking 'Do you want to try this exercise again, take another look at the large set of concordance lines I initially gave you and identify which ones relate Islam to homosexuality in some way.' This time, ChatGPT produced five concordance lines, of which only one (line 8) it had mentioned previously. Again, the examples produced were completely different from the ones identified by the human analyst. ChatGPT divided the five concordance lines into 2 groups, although it did not provide a rationale for this division. This is perhaps understandable, as this was not explicitly included in the second prompt question. The human analyst had (incorrectly) surmised that ChatGPT would provide such a rationale of its own accord.

Strangely, however, with all five of the new concordance lines, ChatGPT's analysis concluded that each line 'was not directly related to Islam'. In this second attempt, ChatGPT therefore did not provide any cases where homosexuality was linked with Islam. Interestingly, it did correctly identify 5 cases where there was no link, although this is not something it was asked to do. It is not clear why ChatGPT singled

out these 5 lines in particular, when there were several dozen other lines which also did not have a link between homosexuality and Islam.

Additionally, ChatGPT's second analysis contained incorrect statements. For example, it claimed that line 8 referred 'to religious organizations employing non-believers and homosexuals', yet this line does not appear to refer to any organisation. It could be the case that ChatGPT was referring to the organisation that created the travel website (www.morocco-travel.com). However, this website is not a religious organisation. Overall, ChatGPT's analysis of concordance lines was deemed to be quite inaccurate, resulting in some unexpected errors such as incorrectly quoting the data in the first attempt and not following the instructions correctly in the second attempt.

2.3 Case study 3: Function-to-form identification and analysis of direct and indirect questions in economics research articles (Curry, 2021)

The third analysis sought to replicate part of Curry's (2021) study of direct and indirect questions, which involved a corpus-based contrastive analysis of questions as reader engagement markers in economics academic writing in English, French, and Spanish. The analysis used the KIAP-EEFS comparable corpus to identify questions and, once identified, the questions were analysed according to their frequency, length, form, function, location within the text, tense and aspect, voice, polarity, and sentence complexity, with a view to comparing approaches to question raising in economics academic writing across languages. A critical facet of this study was the development of a function-to-form approach to question identification whereby, direct questions were identified by searching for question marks as illocutionary force indicating devices, while indirect questions were identified through an iterative, functionally driven approach, reported in detail in Curry (2023). The questions identified were used to varying effects, for example, serving to frame the discourse (through indirect questions) or to organise the text (with direct polar questions), as Table 5 demonstrates.

Table 5: Examples of direct and indirect questions from Curry (2021).

Example type	Question identified by Curry (2021)	Explanatory notes
Example of direct question used to organise the text	1. Do strong FRPs increase the location of knowledge factors outside the source country?	Question 1 is a direct polar question, used to signal what is coming next in the text. The author raises the question to signal that they are moving on to a new topic.
Example of indirect question used to frame the discourse	2. Finally, we briefly characterise some stylized facts regarding our estimated stocks and ask whether there are trends in net foreign assets and shifts in their composition over time.	Question 2 is an indirect question raised using the phrase "ask whether". In the context of the paper, this question is a research question, guiding the focus of the research paper.

This case study focuses solely on the results of the English language analysis of questions. As criteria of evaluation, it focuses on question identification and analysis in terms of question length, form, and function, comparing ChatGPT's analysis of the extracts from the English subcorpus, *engecon,* with the analysis of this dataset reported in the original study.

To conduct the analysis, ChatGPT was given a more detailed prompt than those issued in the previous two case studies (presented in Table 6). The prompt is based on the analysis in Curry (2021), which grouped question length into three categories (<15 words, 16-25 words, and >26 words), question

forms into six categories (content, polar, declarative, alternative, elliptical, and indirect), and question function into seven categories (a modification of Hyland's framework of question functions in academic writing (2002)).

Table 6: Prompt Given to ChatGPT for each extract from the *engecon.*

Analyse the following text and identify all the direct and indirect questions in the text. Then categorise the questions according to:
- Sentence length as <15 words, 16-25 words, or >26 words
- Question form as content, polar, declarative, alternative, elliptical, or indirect
- Question function based on Hyland (2002): Getting Attention and Focusing the Reader; Framing the Discourse; Organising the Text; Creating a Niche; Expressing and Attitude and Counter-Claiming; Setting up Claims and Protecting the Writer; Asking Real Questions and Pointing Forward

[TEXT EXRACT]

As the function-to-form approach often requires contextualisation within the whole text (Curry, 2023), initially full texts were used as input for ChatGPT. However, due the limitations of the tool, ultimately shorter extracts from the subcorpus had to be used. This was because ChatGPT could not process the full texts, citing that they were too long, despite the texts being under the stated wordcount that ChatGPT4's developers claimed it could process on the tool's homepage.[2] Following the above prompt, ChatGPT was given the following five extracts to analyse:

- The section tagged <intro> in *engecon10*;
- The section tagged <mid> in *engecon4*;
- The section tagged <mid> in *engecon18*;
- The section tagged <mid> in *engecon36*;
- The section tagged <intro> in *engecon6.*

These five texts and their extracts were chosen as they contained the highest number of questions in *engecon*, in decreasing order. These extracts were previously identified to contain 47 questions (Curry 2021; see Table 7).

Table 7: Question occurrence in *engecon* extracts, based on Curry (2021).

Extract from text	Direct questions	Indirect questions	Total
engecon4	9	2	11
engecon6	2	3	5
engecon10	10	0	10
engecon18	9	1	10
engecon36	11	0	11
Total	**41**	**6**	**47**

---

[2] While this paper was undergoing peer-review, a beta version of ChatGPT 4 was released which allowed for file upload, though still with limitations. As such, the amount of text that ChatGPT can process is likely to increase. Looking forward, this limitation is not likely to persist as a problem for corpus linguists.

Upon analysing the same extracts with ChatGPT, a total of 42 questions were found, with differing numbers of direct and/or indirect questions identified within each extract, when compared to the original study, as shown in Table 8.

Table 8: Question occurrence in *engecon* extracts, based on ChatGPT analysis.

Extract from text	Direct questions found by ChatGPT	Indirect questions found by ChatGPT	Total
engecon4	9	0	9
engecon6	2	0	2
engecon10	5	1	6
engecon18	9	6	15
engecon36	10	0	10
Total	**35**	**7**	**42**

Though far from ideal, this initial difference may seem marginal. However, this variance becomes more pronounced when overlap is considered, where only 66% of the direct questions, 17% of indirect questions, and 60% of the total number of questions found by ChatGPT account for the questions found in the original study, as can be seen in Table 9.

Table 9: Questions shared by ChatGPT and Curry (2021) analysis in *engecon* extracts.

Extract from text	Direct questions shared with original	Indirect Questions shared with original	Number of Questions shared with original
engecon4	5	0	5
engecon6	2	0	2
engecon10	5	0	5
engecon18	6	1	7
engecon36	9	0	9
Total	**27**	**1**	**28**

Understanding why this variation occurs is important, as it raises questions about the value of ChatGPT for qualitative research. One may wonder whether ChatGPT is identifying questions in the data that were missed in the original analysis. If so, this would signal a value of the tool. However, as Table 10 demonstrates, further investigation of the questions identified by ChatGPT found that of the passages identified by ChatGPT as direct or indirect questions, only 79% actually existed within the data supplied to the tool. The remaining 21% of cases arise as a result of ChatGPT reporting the presence of questions that were not, in fact, present in the extracts.

Table 10: Authentic, fabricated, and duplicated questions identified by ChatGPT.

Extract from text	Questions from ChatGPT that exist in *engecon*	Questions from ChatGPT that do not exist in *engecon*	Questions from ChatGPT that are duplications
engecon4	5	4	0
engecon6	2	0	0
engecon10	6	0	0
engecon18	11	3	1
engecon36	9	1	0
Total	**33**	**8**	**1**

ChatGPT fabricated questions, for example, by adding question marks or question tags to declarative statements and presenting these as questions in the analysis. It also duplicated questions, by presenting one question as two slightly different questions. Examples of each case are given in Table 11.

Table 11: Examples of fabricated and duplicated questions.

Example type	Question identified by ChatGPT	Notes
Fabricated question	3. It is sometimes suggested that regions with high transport costs are disadvantaged as potential locations for establishment of new activities or participation in production networks, true?	Question 3 does not occur in the extracts. The declarative statement, "It is sometimes suggested that regions with high transport costs are disadvantaged as potential locations for establishment of new activities or participation in production networks" does occur in the extracts; however, ChatGPT added the question tag, "true?" and presented it as a question.
Duplicated question	4. It is not clear what role marriage plays. 5. First it is not clear what role marriage plays.	ChatGPT identified both Question 4 and 5 as indirect questions. Question 5 does occur in the extract as an indirect question. However, for Question 4, ChatGPT removed the word "First", capitalised the "It" and presented it as a separate question.

Ultimately, only 28 of the 47 questions identified by ChatGPT were questions that were also identified in the original analysis by Curry (2021). ChatGPT's analysis of these questions in terms of the criteria of evaluation (i.e., question length, question form, and question function) was extracted and compared to the original classification in Curry (2021). As Table 12 shows, ChatGPT's analysis had varying degrees of success for each criterion of evaluation.

Table 12: ChatGPT's analysis of questions in terms of length, form, and function.

Criteria of evaluation	Number of correct categorisations	Percentage of correct categorisations
Question Length	21	75
Question Form	24	86
Question Function	5	18

ChatGPT did reasonably well at grouping questions by their length, with 75% accuracy. However, accurate counting of words is not a challenging task, and 100% accuracy for automated counting is already well within reach, even through the use of less sophisticated tools, such as Microsoft Excel. For example, where the following question has 14 words, ChatGPT categorised it having between 16 and 25 words: 'why are the magnitudes of the gains based upon these two approaches so different?'. In terms of form categorisation, 86% of ChatGPT designations were accurate. This is where ChatGPT performed most effectively. However, ChatGPT labelled Question 1 in Table 5 as a content question, despite it clearly being a polar question (i.e., eliciting a *yes* or *no* answer). Therefore, its inaccuracy appears illogical and its approach to classification is difficult to access and understand, as ChatGPT successfully identified other polar questions in the data. For function, only 18% of ChatGPT's categorisations were accurate, demonstrating the tool's incapacity to effectively identify the function of questions, despite adequate context being provided in the text extracts. Moreover, ChatGPT at

times created new categories, outside of those given, for example labelling a question with the function, 'pointing forward', which was evidently an unexplained modification of the function, 'asking real questions and pointing forward'.

Overall, ChatGPT's approach to function-to-form analysis is arguably quite poor. Not only is the initial identification of the questions fraught, with ChatGPT failing to identify many questions, extracting text that is not performing the act of asking a question, fabricating questions, and duplicating questions, but the analysis of these cases is also poor. ChatGPT fails to accurately count sentence length consistently and appears incapable of accurately interpreting the functions of questions, despite the greater context given with the large text extracts. ChatGPT is most effective at categorising questions according to form. However, when asked to repeat the analysis and when using a lower temperature parameter, ChatGPT identifies different questions, gives different categorisations for the same questions across length, function, and form, and presents its analysis in different formats (e.g., lists, tables, paragraphs of text), and continues to modify the data, demonstrating inconsistency not only in the presentation of results, but also in its analytical approach.

3. Discussion and conclusion

This paper aimed to investigate the affordances of ChatGPT for conducting automated qualitative analyses as part of corpus approaches to discourse studies. In the first case study, ChatGPT was given a set of keywords with a request to semantically group them. It had limited context, with only the prompt and the keywords at its disposal. For this initial task, ChatGPT performed reasonably well, in that the groupings of the keywords made sense, were relevant to the context of the study, and overlapped in some ways with the original study. However, the categories are inevitably surface-level and generic, reflecting parallels with semantic taggers and lacking the granularity required in research on specialised discourses. Taking an iterative approach to the generation of more granular themes, ChatGPT appears to offer more nuanced analyses. However, there remains a need for the human analyst to engage with the data, to ensure that the categorisation of the words takes into consideration their contextual meanings.

In the second case study, ChatGPT was given a little more context to facilitate its qualitative analysis: a set of concordance lines. Through concordance analysis, it emerges that ChatGPT draws on co-textual knowledge to identify meaning that is not obvious at the surface level of the text. However, it gave inaccurate results. The results of ChatGPT's concordance analysis bore little resemblance to the original study, failing to identify similar categories. Moreover, it inferred a hateful and discriminatory relationship between Islam and homosexuality, using evidence to support this claim that made no mention of Islam and had no relation to Islamophobic stereotypes, as its analysis claims. This particular result is problematic, as ChatGPT appears to be drawing on information from other concordance lines and conflating lines of data and evidence to produce results that might be compelling but are, ultimately, fabricated. Nucleus sampling and prompt guides could potentially mitigate some of ChatGPT's inability to effectively analyse each concordance line independently. However, the potential irregularity of concordance lines in any given analysis (e.g. the number of tokens or characters each one contains) means that the required nucleus sampling parameters would need to vary within and across studies. Accounting for this variability could be challenging for analyses involving large sets of concordance lines, and if not done effectively could result in unreliable results. In a similar vein, in other instances, ChatGPT modified the language of the concordance lines when it

included them in the results, a process that is inimical both to corpus linguistics and discourse studies, as well as the commitment to data integrity that characterises empirical linguistics in a broader sense (Brookes & McEnery, 2020; Lin & Adolphs, 2023). As with the keyword analysis, when asked to reconduct the concordance analysis, ChatGPT gave different results, and the results were not any more accurate the second time. Ultimately, the non-deterministic nature of ChatGPT (Qureshi, 2023) and the lack of clarity surrounding its analytical processes pose problems for replicability and repeatability, as the tool's strength in producing different text for the same input becomes a weakness, where transparency and methodological rigor cannot be ensured.

In the third case study, with an even greater context afforded to ChatGPT, the results were equally, if not more, unreliable. In analysing larger extracts from research articles, ChatGPT fails to accurately identify direct and indirect questions. Like the analysis of concordance lines in the second case study, in the third case study, ChatGPT extracted strings of texts that are not acting as questions and presented them as questions in the analysis. It also modified the data to fabricate questions, which it included in the analysis. This practice makes the notion of using ChatGPT for automated qualitative analysis a perhaps problematic prospect, as using the tool (in its current state) is likely to undermine the rigor and integrity of a study. Moreover, when performing simple tasks, including counting the number of words in questions, ChatGPT performs poorly. Likewise, for functional analysis, the tool's inaccuracy, despite the added context and co-text, render it unsuitable for this purpose. So, too, does its non-deterministic nature (Qureshi, 2023), as the creation of new functional categories, the varied outcomes of response regeneration, and the lack of evident logic in the analytical process all serve to undermine its usability for corpus linguistics in general, as well as in its application to discourse studies.

Ultimately, while it has been argued that ChatGPT can produce effective qualitative analysis (Hämäläinen et al., 2023; Rahman et al., 2023), in its current state, the extent of such analytical promise for research based on corpus approaches to discourse studies is materially and theoretically limited. One may argue that a limitation of this study is the assumption that ChatGPT should be used to conduct automated qualitative analysis, given that the generative AI tool is not necessarily designed for that purpose. However, as research elsewhere argues that ChatGPT can be used to conduct such analyses, it is important to reflect on the potential use of generative AI in the context of corpus and discourse studies and to offer empirically grounded guidance and a word of caution for those wishing to use ChatGPT to conduct qualitative analyses. In discourse analysis, analytical approaches range from patterned language analysis using corpus analysis software to whole text analysis. Yet in its current format, ChatGPT does not facilitate such approaches. It is unreliable, as it modifies the data, it is inoperable, as it cannot manage large texts, and it is inimical to contemporary approaches to linguistics analysis, interwoven with open science perspectives, data ethics, and repeatability and replicability (Brookes & McEnery, 2020). From an ontological perspective, ChatGPT is something of a 'black box'. The positioning of the analyst is central to discourse analysis (Dunn & Neumann, 2016); however, it is not possible to access the ontology of ChatGPT or its approach to analysis, if indeed it has one, in part owing to its non-deterministic nature.

In reflecting on how ChatGPT corresponds to a human analyst, the issue of context is of interest. In research using corpus approaches to analyse discourse, context is key. Looking at deeply contextualised language can reveal idiosyncratic features of language use in a range of contexts, and the more context to which a researcher has access, the more nuanced, granular, and arguably accurate

the analysis can be. However, for ChatGPT context appeared debilitating. With limited context for keyword categorisation, ChatGPT performed reasonably well. However, for the second and third case studies, with a greater degree of context afforded through the co-text, the analysis was less accurate. ChatGPT seemed to conflate and modify data, likely owing to its probabilistic approach to synthesis and its goal to produce content, regardless of the input (Antaki et al., 2023). Modifying ChatGPT's parameters may offer future research some direction for retesting the value of the generative AI tool for conducting qualitative research, as lower temperature scores and restricted nucleus sampling could address issues in data modification and concordance analysis. However, for the former, issues of data modification persisted within each study, despite a lower temperature score being applied. Moreover, for a replication study, it is theoretically possible to modify ChatGPT's parameters *ad infinitum* with a view to arriving at findings that more closely reflect the original study. One may argue that in doing so, a set parameters can be identified that could support qualitative analysis. However, one cannot assume those same settings will suit a different qualitative analysis for a different study. Ultimately, the problem of the 'black box' would remain and modifying study parameters to suit anticipated findings could result in cherry-picking (Widdowson 2004) and methodological circularity (Curry, 2021).

Overall, ChatGPT appears not to be a solution to challenges in automated qualitative analysis in research using corpus approaches to analyse discourse – at least for now[3]. As a generative AI tool, ChatGPT is designed to produce human-like text and to emulate, to some degree, human-like intelligence. However, ChatGPT is presently unable to meet the standards of the human analyst (though it might be able to support a human-led analysis, with close supervision and scrutiny of results). Generative AI continues to develop apace, and it is impossible to know exactly what the future holds for tools such as ChatGPT. Notably, though, Korteling et al. (2021) argue that it is unlikely that such tools will be able to truly simulate human intelligence in the near future, meaning that both the language they produce and the tasks they undertake cannot be considered entirely substitutable for human language and intelligence.

These points notwithstanding, with further developments and modifications, it might be possible to develop generative AI technology for conducting high-quality qualitative analysis. However, to do so, developers will need to address issues of integrity, ethics, and transparency, in order to demystify the analytical black box. Crucially, to guide any future development of such tools for corpus linguistics and discourse analysis, due consideration will need to be paid to the current potential for generative AI tools to modify the linguistic data on which analyses are based. In linguistics research, the subtleties of language choices matter. With this in mind, the kinds of changes that ChatGPT made to the concordance lines and questions in the second and third of our case studies would, in the context of a 'real' study, serve to harm the integrity of our data and thus undermine our analysis and results. Moving forward, future attempts to overcome the barriers to using generative AI in linguistics research will need to address the application of such tools across diverse languages and research areas, where different epistemologies may have currency. In this endeavour, it may also benefit us to

---

[3] At the time of writing, OpenAI announced a number of future plans for ChatGPT, including allowing for longer input prompts. With new developments, it will be important to critically reflect on what this means for both data input and analytical approaches.

look beyond linguistics, to learn from the knowledge and practices surrounding generative AI use in other disciplines.

References

Antaki, F., Touma, S., Milad, D., El-Khoury, J., & Duval, R. (2023). Evaluating the performance of ChatGPT in ophthalmology: An analysis of its successes and shortcomings. *Ophthalmology Science*, 1-7. https://doi.org/10.1016/j.xops.2023.100324

Anthony, L. (2021). Programming for corpus linguistics. In M. Paquot, & Gries, S. T. (Eds.)., *A Practical Handbook of Corpus Linguistics* (pp. 181-207). Springer. https://doi.org/10.1007/978-3-030-46216-1_9

Anthony, L. (2023). AntConc (Version 4.2.2) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Agarwal, P., Asif, A., Parida, S., Sekhar, S., Dash, S. R., & Panda, S. (2023). Generative chatbot adaptation for odia language: A critical evaluation. In *2023 1st International Conference on Circuits, Power and Intelligent Systems (CCPIS)* (pp. 1-7). IEEE. https://doi.org/10.1109/CCPIS59145.2023.10291329

Baker, P. (2023). *Using corpora in discourse analysis*. 2nd Edition. Bloomsbury Publishing.

Baker, P., & Brookes, G. (2022). *Analysing language, sex and age in a corpus of patient feedback: A Comparison of approaches*. Cambridge University Press. https://doi.org/10.1017/9781009031042

Baker, P., Gabrielatos, C., & McEnery, T. (2013). *Discourse analysis and media attitudes: The representation of Islam in the British press*. Cambridge University Press. https://doi.org/10.1017/CBO9780511920103

Bin-Nashwan, S. A., Sadallah, M., & Bouteraa, M. (2023). Use of ChatGPT in academia: Academic integrity hangs in the balance. *Technology in Society*, *75*, 102370. https://doi.org/10.1016/j.techsoc.2023.102370

Blanchard, N., Brady, M., Olney, A. M., Glaus, M., Sun, X., Nystrand, M., ... & D'Mello, S. (2015). A study of automatic speech recognition in noisy classroom environments for automated dialog analysis. In *Artificial Intelligence in Education: 17th International Conference, AIED 2015, Madrid, Spain, June 22-26, 2015. Proceedings 17* (pp. 23-33). Springer International Publishing. https://doi.org/10.1007/978-3-319-19773-9_3

Brookes, G. (2018). Insulin restriction, medicalisation and the Internet: A corpus-assisted study of diabulimia discourse in online support groups. *Communication & Medicine*, *15*(1), 14-27. https://doi.org/10.1558/cam.33067

Brookes, G., & McEnery, T. (2019). The utility of topic modelling for discourse studies: A critical evaluation. *Discourse Studies*, *21*(1), 3-21. https://doi.org/10.1177/1461445618814032

Brookes, G., & McEnery, A. (2020). Corpus linguistics. In S. Adolphs, & D. Knight (Eds.), *The Routledge Handbook of English Language and Digital Humanities* (pp. 378-404). Routledge. https://doi.org/10.4324/9781003031758-20

Buruk, O. O. (2023). Academic Writing with GPT-3.5 (ChatGPT): Reflections on practices, efficacy and transparency. In *Proceedings of the 26th International Academic Mindtrek Conference* (pp. 144-153). Association of Computing Machinery. https://doi.org/10.1145/3616961.3616992

Crosthwaite, P., & Baisa, V. (2023). Generative AI and the end of corpus-assisted data-driven learning? Not so fast!. *Applied Corpus Linguistics*, *3*(3), 100066. https://doi.org/10.1016/j.acorp.2023.100066

Curry, N. (2021). *Academic writing and reader engagement: Contrasting questions in English, French and Spanish corpora*. Routledge. https://doi.org/10.4324/9780429322921

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. https://doi.org/10.1075/ijcl.20065.cur

Curry, N., & Riordan, E. (2021). Intelligent CALL systems for writing development: Investigating the use of Write and Improve for developing written language and writing skill. In K. Kelch, P. Byun, S. Cervantes, &. Safavi, S. (eds.), *CALL theory applications for online TESOL education* (pp. 252-273). IGI Global. https://doi.org/10.4018/978-1-7998-6609-1.ch011

Deng, L., & Liu, Y. (eds). (2018). *Deep learning in natural language processing*. Springer. https://doi.org/10.1007/978-981-10-5209-5

Dunn, K. C., & Neumann, I. B. (2016). *Undertaking discourse analysis for social research*. University of Michigan Press. https://doi.org/10.3998/mpub.7106945

Fogel, A. L., & Kvedar, J. C. (2018). Artificial intelligence powers digital medicine. *NPJ digital medicine*, *1*(1), 5. https://doi.org/10.1038/s41746-017-0012-2

Gillings, M., Mautner, G., & Baker, P. (2023). *Corpus-assisted discourse studies*. Cambridge University Press. https://doi.org/10.1017/9781009168144

Hämäläinen, P., Tavast, M., & Kunnari, A. (2023). Evaluating large language models in generating synthetic hci research data: a case study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-19). Association for Computing Machinery. https://doi.org/10.1145/3544548.3580688

Hoorens, S., Beuls, K., & Van Eecke, P. (2017). Constructions at work! Visualising linguistic pathways for computational construction grammar. In *Preproceedings of the 29th Benelux Conference on Artificial Intelligence. Groningen* (pp. 224-237).

Hunt, D. & Brookes, G. (2020). *Corpus, discourse and mental health*. Bloomsbury.

Hyland, K. (2002). What do they mean? Questions in academic writing. *Text & Talk*, *22*(4), 529-557. https://doi.org/10.1515/text.2002.021

Jackson, P. C. (2019). *Introduction to artificial intelligence*. Courier Dover Publications.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J, Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1, 7-36. https://doi.org/10.1007/s40607-014-0009-9

Kocoń, J., Cichecki, I., Kaszyca, O., Kochanek, M., Szydło, D., Baran, J., ... & Kazienko, P. (2023). ChatGPT: Jack of all trades, master of none. *Information Fusion*, 101861. https://doi.org/10.1016/j.inffus.2023.101861

Korteling, J. H., van de Boer-Visschedijk, G. C., Blankendaal, R. A., Boonekamp, R. C., & Eikelboom, A. R. (2021). Human-versus artificial intelligence. *Frontiers in artificial intelligence*, *4*, 622364. https://doi.org/10.3389/frai.2021.622364

Larsson, T., Plonsky, L., & Hancock, G. R. (2020). On the benefits of structural equation modeling for corpus linguists. *Corpus Linguistics and Linguistic Theory*, *17*(3), 683-714. https://doi.org/10.1515/cllt-2020-0051

Lin, P. (2023). ChatGPT: Friend or foe (to corpus linguists)?. *Applied Corpus Linguistics*, *3*(3). https://doi.org/10.1016/j.acorp.2023.100065

Lin, P., & Adolphs, S. (2023). Corpus linguistics. In L. Wei, Z. Hua, & Simpson, J. (eds.), *The Routledge handbook of applied linguistics* (pp. 296-308). Routledge. https://doi.org/10.4324/9781003082644-25

Misra, D. P., & Chandwar, K. (2023). ChatGPT, artificial intelligence and scientific writing: What authors, peer reviewers and editors should know?. *Journal of the Royal College of Physicians of Edinburgh*. https://doi.org/10.1177/1478271523118102

OpenAI (2023). OpenAI. Available online: https://openai.com/ [last accessed 20 November 2023]

Partington, A., Duguid, A. and Taylor, C. (2013), *Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies (CADS)*. John Benjamins. https://doi.org/10.1075/scl.55

Paulus, T. M., & Lester, J. N. (2016). ATLAS.ti for conversation and discourse analysis studies. *International Journal of Social Research Methodology*, *19*(4), 405-428. https://doi.org/10.1080/13645579.2015.1021949

Qureshi, M. A. (2023). ChatGPT: A tool to embrace or ban in Academia?. *Irish Journal of Academic Practice*, *11*(1), 1-3. https://doi.org/10.21427/VNG7-1V17

Rahman, M. M., Terano, H. J., Rahman, M. N., Salamzadeh, A., & Rahaman, M. S. (2023). ChatGPT and academic research: a review and recommendations based on practical examples. *Journal of Education, Management and Development Studies*, *3*(1), 1-12. https://doi.org/10.52631/jemds.v3i1.175

Rayson, P. (2008). From key words to key semantic domains. *International journal of corpus linguistics*, *13*(4), 519-549. https://doi.org/10.1075/ijcl.13.4.06ray

Shoenbill, K. A., Kasturi, S. N., & Mendonca, E. A. (2023). Artificial intelligence, machine learning, and natural language processing. In Daaleman, T.P., & Helton, M.R. (eds) *Chronic illness care: Principles and practice* (pp. 469-479). Springer International Publishing. https://doi.org/10.1007/978-3-031-29171-5_36

Shneiderman, B. (2020). Design lessons from AI's two grand goals: human emulation and useful applications. IEEE Tra*nsactions on* Tech*nology and* So*ciety*, 1, 73–82. https://doi.org/10.1109/TTS.2020.2992669

Siiman, L.A., Rannastu-Avalos, M., Pöysä-Tarhonen, J., Häkkinen, P., Pedaste, M. (2023). Opportunities and challenges for ai-assisted qualitative data analysis: An example from collaborative problem-solving discourse data. In: Huang, YM., Rocha, T. (eds) Innovative Technologies and Learning (pp. 87-96). Springer. https://doi.org/10.1007/978-3-031-40113-8_9

Taufek, T. E., Nor, N. F. M., Jaludin, A., Tiun, S., & Choy, L. K. (2021). Public perceptions on climate change: A sentiment analysis approach. *GEMA Online Journal of Language Studies*, *21*(4), 209-233. http://doi.org/10.17576/gema-2021-2104-11

Widdowson, H. G. (2004). *Text, context, pretext: Critical issues in discourse analysis*. Blackwell. http://doi.org/10.1002/9780470758427

Zappavigna, M. (2023). Hack your corpus analysis: How AI can assist corpus linguists deal with messy social media data. *Applied Corpus Linguistics*, *3*(3). https://doi.org/10.1016/j.acorp.2023.100067

Zhou, W., & Gao, B. (2023). Construction and application of English-Chinese multimodal emotional corpus based on artificial intelligence. *International Journal of Human–Computer Interaction*, 1-12. https://doi.org/10.1080/10447318.2023.2169526