A methodological integration of corpus analysis and content analysis

Sarunya Tarat, Nawarat Siritararatn and Woravut Jaroongkhongdach
Kasetsart University

Abstract

Corpus analysis, a method for analyzing natural language texts using corpus software tools, has been combined with several other research methods, leading to the integrated methods known as corpus-based critical discourse analysis, corpus-driven discourse analysis, and corpus assisted critical discourse analysis. From our extensive searches of databases, however, we found only a few papers that made use of both corpus analysis and content analysis. This paper aims to integrate corpus analysis with content analysis to produce corpus-based content analysis. The integration of the two is achieved by the use of keywords (based on keyness derived from corpus analysis) to create categories and codes (considered as belonging to content analysis). We piloted this integration with a set of data comprised of a hundred LGBTQ-related research articles from SAGE Online, published between 2001 and 2020. The integration of corpus analysis and content analysis contributes a new way of exploring texts.

1. Introduction

A methodological integration involves the integration of two or more research methods or approaches in a single study. An example of a methodological integration that is widely known and employed in research is the use of a questionnaire and an interview. Methodologically, the integration of multiple research methods can be useful for several reasons, one of which is reducing bias. In other words, the integration of multiple methods is "an alternative to traditional measures of reliability and validity, enabling researchers to overcome limitations associated with a single method or their own biases" (Baker & Levon, 2015, p.223). Additionally, integrating multiple methods can provide a better understanding of a particular phenomenon as "the use of multiple methods that do not share the same findings can enhance what is known about a given research question" (Turner et al., 2017, p.244). Also, in a single study, the integration of multiple methods allowed researchers to verify their findings (Yeasmin & Rahman, 2012). For example, Franceschi (2018) integrated traditional ESL/EFL speaking activities with counselling techniques as a pedagogical method for teaching physician-patient communication. With the use of the integrated method, it helps to verify the extent to which the students' linguistic output, communication skills, and rapport-building capabilities could be improved

In the field of linguistics, corpus analysis has been integrated with other research methods or approaches. For instance, the integration of corpus analysis with critical discourse analysis (CDA) known as corpus-based CDA has been used to study social practices based on text-based data sets (Kim, 2014). In addition, corpus analysis has been combined with a pattern-based approach as data-driven learning (DDL) to facilitate learning grammar and vocabulary using texts taken from data sets (Sun & Hu, 2020).

In this paper, we aim to reveal how corpus analysis can be integrated with content analysis to analyze a set of written texts. The aim of integrating corpus analysis with content analysis is to employ the strengths of the two methods by using the analysis of corpus data to create a methodological framework for content analysis.

This paper begins with a justification for corpus-based content analysis, followed by conceptualization of corpus linguistics and content analysis and their applications, the aid of computers in content analysis, and a framework for corpus-based content analysis.

2. Justification for corpus-based content analysis

Extensive database searches which we carried out for "corpus-based content analysis" revealed that relatively few research papers explicitly use this term compared to "corpus-based critical discourse analysis/corpus-based CDA", "corpus-driven discourse analysis", or "corpus-assisted discourse studies". This can be seen in the number of papers (in parentheses) found in searches for "Corpus-based content analysis" from the following databases: Scopus (4), SAGE Journals (60), Google Scholar (110), and Taylor & Francis Journals (0). In contrast, the terms "corpus-based critical discourse analysis" or "corpus-based CDA" feature more frequently, as can be seen from the numbers of papers in Scopus (82), SAGE Journals (340), Taylor & Francis Journals (604), and Google Scholar (882). Similarly, a search for "corpus-driven discourse analysis" uncovers 133 research articles in Google Scholar, 90 research articles in Scopus, two research articles in SAGE Journals, and one in Taylor & Francis Journals. As it is frequently used with discourse studies, the term "corpus-assisted discourse studies" appears in 1,490 research articles in Google Scholar along with 111 in Scopus, 48 in SAGE Journals, and 33 in Taylor & Francis Journals.

From these database searches, we can see a clear distinction between the number of research articles that explicitly mention "corpus-based content analysis" when compared with the others. This suggests that less attention has been given to the integration of corpus analysis with content analysis, implying that there remains a gap in integrating the two methods used in research. The integration of corpus analysis with content analysis is worth pursuing, in order to see if it offers a new way to study texts.

3. Conceptualization of corpus linguistics

Many people may understand corpus linguistics as a method to analyze authentic data using computer software. Nevertheless, corpus linguistics is more than just using software for text analysis. Based on our literature review, corpus linguistics has been characterized in two main ways: theory and approach (or methodology).

Stubb (1993) argues that "a corpus is not merely a tool of linguistic analysis, but an important concept in linguistic theory" (Stubb, 1993, pp. 23-24). Given that corpus linguistics is a theory, it allows linguists to perceive language from a theoretical viewpoint as can be seen in other theories (Tognini-Bonelli, 2001). Butler (2004), for instance, points out the relationships of corpus linguistics to two functional theories: Functional Grammar (FL) and Systemic Functional Grammar. Also, with its theoretical perspective, we can see that corpus linguistics has been applied in other fields (e.g., language acquisition or education). For example, corpus linguistics holds a theoretical status in language studies as well as teaching and learning language (McEnery, Xiao, & Tono, 2006).

Besides having theoretical status, corpus linguistics is defined as an approach or a methodology. Biber and Reppen (2015) propose that corpus linguistics is "a research approach [that] facilitates empirical investigations of language variation and use, resulting in research findings that have much greater generalizability and validity than would otherwise be feasible" (p.1). Viewed from the same perspective, but coined differently, corpus linguistics can "be conceptualized as a quantitative method of analysis" (Baker, 2006, p.8), or as "a methodological basis for studying language" (Tognini-Bonelli, 2001, p.1). Therefore, corpus linguistics has the notable characteristic of dealing with large amounts of text-based data (Vaughan & O'Keeffe, 2015), and has been used as a popular and increasingly widespread method of analysis (Baker, Gabrielatos, & McEnery, 2013; McEnery, Xiao, & Tono, 2006).

4. The application of corpus analysis with other fields

With regards to language analysis through corpus linguistics, the terms 'corpus-driven', 'corpus-based', and 'corpus-assisted' have been discussed as three types of corpus analysis. In corpus-driven analysis, linguists aim to use corpus data to build theory "from scratch", claiming that they are completely free from pre-corpus theoretical premises, and that they base their theories exclusively on the corpus data (Xiao, 2008, p.993). As it is an inductive approach (Gray & Biber, 2015), linguists who employ 'corpus-driven' analysis are committed to using the data as a whole (Tognini-Bonelli, 2001). With the integrity of the whole set of data, the corpus itself is viewed as "more than a repository of examples to back pre-existing theories or a probabilistic extension to an already well-defined system" (Tognini-Bonelli, 2001, p.84). This is the reason that corpus-driven analysis rejects evidence from sources outside of the corpus (Xiao, 2008). Without any pre-existing theories, corpus-driven analysts can look at texts by "discovering facts about language free from the influence of existing theoretical frameworks, which are considered to be based on intuitions, and, therefore, are not comprehensive or reliable" (McEnery & Gabrielatos, 2006, p.36).

For corpus-driven studies, it is suggested that "the corpus itself should be the sole source of our hypotheses about language" (McEnery & Hardie, 2012, p.6). Working on language analysis through corpus-driven analysis, linguists have to "process the raw text directly and the patterns of this uncontaminated text are able to be observed" (Sinclair & Carter, 2004, p. 191). Corpus-driven analysis indicates the descriptions of the whole set of word sequences in a corpus (Gray & Biber, 2015). This type of corpus analysis has shown "strong tendencies for words and grammatical constructions to pattern together in particular ways" (Biber, 2010, p.160). This is because corpus-driven analysis provides lexis, syntax, pragmatics, semantics, and discourse as being only one level of language description (Xiao, 2008). Consequently, it allows linguists to see salient or frequent items or patterns in a corpus (Baker, Gabrielatos, & McEnery, 2013) for language analysis.

Unlike corpus-driven analysis, corpus-based analysis aims to "expound, test or exemplify theories that were formulated before large corpora became available to inform language study" (Tognini-Bonelli, 2001, p.65). With its aims of testing theories, corpus-based analysis "insulates theory from data or standardizes data to fit theory" (Xiao, 2008, p.995). In particular, "corpus-based studies typically use corpus data in order to explore a theory or hypothesis in order to validate it, refute it or refine it" (McEnery & Hardie, 2012, p.6), allowing the corpus data to be used as a method to test existing hypotheses or theories (Baker, Gabrielatos, & McEnery, 2013). Baker et al. (2008) make clear the standpoint of corpus-based analysis:

> Corpus-based analysis does not merely involve getting a computer to objectively count and sort linguistic patterns along with applying statistical algorithms onto textual data. Subjective researcher input is, of course, normally involved at almost every stage of the analysis. The analyst, informed by the quantitative aspects mentioned earlier, has to decide what texts should go in the corpus, and what is to be analysed. He/she then needs to determine which corpus-based processes are to be applied to the data, and what the 'cut-off ' points of statistical significance should be. (p.277)

Defined as a deductive approach (Hubbard, 2010), the use of pre-selected lexical expressions can be explored (Gray & Biber, 2015) and the patterns of registers in a language further investigated (Biber, 2010). In corpus-based analysis, it is important to note that "the explanations must be developed using other methodologies and evidence from other sources, including intuitions" (Xiao, 2008, p.992). That is, corpus-based research is not just a form of quantitative analysis; it includes interpretation requiring researchers to develop interpretation skills (Hunt, 2015).

'Corpus-assisted' was first used in the term corpus-assisted discourse studies (CADS) by Alan Partington in 2004. It is defined as "that set of studies into the form and/or function of language as communicative discourse which incorporate the use of computerized corpora in their analyses" (Partington et al., 2013, p.10). In other words, corpus-assisted discourse analysis is the integration of corpus linguistics applied to discourse analysis (Ancarno, 2020), combing corpus data with discourse studies (Partington, 2006), but not tied to any particular school of discourse analysis (Partington et al., 2013). That is, "corpus techniques were only one sort amongst others and that CADS analysts employ as many [techniques] as required to obtain the most satisfying and complete results, hence corpus-assisted" (Partington et al., 2013, p.10).

CADS uncovers not only particular language or linguistic varieties, but also a particular situation, purpose or function repeatedly occurring in communicative discourse (Lischinsky, 2018, p. 61). The analysis of rhetorical patterns in research funding proposals is an example of CADS (Paltridge, 2012). CADS requires the analysis of data consisting of hundreds of concordance lines by hand to identify broader themes or patterns in the corpus which are not easily spotted through word frequencies, keyword analysis, and collocations (Baker et al. 2008). However, there has been much debate about statistical analyses in which "statistical analyses – frequency and key item lists, lexical and semantic sketches, even concordances – are still very much abstractions" (Partington, 2015, p.223).

With respect to corpus analysis, it might be better to talk about it as being associated with other disciplines, namely discourse analysis, critical discourse analysis, and discourse studies. For example, 'corpus-driven' is used with discourse analysis to examine the lexical and syntactic features of texts (Ikeo, 2019), and to compare two different language texts (Baker & Vessey, 2018). In addition, corpus-driven analysis is applicable to a discourse-historical approach (DHA), for example studying transcripts of television programs (Smith, 2010).

Corpus-based analysis has been integrated into discourse and critical discourse analysis to investigate language usage for identity construction (Afzaal et al., 2019; Kadiri, 2017) and analyze the representations of groups of people in media (Baker, 2020; Strom & Alcock, 2017).

Regarding discourse studies, corpus-assisted analysis is mainly used with discourse studies to examine the construction or representations of things or people in discourse (Brindle, 2016; Jaworska, 2016; Owen, 2014). A list of examples of studies is provided in Table 1.

Table 1. Corpus analysis in other fields

Author (s)	Corpus Analysis Used	Other Disciplines	Other Disciplines
Smith (2010)	Corpus-driven	Discourse analysis and Discourse-Historical Approach (DHA)	Study transcripts of Hugo Chávez's television programs
Baker & Vessey (2018)	Corpus-driven	Discourse analysis	Compare English and French Islamist extremist texts
Ikeo (2019)	Corpus-driven	Discourse analysis	Examine the lexical and syntactic features of present-tense narrative in fictional texts
Efe (2019)	Corpus-driven	Critical Discourse Analysis	Explore representations of Syrian asylum seekers in the Turkish press
Kadiri (2017)	Corpus-based	Critical Discourse Analysis	Investigate language usage and the construction of Muslim identities in the BBC sitcom *Citizen Khan*
Strom & Alcock (2017)	Corpus-based	Critical Discourse Analysis	Investigate the representation of Latin immigrant children in the United States media
Afzaal et al. (2019)	Corpus-based	Critical Discourse Analysis	Examine the construction of the China-Pakistan Economic Corridor in the Pakistani press
Baker (2020)	Corpus-based	Discourse analysis	Analyze representations of obesity in newspapers
Owen (2014)	Corpus-assisted	Discourse studies	Examine the dispute over intellectual property protection and global HIV/AIDS access to medicine
Brindle (2016)	Corpus-assisted	Discourse studies	Examine the discursive constructions of the student movement in English-language newspapers in Taiwan
Jaworska (2016)	Corpus-assisted	Discourse studies	Explore representations of hosts in commercial tourism discourse

5. Conceptualizations of content analysis

There has been debate over whether content analysis, as a research method, is quantitative or qualitative. Considered as a quantitative method, content analysis is seen as "one of the most popular and rapidly expanding techniques for quantitative research" (Neuendorf, 2017, p.19). Drisko and Maschi (2016) point out that content analysis involves related research techniques which make inferences systematic, credible, valid and replicable. Similarly, Krippendorff (2004) sees content analysis as "a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use" (p.18). In fact, the main goal of content analysis has not changed much for over 50 years as it is "a research technique for objective, systematic and quantitative description of the manifest content of communication" (Berelson, 1952, p.18). Due to its systematic procedures, content analysis can make valid inferences from texts (Weber, 1990) which allows analysts to replicate their research results (Riffe et al., 2019).

However, there is another important point to note. Even if researchers are allowed to use a particular framework or perspective to analyze data, it does not mean that "the researcher can anticipate all of the categories before the material is obtained and analyzed. In fact, imposing such constraints on the data could impede the validity of the results" (Downe-Wamboldt, 1992, p.316).

Content analysis is also referred to as a qualitative method. Considered in this way, content analysis is used with qualitative research techniques (Hsieh & Shannon, 2005), particularly as "a method for describing the meaning of qualitative material in a systematic way" (Schreier, 2012, p.1). More specially, qualitative content analysis is increasingly used for analyzing written material (Elo et al., 2014). Sometimes, it is best understood as coding analysis (Walsh & Goldberg, 2020). Regarding its popularity, qualitative content analysis is "one of numerous research methods used to analyze text data" (Hsieh & Shannon, 2005, p.1277). In qualitative content analysis, objectivity cannot be as important a concern as reflexivity (Schreier, 2012). So it is found that qualitative content analysis is under criticism due to questions about trustworthiness in evaluating qualitative content analysis (Elo, 2014).

Having considered these advantages anddisadvantages, this paper employs quantitative content analysis due to its ability to make inferences valid.

6. The application of content analysis to other fields

Content analysis has been used in fields such as psychology, education, and communication. For example, in the field of psychology, content analysis has been used to examine trends and issues at certain periods of time (Bledsoe et al., 2019; Cokley et al., 2014; Daniels et al., 2015). Similarly, in education, it has been used to explore the trends, topics, and theoretical perspectives of research articles (Lee & Taylor, 2013; Lin et al., 2014; Parsons et al., 2016). Additionally, quantitative content analysis used in the field of communication has analyzed the predominance of products in advertisements (Ahn et al., 2020). Meanwhile, qualitative content analysis has explored key concepts and themes in research articles in education (Bond et al., 2019). Also with a qualitative perspective, it has been used to examine the portrayal of violence in newspapers (Lee & Wong., 2020). A summary of these studies is presented in Table 2.

Table 2. Content analysis in other fields

Field	Author (s)	Issues Explored
Psychology	Cokley et al. (2014)	Popular areas of research and publication
	Daniels et al. (2015)	Critical issues and trends in the research and practice of military psychology
	Hargons et al. (2017)	Sexual orientation, sexual identity, and sexual minorities
	Cade et al. (2018)	Author characteristics and article characteristics
	Bledsoe et al. (2019)	Publication, methodological, and topical trends
Education	Lee & Taylor (2013)	Trends in moral education
	Lin et al (2014)	Trends in science education
	Chiang et al. (2016)	The instructional dimension and the patent dimension
	Parsons et al. (2016)	Topics, theoretical perspectives, research designs, and data sources
	Bond et al. (2019)	Key concepts and themes
Communication	Gao (2008)	Gender inequality in employment
	Cox & Ward (2019)	Portrayals of Black women
	Ahn et al. (2020)	Predominance of global fast-food brands
	Lee & Wong (2020)	The portrayal of domestic homicides
	Lucibello et al. (2021)	Positive and negative body image, weight stigma

7. Computers as an aid to content analysis

Researchers have made use of computers in content analysis, especially when texts are in a digital format (Pollach, 2012). With advances in computer software, computer coding for content analysis is now more reliable and efficient than human coding, and the analysis of large data sets is facilitated (Lacy et al., 2015). For example, Cheng et al. (2016) proposed the use of content analysis through the aid of text analysis software (Leximancer) which can help researchers reach detailed conceptual insights. In addition, text analysis software helps to "extract useful information from text data using natural language processing techniques, which can be broken down into document summaries, information retrieval, and trend analysis" (Lee & Sun Hong, 2020, p.1326).

Considered as computer-aided text analysis, corpus analysis "can contribute insights into the analysis of textual data in the social sciences" (Pollach, 2012, p.264). Thus we see that the use of corpus analysis can support human coding processes when working with large amounts of text for content analysis.

However, there are few studies using corpus analysis in combination with content analysis. The focus of these studies has particularly been in content validation and content-impact models. Pan and Qian (2017) adopted a corpus-based approach to investigate the content validity of the grammar section of a high-stakes test, the National Matriculation English Test (NMET) in China. With the use of corpus-based analysis, the grammar items could be classified into the target categories to examine whether the test items covered all the aspects of grammatical knowledge anticipated. Similarly, Roth (2017) applied qualitative content analysis to a corpus of widely adopted textbooks used in teaching morphemic analysis. For content validation, Yasunaga et al. (2019) developed a large, annotated corpus for data-driven models for scientific paper summarization using a content-impact model. The model captured both the papers' content as highlighted by the authors and the impact perceived by the research community (hybrid summarization).

Although there are studies using corpus analysis with content analysis, those studies focus on content validation of tests or textbooks and content-impact models derived from an annotated corpus. Therefore, we can say that there are still gaps in using corpus analysis, in particular corpus-based analysis combined with content analysis.

8. A framework for corpus-based content analysis

Within the domain of integrated methods, application of corpus analysis can complement content analysis. In this paper, using the strength of corpus analysis to extract salient lexical items or patterns (Baker, Gabrielatos, & McEnery, 2013), corpus-based analysis is chosen to extract lexical items from large amounts of data in order to create a methodological framework for content analysis to make content manifest (Riffe et al., 2019). Methodologically, corpus-based and content analysis can be combined into two stages where the first stage refers to the use of corpus-based analysis (Gray & Biber, 2015; Biber, 2010) while the second stage deals with content analysis (Hsieh & Shannon, 2005; Neuendorf, 2017).

In order to visualize the process of corpus-based content analysis, a framework is shown in Figure 1.

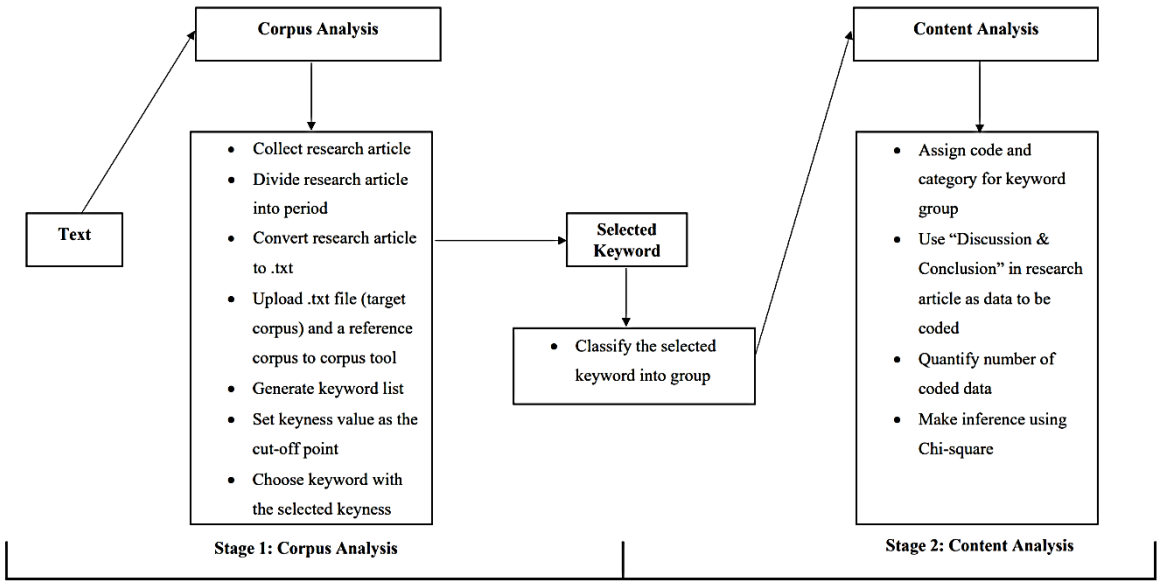

Figure 1. A framework for corpus-based content analysis

To further clarify how we attempted to integrate corpus analysis with content analysis in research, we piloted this integrated method to analyze journal content from the 21st century. A hundred LGBTQ-related research articles in SAGE Journals Online between 2001 and 2020 were divided into four time periods (spanning five years per period). We followed the two stages of corpus-based content analysis. The steps within each stage are described as follows.

Stage 1: Corpus Analysis

Step 1: collecting a hundred LGBTQ-related research articles in SAGE Journals Online 2001-2020
Step 2: dividing the research articles into four time periods spanning five years per period (2001-2005, 2006-2010, 2011-2015, 2016-2020)
Step 3: converting the article file formats from .pdf to .txt
Step 4: uploading the articles in .text and a reference corpus to a corpus tool (AntConc)
Step 5: generating keyword lists
Step 6: setting keyness value as the cut-off point
Step 7: choosing keywords with the selected keyness value

However, it is noted that classification of the selected keywords into groups must be completed before moving to the second stage. When the selected keywords are classified, these are used to create a framework for content analysis.

Stage 2: Content Analysis

Step 1: assigning codes and categories for the groups of keywords
Step 2: using "Discussion & Conclusion" sections in research articles as data to be coded
Step 3: quantifying numbers of coded data
Step 4: making inferences from the quantification by using Chi-square tests

As mentioned, corpus-based content analysis is an integrated method to open up a new way of creating a framework for content analysis. In this paper, we focused on keyword analysis from corpus-based processes. To generate keywords from the LGBTQ-related research articles corpus as the target corpus, we used the BE06 Corpus produced by Paul Baker (2009) as the benchmark corpus. The BE06 Corpus is a one-million-word reference corpus of general written British English (Baker, 2009). With the size of the BE06 Corpus as the reference corpus, it is acceptable to make comparisons with the LGBTQ-related research articles corpus as this target corpus is smaller in size.

Based on keyword analysis from corpus-based processes, we found that there are keywords relevant to different dimensions of LGBTQ. These keywords are assigned as codes aligned with specific categories. For example, Chinese, American, and Latino are keywords assigned as codes in the category of ethnic. In addition, keywords like equality, marriage, and adoption are found to be highly related to LGBTQ in terms of civil rights and law. With regards to keyword analysis, we found that LGBTQ is related to psychological issues in which keywords in this category include esteem, anxiety, respect, inclusive, and internalized.

9. Conclusion

It is acknowledged that each research method has strengths and weaknesses. Like other methods, corpus analysis and content analysis have benefits and drawbacks. In this paper, we attempt to integrate corpus analysis with content analysis to contribute a new and better understanding of possibilities in integrating research methodologies. In terms of methodological integration, corpus-based content analysis is found to be limited compared with other integrated methods, namely corpus-based CDA, corpus-based discourse analysis, and corpus-assisted discourse studies. For content analysis, corpus analysis is useful to reduce selection bias and time-consuming processes in creating a framework for analysis of content. The aid of corpus analysis helps content analysts to have linguistic grounds for developing codes for analytical frameworks. We hope this paper has provided some ideas on how corpus analysis can be integrated with content analysis.

References

Afzaal, M., Hu, K., Ilyas Chishti, M., & Khan, Z. (2019). Examining Pakistani news media discourses about China–Pakistan economic corridor: A corpus-based critical discourse analysis. *Cogent Social Sciences, 5*(1), 1-18. https://doi.org/10.1080/23311886.2019.1683940

Ahn, R. J., Nelson, M. R., & Ferguson, G. M. (2020). Local and standardized strategies: A content analysis of newspaper food and beverage advertising in Jamaica. *Newspaper Research Journal, 41*(2), 179-203. https://doi.org/10.1177/0739532920919828

Ancarno, C. (2020). Corpus-assisted discourse studies. In A. De Fina & A. Georgakopoulou (Eds.), *The Cambridge handbook of discourse studies* (pp. 165-185). Cambridge University Press.

Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.

Baker, P. (2009). The BE06 corpus of British English and recent language change. *International Journal of Corpus Linguistics, 14*(3), 312-337. https://doi.org/10.1075/ijcl.14.3.02bak

Baker, P. (2020). Analysing representations of obesity in the Daily Mail via corpus and down-sampling methods. In J. Egbert & P. Baker (Eds.), *Using corpus methods to triangulate linguistic analysis* (pp. 85-108). Routledge.

Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyżanowski, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society, 19*(3), 273-306. https://doi.org/10.1177/0957926508088962

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

Baker, P., & Levon, E. (2015). Picking the right cherries? A comparison of corpus-based and qualitative analyses of news articles about masculinity. *Discourse & Communication, 9*(2), 221-236. https://doi.org/10.1177/1750481314568542

Baker, P., & Vessey, R. (2018). A corpus-driven comparison of English and French Islamist extremist texts. *International Journal of Corpus Linguistics, 23*(3), 255-278. https://doi.org/10.1075/ijcl.17108.bak

Biber, D. (2010). Corpus-based and corpus-driven analyses. In B. Heine & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis* (pp. 159-192). Oxford University Press Inc.

Biber, D., & Reppen, R. (2015). *The Cambridge handbook of English corpus linguistics*. Cambridge University Press.

Berelson, B. (1952). *Content analysis in communication research*. Free Press.

Bledsoe, K. G., Logan-McKibben, S., McKibben, W. B., & Cook, R. M. (2019). A content analysis of school counseling supervision. *Professional School Counseling, 22*(1), 1-8. https://doi.org/10.1177/2156759x19838454

Bond, M., Zawacki-Richter, O., & Nichols, M. (2019). Revisiting five decades of educational technology research: A content and authorship analysis of the British Journal of educational technology. *British Journal of Educational Technology, 50*(1), 12-63. https://doi.org/10.1111/bjet.12730

Brindle, A. (2016). A corpus analysis of discursive constructions of the sunflower student movement in the English-language Taiwanese press. *Discourse & Society, 27*(1), 3-19. https://doi.org/10.1177/0957926515605957

Butler, C. S. (2004). Corpus studies and functional linguistic theories. *Functions of Language, 11*(2), 147-186. https://doi.org/10.1075/fol.11.2.02but

Cade, R., Gibson, S., Swan, K., & Nelson, K. (2018). A content analysis of counseling outcome research and evaluation (CORE) from 2010 to 2017. *Counseling Outcome Research and Evaluation, 9*(1), 5-15. https://doi.org/10.1080/21501378.2017.1413643

Cheng, M., Edwards, D., Darcy, S., & Redfern, K. (2016). A tri-method approach to a review of adventure tourism literature: Bibliometric analysis, content analysis, and a quantitative systematic literature review. *Journal of Hospitality & Tourism Research, 42*(6), 997-1020. https://doi.org/10.1177/1096348016640588

Chiang, F., Zhu, G., Wang, Q., Cui, Z., Cai, S., & Yu, S. (2016). Research and trends in mobile learning from 1976 to 2013: A content analysis of patents in selected databases. *British Journal of Educational Technology, 47*(6), 1006-1019. https://doi.org/10.1111/bjet.12311

Cokley, K., Awosogba, O., & Taylor, D. (2014). A 12-Year content analysis of the Journal of Black psychology (2000-2011). *Journal of Black Psychology, 40*(3), 215-238. https://doi.org/10.1177/0095798413486157

Cox, V., & Ward, L. M. (2019). A wholistic view of Black women on scripted TV: A content analysis. *Journal of Black Psychology, 45*(6-7), 540-570. https://doi.org/10.1177/0095798419887072

Daniels, J. A., Spero, R. A., Leonard, J. M., & Schimmel, C. J. (2015). A content analysis of military psychology: 2002–2014. *Military Psychology, 27*(6), 366-375. https://doi.org/10.1037/mil0000091

Downe-Wamboldt, B. (1992). Content analysis: Method, applications, and issues. *Health Care for Women International, 13*(3), 313-321. https://doi.org/10.1080/07399339209516006

Drisko, J. W., & Maschi, T. (2016). Content analysis. Oxford University Press.

Efe, İ. (2019). A corpus-driven analysis of representations of Syrian asylum seekers in the Turkish press 2011–2016. *Discourse & Communication, 13*(1), 48-67. https://doi.org/10.1177/1750481318801624

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). Qualitative content analysis: A focus on trustworthiness. *SAGE Open, 4*(1), 1-10. https://doi.org/10.1177/2158244014522633

Franceschi, D. (2018). Physician-patient communication: An integrated multimodal approach for teaching medical English. *System, 77*, 91-102. https://doi.org/10.1016/j.system.2018.02.011

Gao, Z. (2008). Gender discrimination in Chinese recruitment advertisements: A content analysis. *Journal of Asia-Pacific Business, 9*(4), 395-418. https://doi.org/10.1080/10599230802453638

Gray, B., & Biber, D. (2015). Phraseology. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 125-142). Cambridge University Press.

Hargons, C., Mosley, D. V., & Stevens-Watkins, D. (2017). Studying sex: A content analysis of sexuality research in counseling psychology. *The Counseling Psychologist, 45*(4), 528-546. https://doi.org/10.1177/0011000017713756

Hsieh, H., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277-1288. https://doi.org/10.1177/1049732305276687

Hubbard, H. (2010). Stance and style: A corpus-driven perspective on television coverage of the 2009 South African general election. *Language Matters, 41*(1), 3-24. https://doi.org/10.1080/10228191003786268

Hunt, S. (2015). Representations of gender and agency in the Harry Potter series. In P. Baker & A. McEnery (Eds.), *Corpora and discourse studies: Integrating discourse and corpora* (pp. 266-284). Palgrave Macmillan.

Ikeo, R. (2019). 'Colloquialization' in fiction: A corpus-driven analysis of present-tense fiction. *Language and Literature: International Journal of Stylistics, 28*(3), 280-304. https://doi.org/10.1177/0963947019868894

Jaworska, S. (2016). A comparative corpus-assisted discourse study of the representations of hosts in promotional tourism discourse. *Corpora, 11*(1), 83-111. https://doi.org/10.3366/cor.2016.0086

Kadiri, S. M. (2017). *A corpus-based analysis of the construction of identities in the BBC sitcom Citizen Khan* (Publication No. 28277593) [Doctoral dissertation, Lancaster University]. ProQuest Dissertations and Theses Global.

Kim, K. H. (2014). Examining US news media discourses about North Korea: A corpus-based critical discourse analysis. *Discourse & Society, 25*(2), 221-244. https://doi.org/10.1177/0957926513516043

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). SAGE Publications, Inc.

Lacy, S., Watson, B. R., Riffe, D., & Lovejoy, J. (2015). Issues and best practices in content analysis. *Journalism & Mass Communication Quarterly*, 1-21. https://doi.org/10.1177/1077699015607338

Lee, C., & Taylor, M. J. (2013). Moral education trends over 40 years: A content analysis of the Journal of moral education (1971–2011). *Journal of Moral Education, 42*(4), 399-429. https://doi.org/10.1080/03057240.2013.832666

Lee, C., & Wong, J. S. (2020). 99 reasons and he Ain't one: A content analysis of domestic homicide news coverage. *Violence Against Women, 26*(2), 213-232. https://doi.org/10.1177/1077801219832325

Lee, S., & Sun Hong, H. (2020). Text network analysis of research topics and trends on global health nursing literature from 1974-2017. *The Journal of Advanced Nursing, 77*, 1325-1334. https://doi.org/10.1111/jan.14685

Lin, T., Lin, T., & Tsai, C. (2014). Research trends in science education from 2008 to 2012: A systematic content analysis of publications in selected journals. *International Journal of Science Education, 36*(8), 1346-1372. https://doi.org/10.1080/09500693.2013.864428

Lucibello, K. M., Vani, M. F., Koulanova, A., DeJonge, M. L., Ashdown-Franks, G., & Sabiston, C. M. (2021). #quarantine15: A content analysis of Instagram posts during COVID-19. *Body Image, 38*, 148-156. https://doi.org/10.1016/j.bodyim.2021.04.002

McEnery, T., & Gabrielatos, C. (2006). English corpus linguistics. In B. Aarts & A. McMahon (Eds.), *The handbook of English linguistics* (pp. 33-71). Wiley-Blackwell.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Taylor & Francis.

Neuendorf. (2017). *The content analysis guidebook* (2nd ed.). SAGE Publications, Inc.

Owen, T. (2014). The 'Access to medicines' campaign vs. Big Pharma. *Critical Discourse Studies, 11*(3), 288-304. https://doi.org/10.1080/17405904.2014.915860

Paltridge, B. (2012). *Discourse Analysis* (2nd ed.). Bloomsbury Academic.

Pan, M., & Qian, D. D. (2017). Embedding corpora into the content validation of the grammar test of the national matriculation English test (NMET) in China. *Language Assessment Quarterly, 14*(2), 120-139. https://doi.org/10.1080/15434303.2017.1303703

Parsons, S. A., & Gallagher, M. A. (2016). A content analysis of nine literacy journals, 2009-2014. *Journal of Literacy Research, 48*(4), 476-502. https://doi.org/10.1177/1086296x16680053

Partington, A. (2004). Corpora and discourse, a most congruous Beast. In A. Partington, J. Morley and L. Haarman (eds) *Corpora and Discourse* (pp. 11–20). Peter Lang.

Partington, A. (2006). *The linguistics of Laughter: A corpus-assisted study of Laughter-talk*. Routledge.

Partington, A. (2015). Corpus-assisted comparative case studies of representations of the Arab world. In P. Baker & A. McEnery (Eds.), *Corpora and discourse studies: Integrating discourse and corpora* (pp. 220-243). Palgrave Macmillan.

Partington, A., Duguid, A., & Taylor, C. (2013). Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies (CADS). John Benjamins Publishing.

Pollach, I. (2012). Taming textual data: The contribution of corpus linguistics to computer-aided text analysis. *Organizational Research Methods, 15*(2), 263-287. https://doi.org/10.1177/1094428111417451

Riffe, D., Lacy, S., Fico, F., & Watson, B. (2019). *Analyzing media messages: Using quantitative content analysis in research* (4th ed.). Routledge.

Roth, D. (2017). Morphemic analysis as imagined by developmental reading textbooks: A content analysis of a textbook corpus. *Journal of College Reading and Learning, 47*(1), 26-44. https://doi.org/10.1080/10790195.2016.1218807

Schreier, M. (2012). *Qualitative content analysis in practice*. SAGE.

Sinclair, J., & Carter, R. (2004). *Trust the text: Language, corpus and discourse*. Routledge.

Smith, D. N. (2010). *A corpus-driven discourse analysis of transcripts of Hugo Chávez's television programme 'Aló Presidente'* [Doctoral dissertation, The University of Birmingham]. https://etheses.bham.ac.uk/id/eprint/731/

Strom, M., & Alcock, E. (2017). Floods, waves, and surges: The representation of Latin@ immigrant children in the United States mainstream media. *Critical Discourse Studies, 14*(4), 440-457. https://doi.org/10.1080/17405904.2017.1284137

Stubbs, M. (1993). British traditions in text analysis: From Firth to Sinclair. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 1-36). John Benjamins Publishing Company.

Sun, X., & Hu, G. (2020). Direct and indirect data-driven learning: An experimental study of hedging in an EFL writing class. *Language Teaching Research*, 136216882095445. https://doi.org/10.1177/1362168820954459

Tognini-Bonelli, E. (2001). Corpus linguistics at work. John Benjamins Publishing. Turner, S. F., Cardinal, L. B., & Burton, R. M. (2017). Research design for mixed methods. *Organizational Research Methods, 20*(2), 243-267. https://doi.org/10.1177/1094428115610808

Vaughan, E., & O'Keeffe, A. (2015). Corpus Analysis. In K. Tracy, C. Ilie, & T. Sandel (Eds.), *The international encyclopedia of language and social interaction* (pp. 1-17). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118611463/wbielsi168

Walsh, M., & Goldberg, R. M. (2020). Rethinking counseling recruitment for transgender clients: Using content analysis to investigate trends. *Journal of LGBT Issues in Counseling, 14*(3), 210-227. https://doi.org/10.1080/15538605.2020.1790466

Weber, R. P. (1990). *Basic content analysis*. SAGE Publications, Inc.

Xiao, R. Z. (2008). Theory-driven corpus research: Using corpora to inform aspect theory. In A. Ludeling, & M. Kyto (Eds.), *Corpus Linguistics: An International Handbook (Vol. 1). (Handbooks of Linguistics and Communication Science)* (pp. 987-1008). Mouton de Gruyter.

Yasunaga, M., Kasai, J., Zhang, R., Fabbri, A. R., Li, I., Friedman, D., & Radev, D. R. (2019). ScisummNet: *A large, annotated corpus and content-impact models for scientific paper summarization with citation networks* (P06-1005). Cornell University. https://arxiv.org/abs/1909.01716

Yeasmin, S., & Rahman, K. F. (2012). 'Triangulation' research method as the tool of social science research. *BUP JOURNAL, 1*(1), 154-163.