



## OPEN ACCESS

## EDITED BY

Antonio Bova,  
Catholic University of the Sacred Heart, Italy

## REVIEWED BY

Gisela Redeker,  
University of Groningen, Netherlands  
Aleksandra Bagasheva,  
Sofia University "St. Kliment Ohridski", Bulgaria

## \*CORRESPONDENCE

Muhammad Afzaal  
✉ muhammad.afzaal1185@gmail.com

RECEIVED 02 May 2024

ACCEPTED 28 June 2024

PUBLISHED 18 July 2024

## CITATION

Fu Y, Afzaal M and El-Dakhs DAS (2024)  
Investigating discourse markers "you know"  
and "I mean" in mediatized English political  
interviews: a corpus-based comparative  
study.  
*Front. Commun.* 9:1427062.  
doi: 10.3389/fcomm.2024.1427062

## COPYRIGHT

© 2024 Fu, Afzaal and El-Dakhs. This is an  
open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](#). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# Investigating discourse markers "you know" and "I mean" in mediatized English political interviews: a corpus-based comparative study

Yanli Fu<sup>1</sup>, Muhammad Afzaal<sup>2,3\*</sup> and Dina Abdel Salam El-Dakhs<sup>3</sup>

<sup>1</sup>School of Foreign Languages, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu province, China, <sup>2</sup>Institute of Corpus Studies and Applications, Shanghai International Studies University, Shanghai, China, <sup>3</sup>College of Humanities and Sciences, Prince Sultan University, Riyadh, Saudi Arabia

This study employs a corpus-based approach to examine and compare the use of two discourse markers (DMs), "you know" and "I mean", within the context of two mediatized English political interviews. The analysis encompasses frequencies, functions, co-occurrences, and positional distributions of these DMs. The study utilizes specialized corpora from two political interview programs: CGTN's The Point with Liu Xin and BBC's HARDtalk. The frequency analysis reveals that "you know" is statistically more prevalent than "I mean" in both programs, reflecting the spontaneity, interactivity, and need for clarification characteristic of political interviews. Notably, the Chinese interviewer (IR) uses "you know" more extensively, possibly due to a cultural preference for ensuring mutual understanding and engaging the audience, while the British IR employs "I mean" slightly more frequently, likely reflecting a tendency to clarify or reframe statements for precision. Functionally, these DMs serve diverse purposes such as hedging, agreeing, and monitoring across various domains including interpersonal, sequential, and rhetorical. Positional analysis shows "you know" typically appearing medially and "I mean" often in initial positions. These results underscore the distinctive interviewing styles of the two IRs and the pivotal role of these DMs in fulfilling a spectrum of communicative functions. This research offers valuable insights into the interviewer's perspective in political interviews.

## KEYWORDS

corpora, political interviews, discourse markers, frequency, functions, communication

## Introduction

Discourse markers (DMs) are a fundamental aspect of language that help speakers to convey meaning and organize their ideas in conversations, speeches, and written texts. [Verdonik \(2022\)](#) defines discourse markers (DMs) as small words used in spoken language that have non-propositional functions in discourse. DMs can take various grammatical forms, such as adverbs, conjunctions, and multi-word phrases. Researchers agree that individual DMs can serve multiple functions concurrently (multifunctionality) or in various discourse contexts (polyfunctionality) due to their many possible meanings ([Fischer, 2006](#); [Bolden, 2015](#); [Crible, 2017](#)). [Macaulay \(2013\)](#) highlights the numerous spontaneous, interactional, social, sociable, and polite functions that discourse markers (DMs) may simultaneously serve, making it

challenging to allocate each occurrence to a primary function. In recent years, an increasing scholarly interest has been observed in the examination of multifunctionality and the frequency of DM usage across diverse genres and interactive contexts. These investigations have delved into the various ways in which speakers employ DMs to navigate epistemic stances, convey (inter)subjectivity and connectivity, manage information, and regulate interpersonal exchanges in conversational settings. For example, Schourup (1982) investigated DMs in English conversations and focused on “well” and “you know” through recorded and introspective data, and explained why each particle can serve various context-specific functions. House (2013) investigated how ELF (English as a *lingua franca*) speakers strategically reinterpret DMs like “okay,” “so,” “yes/yeah” to express connectedness and subjectivity and improve their pragmatic competence during academic consultation hours. Crible et al. (2019) explained the use of DMs and their multiple functions in translation in TED talks, emphasizing the impact of multilingual contexts on the diverse nature of DMs in varied linguistic environments and underscoring the significance of translation and annotation in the analysis. Gao and Lee (2019), using data from Ohio State University, explored the use of DMs for expressing emotions and intentions, proposing that DMs reflect human emotions within specific linguistic contexts, thus influencing their interpretation.

The study uses political interviews as a genre because they are investigable by their proper functionality in developing institutional discourse, e.g., mediated discourse, media-related communication, and political talks (Furkó, 2020). According to Feldman (2016), political interviews are produced within an institutional setting that describes various political roles of the participants and their underlying political motives. The interviewer (IR) represents a particular media organization that establishes the professional values of neutrality, integrity, and truthfulness. Conversely, the interviewee (IE) embodies the political ideology of a political party, intent on propagating its concepts, activities, messages, and slogans. In essence, political interviews can be delineated as dyadic interactions marked by distinct turn-taking patterns and constraints. The interaction between interviewers (IR) and interviewees (IE) is marked by a mutual implicit understanding and interdependence.

As Šandová (2014) contends, political interviews can be perceived as a distinct genre in which political speakers employ diverse linguistic and discourse strategies to exert influence and persuade their audience regarding their political ideologies. Consequently, it is not surprising to witness a growing number of studies dedicated to investigating the linguistic techniques employed by IEs to achieve specific communicative objectives. For example, Simon-Vandenberg (2000) conducted a study on the parenthetical function of the DM “I think” in political discourse and found that politicians prefer using the phrase “I think” to express their opinions rather than stating facts, thus personalizing the discourse. Additionally, Fetzer (2014) examined the distribution, collocates, and functions of “I think,” “I mean” and “I believe” used by politicians in political discourse, encompassing monologic speeches and dialogic face-to-face interviews. His study disclosed that “I mean” was the second most frequent occurrence in dialogic interviews but never appeared in speeches. The most frequent collocate for “I mean” was found to be “but” indicating a politician’s intention to adopt a different position. Conversely, “I believe” and “I think” tended to collocate with “and,” suggesting a firm backing of the preceding argument. English politicians were found to employ more

interpersonal DMs, such as “you know,” indicating an emphasis on building relationships with IRs and reflecting the conversationalization in political discourse. Both Iranian and English politicians used DMs from cognitive categories, such as “I mean,” to clarify their points and ensure comprehension by the audience. Furkó (2015) investigated the functions of pragmatic markers based on English and Hungarian corpora consisting of political interviews and related online feedback comments, with a specific focus on “of course,” “like,” and “sure.” “Of course,” emerged as the most frequently utilized marker by interviewees (IEs), serving various communicative purposes, such as self-correction and lexical search. Additionally, the study revealed how DMs contributed to politically manipulated content in political discussions through the use of quotation markers, decontextualization, recontextualization, legitimization of viewpoints, and audience categorization. Nevertheless, there is a scarcity of significant studies that explore the use of DMs in political interviews. Even among the studies that have examined DMs, the focus has largely been on how politicians or IEs respond to adversarial questioning from IRs, framing their responses as questions and holding IRs accountable for their statements. Little attention has been paid to the practices of IRs and their strategic utilization of linguistic devices during interviews.

This study, therefore, intends to fill the identified gaps by exploring the following research questions in two kinds of political TV interviews.

- 1 How frequently do the two interviewers employ the discourse markers “you know” and “I mean”? Are there any differences in their usage between the two interviewers?
- 2 What functions do these discourse markers serve as used by the two interviewers? Are there any disparities in their usage between the two interviewers?
- 3 What are the co-occurrence patterns and positional distributions of these discourse markers? Are there any differences between the two interviewers in these aspects?

## Materials and methods

### Corpora of the study

The data utilized in this study is derived from an extensive and ongoing large-scale research project encompassing a total of 120 political interviews conducted in the English language. These interviews were conducted over the period spanning 2020 to 2022, a period characterized by the challenges imposed by the global pandemic. Specifically, 60 of these interviews, constituting 168,095 tokens, were sourced from *The Point with Liu Xin*, a program broadcast on CGTN (China Global Television Network), which is hosted by the accomplished Chinese female IR, Liu Xin. The remaining 60 interviews, totalling 229,762 tokens, were procured from *HARDtalk*, a program broadcast on the BBC (British Broadcasting Corporation), and hosted by the esteemed British male IR, Stephen Sackur. Each of the 60 interviews collected from both programs was meticulously chosen to maintain a balanced representation of two distinct cultural groups. Thirty IEs (IEs) belong to the eastern cultural group, encompassing China, Pakistan, Singapore, and Thailand, while the remaining 30 IEs are from the western cultural group, which

includes the United Kingdom, the United States, Germany, France, and Australia. The selection of these two programs was guided by their institutional similarities, as delineated by Feldman (2016).

Moreover, a subset of twenty interviews, randomly selected from *The Point with Liu Xin*, has been designated as the Chinese Political Interview Corpus (CPIC). Within the CPIC, ten IEs represent the eastern cultural cohort and are denoted as CPIC-E, while the remaining ten interviews comprise IEs from the western cultural context, designated as CPIC-W. Similarly, twenty interviews, randomly extracted from *HARDtalk*, constitute the British Political Interview Corpus (BPIC). In the BPIC, ten IEs are from the eastern cultural group, referred to as BPIC-E, and ten IEs originate from the western cultural cohort, denoted as BPIC-W. The composition of these two corpora for the present study is provided in Table 1.

## Methodology

The present investigation employs Flick's (2014) "one-after-the-other" categorization, commonly recognized as the Sequential Quan—Qual MMR (Mixed Methods Research) design in accordance with the classification outlined by Ivankova and Creswell (2009). In other words, it employs a two-step research approach, commencing with a quantitative methodology and subsequently integrating a qualitative method.

The utilization of corpora in the examination of political discourse by discourse analysts has been a prevalent practice since the early twentieth century. Using corpora enables researchers to analyse usage in context and identify usage patterns, such as collocation with other words, text type, and speaker characteristics. Corpus-based methods have emerged as a valuable technique for analysing distinctions in language usage, particularly concerning speaker identity attributes, such as gender (Baker, 2006). This approach has been used in a variety of linguistic fields, such as discourse analysis, forensics, grammar, language teaching, lexicography, literary studies, pragmatics, sociolinguistics, and translation studies (O'Keeffe and McCarthy, 2010; Afzaal et al., 2021, 2022; Zhang et al., 2023; Barbara et al., 2024). Since the 1980s, corpus-based research has examined discourse markers (DMs) at various levels of complex linguistic structures in written and oral texts. Fischer (2006) notes that such research has become a mainstay of corpus-based research, with numerous publications addressing the topic. Discourse markers are popular in corpus linguistics due to their high frequency and characteristic use in spontaneous conversation (Rühlemann, 2019). This approach is particularly well-suited to studying pragmatic markers, as their usage could be more easily discernible through intuition, grammaticality

judgments, or assessments of speaker characteristics. It is recorded that Corpora of Spoken Language have profoundly influenced the studies of DMs, particularly cross-language research (Aijmer, 2013; Raza et al., 2022; Diab et al., 2023; Mahlous, 2024).

This methodology aligns seamlessly with the present study, given the specific identity attributes of the two IRs. Moreover, the corpus-based method is remarkably conducive to triangulation, offering a means to obtain results of relatively neutral disposition and significantly mitigating the subjective biases inherent in researchers (Baker, 2006). Additionally, it effectively complements and synergizes with both quantitative and qualitative research approaches.

## Analysis procedures

When the database is set for analysis, the frequency statistics for a general quantitative distribution of the two discourse markers (DMs), namely "you know" and "I mean," across the four sub-corpora are computed using LancsBox. The identification of these DMs adheres to Cribble's (2018) operational definition, which characterizes them as grammatically heterogeneous, syntactically optional, polyfunctional, imbued with procedural meaning and discourse-structuring functionality, and contributing to the dynamics of speaker-hearer interaction. Canonical usages are excluded if "you know" is an integral part of the clause's syntactical construction and cannot be omitted without altering its semantics, as exemplified by the abbreviated structure "Do you know..." (Schiffrin, 1987). In practical terms, identifying "you know" as a DM is not a challenging task (Macaulay, 2002). Similarly, instances of non-pragmatic marking with "I mean" are excluded if they are syntactically integrated into the clause, such as the formulaic expression "you know what I mean..." (Beeching, 2016). Utilizing the KWIC tool in LancsBox (Brezina, 2018), all occurrences of 'you know' and 'I mean' are retrieved from the two corpora in the form of concordance lines, and the absolute frequency of these identified DMs is calculated. LancsBox provides an index number and file name for each generated concordance line, greatly facilitating the retrieval of each DM in specific contexts for subsequent functional analysis.

The process of normalization is a commonly employed methodological choice for comparing the frequency of DM usage by the two IRs, which involves assigning numerical values as a basis for comparison (Baker, 2006; Brezina, 2018). The normalized results are then compared to establish frequency rankings. To statistically compare the usage frequency of "you know" and "I mean" between the two IRs, the log-likelihood statistic is employed, using an online tool available on the UCREL website,<sup>1</sup> which provides significant difference tests in frequency between two corpora through the use of four figures (Rayson, 2016). The produced log-likelihood (LL) value must exceed 3.84 for the difference to be deemed statistically significant at a significance level of  $p < 0.05$ . This test is particularly well-suited for corpus research because it accounts for size differences between corpora, as demonstrated by Aijmer (2002), Buysse (2012, 2017), and Öztürk and Durmuşoğlu Köse (2021).

In terms of functional analysis, the current study employs Cribble and Degand's (2019) two-dimensional DM functional taxonomy,

TABLE 1 The composition of the two corpora.

Corpus code	Number of tokens	Number of types	Proportion (%)
CPIC-E	28,622	3,541	42
CPIC-W	40,257	4,309	58
CPIC	68,879	5,999	100.00
BPIC-E	37,748	4,339	46
BPIC-W	43,433	4,673	54
BPIC	81,181	6,939	100.00

<sup>1</sup> <http://ucrel.lancs.ac.uk/llwizard.html>

TABLE 2 Crible and Degand's (2019) two-dimensional DM functional taxonomy.

Domains	Functions
Ideational (IDE): state of affairs in the world, semantic relations between real/external events	Addition (ADD), alternative (ALT), cause (CAU), concession (CCS), condition (CND), consequence (CSQ), contrast (CTR), hedging (HDG), monitoring (MNT), specification (SPE), temporal (TMP), agreeing (AGR), disagreeing (DIS), topic (TOP), quoting (QUO)
Rhetorical (RHE): speaker's meta-comments on the on-going speech	
Sequential (SEQ): structuring of local and global discourse segments: topics and turns	
Interpersonal (INT): interactive management of the exchange & the speaker-hear relationship	

following a top-down corpus-linguistic approach. This model is categorized into four domains: ideational, rhetorical, sequential, and interpersonal, with fifteen functions attributed to DMs, as outlined in Table 2. Furthermore, the documented functions of “you know” and “I mean” in the literature are incorporated into this two-dimensional functional taxonomy following rigorous revisions of previous studies. For instance, the “reformulation” function of the DM “I mean” aligns with the “rhetorical-alternative” function, and the “topic shift” function of the DM “you know” corresponds to the “sequential-topic” function within this model (Fox Tree and Schrock, 2002). It is important to note that the two DMs may not fulfil all fifteen functions presented in this model, but they are included here for the sake of completeness.

## Results

### Frequency of the DMs “you know” and “I mean” in the two corpora

The corpus under investigation, denoted as CPIC, comprises a total of 68,879 tokens dispersed across 20 texts. Within this corpus, the DM “you know” is manifested 141 times, with CPIC-E contributing 47 instances and CPIC-W comprising 94 occurrences, as shown in Table 3. Additionally, CPIC contains 22 tokens of the DM “I mean,” with 9 instances in CPIC-E and 13 instances in CPIC-W, as illustrated in Table 3. Similarly, in BPIC consisting of a total of approximately 81,181 words, comprises 60 instances of the DM “you know,” where 18 instances are found in BPIC-E and 42 in BPIC-W. Moreover, BPIC encompasses 37 occurrences of the DM “I mean,” distributed with 13 instances in BPIC-E and 24 in BPIC-W, as depicted in Table 3. The overall distribution of these two DMs in BPIC closely mirrors that in CPIC, with the two IRs’ use of the DM “you know” being slightly higher than “I mean” in both corpora. Table 3 showcases the statistical difference in the comparative distribution of the use of ‘you know’ and ‘I mean’ in CPIC and BPIC.

The report in Table 3 also presents the normalized frequency, a crucial metric for evidentiary and investigative purposes, in line with Macaulay (2002). Importantly, this metric serves as the exclusive and comprehensive approach to assess the differential utilization of these DMs by the IRs. A comparative analysis of the two corpora shows that

TABLE 3 The frequency of DMs “you know” and “I mean” in corpora.

Corpus	Tokens	R.F. *of “you know”	N.F. *of “you know”	R.F. of “I mean”	N.F. of “I mean”
CPIC	68,879	141	20.5	22	3.2
CPIC-E	28,622	47	16.4	9	3.1
CPIC-W	40,257	94	23.3	13	3.2
BPIC	81,181	60	7.4	37	4.6
BPIC-E	37,748	18	4.8	13	3.4
BPIC-W	43,433	42	9.7	24	5.5

\*The raw frequencies (R.F.) and normalized frequencies (N.F.) (per 10,000 words) in corpora.

the British IR, exhibiting a normalized frequency of 7.4, deploys the DM “you know” less frequently than his Chinese counterpart, with a normalized frequency of 20.5. In contrast, the DM “I mean” is used more frequently by the British IR, evidenced by a normalized frequency of 4.6, in comparison to the Chinese IR’s normalized frequency of 3.2. Additionally, the examination of the frequencies of these two DMs, both within and between the two corpora, involves statistical testing employing the log-likelihood statistic. The results of this analysis reveal that in CPIC, the Chinese IR employs the DM “you know” ( $LL = 96.96, p < 0.05$ ) significantly more often than the DM “I mean”. Similarly, in BPIC, the British IR exhibits a strong preference for the DM “you know” ( $LL = 5.51, p < 0.05$ ) over “I mean.” The comparative analysis demonstrates that the Chinese IR employs the DM “you know” ( $LL = 48.25, p < 0.05$ ) statistically more frequently than her British counterpart. By contrast, the British IR uses the DM “I mean” slightly higher than the Chinese IR, although no statistically significant difference emerges.

### Functional analysis of the DMs “you know” and “I mean” in the two corpora

In the CPIC corpus, an analysis of the DM “you know” reveals the presence of eight functions, namely, hedging (HDG), agreeing (AGR), monitoring (MNT), topic (TOP), specification (SPE), disagreeing (DIS), alternative (ALT), and cause (CAU). These functions are identified across three distinct domains, resulting in twelve domain-function combinations: INT-HDG, INT-AGR, INT-MNT, INT-CAU, INT-SPE, INT-DIS, SEQ-MNT, SEQ-TOP, SEQ-HDG, SEQ-SPE, RHE-ALT, and RHE-AGR. Notably, the function of hedging (HDG) predominates in the CPIC corpus, accounting for a significant proportion of 39.7%, as indicated in Table 4. Among the three domains within the CPIC corpus, the interpersonal (INT) domain stands out, encompassing the majority of its distribution at 74.5%, followed by the sequential (SEQ) domain at 23.4%, and the rhetorical (RHE) domain at 2.1%. A deeper functional analysis reveals that five distinct functions—alternative (ALT), hedging (HDG), agreeing (AGR), topic (TOP), and specification (SPE)—are detected in association with the DM “I mean” in the CPIC corpus. Remarkably, alternative (ALT) emerges as the most frequent function, constituting 40.9% of its usage. These five identified functions manifest themselves across three domains, yielding a total of five combinations: RHE-ALT, INT-HDG, INT-AGR, INT-SPE, and SEQ-TOP, with the interpersonal



TABLE 4 The distribution of functions and domains of the DMs “you know” and “I mean” in CPIC.

Categories		DMs	
		You know	I mean
Function	Hedging (HDG)	56 (39.7%)	8 (36.4%)
	Agreeing (TOP)	41 (29.1%)	2 (9.1%)
	Monitoring (MNT)	27 (19.2%)	0
	Topic (TOP)	5 (3.6%)	1 (4.5%)
	Specification (SPE)	4 (2.8%)	2 (9.1%)
	Disagreeing (DIS)	4 (2.8%)	0
	Cause (CAU)	2 (1.4%)	0
	Alternative (ALT)	2 (1.4%)	9 (40.9%)
Total		141 (100%)	22 (100%)
Domain	Interpersonal (INT)	105 (74.5)	12 (54.5%)
	Sequential (SEQ)	33 (23.4%)	1 (4.5%)
	Rhetorical (RHE)	3 (2.1%)	9 (41%)
Total		141 (100%)	22 (100%)

(INT) domain taking precedence, representing 54.5% of its utilization, as depicted in Table 4.

Conversely, in the BPIC corpus, a set of seven functions are identified for the DM “you know,” encompassing monitoring (MNT), hedging (HDG), agreeing (AGR), specification (SPE), cause (CAU), disagreeing (DIS), and topic (TOP). These functions spread across two domains, resulting in nine domain-function combinations: INT-HDG, INT-AGR, INT-MNT, INT-CAU, INT-DIS, INT-SPE, SEQ-MNT, SEQ-TOP, and SEQ-SPE. Obviously, monitoring (MNT) emerges as the most frequently utilized function, accounting for 28.3% of its occurrence. The distribution of domains associated with the DM “you know” in the BPIC corpus exhibits slight variations compared to CPIC, with two identified domains: INT, representing 70% of occurrences, and SEQ, accounting for 30%, as presented in Table 5. The overall distribution of domain-function combinations for the DM “you know” in the two corpora displays similarities, especially within the top three categories. In CPIC, the most frequent category is INT-HDG, followed by INT-AGR and SEQ-MNT, with proportions of 36.9, 28.4, and 15.6%, respectively. In BPIC, the most frequent categories are INT-HDG, INT-AGR, and SEQ-MNT, with proportions of 26.6, 25, and 23.3%, respectively. However, unlike CPIC, BPIC lacks three domain-function combinations: SEQ-HDG, RHE-ALT, and RHE-AGR. Moreover, an investigation of the DM “I mean” in BPIC reveals the presence of eight functions: hedging (HDG), agreeing (AGR), monitoring (MNT), topic (TOP), specification (SPE), disagreeing (DIS), alternative (ALT), and cause (CAU). These functions are distributed across three domains, resulting in eight unique domain-function combinations: RHE-ALT, INT-HDG, INT-AGR, INT-SPE, SEQ-TOP, INT-CAU, INT-DIS, and SEQ-MNT. It is observed the function of alternative (ALT) dominates in BPIC, constituting 35.2% of its usage, while the interpersonal (INT) domain represents the most frequent category, with 54.1% of its utilization.

The functional distribution of the DM “you know” demonstrates that hedging (HDG) is the predominate function, and interpersonal-hedging (INT-HDG) (see Example [1]) stands out as the most

TABLE 5 The distribution of functions and domains of the DMs “you know” and “I mean” in BPIC.

Categories		DMs	
		You know	I mean
Function	Hedging (HDG)	16 (26.6%)	8 (21.6%)
	Agreeing (INT-AGR)	15 (25%)	3 (8.1%)
	Monitoring (MNT)	17 (28.3%)	1 (2.7%)
	Topic (TOP)	3 (5%)	3 (8.1%)
	Specification (SPE)	4 (6.7%)	1 (2.7%)
	Disagreeing (DIS)	1 (1.7%)	3 (8.1%)
	Cause (CAU)	4 (6.7%)	5 (13.5%)
	Alternative (ALT)	0	13 (35.2%)
Total		60 (100%)	37 (100%)
Domain	Interpersonal (INT)	42 (70%)	20 (54.1%)
	Sequential (SEQ)	18 (30%)	4 (10.8%)
	Rhetorical (RHE)	0	13 (35.1%)
Total		60 (100%)	37 (100%)

commonly observed domain-function combination in both corpora, aligning with its fundamental meaning and core function, as established in prior research (Fox Tree and Schrock, 2002; Lakoff, 2004). In example [1], the Chinese IR (IR) uses the DM “you know” to soften her tone while commenting on the perceived bias against China, a practice recognized as a key mitigating strategy (Östman, 1981).

Example [1]

HE1: There’s an active effort to do damage to China on top of the massive ignorance and on top of the bias which some people might not be aware of you think there’s an active smearing campaign and spreading disinformation. That’s pretty active and pretty strong, um, you know, voluntary action. (CPIC-E)

In addition to the primary HDG function, the INT domain encompasses five other functions, namely, INT-AGR (see Example [2]), INT-MNT (see Example [3]), INT-CAU (see Example [4]), INT-SPE (see Example [5]), and INT-DIS (see Example [6]). For instance, Example [2] illustrates the expression of shared sentiments and concurring opinions of the Chinese IR towards the successful cycling across China by the IE, particularly during the ongoing pandemic, facilitated by the use of the DM “you know.” In Example [3], the final occurrence of “you know” serves the INT-MNT function, acknowledged as a floor-yielding device (Östman, 1981; Fox Tree and Schrock, 2002), signifying that the Chinese IR does not intend to offer further information due to the self-evident nature of the statement (Beeching, 2016). This is based on meta-knowledge regarding communism, and the IR aims for the IE to grasp the information and respond accordingly, in line with previous studies identifying the MNT function (Schiffrin, 1987). In Example [4], the DM “you know” is utilized by the British IR to indicate agreement with the IE and prompt the IE to provide supporting facts. In Example [5], the British IR employs the DM “you know” to deliver an extensive commentary on the current state of women’s rights in Afghanistan, concurrently serving a face-saving function. Example [6] demonstrates a disagreement

between the Chinese IR and the IE regarding the challenges associated with acquiring proficiency in the Chinese language.

*Example [2]*

HE12: Right, well it must have been a challenging because you were doing this. Um, you know against the backdrop of COVID-19, but you managed it [yes]. (CPIC-E)

*Example [3]*

HW16: I'll be interested because it is very interesting, how many people are still being influenced by that kind of you know that kind of fear of about communism? Or you know under the influence of those, those theories you know. (CPIC-W)

*Example [4]*

HW6: Mr. Tong, it just sounds to me as we close, it sounds like you have thrown in your lot with Beijing, and you are gonna defend Beijing. You know I'm just you know pointing out the facts um if you can point to facts which would show to me that Hong Kong treasures will not be able to apply this law properly. (BPIC-E)

*Example [5]*

HE25: You say you are doing what you can from outside, there are some extraordinarily brave women inside the country still delivering the same sort of message I'm thinking for example, who is as I understand it to this very day still active inside Afghanistan on women's rights, but it's dangerous, is not it? And you know it's dangerous because there are at least two serious attempts on your life, including one occasion when you were shot in the arm. (BPIC-E)

*Example [6]*

HW9: But you know again for someone who dares to take up this current this endeavour and manages because it's difficult to learn. Let me tell you this, um I was reading a survey about um what African people view as their preferred language for to learn as a second language, especially for young people, you know how many percent prefer Chinese, just take a wild guess in comparison in comparison to English. (CPIC-W)

In terms of the DM "I mean," although the INT domain accounts for the largest amount, the most predominant domain-function combination in both corpora is RHE-ALT (see Example [7]), followed by the INT-HDG (see Example [8]), with a proportion of 40.9 and 36.4% in CPIC, and a percentage of 35.2 and 21.6% in BPIC, respectively. The functional distribution shows that "I mean" is mainly used for reformulating IR's words or ideas for IE's better understanding (Fox Tree and Schrock, 2002). In Example [7], in which the Chinese IR reformulates the previous metaphor about China's growing economy. In Example [8], the British IR asks the question implicitly (HDG).

*Example [7]*

HW17: This is very interesting, and the metaphor is also very helpful to understand this as we talk about. I mean China has seen a middle-class forming from zero to nowadays numbering over 400

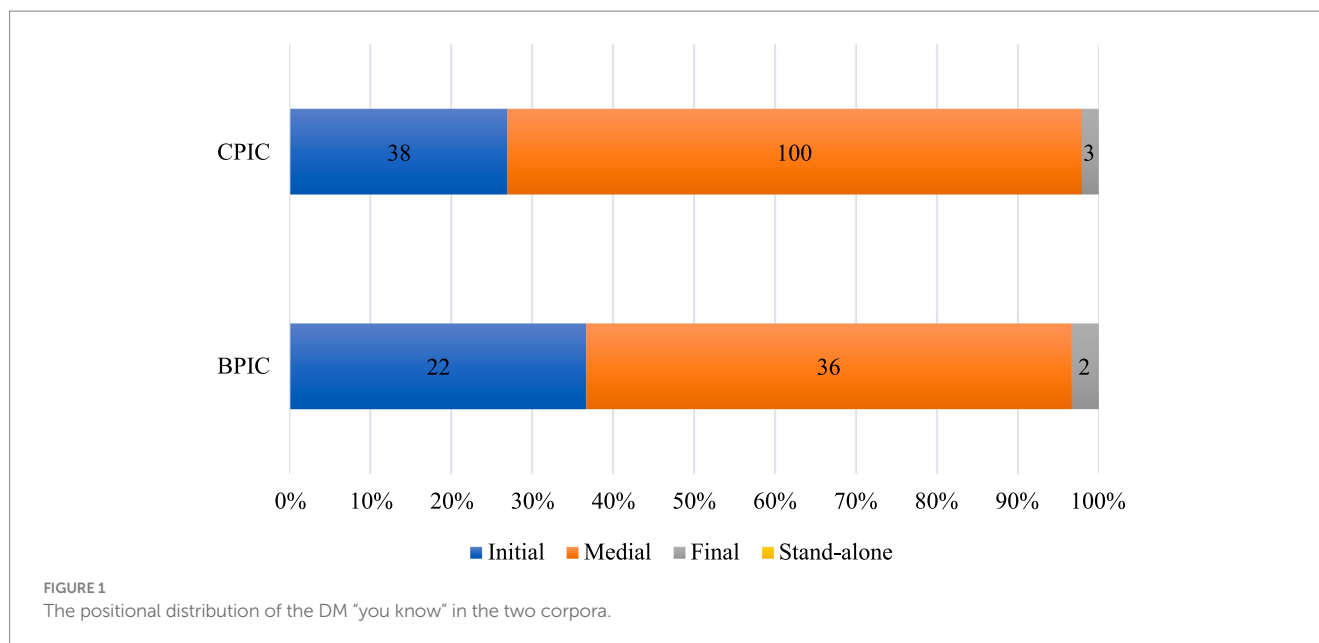
million that number is already bigger than the population of the United States, and that number is expected to continue to rise in the next decade also to come. (CPIC-W)

*Example [8]*

HW21: I've been developing how you have changed and evolved over time both in your comedy and your wider life. I mean, um, you are not a religious chew but you are very clearly um proud. (INT-HDG) (BPIC-W)

DMs often exhibit a high likelihood of co-occurring with specific linguistic elements, which function as contextual cues for their identification and interpretation (Escalera, 2009). Upon analyzing the co-occurrence of the DM "you know" in the two corpora, it becomes evident that certain particles display a stronger tendency to accompany it. A thorough concordance analysis is conducted to explore how "you know" interacts with other linguistic collocations at the utterance level (Borba and Jaeger, 2011). The results demonstrate that the DM "you know" frequently occurs in conjunction with tokens such as "um," "and," "but," "or," "yeah," "because" and "of course." Moreover, these particles are more likely to appear immediately before "you know" rather than after it. In the CPIC and BPIC corpora, there are 32 and 18 instances of words preceding the DM "you know," respectively. Two collocational patterns, "um you know" and "and you know," are particularly prevalent, with 12 (38%) occurrences in CPIC and 8 (44%) occurrences in BPIC. Researchers concur that the particle "um" carries meaning and function. For example, it conveys hesitation, pause, uncertainty, or an attempt to retain the conversational floor (Al-Farag and Mohammed, 2022). Its placement immediately preceding "you know" indicates the IR's hesitation or pause to gain additional time or allow for the subsequent speech segment (Fox Tree and Schrock, 2002). In this context, the use of "you know" reaffirms control because the filler "um" may potentially result in the IR losing the conversational floor, offsetting the adverse effects of a pause (Schourup, 1999). This could serve various purposes, including mitigating the tone of speech (HDG), expressing agreement with the IE (AGR), providing detailed commentary for the IE's comprehension (SPE), or maintaining control over the conversation (MNT). Conversely, when the DM "you know" is preceded by the DM "but," the different collocational pattern signifies a different function, such as denoting the expression of discording opinion (DIS), or a shift in topic (TOP). By contrast, the number of collocational patterns of the DM "I mean" is notably fewer compared to "you know." It is observed that the DM "I mean" is frequently accompanied by the particle "yeah," signaling the IR's conforming opinion (AGR) or contributing to the affirmation of the IE's statements (AGR) (Fetzer, 2014). Additionally, "I mean" is often accompanied by the DM "but," indicating the IR's disagreement with the IE (DIS) or a change in topic (TOP).

The analysis of positional distribution reveals that the two discourse markers (DMs) demonstrate distinct preferences for specific positions within utterances across both corpora. Aligned with Lam's (2007) categorization, a consistent pattern emerges in the positional distribution of these DMs across the two datasets, indicating a stable trend in their usage. Obviously, the most common usage of the DM "you know" by the IRs is in the medial position, whereas the DM "I mean" tends to be used in the initial position, as depicted in Figures 1, 2. For instance, in the CPIC corpus, the majority of instances of "you know" are found



in the medial position (71%), followed by the initial position (22%) and final position (2%). Similarly, in the BPIC corpus, 60% of its occurrences appear in the medial position, 22% in the initial position, and 2% in the final position. Conversely, the positional distribution of “I mean” exhibits an opposing pattern. In the CPIC corpus, 77.3% of instances of “I mean” are in the initial position, followed by the medial position (22.7%). In the BPIC corpus, 81.1% of its appearances are in the initial position, with 18.9% in the medial position. No instances of “I mean” are detected in the final or stand-alone positions in both corpora.

Additionally, the analysis of the association between the positions of “you know” and its functional distribution reveals significant variations in the number of functions across the three identified positions. Specifically, the majority of functions of “you know” are observed in the medial position, with a limited number in the initial position and only one in the final position, aligning with its positional distribution. For example, in CPIC and BPIC, twelve and nine domain-function combinations are found in the medial position, six and four in the initial position, and only one function in the final position. The predominant function of “you know” is INT-HDG, accounting for 37% in the initial position and 39% in the medial position in CPIC. This suggests that the most common function of “you know” does not exhibit a preference for a specific positional distribution, a pattern also mirrored in BPIC. Another noteworthy aspect is the infrequent occurrence of “you know” in the final position in both corpora, where it serves the INT-MNT function, indicating that the IR yields the floor to the IE due to shared meta-knowledge (Schiffrin, 1987). This function appears to be position-specific as it exclusively occurs in the final position.

## Discussion

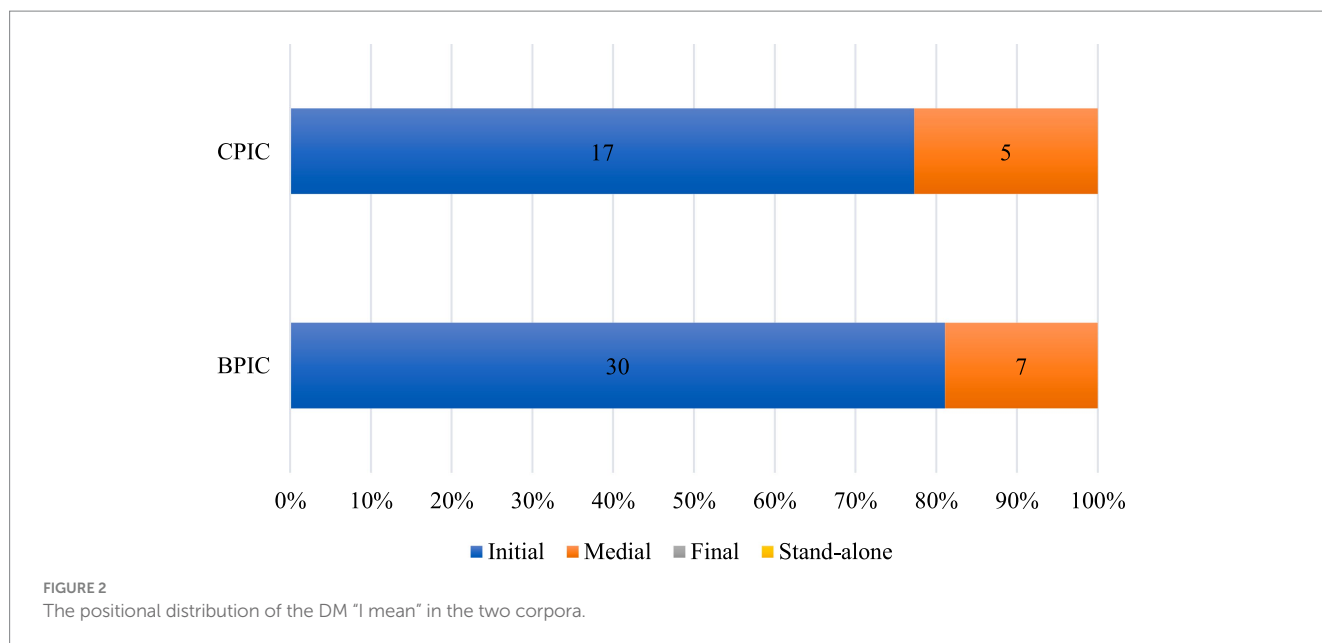
The present study analyses and compares the frequency and functions of the two DMs used by Chinese and British IRs in two televised political interview programs.

The findings in the present study corroborate Macaulay’s (2002) claim that the usage of the DM “you know” displays variability among

individuals of similar backgrounds, and align with assertion that communicative behavior in political discourse is influenced by gender, in relation to the two IRs’ professional background and gender identity.

The frequency comparison within the two corpora reveals a consistent pattern where the DM “you know” is used more frequently than “I mean” in both corpora. The comparative infrequency in the utilization of the DM “I mean” within the two examined corpora reflects the characteristics of political interviews which signifies the limited scope for IRs to adjust their statements during political interviews due to temporal constraints. Consequently, IRs tend to adhere more steadfastly to their predetermined agendas and pre-allocated sequence of turns and topical units (Furkó and Abuczki, 2014) with institutionalized preferences and fixed time frames acting as constraining factors on the employment of the DM “I mean” for making adjustments within political interviews, as suggested by Rangraz (2014).

A statistically significant difference is observed in the frequency of the DM “you know” between the Chinese and British IRs, with the Chinese IR employing it more frequently. This discrepancy can be attributed to various factors, including gender-specific differences in DM usage (Erman, 1992), and the identification of “you know” as a marker more commonly used by women (Lakoff, 2004; Laserna et al., 2014), or women prefer to use it slightly over men (Macaulay, 2002; Beeching, 2016). In terms of the Chinese IR’s language background, it may indicate that the non-native speaker (Chinese IR) utilizes “you know” more frequently than her native counterpart (British IR), which aligns with Buysse’s (2017) and Kwon’s (2020) observation. The Chinese IR’s frequent use of “you know” may serve to convey a sense of confidence and familiarity to the audience. IRs may possess distinct communication styles and individual speech habits that are challenging to alter even in professional contexts. The consistent use of “you know” by the Chinese IR might be a part of her speaking style, aimed at engaging the audience, establishing rapport, and clarifying her points. In contrast, the British IR uses the DM “I mean” slightly more frequently than the Chinese IR, although this difference is not statistically significant. This relatively heightened use of “I mean” aligns with the claims made by proponents



of critical discourse analysis regarding an ongoing process of conversationalization in British institutional discourse (Fetzer, 2014). Different IRs possess distinct styles and approaches to conducting interviews, and their use of specific DMs like “I mean” may be influenced by individual interviewing techniques or personal speaking habits. It is possible that the British IRs frequent use of “I mean” is linked to men’s preference for using DMs signalling repair-work (Erman, 1992). The comparative frequency results highlight the intricate relationship between language and gender identity. To delineate the fragmentation and nuances in identity construction, we should look beyond the use of a single word and consider the extralinguistic and intralinguistic factors of gender influence (Borba and Jaeger, 2011).

Schiffrin (1987) views that “you know” can be interpreted as overdependence on the hearer, while “I mean” suggests overinvolvement with the self, indicating linguistic elements can reveal insights into a speaker’s communicative style and interpersonal involvement in a conversation. This discrepancy in the use of the two DMs may reflect the contrasting interviewing styles of the two IRs. The Chinese IRs frequent use of “you know” may indicate her reliance on this DM to engage with her IEs and ensure their comprehension of her points, which can be seen as overdependence on the listener. By using “you know,” she might be trying to connect with her viewers, establish a sense of shared understanding, and make her message more relatable. It’s a way of checking in with the guests to ensure they are following her train of thought. In contrast, the British IRs repeated use of “I mean” may suggest overinvolvement with the self, as this DM often introduces clarifications or rephrasing of the speaker’s prior statements.

DMs can have various functions when they are used in different contexts (Farahani and Ghane, 2022). The identified number of functions these two DMs perform in both corpora indicate that they are multifunctional (Borba and Jaeger, 2011), and can work on different domains (Erman, 2001; Müller, 2005; Farahani and Ghane, 2022) in political interviews. The prevalent function of “you know” observed in both CPIC and BPIC affirms its role as a hedging (HDG) marker, serving to establish intimacy and shared understanding

(Östman, 1981). This aligns with Borba and Jaeger’s (2011) findings, indicating that both Chinese and British IRs predominantly employ “you know” to manage the IR-IE relationship in political interviews. Similarly, “I mean” predominantly assumes the alternative (ALT) function in both corpora, synonymous with “reformulation” as studied by Fox Tree and Schrock (2002). This reflects the core purpose of “I mean” and is consistent with other research (Pettersson-Traba, 2018; Farahani and Ghane, 2022), indicating that IRs primarily employ this marker to reformulate their statements for the IE’s comprehension (Fetzer, 2014).

Analysis of collocational patterns of “you know” in the two corpora, such as “and you know,” “but you know,” and “um you know,” is consistent with Borba and Jaeger’s (2011) assertion that “you know” exhibits flexibility in collocating with other linguistic elements, with conjunctions being the most common collocates in the left periphery. Conversely, the infrequency of the collocational pattern “and I mean” corresponds to Šandová’s (2015) observation that “I mean” rarely associates with “and” in political interviews. Notably, the primary collocate for “I mean” in political interviews is “but,” forming the “but I mean” cluster, indicating the speaker’s intent to present a divergent perspective, in line with Fetzer’s (2014) findings. The inclusion of the particle “yeah” alongside “I mean” indicates affirmation of the speaker’s reformulation (Fetzer, 2014).

The highest frequency of “you know” appearing in the medial position in the two corpora corresponds to Erman’s (1987) observation that it is frequently used to connect two main clauses. The rare occurrence of “you know” in the final position partially supports Fox Tree and Schrock’s (2002) claim that its final position usage varies, i.e., it can be frequent or infrequent in the final position. However, Crible’s (2018) study contradicts this, showing that “you know” is often found in the final position, which contrasts with the current study where it rarely occurs at the end of utterances. Notably, “I mean” does not appear in the final position in the two corpora, aligning with Fox Tree and Schrock’s (2002) assertion that “I mean” infrequently occurs in this position. While the ranking of the positional distribution of “you know” and “I mean” is similar in the two corpora, the applicability of



these findings to other genres remains uncertain, as the frequency of DM position remains a subject of debate (Fox Tree and Schrock, 2002).

The functional distribution of the two DMs is found to be less constrained by their positional distribution, supporting Erman's (2001) argument that a speaker's turn position does not dictate DM functions. Despite "you know" primarily introducing pre-existing or shared knowledge and "I mean" introducing new information (Erman, 1987), the findings reveal associations between these two markers in terms of their functions. As noted by Pettersson-Traba (2018), "you know" and "I mean" occasionally exhibit similar behavior in discourse, with some of their functions falling into the same categories of use. Functional similarities, such as turn management (MNT), repairing (ALT), monitoring (MNT), and organizing (SEQ), highlight their historical affinity (Fox Tree and Schrock, 2002). Moreover, both "you know" and "I mean," like the identified hedging (HDG) function, operate within the interpersonal (INT) domain, forming the INT-HDG combination.

## Conclusion

This study constitutes an endeavour to disclose and compare the frequency and functions associated with the usage of "you know" and "I mean" by the two IRs within the context of political interviews. By doing so, the investigation takes into account the co-occurrences of the two DMs with other particles, as well as their positional distribution. The study demonstrates different interviewing styles of the Chinese and British IRs through analysing their utilization of these two DMs. Additionally, it also shows that the functional distribution of these two markers is less constrained by their positional placement in political interviews, which provides valuable evidence regarding the distinctiveness and overlapping functions of the two DMs in political interviews. Despite their similarities, they cannot be mutually substituted, particularly within the highly contentious context of political interviews.

The findings obtained from this study contribute to our comprehension of how these DMs are researched and employed by the IRs in the context of mediated political interviews. Consequently, they offer significant and intriguing insights to the existing literature. Furthermore, these efforts have practical implications, incentives, and insights for practitioners interested in conducting similar research.

However, despite the constructive outcomes, there are certain limitations that necessitate attention from future researchers. Firstly, the presented findings must be interpreted cautiously due to the limited scope of this study. For instance, the study solely focuses on discourse produced by the IRs and the two DMs. It does not account for various social parameters such as personality traits and the relationship between IR and IE. Factors like the age, gender, and social status of the IEs have not been taken into consideration. The study is

limited to the discourse markers in political interviews taking discourse markers "you know" as noted since Östman's (1981) seminal work. To address this limitation in future work, researchers could broaden the scope by incorporating participants from both sides and exploring a wider range of DMs beyond those identified in media discourse. Additionally, considering prosodic and visual features within this genre could provide a more detailed understanding of the interaction between variables and discourse markers. Moreover, conducting comparative or cross-linguistic studies that compare English DMs with those in other languages would be beneficial.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Author contributions

YF: Data curation, Formal analysis, Methodology, Writing – review & editing. MA: Investigation, Methodology, Resources, Validation, Writing – review & editing. DE-D: Project administration, Software, Supervision, Validation, Conceptualization, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The researchers thank Prince Sultan University for funding this research project through the Applied Linguistics Research Lab (RL-CH-2019/9/1).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Afzaal, M., Ilyas Chishti, M., Liu, C., and Zhang, C. (2021). Metadiscourse in Chinese and American graduate dissertation introductions. *Cogent Arts Hum.* 8:879. doi: 10.1080/23311983.2021.1970879
- Afzaal, M., Imran, M., Du, X., and Almusharraf, N. (2022). Automated and human interaction in written discourse: a contrastive parallel Corpus-based investigation of Metadiscourse features in machine-human translations. *SAGE Open* 12:215824402211422. doi: 10.1177/21582440221142210
- Aijmer, K. (2002). English discourse particles. Evidence from a Corpus. Amsterdam, the Netherlands: John Benjamins Publishing Company.
- Aijmer, K. (2013). *Understanding pragmatic markers: A variational pragmatic approach*. Edinburgh University Press.
- Al-Faragy, R. F. H., and Mohammed, F. J. (2022). Gender and native-ness differences in the use of speech fillers in political interviews. *Al-Adab J.* 2, 65–78. doi: 10.31973/aj.v2i140.3634

- Baker, P. (2006). Using corpora in discourse analysis. London: Continuum.
- Barbara, S. W. Y., Afzaal, M., and Aldayel, H. S. (2024). A corpus-based comparison of linguistic markers of stance and genre in the academic writing of novice and advanced engineering learners. *Hum. Soc. Sci. Commun.* 11, 1–10. doi: 10.1057/s41599-024-02757-4
- Beeching, K. (2016). *Pragmatic markers in british english: Meaning in social interaction*. Cambridge University Press.
- Bolden, G. (2015). “Discourse markers” in The international encyclopedia of language and social interaction. ed. T. Karen (Hoboken, New Jersey, U.S.: John Wiley & Sons, Inc.), 1–7.
- Borba, R., and Jaeger, A. (2011). “They never realized that, you know”: linguistic collocations and interactional functions of you know in contemporary academic spoken English. *ESpecialist* 32, 195–215.
- Brezina, V. (2018). Statistics in Corpus linguistics. Cambridge, Cambridgeshire, England: Cambridge University Press.
- Buyse, L. (2012). So as a multifunctional discourse marker in native and learner speech. *J. Pragmat.* 44, 1764–1782. doi: 10.1016/j.pragma.2012.08.012
- Buyse, L. (2017). English so and Dutch dus in a parallel corpus: An investigation into their mutual translatability. In *Yearbook of corpus linguistics and pragmatics*, (Vol. 4). Ed. Romero-Trillo (Springer International Publishing AG), 33–61.
- Crible, L. (2017). Towards an operational category of discourse markers: a definition and its model. Amsterdam, the Netherlands: John Benjamins.
- Crible, L. (2018). Discourse markers and (dis)fluency: forms and functions across languages and registers. Amsterdam, the Netherlands: John Benjamins Publishing Company.
- Crible, L., Abuczki, A., Burkšaitienė, N., Furkó, P., Nedoluzhko, A., Rackevičienė, S., et al. (2019). Functions and translations of discourse markers in TED talks: a parallel corpus study of underspecification in five languages. *J. Pragmat.* 142, 139–155. doi: 10.1016/j.pragma.2019.01.012
- Crible, L., and Degand, L. (2019). Domains and functions: a two-dimensional account of DMs. *Discours* 24, 3–35. doi: 10.4000/discours.9997
- Diab, A., Marie, M., Elgharabawy, A., and Elbendary, I. (2023). The effect of political risk and corporate governance on bank stability in the MENA region: Did the Arab Spring uprisings matter? *Cogent Bus. Manag.* 10:2174207.
- Erman, B. (1987). Pragmatic expressions in Swedish: a study of you know, you see, and I mean in face-to-face conversation. Stockholm, Sweden: Almqvist & Wiksell International.
- Erman, B. (1992). Female and male usage of pragmatic expressions in same-sex and mixed-sex interaction. *Lang. Var. Chang.* 4, 217–234. doi: 10.1017/S0954394500000764
- Erman, B. (2001). Pragmatic markers revisited with a focus on you know in adult and adolescent talk. *J. Pragmat.* 33, 1337–1359. doi: 10.1016/S0378-2166(00)00066-7
- Escalera, E. (2009). Gender differences in children’s use of discourse markers: separate worlds or different contexts? *J. Pragmat.* 41, 2479–2495. doi: 10.1016/j.pragma.2006.08.013
- Farahani, M. V., and Ghane, Z. (2022). Unpacking the function(s) of DMs in academic spoken English: a corpus-based study. *Australian J. Lang. Liter.* 45, 49–70. doi: 10.1007/s44020-022-00005-3
- Feldman, O. (2016). Televised political interviews: a paradigm for analysis. *Asian J. Public Opin. Res.* 3, 63–82. doi: 10.15206/ajpor.2016.3.2.63
- Fetzer, A. (2014). I think, I mean and I believe in political discourse: collocates, functions and distribution. *Funct. Lang.* 21, 67–94. doi: 10.1075/fo1.21.1.05fet
- Fischer, K. (2006). Approaches to discourse particles. Amsterdam: Elsevier.
- Flick, U. (2014). Challenges for qualitative inquiry as a global endeavor: introduction to the special issue. *Qual. Inq.* 20, 1059–1063. doi: 10.1177/1077800414543693
- Fox Tree, J., and Schrock, J. (2002). Basic meanings of you know and I mean. *J. Pragmat.* 34, 727–747. doi: 10.1016/S0378-2166(02)00027-9
- Furkó, P. (2015). From mediatized political discourse to The Hobbit: The role of pragmatic markers in the construction of dialogues, stereotypes and literary style. *Language and Dialogue*, 5, 264–282. doi: 10.1075/ld.5.2.04fur
- Furkó, P. (2020). *Discourse markers and beyond: descriptive and critical perspectives on discourse-pragmatic devices across genres and languages*. Palgrave Macmillan.
- Furkó, P., and Abuczki, A. (2014). English discourse markers in mediatized political interviews. *Brno Stud. Engl.* 40, 45–64. doi: 10.5817/BSE2014-1-3
- Gao, X., and Lee, S. (2019). How do discourse markers indicate emotions? In the 30th north American conference on Chinese linguistics. The Ohio State University.
- House, J. (2013). Developing pragmatic competence in English as a lingua franca: using discourse markers to express (inter)subjectivity and connectivity. *J. Pragmat.* 59(Part A), 57–67. doi: 10.1016/j.pragma.2013.03.001
- Ivankova, N., and Creswell, J. (2009). “Mixed methods” in Qualitative research in applied linguistics: A practical introduction. eds. J. Heigham and R. Croker (London: Palgrave Macmillan), 135–161.
- Kwon, Y. (2020). Pragmatic discourse markers you know and sort of in Korean EFL teacher corpus. *Stud. Foreign Lang. Educ.* 34, 291–322. doi: 10.16933/sfle.2020.34.1.291
- Lakoff, R. (2004). “Language and woman’s place” in Language and woman’s place. Text and commentaries. ed. M. Bucholtz (Oxford: Oxford University Press). (Original work published 1975)
- Lam, P. (2007). Discourse particles in an intercultural corpus of spoken English. [Doctoral thesis, The Hong Kong Polytechnic University]. *PolyU Electronic Theses*. Available at: <https://theses.lib.polyu.edu.hk/bitstream/200/4891/1/b22338160.pdf>
- Laserna, C. M., Seih, Y.-T., and Pennebaker, J. W. (2014). Um...Who Like Says You Know: Filler word use as a function of age, gender, and personality. *J. Lang. Soc. Psychol.* 33, 328–338.
- Macaulay, R. (2002). You know, it depends. *J. Pragmat.* 34, 749–767. doi: 10.1016/S0378-2166(01)00005-4
- Macaulay, R. (2013). “Discourse variation” in The handbook of language variation and change. eds. J. Chambers and N. Schilling-Estes (Hoboken, New Jersey, U.S.: John Wiley and Sons, Inc.), 220–236.
- Mahlous, A. R. (2024). The impact of fake news on social media users during the COVID-19 pandemic, health, political and religious conflicts: A deep look. *Int. J. Psychol. Relig.* 5, 481–492.
- Müller, S. (2005). *Discourse markers in native and non-native english discourse*. John Benjamins Publishing Company.
- O’Keefe, A., and McCarthy, M. (2010). The Routledge handbook of Corpus linguistics. New York: Routledge.
- Östman, J.-O. (1981). You know: discourse functional approach. Amsterdam, the Netherlands: Benjamins.
- Öztürk, Y., and Durmuşoğlu Köse, G. (2021). “Well (er) you know ...”: discourse markers in native and non-native spoken English. *Corpus Pragm.* 5, 223–242. doi: 10.1007/s41701-020-00095-9
- Pettersson-Traba, D. (2018). Revisiting you know and I mean: some notes on the functions of the two pragmatic markers in contemporary spoken American English. *Res. Corpus Linguist.* 6, 67–81. doi: 10.32714/ricl.06.06
- Rangraz, M. (2014). The uses of the DMs ‘well,’you know’and ‘I mean’in news interviews. [Master’s Thesis. Linköping, Sweden: Linköping University.
- Rayson, P. (2016). Log-likelihood and effect size calculator. Lancaster: Excel spreadsheet.
- Raza, A., Rashid, S., and Malik, S. (2022). Cultural violence and gender identities: A feminist post-structural discourse analysis of this house of clay and water. 3L: Southeast Asian. *J. Engl. Lang. Stud.* 28, 124–136.
- Rühlemann, C. (2019). How long does it take to say ‘well’? Evidence from the audio BNC. *Corpus Pragmat.* 3, 49–66. doi: 10.1007/s41701-018-0046-y
- Šandová, J. (2014). Speaker involvement in political interviews. Lausanne, Switzerland: Peter Lang.
- Šandová, J. (2015). On the use of cognitive verbs in political interviews. *Brno Stud. English* 41, 41–59. doi: 10.5817/BSE2015-1-3
- Schiffrin, D. (1987). Discourse markers. Cambridge, England: Cambridge University Press.
- Schourup, L. (1982). Common discourse particles in English conversation. [doctoral dissertation, the Ohio State University].
- Schourup, L. (1999). Discourse markers. *Lingua* 107, 227–265. doi: 10.1016/S0024-3841(96)90026-1
- Simon-Vandenberg, A.-M. (2000). The functions of I think in political discourse. *Int. J. Appl. Linguist.* 10, 41–63. doi: 10.1111/j.1473-4192.2000.tb00139.x
- Verdonik, D. (2022). Discourse markers and dialogue act annotation for computational dialogue systems. In *discourse markers in interaction: From production to comprehension*. Eds. M. Cuenca and L. Degand (De Gruyter Mouton), 191–214.
- Zhang, C., Afzaal, M., Omar, A., and Altohami, W. M. (2023). A corpus-based analysis of the stylistic features of Chinese and American diplomatic discourse. *Front. Psychol.* 14:1122675. doi: 10.3389/fpsyg.2023.1122675