

Revisiting *you know* and *I mean*: some notes on the functions of the two pragmatic markers in contemporary spoken American English

Daniela Pettersson-Traba
University of Santiago de Compostela / Spain

Abstract – This article presents a corpus-based study of the pragmatic markers *you know* and *I mean* in contemporary spoken American English. Previous research indicates that *you know* and *I mean* are polysemous in their discourse roles, serving various functions in speech. By drawing on tokens extracted from the *Corpus of Contemporary American English*, the *Corpus of American Soap Operas* and the *Corpus of Spoken, Professional American English*, which include data from text types differing on the scales of formality and spontaneity, the main aims are 1) to compare the use of these two pragmatic markers and 2) to explore whether and how their behavior differs in three text types: TV and radio programs, soap operas, and White House press conferences and faculty/committee meetings. The results demonstrate that, despite overlapping in some of their functions, *you know* and *I mean* cannot be used interchangeably in discourse. Additionally, the functions of the two pragmatic markers vary significantly depending on the corpora, which is due to the particular characteristics of the speech situations in which they are used.

Keywords – pragmatic markers, *you know*, *I mean*, American English, spontaneity, formality

1. INTRODUCTION¹

The Pragmatic Markers (PMs) *you know* and *I mean* are especially prolific in spoken discourse. Consequently, and much like other PMs (e.g. *actually* and *like*), they have been and continue to be considered by some speakers as characteristic of careless speech, being used mainly as fillers in spontaneous discourse (see discussion in Schifffrin 1987: 310 and Brinton 2017: 4). Nevertheless, *I mean* and *you know* are also persistent in formal, less spontaneous discourse, where they serve other functions apart from the aforementioned ones, as illustrated in example (1) from the *Corpus of Contemporary American English* (COCA) (Davies 2008–), in which *I mean* clarifies and expands on the idea presented and *you know* adds extra information while simultaneously appealing to the hearer:

¹ This research was conducted with the financial support of the European Regional Development Fund and the following institutions: Regional Government of Galicia (grants ED481A-2016/168, ED431D 2017/09, and ED431B 2017/12) and the Spanish Ministry of Economy and Competitiveness (grant FFI2017-86884-P). I am indebted to Iván Tamaredo and Mario Serrano-Losada for their most helpful comments on earlier versions of this article. Finally, thanks are also due to the anonymous reviewers and the editors of *Research in Corpus Linguistics*, Javier Pérez-Guerra and María José López-Couso, for their time and consideration, as well as for their invaluable suggestions.



- (1) BROOKE-SHIELDS: Unbelievable company to walk into. *I mean*, I feel so supported and, *you know*, they've been doing the show for twenty years. You think you're going to go on to that set and they're just going to rush through everything.
HODA-KOTB: Yeah. (COCA, 2017, NBC: Today Show)

As PMs, *you know* and *I mean* are characterized by their complex functionality and positional mobility. Over the last thirty years, a great body of research has analyzed their development and use (e.g. Goldberg 1980; Erman 1987, 2001; Fox Tree and Schrock 2002; Brinton 2007; Beeching 2016). However, although extensively researched, there is still no consensus as to their central functions and, what is more, since both PMs perform similar functions, some scholars go as far as to contend that they are used indistinctively (see discussion in Fox Tree and Schrock 2002: 729). Additionally, as claimed by Tagliamonte (2016: 98), most studies on PMs “focus on the salient and interesting forms without considering [...] the other forms with which they vary” (but see Lam 2010 for a study which compares *well* and *so*). Therefore, further research is still needed to (a) clarify the multiple functions they serve in different discourse types and (b) demonstrate that, despite their functional similarities, they do exhibit different behavior in discourse.

Against this backdrop, this article presents a corpus-based analysis of the two PMs *you know* and *I mean* across three different discourse types in American Present-day English (PDE): radio and TV programs, soap operas, and White House press conferences and committee or faculty meetings.² The aim is to discover possible parallelisms and differences concerning the functions of these PMs in the three distinct discourse types, thus shedding further light on their use. While one important motivation for considering the two PMs together in the present study is their functional similarity, another essential reason is the direct opposite, to wit, to demonstrate their autonomy as independent discourse devices.

The remainder of this article is structured as follows: Section 2 presents an overview of PMs in general and of *you know* and *I mean*. In Section 3, the data and methodology employed are described, followed by a discussion of the results (Section 4) and some concluding remarks and suggestions for further research (Section 5).

2. PRAGMATICS MARKERS: *YOU KNOW* AND *I MEAN*

PMs have been widely analyzed since the 1980s, with numerous studies being devoted to defining this word class and explaining their multiple functions in discourse (e.g. Schourup 1985; Schiffrin 1987; Crystal 1988; Watts 1989; Fraser 1990, 1999; Redeker 1990; Brinton 1996, 2017; Blakemore 2002; Aijmer and Simon-Vandenberg 2011; Beeching 2016). Nonetheless, despite extensive research, this field of pragmatics is highly controversial: multiple labels are used in the literature to refer to the forms and scholars have claimed that PMs do not easily fit into any traditional word class (e.g. Brinton 1996; 2017).³ This is because PMs have little or no referential meaning, although they provide ‘procedural’ meaning, making them important for information processing (Blakemore 2002: Chapter 4). Consequently, rather than being superfluous, they are crucial for communication as they indicate speakers’ commitment to their statements, as well as their connection with previous utterances (Fraser 1999: 22). Despite the existing controversies, most scholars concur that PMs display the following features. Firstly, they occur outside the syntactic structure or are only slightly attached to it. Secondly, they are optional rather than obligatory grammatical elements. Thirdly, they are multifunctional and the purpose they serve is highly context-dependent (Brinton 1996: 33–35; Aijmer and Simon-Vandenberg 2011). Additionally, PMs overlap in some of their uses – *I mean* and *you know* are clear examples of this (see Section 4.1). However, as will become evident, this does not mean that they can be used indistinctively in discourse.

Much research has been conducted on *you know* and *I mean*, either separately (Östman 1981; Erman and Kotsinas 1993; Stubbe and Holmes 1995; He and Lindsay 1998; Jucker and Smith 1998; Erman 2001; Macaulay 2004; Fox Tree 2007; Beeching 2016: 97–125 for *you know*, and Brinton 1996, 2007; Fernández-Polo 2014; Beeching 2016: 183–209 for *I mean*), or by accounting for the two together (e.g. Goldberg 1980; Erman 1987; Fox Tree and Schrock 2002; Koczogh and Bálint 2011; Szczyrbak 2014). Some scholars have argued that *you know* and *I mean* are used with no distinction (Fromkin 1973: 42;

² The choice of American English rather than any other variety of English was ultimately a matter of convenience, namely the availability of three corpora representing different degrees of spontaneity and formality in this variety and text types (e.g. soap operas) which, to my knowledge, have not been analyzed in depth as concerns the functions of these two PMs.

³ I have opted for the term ‘Pragmatic Marker’, as it is one of the most general and widely accepted terms to refer to this class of expressions.

O'Donnell and Todd 1991: 69). However, a much more common claim is that they cannot be used interchangeably, even though some of their functions fall into the same categories (clarification, hesitation, and repair). Several functional classifications have been postulated for these PMs (e.g. Fox Tree and Schrock 2002), but one of the most recent and comprehensive categorizations is that proposed by Beeching (2016: Chapters 5 and 8), whose approach is employed here. She puts forward the following categories, which will be briefly explained throughout this section:

- *I mean*: Hedging, (self-)repair, hesitation, clarification, justification, and concession.
- *You know*: Common ground, repair, hesitation, word-search, clarification, attention-getting, and imposition.

Because of the explicit mention of the second person pronoun *you* and the use of the cognitive verb *know* 'to have knowledge', *you know* is more interpersonal than *I mean*, functioning intersubjectively. 'Intersubjectivity' refers to the acknowledgement of the needs and desires of addressees: speakers adapt their messages to account for what they believe are the mental models or presuppositions of their interlocutors about the content presented in utterances (see Traugott 2003; Fitzmaurice 2004; López-Couso 2010). Consequently, one of the main functions of *you know* is the establishment of common knowledge among interlocutors (e.g. Östman 1981: 17; Fox Tree and Schrock 2002). Conversely, *I mean* functions subjectively: because of the use of the first-person pronoun *I*, combined with the communication verb *mean* 'to convey or indicate', this PM serves to transmit speakers' standpoints and beliefs about their own utterances. Thus, *I mean* primarily serves as an indicator of speakers' adjustment in the production of speech (Schiffrin 1987: Chapter 9).

At the (inter)subjective level, both markers can be used as politeness strategies (see Brown and Levinson 1987). For instance, *I mean*, as a negative politeness strategy, is used as a hedging device to predict a less face-threatening utterance either by reducing the strength of evaluative claims and speakers' commitment to them or by diminishing their assertive force. In both cases, *I mean* creates or reestablishes harmony between the interlocutors (Brinton 2007: 41). PM *you know* is used as a positive politeness strategy, appealing to hearers' shared understanding, either because a preceding statement is part of general knowledge or because both interlocutors already share that piece of information, thus appealing to their common ground (Fox Tree and Schrock 2002). Some scholars (e.g. Jucker and Smith 1998), however, claim that more than directly pointing towards the shared knowledge of interlocutors, speakers frequently use *you know* to draw attention to utterances they estimate particularly relevant for the current discourse.

Both PMs are frequent in repairs and hesitations, to stall for time when faced with lexical indecisions or due to ongoing planning problems (e.g. Erman 1987: 144; Fox Tree and Schrock 2002). In addition, *you know* can be used in word-searching (Beeching 2016: 99–100). These functions are especially important for the online production of speech, as they represent common disruptions of language on the go. One example is illustrated in (2), which provides an instance of a repair preceded by *you know*:

- (2) However, if there is the support of how to do the corrections, *you know*, the corrective action (CSPAE, COMM697)

In addition to functioning interpersonally or as staller, *you know* and *I mean* also serve at the textual level as clarifying devices. Both PMs thus make hearers pay special attention to what has been said and/or what is to be said, functioning to alert addressees about modifications of previous statements. They can be used in this sense, for instance, to elaborate on, rationalize, justify, or reformulate previous utterances as illustrated in (1) (see Section 1). Consequently, both PMs are said to be 'forward-looking', as they forewarn addressees about additional information being introduced (e.g. Erman 1987: Chapter 5; Fox Tree and Schrock 2002; Brinton 2007). Additionally, *you know*, placed at left periphery, can serve as an attention-getting device by presenting new information (e.g. Östman 1981: 24–26; Beeching 2016: 101). As such, the PMs performs a highlighting function, signaling content which is highly relevant, as in (3):

- (3) Joni: [...] We don't even have nuns in my church.
Greenlee: *You know*, let's leave religion out of this and focus on the task at hand. (SOAP, 2002, AMC)

However, *you know* can also be placed at the right periphery with quite a dissimilar meaning, namely, to point out a self-evident truth, especially when uttered with a falling pitch.⁴ In many cases, this use of the PM can be stated by the speaker with a condescending or impositional attitude (Beeching 2016: 102–104). One example of such an occurrence of *you know* is the following:

- (4) Jake: [...] you're not looking me in the eye and – I'm here by myself. I can only start assuming the worst, *you know*.
Amanda: You're not the only one with a job. If J.R. wants me to work, I can't exactly just blow him off.
(SOAP, 2011, AMC)

Finally, *I mean* can function to put forward a concession in which speakers modulate their arguments (Beeching 2016: 188–189). In (5) the speaker agrees that the ideal situation would be to have only 30 items on the test, but then contradicts that idea, by stating that it would pose reliability difficulties:

- (5) *I mean*, again, I wish we could go back and only have 30 items, *but* I think that in terms of the reliability, we've got to have enough so that there are like 8 items per strand (CSPAЕ, COMM597)

In sum, the uses of the two PMs fall into the aforementioned categories. Nevertheless, it must be borne in mind that these functions are closely related and often intertwine. Both PMs are commonly used to serve different functions simultaneously and it is not always straightforward which function is most salient. Beeching (2016: 6) claims that “[p]ragmatic markers are notoriously both polysemous and multifunctional. In other words, not only does each pragmatic marker fulfil a variety of functions, it can do so simultaneously. This poses problem of interpretation”. In this analysis, I have decided to respect the PMs multifunctionality by considering all the functions each of the instances of *you know* and *I mean* perform in discourse.

3. DATA AND METHODOLOGY

The data used for the analysis of *you know* and *I mean* is drawn from three corpora of contemporary American spoken English: the spoken component of COCA, the *Corpus of American Soap Operas* (SOAP) (Davies 2011–), and the *Corpus of Spoken, Professional American English* (CSPAЕ) (Barlow 2000). Although they are all representative of Present-day American English, they differ concerning the types of discourse they contain, which vary along two continuums, spontaneity and formality, as shown in (a) and (b):

(a) spontaneity:	+	COCA (spoken)	CSPAЕ	SOAP	–
(b) formality:	+	CSPAЕ	COCA (spoken)	SOAP	–

The spoken component of COCA contains transcripts from radio and TV programs which include a high amount of informal face-to-face interaction. The data in SOAP, retrieved from 10 American soap operas, although reflecting highly informal speech, is non-spontaneous, as it comprises scripted language. Finally, the data in CSPAЕ is strongly inclined towards the opposite end of the spectrum of formality, as it includes formal speech from White House press conferences, as well as faculty and committee meetings. CSPAЕ occupies the intermediate position on the scale of spontaneity, since, despite dealing with topics known beforehand, it represents only real-time speech, comprising question-and-answers sessions and discussions.

The huge differences in the size of the corpora, with COCA (spoken component) amounting to approximately 80 million words, SOAP containing around 100 million words, and CSPAЕ including roughly 2 million words, is not an issue in this analysis. I performed random searches of the PMs, using the random search option in COCA and SOAP, and the concordance software *WordSmith Tools 6* (Scott 2012) for the CSPAЕ data. I limited my study to 100 examples of *you know* and 100 examples of *I mean* in each corpus, thus comprising a total of 300 randomly selected examples of *you know* – 100 from

⁴ The position of the PMs in the sentence, particularly in the case of *you know*, is highly relevant when it comes to the functions in which they are used, as we can infer from the examples provided in this article as well as from previous research (see, for instance, Crystal and Davy 1975: 92–95; Macaulay 2004: 757–760; Beeching 2016: 105–106; Brinton 2017: 757–760). In other words, specific positions in the sentence (i.e. initial, medial, or final) of the PMs are often associated with one or several particular functions. However, position of the PMs in the sentence is not the focus here and thus lies beyond the scope of this article.

COCA, 100 from SOAP, and 100 from CSPAE – and 300 examples of *I mean* chosen at random – again 100 from COCA, 100 from SOAP, and 100 from CSPAE. This rendered a final database of 600 examples, which is balanced for the three corpora, and consequently also for the different text types represented by them, as well as for the two PMs. Nevertheless, as there was no way of directly searching for the two expressions functioning as PMs, it was indispensable to manually discard occurrences of canonical uses, such as those exemplified in (6) and (7):

- (6) McCurry: We'll let *you know* tomorrow. No need to worry about that until this time tomorrow. (CSPAE, Wh97a)
 (7) Rosanna: These are good.
 Parker: Oh, please.
 Rosanna: No, *I mean* it. You look great. (SOAP, 2009, ATWIT)

Moreover, so-called 'in-between uses', which can neither be said to be canonical nor pure PM uses such as the phrases (*do*) *you know what I mean* and *you know what*, exemplified in (8) and (9), were also excluded. In the tag form (*do*) *you know what I mean*, *I mean* is not syntactically detached and cannot be taken as a prototypical case of a PM use. Prefatory *you know what* has also been left out because it can only function as an attention-getting device and cannot serve any of the other six functions of PM *you know*.⁵

- (8) Brooke: *You know what I mean*, you smart aleck. Are you in love?
 Rick: Could be. Probably. (SOAP, 2004, BB)
 (9) Casey: I think you'd do about anything to get her back!
 Gwen: Okay, *you know what?* He doesn't mean that, he's not thinking straight right now. (SOAP, 2006, ATWT)

Consequently, I was left with a total of 83, 60, and 72 PM uses of *you know* and 96, 93, and 98 uses of *I mean* in COCA, SOAP, and CSPAE, respectively (Section 4). These examples were then categorized according to Beeching's (2016) functional classification (see Section 2). However, the clarification and justification functions of *I mean* were merged in my analysis, due to the difficulty in teasing apart the two uses, as Beeching (2016: 187–188) herself recognizes. Additionally, I added the category 'multi' for those instances in which the PMs realized more than one function simultaneously.⁶

Seeing that the three corpora contain spoken American English data from three clearly distinct spoken discourse types, they are perfectly suitable to compare the use of the two PMs in a) different registers of the spoken medium, b) diverse genres/text types, and c) varying degrees of spontaneity.

4. RESULTS AND DISCUSSION

As mentioned in Section 3, a total number of 300 tokens of *you know* and 300 tokens of *I mean* were analyzed, adding up to 600 occurrences. The distribution of the canonical, in-between, and PM uses of *you know* and *I mean* is shown in Figures 1 and 2.⁷ As can be seen in these figures, there are important differences between the two expressions, although their use as PMs is the most pervasive function in both cases. Whereas PM *I mean* accounts for more than 90 percent of the uses in the three corpora, with exceptionally few canonical and in-between uses, the proportion of PM *you know* is much lower, especially in CSPAE and SOAP, in which it makes up for solely 72 and 60 percent, respectively. Additionally, whereas the chi-square test reveals that the amount of use of PM *I mean* does not vary significantly between the three corpora, the results for *you know* do exhibit a significant difference.⁸

⁵ In a recent study, Tagliamonte (2016: 188–193) proposes that *you know what* is undergoing a significant change, from being a fully-fledged question to becoming a PM, as evidenced by teenagers' and young adults' current use of this expression (from the 1990s onwards). In this new use it functions as a rhetorical question which requires no answer on the part of the addressee, but which rather allows speakers to comment on their own discourse, forming a new type of narrating style, as exemplified in (9). An analysis of this new PM, however, lies beyond the scope of this article.

⁶ It is necessary to acknowledge that, despite the importance of extra-linguistic content such as prosody, tone of voice, gestures, and facial expressions for an exhaustive analysis of PMs, these features could not be analyzed due to lack of access to audio and video recordings.

⁷ Note that the number of cases extracted was 100 per corpus and, consequently, the raw frequencies and the percentages coincide.

⁸ In these analyses, when conducting a chi-square test, only differences above the 95 percent confidence threshold (i.e. $p < 0.05$) will be considered as statistically significant.

These findings suggest that, whereas PM *I mean* is equally frequent in the distinct text types represented by COCA, CSPAE, and SOAP, the proportion of use of PM *you know* depends much more on the genre. This, in turn, might indicate that there are more differences in the way PM *you know* is employed by speakers in different speech situations in contrast to *I mean*. Nevertheless, to test this claim it is necessary to compare the functions of the PMs separately in the different text types, which is done in Section 4.2, with *I mean* being the focal point in 4.2.1 and *you know* in 4.2.2. First, however, the focus lies on the comparison of the two PMs by accounting for the functions they share.

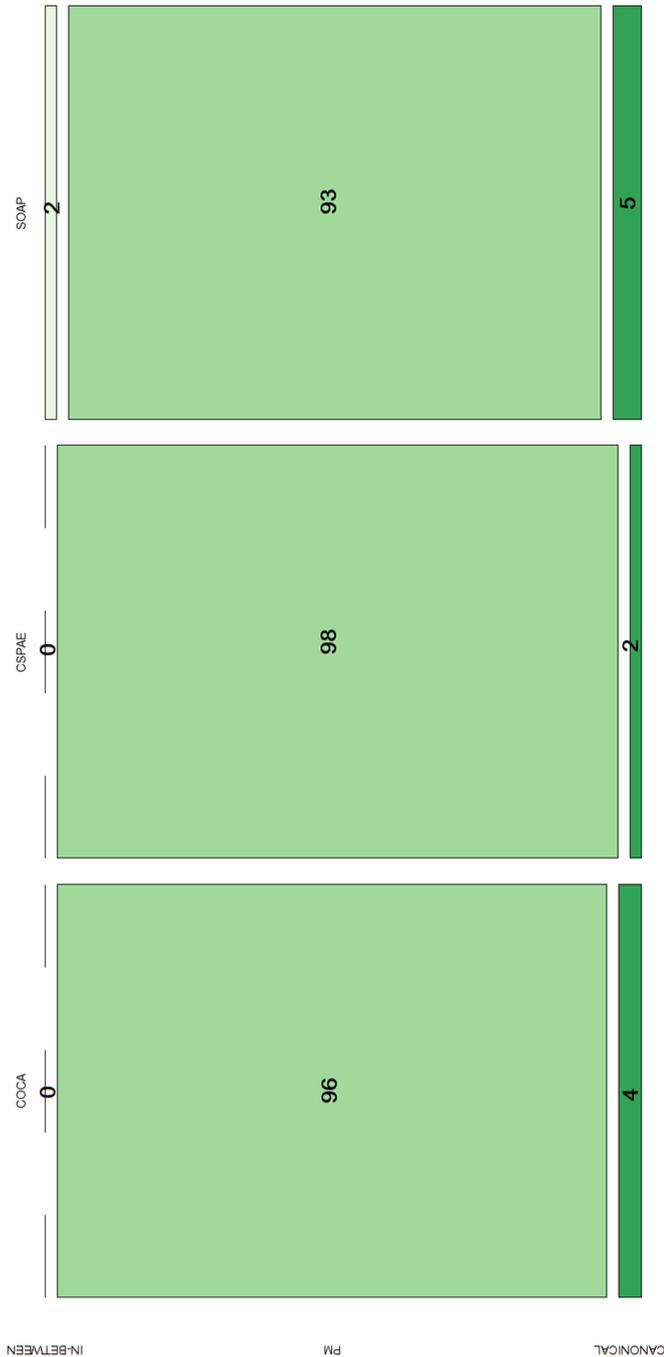


Figure 1: Uses of *I mean* in the three corpora
($\chi^2=5.405$, $df=4$, $p>0.05$)

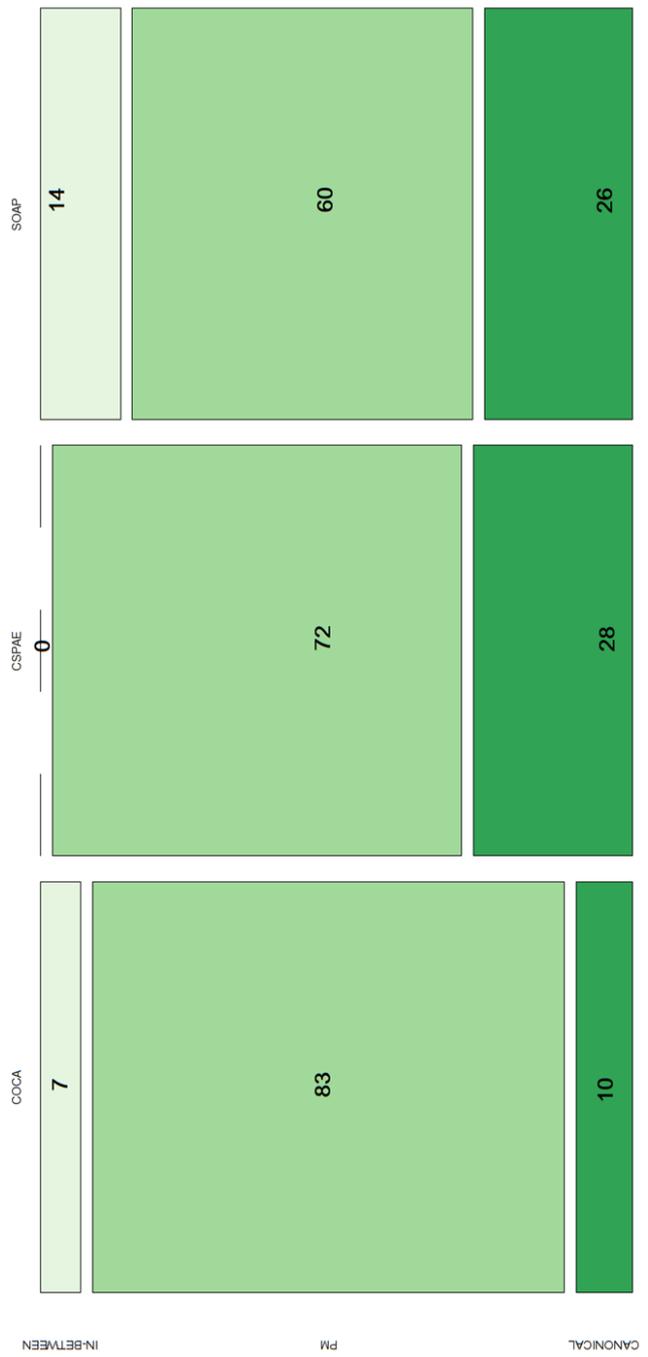


Figure 2: Uses of *you know* in the three corpora
($\chi^2=26.818$, $df=4$, $p<0.001$)

4.1. Comparing *I mean* and *you know*

As discussed in Sections 1 and 2, *I mean* and *you know* sometimes behave similarly in discourse as some of their functions fall into the same categories of use. According to Beeching's (2016) classification, which I have adopted here, both PMs can be used for the following three purposes: to clarify or explain a preceding statement, thus serving as a link between segments of discourse; to hesitate or stall for time to prevent the collapse of fluency of articulation; and to repair or reformulate ideas previously presented. In the analysis provided in this section, only the functions which the two PMs share are considered, to wit, clarification, hesitation, and (self-)repair. However, as shown in Figure 3, the proportion of uses of these three functions differs significantly between the two PMs.

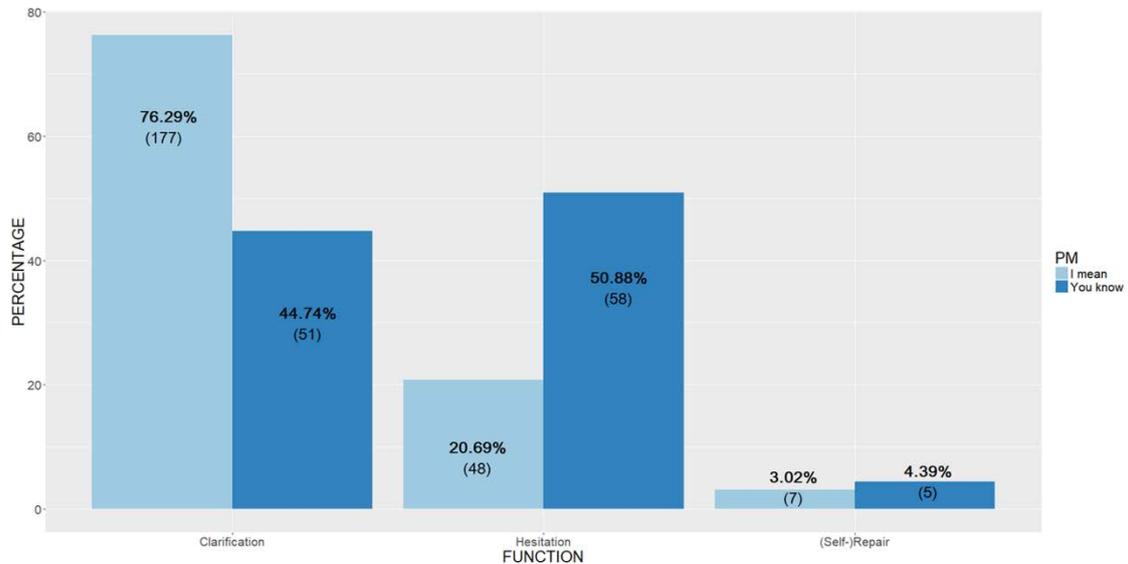


Figure 3: Distribution of *I mean* and *you know* across the functions they share ($\chi^2=34.7016$, $df=2$, $p<0.001$)

Whereas most instances of *I mean* (76.29 percent) serve as clarifying devices, as illustrated in (10), only 44.74 percent of *you know* pertain to this category. This might not come as a surprise given the fact that it is precisely this function of *I mean* which retains most of the semantic load from its canonical sense 'to signify' (see Beeching 2016: 190). In (10), former White House Press Secretary Dee Dee Myers is discussing a decision to be made by Clinton when a journalist questions whether he might decide on the matter earlier than expected. The journalist directly concludes that the enquiry is not precise enough and uses *I mean* to further clarify their intentions. Not only is this use much more frequently introduced by *I mean*, but there also seem to be some rather different implications of choosing *you know* instead of *I mean* as a clarifying mechanism. In light of the many instances of the type of example (11), *you know* appears to be used not only with a clarifying function, but also to get the hearer to see things from the speaker's perspective. In (11), Kendall tries to come up with an excuse for treating Ryan badly and introduces *you know* to clarify why he was 'so freaked out' while simultaneously showing that he was scared and, therefore, acted impulsively without thinking clearly about the future consequences of his actions. Hence, *you know* serves to signal to the hearer that the marked utterance should be considered as highly relevant for the argument put forward, thus functioning intersubjectively as well (see also Jucker and Smith 1998).

(10) Myers: [...] And he's [President Clinton] right now reviewing the situation and we'll make a decision on it by the 3rd.

Voice: Is there any chance that he would make the decision as we're heading to Europe? *I mean*, would he decide this as late as the 1st or the 2nd? (CSPA, WH94)

(11) Kendall: I never meant to hurt you. I just – I got so freaked out, *you know*, I was so sure that I was gonna get sent away to prison for life, and the idea of being away from my boys and everyone I care about. (SOAP, 2009, AMC).

Concerning the PMs use to stall for time when in doubt, Figure 3 makes evident that *you know* is substantially more frequent than *I mean*, as this function totals about half the uses of *you know* and only about a fifth of *I mean*. These findings coincide with Erman's (1987) and Beeching's (2016), which also display a higher use of hesitation with *you know*. As the hesitation function is one of the uses where the

literal meanings of PMs are most semantically bleached, there is no clear reason why this category is more pervasive with *you know*. However, it is important to highlight that in a considerable number of cases when *you know* performs this function, it invites the hearer to collaborate in the composition of the message by drawing the necessary inferences to complete the gaps left by the speaker. Thus, as in the case of the clarifying use discussed above, *you know* often functions intersubjectively in hesitations as well. Such a case is exemplified in (12):

- (12) Bob: [...] Things okay at the hospital?
 Chris: Fine.
 Bob: Personal life?
 Chris: Well, you and Mom will be happy to know that Alison and I – well, I slowed things down considerably.
 Bob: Oh, I see. *You know*, I like Alison. I really do.
 Chris: But?
 Bob: Well, we both know that she's her own worst enemy. It is difficult to be involved with somebody who has so many problems. (SOAP, 2008, ATWT)

In (12), Bob's answer to his son's statement about him and Alison slowing things down is a clear hesitation, as made evident by the introduction of three consecutive PMs: *Oh, I see* and *you know*. However, although *you know* indicates hesitation about what Bob should say, it could simultaneously invite his son to further elaborate on the reason of the 'slowing down' with Alice to assure Bob of what he should say next. Similarly, *I mean*, although sometimes serving exclusively as a hesitation, as in (13), exhibits a double function in many cases, since it usually also expands on previous ideas, as illustrated in (14).

- (13) Gabi: Wow. You're really gay. How do you feel about this?
 Will: Oh, um... *I mean*, I'm at peace mostly, you know? (SOAP, 2012, DAYS)
 (14) Van-Susteren: But I mean, I hear so many people say Trump has got such great kids. You know, *I mean*, it's not like you know, you have got the one side where people, you know, are complaining about but then they point to the fact that, you know, he is a good father. (COCA, 2016, Fox: On The Record: Susteren)

In (13), Will clearly uses *I mean* as a time staller while he searches for an appropriate answer to Gabi's question. In (14), however, besides indicating hesitation, which is obvious from the many instances of PMs used, *I mean* is also employed to further elaborate on the utterance *I hear so many people say Trump has got such great kids*, thus serving to connect the said utterance to the following segment.

Finally, regarding *you know* and *I mean* as repairing devices, they are not frequently used in this function in my data, and do not seem to differ in this category. These findings are in line with those obtained by Beeching (2016: 105, 191): in the case of both *you know* and *I mean*, (self-)repair is the second least frequent function in her data. The low frequency of this (self-)repair strategies in the present dataset could in part be due to the fact that the corpora examined represent rather careful –the radio and TV programs extracted from COCA and the White House press conferences and faculty and committee meetings in CSPAE – and scripted language – the soap operas represented in SOAP –, as these types of speech unsurprisingly contain a limited amount of disfluencies and on-line production errors.

4.2. I mean and you know in different registers and degrees of spontaneity

4.2.1. The case of *I mean*

The result of the chi-square test computed for the distribution of the various functions of *I mean* across the three corpora is statistically significant, as shown in Figure 4, and suggests some variation concerning how *I mean* is employed in the different text types, although there are similarities as well.

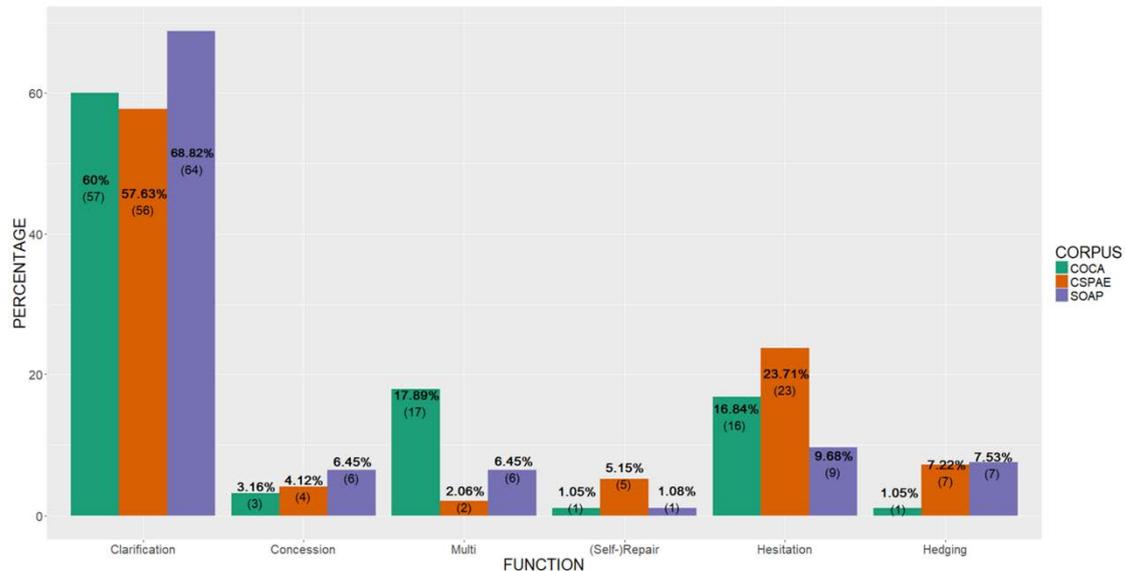


Figure 4: Distribution of *I mean* per function in the three corpora ($\chi^2=31.5042$, $df=10$, $p<0.001$)

Starting with the similarities, the most predominant function of this PM is as a clarifying device, since it accounts for more than half of the cases in the three corpora: 60, 57.63, and 68.82 percent in COCA, CSPAE, and SOAP, respectively. This was expected since it is precisely this function which lies closest to the semantics of canonical *I mean*. A typical example of *I mean* in this category was discussed in (10) above (Section 4.1). The corpora also coincide regarding the least prominent functions of *I mean*, namely (self-)repair and concession. These functions add up to no more than 7 percent of its uses in any of the corpora and there are few instances in the data analyzed to further discuss these functions. However, it is worth highlighting that in the case of self-repairs, my results correspond with findings obtained in previous studies (e.g. Beeching 2016: 185), as this use of *I mean* has been shown to be quite infrequent. However, the same cannot be said about concessions, exemplified in (5) above (Section 2), where the speaker employs the *I mean...but* structure to nuance the argument to indicate that they do not disapprove of other ideas presented concerning the elaboration of the tests discussed, but then introduces *but* to clarify their own viewpoint. In Beeching's (2016: 191) analysis, for instance, this function has been demonstrated to be fairly frequent while it is rare in my data. However, in Beeching's data only a few of the genres analyzed, namely role-play, spoken demographic, and business, account for the rather high frequency of concessions, thus suggesting that this function may just be frequent in specific text types.

Turning now to the differences between how this PM is used in the various text types, Figure 4 shows that *I mean* is more often employed to indicate hesitation in CSPAE than in COCA and SOAP, especially to stall for time or as a resource to plan what to say next. One such case retrieved from CSPAE is given in (15):

- (15) Voice: [...] which of the activities listed in the documents that we got today would he [President Clinton] see done away with, eliminated?
 McCurry: Well, that's – *I mean*, you go back and look at McCain-Feingold, those activities directed to the raising of soft money contributions wouldn't be necessary because you wouldn't have soft money donations. (CSPA E, WH97A)

At first glance, this finding might seem surprising given that CSPAE can be said to contain the most formal speech and less spontaneous discourse than the discourse in COCA, as the topics discussed are known beforehand. However, given that the CSPAE data for the White House Press conferences includes solely the question-and-answer sessions, this genre cannot be classified as purely planned speech. One reason for this high occurrence of hesitation in CSPAE might be the relevance and importance of the content transmitted by speakers in these situations. For instance, in (15), McCurry, the Press Secretary during the Clinton administration, is speaking on behalf of the government of the United States, which means that not only what he says but also how he presents ideas is going to be highly scrutinized. Consequently, he needs to think before he answers the questions posed to opt for an unambiguous and precise response. *I mean* as a hesitation device is thus very useful in these contexts to allow for some planning time. These findings are consistent with those obtained by Szczyrbak (2014: 292) for police interviews and Fernández-Polo (2014: 62–63) for conference presentations, which also exhibit a high frequency of *I mean* for this purpose (see also Fox Tree and Schrock 2002: 741).

Regarding *I mean* as a marker of subjectivity when hedging upcoming statements, there are more cases in CSPAE and SOAP than in COCA: 7.22, 7.53, and 1.05 percent, respectively. In the case of SOAP, this might be explained by the fact that many of the topics included are highly sensitive in the sense of being possibly offensive, as in (16):

- (16) So I decided to check out one of those meetings. And you know, sitting there listening to those stories [...] I realized what my limitations and boundaries were. And, Lulu, you should go. *I mean*, especially you. It's your father and your brother, they're both addicts. (SOAP, 2011, GH)

Here, the speaker is recommending that Lulu attend a meeting for addicts since two members of her family have problems with substance abuse. This is clearly a delicate topic and, therefore, the speaker introduces *I mean* to soften the suggestion. Moreover, giving advice, especially in such a sensitive speech situation, is considered a face-threatening act, making it even more necessary to hedge the statement.

With respect to the hedging function of *I mean* in CSPAE, most instances are employed to save the face of the speaker in the sense of making themselves less committed to the statement uttered. One such example is (17), in which Kapinus reformulates what he is about to say to distance himself from the idea that exams are not the only way of assessing students' knowledge, since no exam is exhaustive enough to test everything they should have learned. He begins by saying *I will just say*, hence referring to himself, but then introduces *I mean* and changes to *you said*, thus removing himself from the statement.

- (17) Kapinus: Okay. I will just say, first of all, and Marsha knows this, but the test can't be everything. It can't give teachers every – *I mean*, you said it won't give teachers everything they need to know. No test can give teachers everything they need to know. (CSPAE, COMR6A97)

Finally, there are cases in which the PM performs various functions simultaneously; these are categorized as 'multi' in Figure 4. A majority of these multi-functional examples combine a hedging use with some other function: clarification (9 instances), staller (6), and self-repair (4). As Figure 4 demonstrates, COCA displays a particularly high number of such cases. Example (18) illustrates a hedging plus clarifying use of *I mean*:

- (18) Gifford: You don't want to be there when that happens because that's just heartbreaking. *I mean*, for Philadelphia fans, they're thrilled. (COCA, 2010, NBC Today)

Frank Gifford, a sports commentator, when narrating the results of an American football match between the New York Giants and the Philadelphia Eagles, introduces *I mean* to elaborate on his previous comment, which could be interpreted as taking the side of the Giants by qualifying their loss as *heartbreaking*, while simultaneously trying to appear more objective.⁹

In brief, while serving several different functions, the vast majority of instances of *I mean* perform a clarifying or justifying function, ranging from 57 to 69 percent of the total uses of this PM in the three corpora. Contrariwise, the occurrences of *you know* are more evenly distributed across the different functions, as will become clear in Section 4.2.2.

4.2.2. The case of *you know*

The distribution of the distinct functions of *you know* in COCA, CSPAE, and SOAP is marginally significant according to the results of the chi-square test, since it is rather close to the 95 percent ($p < 0.05$) confidence threshold, namely $p = 0.0582$, as shown in Figure 5.

⁹ Frank Gifford, before becoming a sports commentator, played for the New York Giants for 12 years.

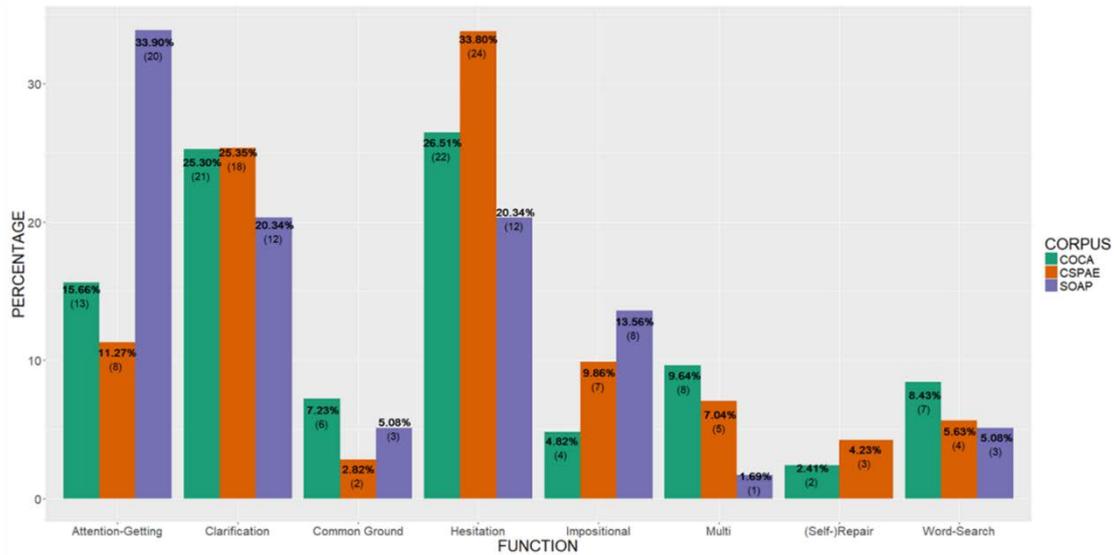


Figure 5: Distribution of *you know* per function in the three corpora ($\chi^2=23.129$, $df=14$, $p=0.0582$)

The frequency of *you know* in each function points to several differences with respect to how it is used in the three corpora, especially between COCA and CSPAE, on the one hand, and SOAP, on the other. First, as concerns the functions of (self-)repair, hesitation, and word-search, *you know* is much less frequent in SOAP than in the other two corpora. This is not surprising if we take the scripted nature of SOAP into account: even though the language used tries to depict natural speech, as evidenced by the high frequency of informal words and phrases, as for instance, *right?*, *okay?*, *creep out*, and *freak out*, these three uses of the PM are typical of online production as they all represent common types of disruptions of language processing.¹⁰ Two such examples are given in (19) and (20):

- (19) Dossey: There was a reaction that the long – *you know*, the long response is probably inappropriate for the students at this grade level. (CSPA E, COMM8A97)
- (20) Patrick-Monahan That’s really the goal, *you know*. Like, to be like, see? Look, they keep – keep getting better, this music is. And so that’s really the pressure. (COCA, 2014, NBC Today)

In (19), *you know* is introduced by the speaker since they get tongue-tied and do not find the word they are searching for, namely *response*. In (20), the interlocutor is hesitating throughout the statement, as it is brimmed with PMs (*you know*, *like*, *see*), exclamations (*look*), and repetitions (*like*, *keep*). Moreover, Patrick does not come to any coherent conclusion as he starts various utterances without finishing any of them. It also seems as if he wishes to invite his hearer/s to fill in the gaps or at least to somehow signal that they understand the content, thus enjoining them to co-construct the message. It is necessary to point out the difficulty in teasing apart the two said functions without access to more information, for instance, prosody, gestures, and/or facial expressions. Nonetheless, speech situations expressing either hesitation, word-search, and/or repair are typical contexts where PMs appear due to their usefulness for planning and modification on the go. As Beeching (2016: 100) correctly and pertinently points out, “[p]ragmatic markers are perfectly adapted to the linear online editing which is required in spontaneous speech, providing the cement which allows the discourse to build in a cyclic manner, phrase by phrase”. In relation to hesitation, it is worth mentioning that this function is especially frequent in CSPAE, just as in the case of *I mean*. This fact is probably related to the importance, in those text types (i.e. committee meetings and political press conferences), of precision and prevention of misunderstanding, as discussed in Section 4.2.1.

¹⁰ For more details on the design of the SOAP corpus and the type of language it represents, see the help section available at <https://corpus.byu.edu/soap>.

As concerns the attention-getting and impositional uses of *you know* (see examples (21) and (22) below, respectively), these functions are more frequent in SOAP than in CSPAE and COCA.¹¹ An explanation for these findings might be found in the very nature of soap operas, a genre full of melodrama, the purpose of which is to entertain the audience by depicting provocative and dramatic scenes. In fact, *soap opera* is defined in Merriam-Webster as “a serial drama [...] characterized by tangled interpersonal situations and melodramatic or sentimental treatment” (*M-S* s.v. *soap opera* 1a). The explanation postulated here is that attention-getting and impositional *you know* are used as a resource in soap operas to aid in the creation of sensational arguments between characters and exaggerated situations, to engage and retain the audience. These contexts are precisely where the two aforementioned uses of *you know* predominate, as exemplified in (21) and (22):

- (21) Carly: If you want to save your marriage, fight for it.
 Holden: *You know*, I'm trying. The problem is all I can think about is you. (SOAP, 2008, ATWT)
- (22) Carly: You know what? Your friends aren't gonna act like that. And if they do, it's just because they're worried about you.
 Liberty: I'm tired of everybody being worried about how sick I am, *you know*? (SOAP, 2010, ATWT)

In both cases, we have instances of dramatic and sensationalist situations. (21) is a case of the PM as an attention-getting marker, in which Holden emphasizes his subsequent statement, that is, his attraction to Carly, in a defensive way as a response to her previous suggestion of fighting for his marriage. If *you know* was removed from the utterance, Holden's statement would sound considerably less emotional. Similarly, in (22), impositional *you know* charges the discourse with emotion: Liberty treats her previous statement as a self-evident fact, thus urging Carly and everybody else to stop worrying about her sickness and to refrain from commenting on it further.

As discussed earlier (Section 2), *you know* always functions intersubjectively due to its explicit mention of the second person pronoun *you*, even though it often serves other functions simultaneously (e.g. attention-getting or impositional). However, as Figure 5 demonstrates, instances in which this PM directly refers to the interlocutors' shared knowledge are scarce in my data. These findings could indicate that *you know* many times signals that the hearer should consider the implications and relevance of interlocutors' shared information for the correct interpretation of the current message, rather than pointing directly towards the shared knowledge of speakers (see (11) in Section 4.1). This low frequency of *you know* as a direct appeal to shared knowledge matches that obtained by Beeching (2016: 105). One such example is the following:

- (23) Gayle-King: [...] Tea, look, first let's say happy birthday. Yesterday was your birthday.
 Tea-Leoni: Thank you.
 Gayle-King: Did you stay up late celebrating?
 Tea-Leoni: No, I didn't because I had to come here and be all, *you know*, morning happy. (COCA, 2015, CBS ThisMorning)

In (23), Gayle King, the host of the American morning show *This Morning*, wishes her guest Tea Leoni happy birthday, followed by a question of whether she celebrated the night before. Leoni answers that she did not celebrate because she had to be at the show, a fact which is well-known by King. PM *you know* in this example hence appeals directly to the common ground both women share.

Finally, it is necessary to comment on the category 'multi'. Contrarily to PM *I mean*, most instances of *you know* classified into this category are examples in which it was too difficult to tease apart the hesitation and word-search uses. These instances were consequently classified as belonging to both functions, due to lack of more context, be it extra- or intralinguistic.

¹¹ Prefatory *you know what*, which is similar to attention-getting *you know*, is also far more frequent in SOAP, in which it amounts to 13 of the 14 examples of in-between uses. The fact that this phrase is particularly common in SOAP ties in with Tagliamonte's (2016: 192) claim that *you know what* is on the rise among the younger generations, as a large number of characters in the soap operas included in SOAP are teenagers or young adults. Analyses devoted to this expression and its undergoing development could thus constitute an interesting avenue for further research along the lines of Torgersen et al.'s (2011) study on the evolution and innovation of nine pragmatic markers (e.g. *innit*, *you get me*, and *you know what I mean*) in London English.

5. SOME CONCLUDING REMARKS

The current article has presented three different sets of results: differences and similarities between *I mean* and *you know*, how *I mean* is used in the three text types analyzed, and the different functions of *you know* across the three corpora. First, as concerns the comparison of the two PMs, an important difference between them is their frequency as PMs as opposed to canonical and in-between uses. Whereas there are hardly any instances of canonical or in-between uses of *I mean*, the frequency of PM *you know* is much lower. With respect to the three functions the two PMs share, namely clarification, hesitation, and repair, *I mean* is much more predominant as a clarifying device and *you know* as a marker of hesitation. Additionally, using *you know* or *I mean* in these functions has rather different implications: *you know* nearly always adds an intersubjective facet in these cases, thus directing addressees to draw the necessary inferences from the message.

Second, as regards the use of *I mean* in the corpora analyzed, the findings demonstrate that, despite being predominantly used as a clarifying device, its frequency in other functions varies depending on the text type. On the one hand, hesitating *I mean* is much more common in CSPAE, probably due to the utmost importance of preciseness and accuracy in the speech situations present in this corpus. On the other hand, hedging *I mean* is far more frequent in CSPAE and SOAP. This is due, in the former case, to a need for speakers to make themselves less committed to their message, thus serving to protect their own face, and, in the latter, to the delicate topics dealt with, which make it necessary for characters to soften their claims.

Third, concerning the functions of *you know* across the text types analyzed, the results display a more even distribution than in the case of *I mean*, since there is not one dominant category. Moreover, the main differences lie between COCA and CSPAE, on the one hand, and SOAP, on the other. Repairs, hesitations, and word-searches occur far less frequently in SOAP due to the scripted nature of the language this corpus depicts. Contrariwise, attention-getting and impositional *you know* are more common in SOAP than in the other two corpora. A possible explanation to account for this pervasiveness of these two uses on the basis of the specific characteristics of the genre of soap operas, namely its melodramatic nature, was proposed in Section 4.2.2. However, this claim is yet a tentative one and further research is needed to clarify this issue. Therefore, a suggestion for future studies on *you know* is to conduct a more comprehensive analysis of the use of the PM in soap operas or similar text types to throw further light on its use in this type of speech.

Notwithstanding the clear patterns of use regarding the two PMs uncovered in the present article, one potential weakness of the study should however be mentioned. The database examined comprises 300 examples of *I mean* and 300 examples of *you know*, which could question the validity of the conclusions drawn from the findings. This rather restricted dataset is partially due to the overall qualitative nature of the study presented, which consists in a thorough and in-depth analysis of the examples extracted from the corpora. In spite of this limitation, I firmly believe that the results obtained shed valuable light on the behavior of these two PMs in different and, as to now, under-researched text types. Additionally, many previous qualitative studies on these and other PMs (e.g. Fernández-Polo 2014 and Serrano-Losada 2015) have dealt with similarly limited samples, but still contributed interesting and significant findings to this field of research. However, to fully assess the generalizability of the findings presented in this preliminary study, a larger and thus more representative database could be examined in future works.

In conclusion, this article has provided valuable evidence on the distinctiveness and autonomy of the PMs *I mean* and *you know*, which, despite overlapping in some functions, show a clearly different distribution across functions and genres. Additionally, this study has demonstrated how *I mean* and *you know* vary in use depending on the characteristics of the speech situations in which they are used, to wit, the degree of spontaneity and formality. In spite of the amount of literature devoted to the controversial topic of PMs in general, and *I mean* and *you know*, in particular, much remains to be done, especially to shed light on how PMs that overlap in certain functions vary when it comes to aspects such as frequency and/or the implications to be drawn from using one instead of the other/s.

REFERENCES

- Aijmer, Karin and Anne-Marie Simon-Vandenberg. 2011. Pragmatic markers. In Jan Zienkowski, Jan-Ola Östman and Jef Verschueren eds. *Discursive pragmatics*. Amsterdam: John Benjamins, 223–247.
- Barlow, Michael. 2000. *Corpus of Spoken, Professional American English*.

- Beeching, Kate. 2016. *Pragmatic markers in British English: meaning in social interaction*. Cambridge: Cambridge University Press.
- Blakemore, Diane. 2002. *Relevance and linguistic meaning: the semantic and pragmatics of discourse markers*. Cambridge: Cambridge University Press.
- Brinton, Laurel J. 1996. *Pragmatic markers in English: grammaticalization and discourse functions*. Berlin: Mouton de Gruyter
- Brinton, Laurel J. 2007. The development of *I mean*: implications for the study of historical pragmatics. In Susan Fitzmaurice and Irma Taavitsainen eds. *Methods in historical pragmatics*. Berlin: Mouton de Gruyter, 37–80.
- Brinton, Laurel J. 2017. *The evolution of pragmatic markers in English*. Cambridge: Cambridge University Press.
- Brown, Penelope and Stephen Levinson. 1987. *Politeness: some universals in language usage*. Cambridge: Cambridge University Press.
- Davies, Mark. 2011–. *Corpus of American Soap Operas*. <http://corpus.byu.edu/soap>
- Davies, Mark. 2008–. *Corpus of Contemporary American English*. <http://corpus.byu.edu/coca>
- Crystal, David. 1988. Another look at *well, you know*.... *English Today* 1: 347–349.
- Crystal, David and Derek Davy. 1975. *Advanced conversational English*. London: Longman.
- Erman, Britt. 1987. *Pragmatic expressions in English: a study of you know, you see and I mean in face-to-face conversation*. Stockholm: Almqvist & Wiksell.
- Erman, Britt. 2001. Pragmatic markers revisited with a focus on *you know* in adult and adolescent talk. *Journal of Pragmatics* 33: 1337–1359.
- Erman, Britt and Ulla-Britt Kotsinas. 1993. *Pragmaticalization: the case of ba' and you know*. Stockholm: Almqvist & Wiksell.
- Fernández-Polo, Francisco Javier. 2014. The role of *I mean* in conference presentations by EFL speakers. *English for Specific Purposes* 34: 58–67.
- Fitzmaurice, Susan. 2004. Subjectivity, intersubjectivity and the historical construction of interlocutor stance: from stance markers to discourse markers. *Discourse Studies* 6/4: 427–448.
- Fox Tree, Jean E. 2007. Folk notions of *um* and *uh, you know* and *like*. *Text & Talk* 27/3: 297–314.
- Fox Tree, Jean E. and Josef C. Schrock. 2002. Basic meanings of *you know* and *I mean*. *Journal of Pragmatics* 34: 727–747.
- Fraser, Bruce. 1990. An approach to discourse markers. *Journal of Pragmatics* 14: 383–395.
- Fraser, Bruce. 1999. What are discourse markers? *Journal of Pragmatics* 31: 931–952.
- Fromkin, Victoria A. 1973. *Speech errors as linguistic evidence*. Berlin: Mouton de Gruyter.
- Goldberg, Julia A. 1980. *Discourse particles: an analysis of the role of y'know, I mean, well, and actually in conversation*. Phd. University of Cambridge.
- He, Agnes Weiyum and Brian Lindsay. 1998. 'You know' as an information status enhancing device. Arguments from grammar and interaction. *Functions of Language* 5: 133–153.
- Jucker, Andreas H. and Sara Smith. 1998. And people just you know like 'wow': discourse markers as negotiating strategies. In Andreas H. Jucker and Yael Ziv eds. *Discourse markers: descriptions and theory*. Amsterdam: John Benjamins, 171–201.
- Koczogh, Helga Vanda and Furkó Bálint Péter. 2011. Gender differences in the use of the discourse markers *you know* and *I mean*. *Argumentum* 7: 1–18.
- Lam, Phoenix W.Y. 2010. Toward a functional framework for discourse particles: a comparison of *well* and *so*. *Text & Talk* 30/6: 657–677.
- López-Couso, María José. 2010. Subjectification and intersubjectification. In Andreas H. Jucker and Irma Taavitsainen eds. *Historical Pragmatics*. Berlin: Mouton de Gruyter, 127–163.
- Macaulay, Ronald. 2004. *You know, it depends*. *Journal of Pragmatics* 34: 749–767.
- M-S = Merriam-Webster Dictionary and Thesaurus*. 2017–. <https://www.merriam-webster.com>
- O'Donnell, William and Loreto Todd. 1991. *Variety in contemporary English*. London: Routledge.
- Östman, Jan-Ola. 1981. *You know: a discourse-functional approach*. Amsterdam: John Benjamins.
- Redeker, Gisela. 1990. Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics* 14/3: 367–381.
- Schiffrin, Deborah. 1987. *Discourse markers*. Cambridge: Cambridge University Press.
- Schourup, Lawrence C. 1985. *Common discourse particles in English conversation*. New York: Garland.
- Scott, Mike. 2012. *WordSmith Tools* version 6. Liverpool: Lexical Analysis Software.
- Serrano-Losada, Mario. 2015. Element-final *like* in Irish English: notes on its pervasiveness, incidence and distribution. In Cristina Suárez-Gómez and Elena Seoane eds. *Englishes today: multiple varieties, multiple perspectives*. Newcastle Upon Tyne: Cambridge Scholars, 9–31.

- Stubbe, María and Janet Holmes. 1995. *You know, eh* and other 'exasperating expressions': an analysis of social and stylistic variation in the use of pragmatic devices in a sample of New Zealand English. *Language & Communication* 15/1: 63–88.
- Szczyrbak, Magdalena. 2014. Pragmatic marker use in police interviews: the case of *I mean* and *you know*. *Studia Linguistica Universitatis Jagellonicae Cracoviensis* 131: 371–379.
- Tagliamonte, Sali. 2016. *Teen talk: the language of adolescents*. Cambridge: Cambridge University Press.
- Torgersen, Eivind Nessa, Costas Gabrielatos, Sebastian Hoffman and Susan Fox. 2011. A corpus-based study of pragmatic markers in London English. *Corpus Linguistics and Linguistic Theory* 7/1: 93–118.
- Traugott, Elizabeth C. 2003. From subjectification to intersubjectification. In Raymond Hickey ed. *Motives for language change*. Cambridge: Cambridge University Press, 124–140.
- Watts, Richard J. 1989. Taking the pitcher to the 'well': native speakers' perception of their use of discourse markers in conversation. *Journal of Pragmatics* 13: 203–237.

Corresponding author

Daniela Pettersson-Traba
Campus Norte · Av. de Castelao s/n
15782 Santiago de Compostela
e-mail: daniela.pettersson@usc.es

received: November 2018
accepted: December 2018