การแปรของหน้าที่ของดัชนีปริเฉท “you know” ที่ใช้ในผู้พูดชาวอังกฤษ จากคลังข้อมูลภาษาอังกฤษแห่งสหราชอาณาจักร (BNC)

Variation of the Pragmatic Expression Functions of “you know” as Used by British English Interlocutors in the British National Corpus (BNC)

โกสินทร์ ปัญญาอธิสิน
Kosin Panyaatisin
ภาควิชาภาษาอังกฤษ คณะอักษรศาสตร์
Department of English, Faculty of Arts,
มหาวิทยาลัยศิลปากร
Silpakorn University

บทคัดย่อ

ในการใช้ภาษาอังกฤษแบบอังกฤษในการสนทนาทั่วไป “you know” จัดว่าเป็นดัชนีปริจเฉทที่มีความกำกวมว่าควรจะมีหน้าที่อะไรในการวิเคราะห์ข้อความ ไบเบอร์และคณะ (1999) เลคอฟ (1973) ให้ความเห็นว่า “you know” มักใช้ในเพศหญิง เนื่องจากเพศหญิงอาจจะมีความอ่อนไหวและไม่มั่นคงเกี่ยวกับการใช้ภาษามากกว่าเพศชาย ในประเด็นเรื่องหน้าที่ “you know” ส่งสัญญาณให้กับผู้ฟังทราบว่า ผู้พูดได้ตระหนักว่า ผู้ฟังน่าจะมีความรู้ หรือ ประสบการณ์ที่อ้างถึงร่วมกันแม้ว่าผู้ฟังจะทราบมาก่อนหรือไม่ก็ตาม อนึ่ง สิ่งนี้แสดงให้เห็นถึงความไม่มั่นใจในการให้ข้อมูลแก่ผู้ฟังเช่นกัน มีนักภาษาศาสตร์จำนวนมากได้วิเคราะห์หน้าที่ของดัชนีปริเฉทดังกล่าว อาทิ เลคอฟ (1973), โฮล์ม (1986), ชริฟฟิน (1987, rev. 2012), โรเมน และ แลง (1991), เออร์เมน (1992) ผู้วิจัยเห็นว่าแนวทางการวิเคราะห์หน้าที่ของดัชนีปริจเฉทของเออร์แมน (1992) มีความครอบคลุมและน่าจะอธิบายพฤติกรรมการใช้ภาษาในภาษาอังกฤษแบบอังกฤษได้ดีที่สุด

ดังนั้นงานวิจัยนี้มุ่งเน้นการวิเคราะห์หน้าที่ของปริเฉท “you know” โดยอาศัยแนวการวิเคราะห์ของเออร์แมน (1992) จากคลังข้อมูลภาษาอังกฤษแห่ง สหราชอาณาจักร (BNC) โดยมีการสร้างคลังข้อมูลขนาดเล็กขึ้นใหม่เพื่อให้สอดคล้องกับการวิเคราะห์ การศึกษานี้วิเคราะห์การแปรของหน้าที่ของดัชนีปริเฉท “you know” และการแปรของหน้าที่ของดัชนีปริจเฉทนี้กับตัวแปรทางสังคม ได้แก่ เพศ และ อายุของผู้พูด จากการวิจัยพบว่า ผู้พูดที่มีเพศต่างกันมีแนวโน้มในการใช้ดัชนีปริเฉทดังกล่าวที่แตกต่างกันอย่างมาก ในด้านตัวแปรอายุพบว่า ผู้พูดสูงวัย และกลุ่มวัยรุ่นใช้ดัชนีปริเฉทในลักษณะที่ตรงข้ามกัน นอกจากนี้ยังพบว่า งานวิจัยชิ้นนี้และงานวิจัยของเออร์เมนคล้ายคลึงกันในประเด็นเกี่ยวกับผู้พูดเพศหญิง โดยสรุปแล้ว อาจตีความได้ว่าผู้ฟังไม่ควรเหมารวมการใช้ดัชนีปริเฉทของผู้หญิงว่ามีหน้าที่เพียงอย่างใดอย่างหนึ่งหรือ เป็นการใช้ภาษาที่ด้อยกว่าผู้ชาย ผู้หญิงมีแนวโน้มการใช้ดัชนีปริเฉทเพื่อบ่งชี้หน้าที่ต่างๆ ในบทสนทนา ได้ค่อนข้างครอบคลุม และมีการกระจายตามหน้าที่ได้ดีกว่า ผู้ชายมีแนวโน้มการใช้ดัชนีปริเฉทเพื่อบ่งชี้หน้าที่ได้จำกัดกว่าซึ่งไม่ได้สอดคล้องกันกับผู้หญิงมากนัก

คำสำคัญ : ดัชนีปริเฉท, คลังข้อมูลภาษาอังกฤษแห่งสหราชอาณาจักร, การแปรภาษา, เพศ

Abstract[1]

In British English vernacular, “you know” can be regarded as a pragmatic expression (PE) (Biber. et. al, 1999) containing controversial functions in conversation. Lakoff (1973) suggests that it is more frequently used by linguistically insecure females (Labov, 2001) rather than by males. In terms of function, this filler is to some extent interpretable as a signal and reassurance for the hearers to share experiences, whether they know it or not. It reveals that the speakers are insecure by showing hesitation to their interlocutors. Lately, the study of PE has caught the attention of several linguists (Lakoff, 1973; Holmes, 1986; Schiffrin, 1987 rev. 2012; Romaine & Lange, 1991; Erman, 1992). The analytical approach to PE function, invented by Erman (1992), appears to be the most interesting due to its explanatory power and validity for PE function analysis, thus shedding light on how English native speakers use PE.

Therefore, this study aims at analysing the functions of the PE “you know” in the British National Corpus (BNC) by adopting Erman’s approach. From a sociolinguistic point of view, the gender difference and the age of speakers pertaining to the use of PE was investigated. The study indicates that men and women exhibit quite a clear trend of using this marker differently. The elderly and adolescents show a reverse trend of PE function usage. It was found that there was a similarity with regard to the issue of gender between this current study and Erman’s (1992) study. To conclude, it can be implied that female speech should not be stereotyped or seen as inferior to that of males. The distributions of the functions of “you know” in female speakers are better

[1] I would like to express my gratitude to the valuable and thought-provoking comments and sug gestions from the two peer-reviewers. Special thanks go to Fliss Rich from Essex, UK, who proofread this manuscript after the paper was accepted.

than in males, with male scores being relatively skewed. Therefore, the functions of “you know” in males do not correspond to female usage and practice.

Keywords: pragmatic expression, BNC, linguistic variation, gender

1. Introduction

Erman (1992: 217) states that research on differences between female and male speech demeanors and competence has been biased towards viewing women’s language as the deficient variant. Some female speech attributes have been characterised as deviating forms from the male linguistic norm. The notion of women’s language as an inferior variant of the dominance of the male standard, is regarded as the “dominance approach” (Sunderland, 2006). The dominance approach was historically much in use from the early and mid-1970s, mostly employing introspection and superficial analysis, rather than conducting empirical research. However, after the deficit of the dominance approach, Coates & Cameron (2014) noted that there were increasing numbers of empirical research papers on gender[2] and language (Homes, 1986) in the “difference approach”. These included studies in interruptions, tag questions, and hedging devices. These previous studies pointed towards male and female differences in communicative behaviour when sharing and exchanging characteristics and identities via linguistic representation.

From the variationist viewpoint of the “difference approach” in gender and language study, Erman (1992: 218) states that some researchers take a more neutral and objective attitude based on the assumption that women and men speak differently because they are members of a different sociolinguistic subculture, and they have internalised different norms for conversational interaction. Erman observes that certain linguistic features such as Pragmatic

[2] Cheshire (2004: 426) defines the differences between two terms: “sex” as the biological or physiological differences between men and women, whereas “gender” pertains to the psychological, social roles and cultural differences between males and females. However, this paper uses both terms, “sex” and “gender”, in the sense of physiological difference.

Expressions (PEs) are seen as markers of inferiority and powerlessness. Lakoff (1973) suggests that PEs, such as "you know" appearing in female language use, lead to associations of linguistic insecurity and inferiority being linked to women. Pertaining to the functions of PE in general , Biber, et al., (1999: 1086) states that PEs[3] are a kind of insertion tending to occur at the beginning of the turn or utterance such as "oh", "well", "you know" and "I mean", etc. Their roles are to signal a transition in the development of a conversational speech (Fraser, 1990) or to signal an interactive relationship between speaker, hearer and message (Schiffrin, 2012) meaning that they function as fillers, topic changes, hedges or back-channelling. To note, this topic interests both theoretical and variationist linguists (Schiffrin, 1987 rev. 2012; Clancy & O'Keeffe, 2015; Erman, 2001; Laserna, Seih & Pennebaker, 2014; Xue & Lei, 2016 and so forth). Therefore, an investigation of the function of the PE "you know," will show that it has several functions rather than only a single function, as well as implications of social meaning as markers of inferior or powerless speech, (O'Donnell & Todd, 1980; Regan, 1983).

To indicate each function of the PE "you know" is not simple. This is because on certain occasions, the single PE "you know" can carry several functions, many of them at a time and not singly in the discourse (please see §2.). However, to solve the polyfunctionality of the PE of "you know," it is plausible to seek for the most salient function of that PE by considering its specific context in the provided sub-corpora.

[3] Biber et al., (1999) named this linguistic expression as a discourse marker or a pragmatic marker. However, I prefer to call it a pragmatic expression, in accordance with Erman's work (1992).

A single PE can perform several functions and each function can compete with one another to a varying degree. This issue raises a problem in statistical analysis, which can be simply solved by counting all possible and additional function occurrences, accordingly. However, this is not the intention, for to do so would lead to another intriguing question investigation.

If in trying to generate other functions of “you know”, it would be relevant to query the certainty of the number of functions, and to what degree each function found would be valid and accurate (some functions of the PE might be stronger than others, to varying degrees)? I think that this step ought to undergo a prototype theory analysis (Rosch, 1973) for ascertaining the degree of salience of each function of “you know,” and what number of its functions are more falsifiable and accurate. However, this paper does not intend to deal with those two issues. I aimed to analyse the PE’s functions according to Erman’s (1992) general protocol. I limited each PE of “you know” to having only one function based on the five dichotomies (see §2.) in a single discourse, and that the function should be the most outstanding function according to the protocol.

Referring to the social factors with regards to the sociolinguistic/ variationist aspect, the difference between genders might identify something more interesting than the overgeneralisation that female speech tended to be inferior to that of males. In addition, the factor of age difference can be investigated, even though identifying the age of the speakers in the BNC sub-corpus is difficult.

I used the (spoken-restrictions) spoken sub-corpora from the British National Corpus (BNC) to investigate this issue of gender and age differences. This spoken sub-corpus accounts for around 10% of the entire BNC (the written one is 90%). Most previous PE studies were based on interviews or some small-scale corpora that focused on particular linguistic communities. To

investigate the other type of data could reveal a general trend of language use in an Anglo context, and this study aims to investigate the two following issues, namely:

1) to categorise the functions of "you know" from the BNC corpora by using the approach of Erman (1992), for showing the distribution of the data which leads to the generalisation; and

2) to compare the functions of "you know" across two social factors, namely the gender and age of speakers.

2. Research methodology

This section comprises several steps of research methodology and the related theoretical backgrounds are also discussed as follows:

Functions of Pragmatic Expression (PE)

I adopt the framework of Erman (1992)[4]. There are five functions of the PE "you know," ranging from 'lexical end' to 'pragmatic end', as shown below.

Figure 1. Functions of the PE "you know" on a level of 2 planes (adapted from Erman, 1992[5])

[4] We did not look at the same-sex and the mixed-sex talk because it was difficult to control in the BNC corpus. Many occasions we found that the numbers of the speaker are more than two as well as the gender of the interlocutors were unknown.

[5] Erman has then developed her theory of PE (Erman, 2001). However, I think that this current approach is still appropriate to use since it is valid with the strong explanatory power.

Referring to figure 1, the function of each PE can occur at any point on this continuum linear plane with two ends. The functions of PE can range from the textual level (with the lexical end) to the interactional level (with the pragmatic function end). Of course, the PE's functions are dependent on the context and grammatical cues. The first three PE's functions, ranging from decoding information (DEC), orientation in the discourse (ORI) and turn-taking (TURN), belong to the textual level, while the hesitation (HES) and repair/ appeal marking (REP) are included in the interactional level, respectively.

To elaborate this, Erman (1992) proposed five dichotomies of the PE's functions, and exemplifies how they operated during the conversations as followed.

2.1 Decoding information (DEC), refers to the speaker using the PE to urge the hearers to interpret the information conveyed in a certain way – in other words this function focuses on the decoding of the message (mostly occurring in the clause boundaries) and the expectation that the hearer will. It functions to draw the attention of the addressees and to ensure that they are following the addressed issue.

(1)[6]

B: But I said. Personally, I'm sorry I haven't replied but I would. I'm going to because I would like to come and…then he said well, don't bother now… You've told me. <u>You know</u>. You just come.

A: that was this term.

B: Um.

[6] These examples are adjusted from Erman (1992: 222-226) and from the spoken BNC, spoken sub-corpora. Some special annotations and symbols were disregarded and simplified.

In example 1, B speaker uses "you know" to function in the thematic structure: that is, the speaker uses "you know" as a signal for the listener to interpret part of the information as common ground, the PE typically appearing after the thematic or thematised parts of the message. Not surprisingly, this is a common function of "you know".

(2) from BNC

A: What street was Fred…on?

B: He was on the … Street.
It was that…Street then. And er just a small three roomed shop and they had er they had cabinet makers and French polishers you know? But he had worked with a big er in a great big firm and then he branched out on his own you know.
But he was a very very hard worker **you know**.
And er my chum's mother, she was a forewoman French polisher and he asked her to come and er work along with him you know, but it was I mean that was during the depression years,

Referring to example 2, B speaker tries to emphasise and repeat the content with regard to the actor "he" whom the speaker B refers. Speaker B attempts to reassure his argument and draw attention of the hearer several times according to the "you know" marker frequently appears.

2.2 Orientation in the discourse (ORI), is very similar to Schifrin's (1987 rev. 2012) discourse marker. The PE mark moves in the conversation and points out the boundaries between various stages in the conversation. It signals the hearers regarding the conversation orientation and organization.

In general, PE in this function serves as boundary markers between structural components of the following types: foreground and background, direct and indirect speech, consecutive events, and consecutive arguments.

(3)

A: I think she thought I was a career woman who would be only too glad or would say oh well,
he’s got to go into a hospital. You know. So she made the decision for me and then said, it’s too late now to put him into a…an isolation hospital. I would have had to do that a few days ago which I thought I didn’t want her to do. (laughs)

In example 3, PE may also be used to mark the transition between two modes of speech, viz. from direct to indirect speech or vice versa. This example shows that “you know” marks a transition from indirect to direct speech.

(4) from BNC

A: <unclear> so what are we going to do?
<pause> Lots of fun.

B: We just did what we were supposed to do really <pause> ha ha <pause> we actually read the first one of the quests together, so it was good <pause> you know, we decided what to do<pause> how to do it.
<pause dur=”15”> They all <unclear> <pause> one kind or another.

A: Is this the new library now, or is it <pause> the new extended <un clear> ?

B: Well.

As seen in the example, speaker B tried to connect several consecutive events. Thus, this marker functions as the discourse marker as addressed in Schiffrin (2012). This PE marker, to some extent plays a role as a common coordinating conjunction.

2.3 Turn-taking[7] (TURN), this PE is used in regulation of the conversation, namely, to mark turn taking/ beginning and turn yielding, and typically occurs at the beginning and end of turns.

(5)

B: you know, Pen. I thought that. Did I tell you when we were in this African village and they all out in the field…the

A: Yes, you did. Yes…Yes.

"You know" has the function of signaling not only turn taking, but also topic switching, whereby speaker B introduces a new topic into the discourse.

(6) from BNC

A: Does she, you, you said she was having vitamins for a while didn't you?

B: Not having now, god she, I mean she doesn't eat a lot, but she eats so many different things, I mean

A: Mm <-|-> having a good variety <-|->

B: <-|-> she's on the organic crisps at school <-|-> and er

A: well

B: fruit and she eats a lot of salad cos she loves salad you know

[7] regulation of turns is used by Erman, but this study prefers to use turn-taking because it looks more simple

A: Mm

B: if I had a salad I mean

A: it’s unusual that <-|-> for children really cos they don’t usually go for salad <-|->

The example reveals that speaker B used “you know” to the mark the turn, signal the next turn, and send the next to his interlocutor. The position of “you know” is placed at the end of the clause.

2.4 Hesitation (HES), according to Erman (1992), this expression shows the hesitation of the speaker. Pausing during the speech can be also the linguistic cue of this marker. “You know” can function as the filler or hedging device to some extent.

(7)

A: and…um (clear throat) I mean…when you get used to that beer which at its best is simply.
<u>You know</u>. Superb. Really. I really is.

B: Um.

A: <u>You know</u>… I ‘ve really got it now. Really. <u>You know</u>… Got it to a…

B: Yeah.

“You know” here serves the function of hesitation. This example shows how a speaker may use one PE for the sake of doing word search and try to continue the speech. It is clear that what is uppermost in the speaker’s mind here is not the decoding of the message (one of the textual functions of PEs), but rather its encoding. To note, multiple uses of the PE from the speaker and the long pause prior or after using the PE are the linguistic cues.

(8) from BNC

A: And we had a during the periods you mentioned we had er lived in a village, and we had a very very honest er grocer.
Who one day we thought well about time we had something so I think we said to him er about these things off the ration can't can't get, what happens to them?
Well we have them he said <laugh><unclear>.
And I think that that shows that **er you know** the distribution of er non rationed food was not quite what

B: Well it was the black market as well wasn't there?

C: Yes.

D: I was going to say there was a flourishing black market wasn't <unclear>

According to example 8, on one hand, the speaker A uses "er" several times. The uncertainty of the speaker can be implied by the use of marker "er". On the other hand, speaker A might need time to process/ construct the complete concept in his cognition and try to utter it out then. Thus, this PE function can be either categorised as hesitation and uncertainty at the same time in some incidents like this.

2.5 Repair/ appeal marking (REP), according to Erman (1992), the speaker is more concerned with the encoding of the message to the hearer. In connection with repair and appeal, the speaker signals to the hearer that something has gone wrong and needs to be corrected. The PE is then frequently accompanied by other PEs such as "I mean". In my view, the denotative meaning of "I mean" refers to the speaker trying to readjust/ clarify his idea that

he previously addressed. The speaker wishes to convey the message to hearers as precisely as possible. Thus, the elaboration of the idea in the discourse and the PEs, “I mean” and “you know” were used as the markers. These two PEs mostly co-occur (in-) between the boundaries of clauses or in the middle of the utterance.

(9)

C: oh this is fame lads. Fame.

B: I know. Um.

C: well your, your future is assured.

A: Fame and fortune.

C: if you ever want a job **you know**, (I mean).

B: yes. It’s like a recording studio. Isn’t it.

According to example 5, the two functions of repair and appeal are shown .The speaker does not seem to know how to go on and therefore simply gives up without finishing the utterance.

(10) from BNC

A: Look Ralph erm if, if you don’t mind me saying can you remember if you do this could you just give me a ring and say it’s done, if not Sid come down in the morning.
Alright?
I mean you got to get yourself in there round the back tuck yourself in all your body.
<pause> Cos we got to erm, you got a magazine <unclear>

B: He can see in the fridge but I can’t, I can’t.

A: **You know** but I mean when it comes to a screw and your, I mean we’re talking about coming like this to undo the screw.

B: Yeah round the back.

A: You, you need somebody really to hold the erm

As the excerpt in example 10 shows, speaker A uses "you know" and consecutively uses "I mean" immediately to show that he aimed to clarify something which he thinks might not be clear for the hearer. He then uses the conjunction "but" in order to show the orientation of the idea he wishes to convey to the hearer. Then, he repairs the previous utterance (which is plausibly misunderstood by the hearer) and appeals to the hearer again by using "I mean" in the next clause to confirm his ground argument, respectively.

3. The spoken sub-corpora retrieved from the BNCweb and the analysis

To investigate the corpora in the BNCweb, I delimited my data to the spoken data corpora. The PE of "you know" can be searched and annotated by_PNP (personal pronoun) and_VVB (the finite base form of lexical verb, comprising the indicative imperative and present subjunctive), respectively. I disregarded the denotative meaning of "you know" which senses "comprehension" or "understanding" and so forth. Also, the construction of interrogative statement such as "do you know...?" was eliminated. Any contracted or deviated forms such as "y' know" were discarded.

Even though Erman's work (1992) also included the contraction form "y' know", I feel it is unnecessary to cover these contracted forms that she did. This is because this paper focuses on the variations of the function of "you know", and does not aim to compare the variants of the form of the underlying /you know/[8] which can consist of many variants. It is certain that the

[8] The slant brackets "//" refer to the underlying form, similar to a phonemic underlying form (Labov, 2001).

associations between each variant (form) of /you know/ with variation of their functions (according to the variant forms of /you know/) will be varied as well.

In addition, the number of tokens/ hits in both full form and contracted form are numerous in the BNC spoken sub-corpora. I randomised the data and selected the tokens carefully. In Erman’s work (1992), she used all kinds of the (variants) forms. This might derive from the size of her data tokens being relatively small. Thus, all types of variants of “you know” need to be collected (see table 3.). In my view, these phonological variants, “you know” and “y’ know” still refer to the same thing pragmatically and grammatically[9]. This study did not focus on the variants of the phonological variation of “you know” associating with the variation of the functions of “you know,” even though it is interesting that the contraction form “y’ know” will show a different pattern of use when compared to the full form.

In terms of the spoken sub-corpus data management in BNC, the size of this spoken sub-corpus is 300 tokens. I employed “thinning” mode first for randoming and searching for the first 100 hits/ tokens for each group in order to make spoken sub-corpora for this study. The tokens were stratified by two genders and three age groups. To elaborate this, the demographical data stratification included 6 groups (or 6 sub-corpora), namely the young-male group, the adult-male group, the elderly-male group, the young-female group,

[9] Thepkanjana (2017) addressed that the PE markers have undergone grammaticalised processes. Their surface constructions are shown as clausal entities and they evolve as a grammatical category unit. To be specific, these PEs have undergone the pragmaticalisation process (Erman, 1989). Thus, they finally become the pragmatic markers/ expressions (PEs). Thus, their genuine denotative meanings might be bleached.

the adult-female group and finally the elderly-female group. Then, each group's tokens (100 tokens) were randomised by using "show in random order mode". The token frequency of each group was 100 tokens at first, according to the first query. Then, there were 600 hits in total. After considering the grammatical (not the denotative meaning of "perceiving/ understanding something") and the pragmatic criteria as introduced by Erman (1992), the first 50 tokens of each group were analysed. Therefore, the 6 spoken sub-corpora from the BNC web consisted of 300 hits (n = 300).

In Erman's study (2001), she also looked at ongoing linguistic change in the use of the functions of "you know". This was aimed to see the linguistic shifts. It can reveal how age difference affects the use of PEs in speakers as well. However, it should be borne in mind that this current study cannot conduct the change in apparent time (Labov, 2001) prediction in the BNC. This is because it is not only difficult to identify the time of the speakers' conversations, nor to compare each age group. The recorded time is hard to trace back to a specific point of time. The BNC was created around 1991, completed in 1994 and has been revised in 2001, 2007 and 2014[10], respectively. The spoken corpus was gathered from several researches and accumulated from different times of data recordings. Thus, both the age of the speakers and the recorded time point identified in the BNC, cannot be an appropriate representation. Therefore, the change in apparent time implication (Labov, 2001) cannot be claimed (neither panel study nor trend study can be done in the BNC in general). However, it is still interesting to see the distribution of the function of

[10] http://www.natcorp.ox.ac.uk[Accessed 1 September 2017].

“you know” across three age groups and two genders. These interactional analyses are still beneficial in order to understand the trends of the function of “you know” in general.

This paper selected an emic approach (Milroy and Gordon, 2008) for deciding the age groups of speakers. It consisted of three interval age groups: 15-24, 35-44 and 60+ (above) from the vernacular/ spoken data corpus from the BNC (see table 2.). This emic approach was less arbitrary since it categorised speakers into three groups based on their shared experiences in life at certain periods of time. Another social factor is gender. The gender here refers to a binary physiological difference of speakers only, (male and female) in general, not multiple shades of sexual orientations in speakers (Cheshire, 2004).

4. Findings

Table 1 General distribution of PE “you know” over spoken texts with sex as a factor

Category	No. of words	No. of hits	Frequency per million words
Female	3,290,569	11256	3,420.68
Male	4,949,938	13200	2,666.7
Total	8,240,507	24456	2994.19[11]

[11] http://www.natcorp.ox.ac.uk[Accessed 1 September 2017].

Table 2 General distribution of PE "you know" over spoken texts for age as a factor

Category	No. of words	No. of hits	Frequency per million words
60+	1,137,433	4,859	4271.9
45-59	1,636,743	5,101	3116.56
35-44	1,075,749	3,185	2960.73
25-34	1,120,307	3,957	3532.07
15-24	594,400	1,874	3152.76
0-14	385,234	1,069	2774.94
Total	5,949,866	20,045	3368.98

According to table 1 and table 2, the distribution is shown of the PE "you know" across two genders and several age groups, categorised by decade. There are 24,456 hits[12] over 451 texts in this sub-corpus. In table 1, even though females have sparingly lower frequencies (no. of hits) compared to males in the spoken text, females favour the PE "you know" more than males (3,420.7 vs. 2,666.7, according to the freq. per million words). In table 2, the age stratification of speakers is presented by decade in the spoken corpus. According to the age-emic stratification, these three age groups represent three generations in these sub-corpora (dark bars). The elderly group (+60 yrs old) gains the highest frequencies including the number of word-counts and hits, and is followed by the adult group (35-44 yrs old) with its word-count frequency being slightly similar to the elderly group. To note, the youngest group in the corpus (0-14 yrs old) shows a very small number of both word-counts and hits;

[12] "Hits" refer to total the frequency of a particular searched token which is retrieved from the corpora (BNCweb).

it accounts for around one-third of other groups. However, the frequencies of hits of these three age groups are still high enough to operate the random sampling.

In the next section, the analyses are demonstrated and discussed, starting with (4.1) the distribution of function of “you know” with gender, (4.2) the distribution of function of “you know” with age factor, (4.3) the interaction between males across several age groups, and finally, (4.4) the interaction between females with age factors.

4.1 The distribution of function of “you know” with gender factor

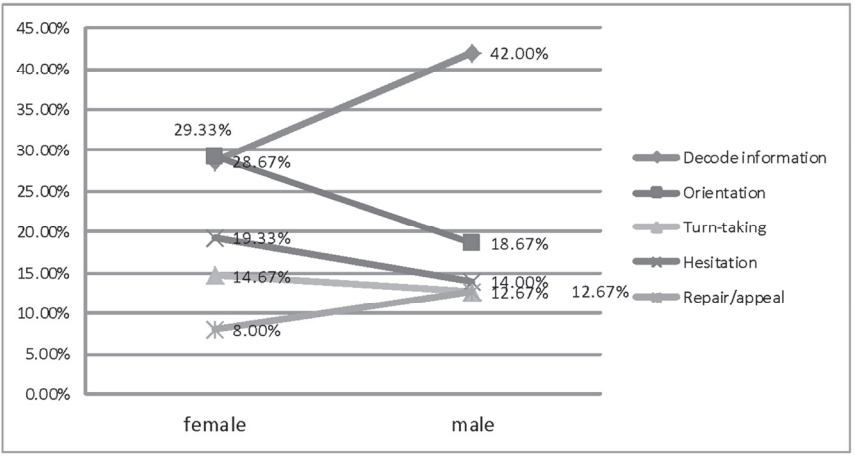

Figure 3 The distribution of PE’s functions in use across two genders[13]

[13] In fact, the difference between the genders with regard to each PE’s function has already tested pairwise by using the chi-square test. However, it was found that all pairwise interactions between the two genders and three age groups were not statistically significantly different (please see table 2 in the appendix).

This reveals that males gain the highest percentage in decoding information (DEC), followed by orientation in discourse (ORI) accounting for 42% and 18.67%. The distinction of DEC is about twice as much as in ORI. However, females show a contrasting result to males in that the first two functions' percentages (in DEC & ORI) are very close (29.33% and 28.67%, respectively).

It is obvious that females employ all functions of "you know" (out of five) in a more well-distributed way than males. Nevertheless, in males, the rates of use of each function look bundle-like in the last three functions, (hesitation (HES) 14%, turn-taking (TURN), 12.67%, and repair/appeal (REP), 12.67%) and this dramatically increases the rate of use of the decoding information (DEC).

Interestingly, (figure 3) females gain lower percentages in the repair/appeal (REP) than males (8% vs. 12.6%). This can imply that males tend to repair the misunderstanding or negotiate arguments/issues than females. However, the percentage distinction between these two genders is sparingly low. In addition, men perform contrasting behaviors by doing both (DEC) and (REP) to the hearers as well. However, the REP found in males is around one-third of the DEC. To some extent, females are assumed to be linguistically vulnerable, revealing a slightly lower score in REP (repair and appeal) than males. In addition, the rates of use in hesitation/ hedging (HES) in both genders are not that much different, accounting for 19.33% by females and 14% by males. Thus, it is not really valid to presume that only females show much reluctance during discourses. Both genders show hesitation but females are prone to be slightly higher.

4.2 The distribution of function of “you know” with age factor[14]

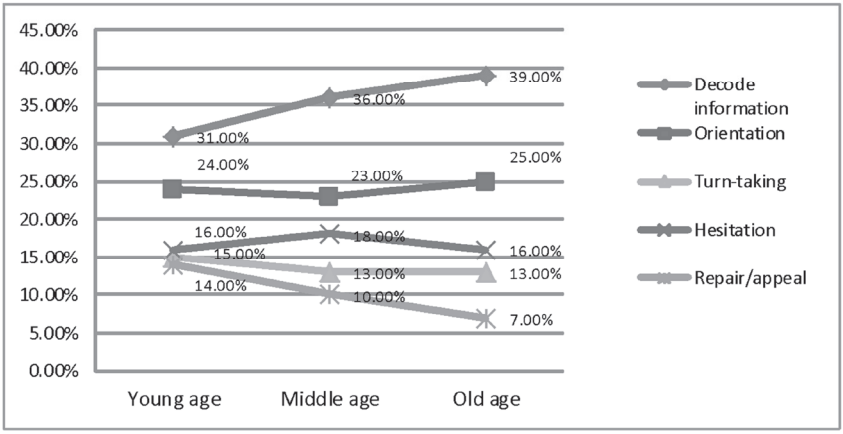

Figure 4 The distribution of PE’s functions in use across three age groups

Referring to the previous finding decoding information (DEC) plays a dominant role in both genders, accordingly. This function accounts for around one-third of the entire data (33% in young speakers, 36% in adult speakers, and 39% in the elderly speakers). In terms of trend of use across three generations, DEC gradually rises from the youngsters to the elderly. This might generally suggest that the more people age, the more they use DEC. It is surprising that the elderly gain the highest score amongst other groups. By contrast, the old aged speakers use REP a great deal less than other generations, by around 7%.

[14] The raw frequency and percentage tables are shown in the appendix.

In the adult group, the score gaps/ ranges in each function are not wide, but quite evenly distributed. However, young speakers reveal the mixed trend as bundle-like for some functions. The youngsters show similar proportions of use in TURN, HES and REP while showing large gaps/ranges from the previously three stated groups, DEC and ORI. Finally, rates of use of orientation to discourse (ORI) function are very similar across three generations (24%, 23% and 25%); appearing to be stable in all groups.

As remarked earlier, the elderly group might possess unique linguistic demeanors compared to the other groups[15]. It makes sense that the function of the DEC and the REP revealed a negative correlation. The function of the DEC is to reassure that people will follow and understand something according to the speaker's argument, or draw people's attention. By contrast, the function of REP signals the conciliation and negotiating with regard to the (misleading) argument.

[15] The possibility of maintaining or losing function in the PE "you know" can occur in this way to some extent. When some functions of PE "you know" have finally demised, other PEs will compete and replace them by gaining new functions.

4.3 The interaction between males across several age groups

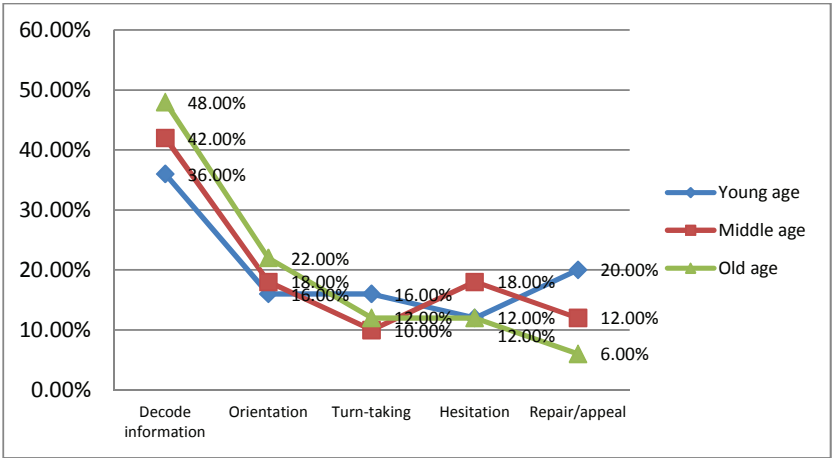

Figure 5 The distribution of PE’s functions in use across three age groups in male speakers only (n=150)

Referring to figure 5, it is found that the DEC function is dominant (around 40%) with a percentage far greater than the other score averages. The DEC function is a primary function of “you know”.

In men, each function of the PE “you know” shows a very similar rate of use across three generations, regardless of DEC and REP function. Even though the DEC use rate is outstanding amongst other functions, the proportion across the three age groups is still very close. However, the REP function is different. Adolescents favour it a great deal (20%) while the two remaining older groups use it infrequently, 12% and 6%, respectively.

To reiterate, DEC gains the highest score in male speakers across three generations. However, it is not true in females across three generations, as discussed below (in figure 6). Also, the trend of PE use over time, DEC and ORI

found in males tend to gradually increase when the people get older (36%, 42% and 48%) in general. By contrast, the function of REP tends to decrease over time as people get older (20% in youngsters, 12% in adults and 6% in the elderly). However, the apparent time change estimation cannot be claimed here. Lastly, the rates of use of turn-yielding (TURN) and the hesitation (HES) fluctuate, although the proportions (%) of each category is quite similar.

4.4 The interaction between females with age factors

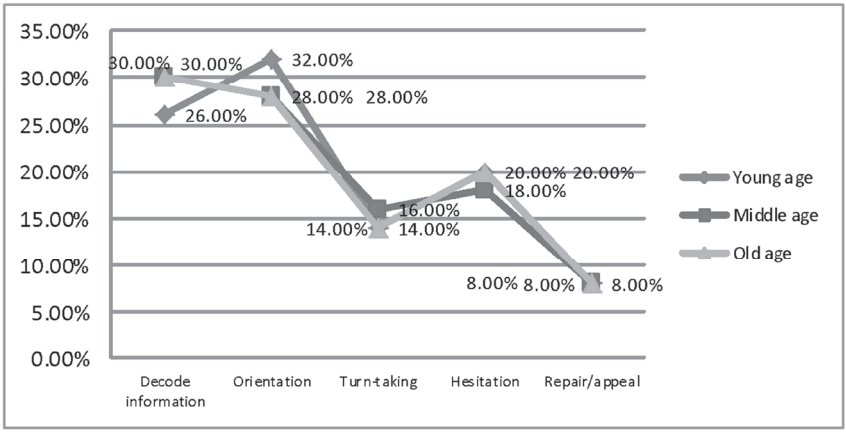

Figure 6 The distribution of the PE's functions in use across three age groups in female speakers only (n= 150)

According to figure 6, the trend of PE use in females is much in line with males, across the three generations. The proportion (%) of each PE's function in use by females is very similar. This referred result in female speakers seems to be slightly better off than the male's groups (figure 5).

However, there are some differences. In males, DEC is the only predominantly favoured function while in females, the DEC and the ORI show the highest rate of use (28% in average). The DEC's percentage across three age groups in females are much lower than males' ones. In terms of the function of TURN, the percentages of use across the three generations are intermediate and similar (14%, 16%, 14%) in females. This incidence is alike in RES as well. They share identical use rates (8%, 8%, and 8%) across all three generations, even though the percentage in the RES function is less than the TURN function (+14%).

Interestingly, in the hesitation (HES) function, females across the three age groups use it a great deal, with similar proportions (20%, 18%, 20%). Compared to the male groups, the HES function tends to fluctuate in scores across the three generations, with lower scores, around 12% in the adolescents and the elderly (see figure 5), regardless of the adult men group. Thus adult men use HES at a similar rate to that found in females (18%).

These results imply that, generally, females are more likely than males to show signs of hesitation or reluctance during conversation, except for the adult men who gain high scores in HES. This can be explained by the work life context in adults, especially in the case of adult men. Much use of hesitation might be associated with a working lifetime. People need to be more compromising and flexible during this life stage. Showing less face-threatening acts (Levinson, 2013) and seeking approval from their peers might be more beneficial for their careers and marriage life.

To sum up, gender difference has an effect on how people use the function of “you know”. It is found that male speakers are salient in terms of the DEC and REP function of PE, while female speakers show similar rates of use in DEC and ORI, but a higher rate than males in HES.

5. Conclusion and discussion

This study investigated the function of the PE "you know" as retrieved from the spoken sub-corpus of the BNCweb. I adopted Erman's pragmatic expression approach to analyse this data. Holistically, there are gender different tendencies in the use of function of "you know".

To elaborate, some of the most salient differences were that women tend to use the PE "you know" between complete propositions to connect consecutive arguments, which can be referred to as orientation in the discourse and markers of hesitation, whereas male speakers prefer to use "you know" either as attention-drawing/seeking devices (or decoding information) or to signal repair work (or repair/appeal function).

Erman (1992) claims that males are prone to make the hearer accept their arguments and to position themselves to make their speech impressive, while women may feel obliged to keep the conversation going smoothly (orientation in discourse and hesitation) and politely. However, females do show that they also favour the decoding information function as well, but in a lower proportion compared to males.

In addition, it is still true that the hesitation (HES) function is marked for females. Both Erman's (1992) and this study exhibit the same trends[16]. Nevertheless, the proportion in Erman's work is much higher than this current study (see table 3.). However, the behavior of women pertaining to the polyfunctions of "you know" are more varied and widespread than men, as based on this limited sub-corpus[17].

[16] The adult male group still favours the HES function as well in this study. This is the exception.

[17] However, this latter claim does not imply that women are more over-generalised than men with regard to the function of PE apart from "you know". That is to say, women might lack other PE use, and might mostly enjoy "you know" on a basis of using only one PE "you know" for many functions, out of the five or more.

Table 3 Comparing the percentages of “you know” in the function of HES between this study and Erman’s study (1992: 228)

The function of “you know” as the hesitation filler or hedging device (HES) (the scores are averaged)		
Research	male	female
This study (2017) (n = 300) Sub-corpus size from BNC = 8,240,507 words	14.00%	19.33%
Erman (1992) (n =346) corpus size = 65,000 words	20.00%	28.00%

Labov (2001) addressed that females might be more vulnerable and lack linguistic security regarding their language use. Thus, the claims in Lakoff (1972) about gender dominance, and Labov (2001) pertaining to female speech behavior, appear to be valid. Nevertheless, it should not be overgeneralized since the function of HES is not the dominant trend, with the domination being in decoding information and orientation of discourse.

The other main finding is age as a social factor. Even though gender differences show a number of contrasting results, it is the age factor that mainly shows homogenous findings. Most functions of “you know” are relatively similar and show similar trends across entire age groups in general. However, some exceptions show in the old aged speakers, and the youngsters present a slightly reversed trend of the function. The adolescent usage is more inclusive/ close together (more variation), while the elderly trend is more dispersed/ widespread (less variation), please see figure 4.

Finally, this study can support the claim that female speech should not be stereotyped as worse than or inferior to males. Also, the marker of reluctance and hesitation is not only employed by the females, but it is also favored by males in specific groups, especially in adults. This PE “you know” is attributed to polyfunctionality, and tends to reveal that female speakers are likely to employ PE in a more balanced and well-distributed way than males. Males seem to favour certain PE functions in the discourses.

There are some limitations in this current study. In fact, the BNC corpus comprises several social factors, not only these two focused predictors, such as interactional type, range of context situation (formal/ informal), (broad) social class, genre, dialect of the speaker, domain of the context, educational level and so forth. These confounding factors are very interesting to examine and can definitely affect the trend of PE use. Thus, in the further study, these other factors can be conflated and categorised into many sub-corpora for use in the appropriate statistical calculation in multivariate analyses.

The result of this current study cannot be compared with Erman’s study (1992) directly. In Erman’s work, her findings included only the interview and conversational style of 12 speakers with a high level of education. Thus, these speakers are probably from the middle social class. Although the BNC can provide numerous social factors, the BNC still does not have some of the crucial demographic factors which featured more specifically in Erman’s work. In addition, the problem is about the genre of the spoken content in the BNC. In this current data, it is quite challenging to manage and control more than 15 genres. I would have classified and created the sub-corpora to conflate them in order to be able to compare them with Erman’s work. However, another problem is that the proportion of each sub-genre does not equate. Some might gain a much higher number of tokens but some might be minimal.

Another limitation is that in the original paper (Erman, 1992) compared the function of the PE “you know” by two types of dyadic speakers, namely same-sex speakers and mixed-sex speakers. However, this current work cannot control this factor. This is because some of the discourse excerpts from spoken BNC sub-corpora can consist of many people conversing with one another, and not only two. Nevertheless, this paper can ensure that the spoken sub-corpus I used has at least two speakers, and not a long narrative addressed by a single person.

To reiterate the issue of the contraction form, “y’ know” (see 3.), its pragmatic functions might be associated with the lower social class speakers, the high level of linguistic insecurity and the adolescent speakers to some extent, in the casual style. However, it will never been verified until further analysis. Thus, it is quite interesting to examine the distribution of functions of the PE “you know” in both the full form (“you know”) and the contracted form (“y’ know”) as discussed earlier. In an example shown below, David Beckham, a world famous British football player also uses “you know” during his interview. The lower social class and the interview context factors might not always be the case for supporting this claim, and other confounding factors and other types of data require more investigation. In addition, according to the urban dictionary[18], the PE “you know” tends to have a notorious connotation as a non-informative, non-academic and unfavourable linguistic form.

[18] http://www.urbandictionary.com/define.php?term=you%20know[Accessed 1 Sep 2017].

(11)[19]

Beckham: I don't usually get nervous in games and, you know, taking free kicks and penalties but that's the first time I've been that nervous in, in a game, where I have to, I'm in a situation where I'm, you know...

These PEs "you know" signify pauses of hesitation or reluctance used by David Beckham during his interview with BBC.

The last point of limitation is that all chi-square pairwise results (between two genders and five types of function of "you know") are all negative (see table 5 in appendix). Thus, the claim I proposed here is not so strong. However, it is still interesting and can reveal the variation distributions across several social factors and the function of "you know's" variants. I think that if the size of this corpus is larger, as well as with an input of more and other social variables into the models of analyses, some of the pairwise results might be significantly different. In this study, I have not compared the interaction amongst the gender, age of speakers and the function of the PE. In that we cannot claim for the change in apparent time, some of the interactions might show statistically significant differences when using a chi-square test (see table 4 in appendix). However, as addressed earlier, this BNC spoken sub-corpora cannot be traced back, and cannot indicate the specific point of time of the speakers. It is not able to refer to the actual age of the speakers at the present time.

[19] Rich (2017) p.c. https://en.wiktionary.org/wiki/you_know[Accessed 1 Sep 2017]. https://www.youtube.com/watch?v=Fe8_x4HwV7w[Accessed 1 Sep 2017].

References

Baker, P. (2010). **Sociolinguistics and corpus linguistics**. Edinburgh: Edinburgh University Press.

Biber, D. (1999). **Longman grammar of spoken and written English**. Harlow, England; [New York]: Longman.

Clancy, B., & O'Keeffe, A. (2015). Pragmatics. In D. Biber & R. Reppen (Eds.), **The Cambridge Handbook of English Corpus Linguistics** (pp. 216-234). Cambridge: Cambridge University Press.

Cheshire, J. (2004). Sex and gender in variationist research. **The handbook of language variation and change**. J. K. Chambers, P. Trudgill and N. Schilling-Estes. Malden, MA, Blackwell Publishers: 423-443.

Coates, J. (1986). **Women, men, and language : a sociolinguistic account of sex differences in language**. London; New York: Longman.

Coates, J., & Cameron, D. (2014). **Women in their speech communities : new perspectives on**

Languageand sex. from http://public.eblib.com/choice/publicfullrecord.aspx?p=1791052

Dubois, B. L., & Crouch, I. (1975). The question of tag questions in women's speech: they don't really use more of them, do they? **Language in Society**, *4*(3), 289-294.

Erman, B. (1987). Pragmatic Expressions in English. A Study of You know, You see and I mean in

Face-to-Face Conversation. **Stockholm studies in English**, *69*, 1-238.

Erman, B. (1989). From lexical to pragmatic meaning. In **Proceeding from the Forth Nordic Conference for English Studies**. Copenagen: Copenhagen University.

Erman, B. (1992). Female and male usage of pragmatic expressions in same-sex and mixed-sex interaction. **Language Variation and Change**, *4*(2), 217-234. doi: 10.1017/S0954394500000764

Erman, B. (2001). Pragmatic markers revisited with a focus on you know in adult and adolescent talk. **Journal of Pragmatics Journal of Pragmatics**, *33*(9), 1337-1359.

Holmes, J. (1986). Functions of you know in women's and men's speech. **Language in Society**, *15*(1), 1-21.

Labov, W. (2001). **Principles of linguistic change**. *Vol. 2, Vol. 2*. Oxford: Blackwell.

Lakoff, R. (1973). Language and woman's place. **Language in Society**, *2*(1), 45-79.

Laserna, C. M., Seih, Y.-T., & Pennebaker, J. W. (2014). Um... who like says you know: Filler word use as a function of age, gender, and personality. **Journal of Language and Social Psychology**, *33*(3), 328-338.

Levinson, S. C. (2013). **Pragmatics**. Cambridge: Cambridge University Press.

Litosseliti, L. I. A. (2016). **Gender and language: theory and practice**. [Place of publication not
identified]: ROUTLEDGE.

Milroy, L. and M. Gordon (2008). **Sociolinguistics Method and Interpretation**. New York, NY, John Wiley & Sons.

O'Donnell, W. R., & Todd, L. (2013). **Variety in Contemporary English**. Florence: Taylor and Francis.

Ragan, S. L. (1983). Alignment and conversational coherence. **Conversational coherence: Form, structure, and strategy**, 157-171.

Romaine, S., & Lange, D. (1991). The use of like as a marker of reported speech and thought: A case of grammaticalization in progress. **American Speech**, *66*(3), 227-279.

Rosch, E. H. (1973). Natural categories. **Cognitive psychology**, *4*(3), 328-350.

Schiffrin, D. (2012). **Discourse markers**. Cambridge [u.a.]: Cambridge Univ. Press.

Sunderland, J. (2006). **Language and gender : an advanced resource book**. London; New York: Routledge.

Tannen, D., Simon, & Schuster, A. (2011). **You just don't understand**. from http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=375549

Thepkanjana, K. (2017). **Grammaticalization**. Chulalongkorn University, Bangkok, Thailand, Chulalongkorn University Press.

Xue, W., & Lei, L. (2016). Functions of you know and I mean in academic seminar discussions by ELF speakers. **Chinese Journal of Applied Linguistics**, *39*(3), 288-303.

Appendix

Table 4. Cross-tabulation of PE's functions across gender and age of speaker

Gender	Age of speaker	Decoding information (DEC)	Orienta-tion (ORI)	Turn-taking (TURN)	Hesitation (HES)	Repair/ appeal (REP)	Grand total
Female speakers	**Adolescent**						
	n	13	16	7	10	4	50
	%	26.00%	32.00%	14.00%	20.00%	8.00%	100.00%
	Adult						
	n	15	14	8	9	4	50
	%	30.00%	28.00%	16.00%	18.00%	8.00%	100.00%
	Elderly						
	n	15	14	7	10	4	50
	%	30.00%	28.00%	14.00%	20.00%	8.00%	100.00%
	Total n	43	44	22	29	12	150
	Total %	28.67%	29.33%	14.67%	19.33%	8.00%	100.00%
Male speakers	**Adolescent**						
	n	18	8	8	6	10	50
	%	36.00%	16.00%	16.00%	12.00%	20.00%	100.00%
	Adult						
	n	21	9	5	9	6	50
	%	42.00%	18.00%	10.00%	18.00%	12.00%	100.00%
	Elderly						
	n	24	11	6	6	3	50
	%	48.00%	22.00%	12.00%	12.00%	6.00%	100.00%
	Total n	63	28	19	21	19	150
	Total %	42.00%	18.67%	12.67%	14.00%	12.67%	100.00%
	Grand total n	106	72	41	50	31	300
	Grand total %	35.33%	24.00%	13.67%	16.67%	10.33%	100.00%

Table 5 Chi-square test between male and female speakers (at the level of $p < 0.05$)

No.	Female vs. Male	result	The chi-square score and its interpretations ($p < .05$)
1	Decoding information (DEC)	not sig	The χ^2 is 3.2096. The *p*-value is .073208.
2	Orientation (ORI)	not sig	The χ^2 is 3.1764. The *p*-value is .074709.
3	Turn-taking (TURN)	not sig	The χ^2 statistic is 0.2055. The p-value is .650337.
4	Hesitation (HES)	not sig	The χ^2 statistic is 1.1817. The p-value is .277007.
5	Repair/appeal (REP)	not sig	The χ^2 is 1.5032. The p-value is .220183.