Functions of discourse marker you know in a Saudi learner corpus

Sami Algouzi (Najran University, Saudi Arabia)

This paper examines the extent to which *you know* is used as a discourse marker by Saudi native speakers of Arabic. The Saudi learner corpus (henceforth LINDSEI –AR) employed in this study is an official component of the Louvain International Database of Spoken English Interlanguage (LINDSEI) and as such, the design is determined by the general LINDSEI model. This corpus is the first part of LINDSEI representing English used by Saudi learners. It consists of 50 informally recorded interviews in English between Saudi students studying English as foreign language. In total the corpus comprises almost 68,100 words of interviewee speech. The participants in this study are all third and fourth year university students (males and females, aged 20–25 years) majoring in English linguistics/ translation. The level of study is a criterion for this corpus (De Cock, 2004; Gilquin, 2008; Gilquin et al, 2010).

Not only is the use of *you know* in this learner corpus examined from a quantitative and a qualitative perspective, it is also compared with native speaker corpus, the Louvain Corpus of Native English Conversation (LOCNEC) (De Cock 2004). In fact, following the standard format of LINDSEI gives this learner corpus the advantage of using native speaker of English data, LOCNEC, compiled by De Cock (2004), which also used the same structure as LINDSEI. Native speakers in LOCNEC were British students majoring in English language and/or linguistics at the University of Lancaster.

Qualitatively, the functions of *you know* are analysed with a moderately corpus-based approach, implementing functions of *you know* from previous studies (e.g. Brinton, 1996; Fraser, 1990; Fraser, 1999; Fraser, 2006; Schiffrin, 1987). In order to distinguish *you know*'s discourse marker functions from its other uses, key characteristics such as optionality, orality, multi-functionality identified in previous researches are considered (e.g. Brinton, 1990; Brinton, 1996; Schiffrin, 1987; Schourup 1999). Regarding the quantitative part of the study and the frequency of discourse markers, the tokens are counted manually and then double-checked via a computer search. The results of frequency of *you know* is normalized, that is, standardized according to a consistent text length of 10,000 words. With this information, the number of *you know* tokens is calculated and multiplied by 10,000 and divided by the total word count.

The investigation shows that the six discourse marker functions of *you know*, following Brinton's (1996) classification three textual and three interactional, are used by Saudi speakers and native speakers as can be seen in Table 1.1 and Figure 1.1. Most of the functions are used by Saudi speakers more than native speakers except *you know* for lexical or content search. The most frequent function in both corpora is *you know* for speaker/hearer shared knowledge. The difference between the two corpora for this function achieved statistical significance $p < 0.0001$.

Table 1.1 Distribution of you know in the LINDSEI-AR and LOCNEC (frequencies per 10,000 words in parentheses)

	Search	Restart & Repair	Explain	General Knowledge	Shared Knowledge	Narrative
SLC (68,100 words)	52 (8)	49 (7)	11 (2)	33 (5)	161 (24)	52 (8)
LOCNEC (68,000 words)	63 (9)	32 (5)	20 (3)	20 (3)	91 (13)	47 (7)

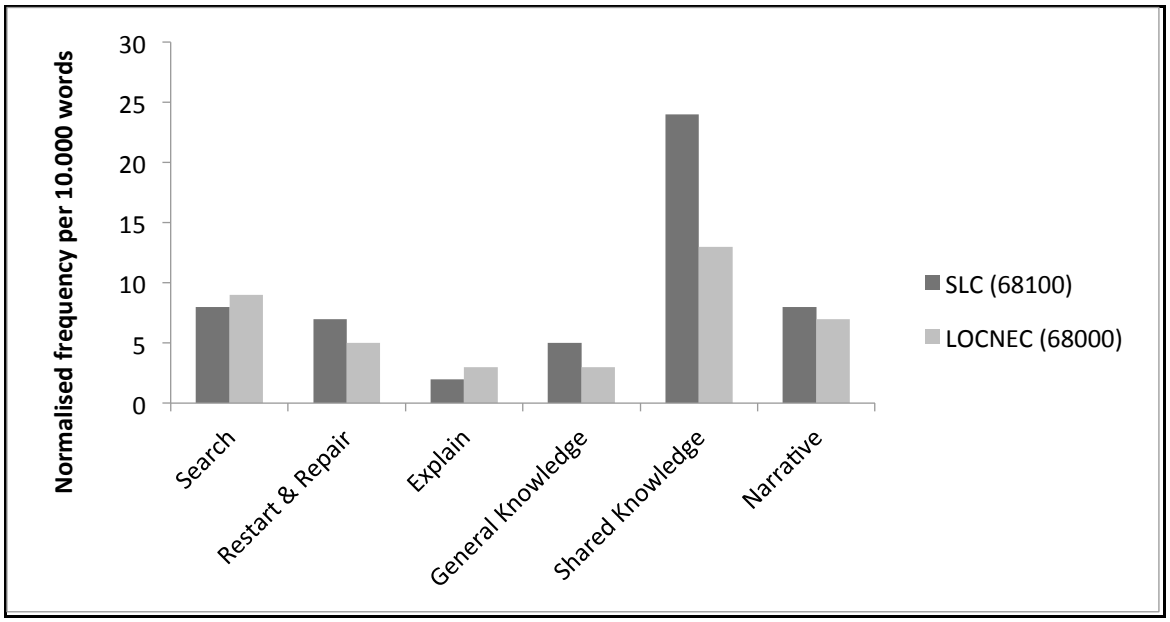

Figure 1.1 Distribution of DM functions of you know in LINDSEI-AR according to gender

Regarding gender in the learner corpus, similar to results in Macaulay (2002), it is found as in Table 1.2 and Figure 1.2 that the female speakers used a higher rate 'raw frequency' of *you know* with 200 tokens or 65 per 10,000 words compared to the male speakers who used it 228 times, or 61 per 10,000 words. The frequency of the non- discourse marker and the unclear instances was higher for the male speakers than the females. However, neither frequency difference was statistically significant.

Table 1.2 Number of tokens of you know in the LINDSEI-AR according to gender (frequencies per 10,000 words in parentheses)

	Total Frequency	Non-DM use	DM use	Unclear
Male (37,108 words)	228 (61)	23 (6)	183 (49)	21 (6)
Female (30,900 words)	200 (65)	14 (5)	175 (57)	11 (4)

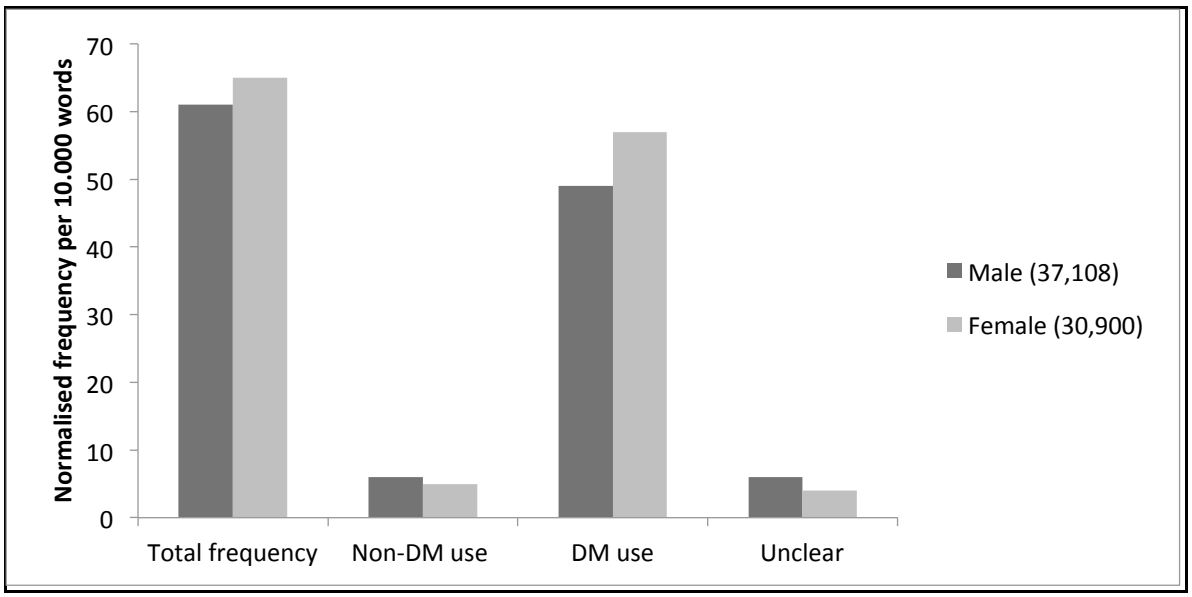

Figure 1.2 Distribution of the linguistic you know in the LINDSEI-AR according to gender

The analysis of the discourse marker use of *you know* shows that the gender has some effect. Even though the number of tokens is quite small, it is worth discussing. Table 1.3 and Figure 1.3 illustrate in detail the distributional pattern between male and female speakers. For both gender groups, *you know* is used most frequently as a speaker/hearer shared knowledge marker. The female speakers used 95 tokens, or 31 per 10,000 words in their speech, whereas the males used 66 tokens, or 18 per 10,000 words. This difference is highly significant ($p < 0.0001$). The female speakers used *you know* for speaker/hearer shared knowledge more than the total of all other functions (80 versus 95). My results are similar to those of Macaulay (2002) who found that women are more likely to use *you know* than men. In another study by Stubbe and Holmes (1995), *you know* was found to be more frequent in the speech of the young male than in the speech of the young female, both from a working class background. However, because LINDSEI does not provide information about the social background of the participants, the results of this study are not compared to those of Stubbe and Holmes (1995).

Table 1.3 Functions of you know in the LINDSEI-AR according to gender (frequencies per 10,000 words in parentheses)

	Search	Restart & Repair	Explain	General Knowledge	Shared Knowledge	Narrative
Male (37,108 words)	41 (11)	15 (4)	6 (2)	17 (5)	66 (18)	38 (10)
Female (30,900 words)	11 (6)	34 (11)	5 (2)	16 (5)	95 (31)	14 (5)

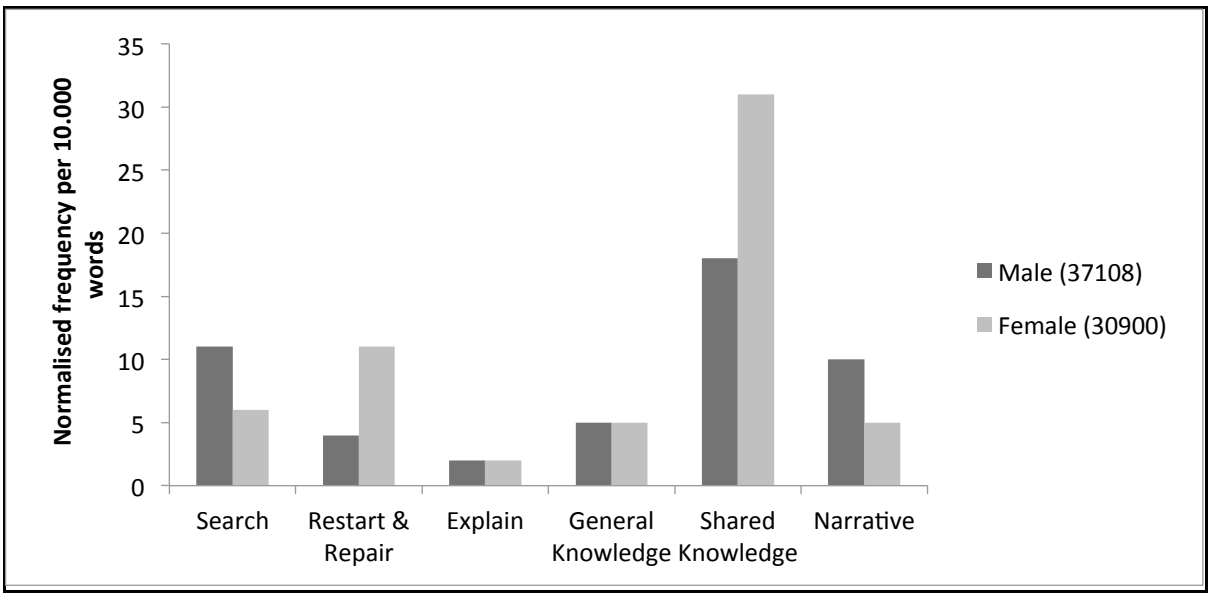

Figure 1.3 Distribution of DM functions of you know in LINDSEI-AR according to gender

In fact, there are considerable individual differences in the LINDSEI –AR in terms of the rate of occurrence of the discourse marker *you know*. There are some speakers who used *you know* at a markedly higher rate (one speaker 30 tokens and another speaker 34 tokens) in an average of 1,400 words per recording, which is a clear contrast to the other speakers' use which ranged from two to maximum 20 tokens. On the other hand, the analysis of the native speakers data showed that the highest frequency of *you know* use was 35 tokens in one interview, interviewee E11. However, the average total number of words in E11 interview (B turns) was 2,800 words. So, in this situation, I argue that high use of *you know* which occurred in the LINDSEI –AR is not necessarily due to the lack of fluency of speakers but rather occurred as 'speech habit' (Östman, 1981, p. 27) or what Macaulay calls 'part of the speaker's discourse style' (2002, p. 765). Macaulay (2002) points out that the discourse marker *you know* does not appear to mark 'assumptions of shared knowledge, but rather to form part of the speaker's discourse style and the rhythmic organization of utterances' (2002, p. 765). Even if Macaulay's argument is valid, it is still necessary to classify tokens of *you know* which conform to shared knowledge as a discourse marker use and this is what has been done in this study.

References

Brinton, L. J. (1990) The development of discourse markers in English. In Fisiak, J. (ed) *Historical Linguistics and Philology*. Berlin, Mouton de Gruyter, pp. 45–71.

Brinton, L. J. (1996) *Pragmatic Markers in English*. *Grammaticalization and Discourse Functions*. Berlin, Mouton de Gruyter.

De Cock, S. (2004) Preferred sequences of words in NS and NNS speech. *Belgian Journal of English Language and Literatures (BELL)*, New Series, 2(2), 225–246.

Fraser, B. (1990) An approach to discourse markers. *Journal of Pragmatics*, 14, 383–395.

Fraser, B. (1999) What are discourse markers in English? *Journal of Pragmatics*, 31, 931–952.
Fraser, B. (2006) Towards a Theory of Discourse Markers. In: Kerstin Fischer (ed) *Approaches to Discourse Particles*. Amsterdam, Elsevier. pp. 189–204.
Gilquin, G. (2008) Hesitation markers among EFL learners: Pragmatic deficiency or difference? *Pragmatics and Corpus Linguistics: A mutualistic entente*, 2, 119.
Gilquin, G., De Cock, S. &Granger, S. (eds) (2010) *The Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM*. Louvain-la-Neuve, Presses Universitaires de Louvain.
Macaulay, R. (2002) You know, it depends. *Journal of Pragmatics*, 34(6), 749–767.
Östman, J. O. (1981) *'You* Know': A discourse-functional study. Amsterdam, John Benjamins
Schiffrin, D. (1987) *Discourse Markers*. Cambridge, Cambridge University Press.
Schourup, L. (1999) Discourse markers. *Lingua*, 107, 227–265.