UNIVERSITETET I AGDER

"You Know, It's Like Everywhere. Nothing's Doing what it's Supposed to Anymore"

A Corpus-based Study on Sociolinguistic Distribution of the Discourse Markers *Like* and *You Know*

SIRI JACOBSEN

SUPERVISOR
Ingrid Kristine Hasund

University of Agder, 2019
Faculty of Humanities and Education
Department of Foreign Languages and Translation

Abstract

Discourse markers are sometimes socially stigmatized by people, and researchers conducting corpus-based studies report contradictory results about sociolinguistic distribution of such items. This study focuses on the two discourse markers *like* and *you know* and comparisons of distribution of these discourse markers among men and women, and among people of higher and lower education have been conducted. The study uses American speakers aged 20-59 to answer the two following questions: (1) can social differences regarding frequency of employing DMs be detected? and (2) do different social classes use discourse markers for different purposes? Mixed methods were employed to answer the questions. The qualitative part of the study aimed to identify functions which the two discourse markers serve in order to answer whether there are differences regarding how social classes use *like* and *you know*. It was found that *like* can be used for signaling an upcoming approximation, introducing one of several examples, drawing focus to particular elements in the discourse that are of importance, and for signaling an upcoming quote. *You know* can be used for introducing previous knowledge as background information to an upcoming main point, reassuring the addressee of the validity or the speaker's conviction of what s/he says. *You know* can also be used when a message has been unclearly delivered where the token either signals that the speaker transitions into an elaboration, or where the speaker gives the addressee an opportunity to ask the speaker for an elaboration. Lastly, *you know* could be used when the speaker wishes to go back and repair part of the discourse. It was found that neither men compared to women, nor people of higher education compared to people of lower education use the two discourse markers for different purposes. *Like* was not used with significantly different frequencies neither when comparing gender nor educational level. Men and women did not differ in how frequently they employed *you know*, but it was found that people of lower education use *you know* significantly more than people of higher education.

Acknowledgements

They say it takes a village to raise a child—the same could be said about raising a linguist. First and foremost, I would like to thank my supervisor Associate Professor at the University of Agder, Ingrid Kristine Hasund. Your guidance, expertise, and the way you have challenged me throughout this project have resulted in me acquiring a level of knowledge about this field that is beyond anything I could have imagined at the onset of this wild journey.

I also want to thank Elin Berg, Jovana *Dašić*, Marikken Auensen, and Veronika Hamann with whom I have been in a group of corpus linguists through this project. It has been a tremendous help to discuss my work with you; you have helped to conquer every methodological issue I have met, and it feels like we can move mountains when we come together. I also want to thank you for the support and courage which has motivated me to carry out this study.

In addition, without mentioning any names, I would like to express my gratitude to my friends, who have cooked me meals when I have been too busy to be in the kitchen, allowed me to take some much needed breaks between all the hard word, and who have cheered me through the process of writing this thesis.

Lastly, I would like to thank my mama, who never discouraged me from asking questions, which has kept my curiosity and urge to search for answers alive, and my dad, who passed away only four weeks after I began my English studies. You encouraged me to make a career out of my language skills, and I hope you are looking down on me with pride.

Siri Jacobsen
Kristiansand, May, 2019

Contents

1 Introduction 1
1.1 Aims and Purpose 3
2 Theoretical Background 4
2.1 Definitions of DMs 4
2. 2 Previous Accounts of Functions 6
2.2.1 *Like* 6
2.2.2 *You know* 10
2.3 Sociolinguistics 11
2.3.1 Like 12
2.3.1.1 Studies Focusing on the DM *like* at Large 12
2.3.1.2 *Studies Focusing Exclusively on Quotation Markers* 14
2.3.2 *You know* 15
2.3.3 Summary 17
3 Methodology 19
3.1 The Corpus 19
3.2 Categorizing the DMs 20
3.2.1 Selecting the Term *Discourse Marker* 23
3.3 Defining Social Groups 24
3.4 Statistics 24
4 Results 26
4.1 Functions 26
4.1.1 Superordinate Functions 26
4.1.1.1 Hedges and Boosters 26
4.1.1.2 Turn Management 27
4.1.1.3 Verbal Fillers 28
4.1.2 Subordinate Functions 29
4.1.2.1 *Like* 29
4.1.2.1.1 Basic Meaning 29
4.1.2.1.2 Approximator 29
4.1.2.1.3 Exemplifier 30
4.1.2.1.4 Focuser 31
4.1.2.1.5 Quotation 33
4.1.2.2 *You Know* 34
4.1.2.2.1 Basic Meaning 34
4.1.2.2.2 Previous Knowledge 35

4.1.2.2.3 Reassuring the Addressee 37
4.1.2.2.4 Elaboration 37
4.1.2.2.5 Self-repair 39
4.2 Statistics 40
4.3 Summary 48
5 Discussion 50
5.1 The Qualitative Part of the Study 50
5.2 The Quantitative Part of the Study 52
5.2.1 *Like* 54
5.2.2 *You Know* 56
5.3 Limitations 58
6 Conclusions 58
6.1 Suggestions for Further Research 60
7 References 61
Appendix I 66
Appendix II 70
Appendix III 73
Approximator (*like*) 73
Exemplifier (*like*) 74
Focuser (*like*) 76
Quotative (*like*) 79
Rag bag (*like*) 81
Non-DM (*like*) 83
Previous knowledge (*you know*) 83
Reassuring the addressee (*you know*) 86
Elaboration (*you know*) 88
Self-repair 90
Rag bag (*you know*) 92
Non-DM (*you know*) 94

List of Tables and Figures

Table 4.1	Distribution of *like* according to gender	40
Table 4.2	Distribution of *you know* according to gender	41
Table 4.3	Distribution of *like* according to educational level	43
Table 4.4	Distribution of*you know* according to educational level	44
Table 4.5	Distribution of *like*	*45*
Table 4.6	Distribution of *you know*	*47*

Figure 4.1	Relation between *you know* and knowledge	35
Figure 4.2	Mean distribution of *like* according to gender	41
Figure 4.3	Mean distribution of *you know* according to gender	42
Figure 4.4	Mean distribution of *like* according to educational level	43
Figure 4.5	Mean distribution of *you know* according to educational level	44
Figure 4.6	Mean Distribution of *like*	46
Figure 4.7	Mean Distribution of *you know*	*47*

1 Introduction

The title of this thesis, "You know, it's Like Everywhere. Nothing's Doing what it's Supposed to Anymore", is a quote from Marilyn, one of the participants in this corpus-based study. What she refers to is the many instances of atypical weather found around the globe. The quote could just as well have been used in reference to the concept of discourse markers (henceforth written DMs); they are found *like everywhere* in informal discourse, and due to heavy semantic bleaching, they serve very different functions from the words from which they derive. Moreover, DMs such as *like* and *you know* are highly versatile given they serve a range of pragmatic functions. Although they rarely affect neither the truthfulness, nor add to the propositional content of an utterance (Höcker, 1991 in Jucker & Ziv, 1998, p. 3), they often create local or global coherence in discourse (Aijmer, 2013, p. 5), and they can be instructional cues to how an utterance is to be understood (Fraser, d.u., p. 186).

DMs are sometimes socially stigmatized and negatively evaluated (Watts, 1989; Stubbe & Holmes, 1995; Buchstaller, 2013). Brinton (1996) writes there are controversial suggestions claiming DMs are characteristic of women's speech (p. 35) and attitudinal surveys regarding *like* reveal the majority of people surveyed believe women are the dominant users of this DM (Dailey-O'Cain, 2000; Blyth, Reckenwald & Wang, 1990; Lange 1986 in Romaine & Lange, 1991). Buchstaller (2014) concluded that both her American and British informants in her survey associate quotative *like* with people of lower, rather than higher education. She also found that the American informants do not believe there is a class bias towards quotative *like*. However, the British informants, she found, associate it with working-class people rather than those belonging to the middle-class (p. 228, 230). Dailey-O'Cain's (2000) survey revealed strong negative opinions towards *like*, and common reasons for the dislike were that the people surveyed believed it makes the speaker sound lazy or uneducated.

Less work has been conducted to test stigma and attitudes towards *you know*, but the following has been found in previous literature: Watts (1989) recorded a conversation between him and some of his family members where he steered the conversation towards their attitudes towards DMs. He found that one family member believed using *you know* was terrible, although another family member believed there were times where employing this token was unproblematic. Beeching (2016) conducted a survey asking people to judge an example of *you know* in utterance-final position. They had to answer whether they believed the presence of *you know* made the speaker appear polite, direct, educated, and friendly. The

results revealed neutral attitudes, with the exception of older subjects, who considered the example to sound uneducated (p. 113-115). In addition to this limited survey, some thoughts regarding the employment of *you know* comes from Lakoff (2004) and her study "Language and Women's Place", which was first published in 1975. She claimed certain DMs could appear in utterances where they were not needed, and that this was a trait of women's language. She concluded women use tokens such as *you know* as an apology for making an assertion because they are inferior in society. Her study proved to be influential on how certain linguistic traits are judged to be masculine or feminine, even decades after it was first published. Livia (2004) argues the ideas in Lakoff's work are found in popular culture varying from children's books to situational comedy, and in classes where transsexuals are taught to speak like women. This, undoubtably, helps reinforce women as stereotypical users of what Lakoff called women's language.

Lakoff's study was central in forming a new tradition of conducting research on gender-based linguistics. However, her methodology, which was rooted in introspection, was soon judged inappropriate, and new corpus-based, empirical methods formed the base of subsequent studies. Small-scale, corpus-based studies have been conducted to test whether these assumptions about the sociolinguistic distribution of DMs reported above reflect reality. Results from these studies have shown great divergence with regards to gender; Macaulay's (2002) study on Scottish speakers' frequency of DM *you know*, Ferrara & Bell's (1995) study on Texans' frequency of quotative *like* from 1990, Andersen's (2001) and Hasund's (2003) studies of teenage Londoners' frequency of DM *like* and Beeching's study on British speakers' use of *like* all revealed women in their studies are more frequent users of the DMs they targeted. Blyth et al.'s (1990) study on quotative *like* produced primarily by New York residents concludes men are more frequent users. Yet Holmes' (1986) study on New Zealanders use of *you know*, Stubbe & Holmes' study on New Zealanders use of *you know, eh*, I *mean*, *I think* and *sort of* /*kind of*, Koczogh & Furkó's (2011) study on Americans' use of *you know* and *I mean*, Ferrara et al.'s (1995) studies from 1992 and 1994 on Texans' frequency of quotative *like*, and Beeching's study on British speakers' use of *you know* report men and women participating in these studies do not vary in the overall frequency of the DM(s) which each study focuses on.

The social class of the speaker has been given less attention in previous studies compared to gender, and the studies which have considered social class vary in their conclusions, as was the case regarding gender. Beeching (2016) found that *you know* is used

most frequently among middle-class speakers and least frequently among working-class speakers, while *like* was used least by upper-class speakers and most frequently employed among lower middle-class and working-class speakers (p. 107, 138). Andersen (2001), contrary Beeching, found that speakers from the highest social class were the most frequent users of DM *like*. Hasund (2003), who also focused on *like*, found no difference with regards to social class among young speakers. Neither Macaulay (2002) found variances regarding social class when comparing middle-class and working-class speakers use *you know*. However, Stubbe et al. (1995) found that *you know* is more common among working-class speakers.

1.1 Aims and Purpose

Although many sociolinguistic studies on DMs have emerged over several decades, the field is far from saturated. As displayed above there are indications that people associate DMs with particular social groups although corpus-based studies report varying results. Most of these studies are conducted on small sets of corpus data, thus it is important to continue research of this kind to collect a larger body of research to give a fuller picture of social variance within the field of DMs. Moreover, it is important to conduct this type of research on people from different geographical areas, which can be helpful in comparing the developmental use of DM in different regions and countries.

This corpus-based study on American speakers aged 20-59 focusing on the DMs *like* and *you know*, and the social groups gender and educational level seeks to answer the following research questions:

- Can social differences regarding frequency of employing DMs be detected?
- Given their versatility, do different social groups use DMs for different purposes?

The first question will be answered by conducting quantitative research. However, in order to say anything about whether there are social differences with regards to the purpose for which DMs are employed, mixed methods research must be conducted (Hashimi & Babaii, 2013). Based on the corpus material, the qualitative part of the research serves to identify specific functions which the two DMs in focus serve. The quantitative part of the research, on the other hand, allows an insight to how frequently these different functions are used in order to compare the social groups.

This thesis is limited in scope and will, as mentioned above, only focus on *like* and *you know*. This is because these two DMs are highly frequent in informal discourse, which probably results in negative evaluations and stereotyping of certain people. By comparison, less frequently used DMs such as *you see* and *anyway* do not seem to carry the same connotations. *Like* and *you know* are highly versatile, and it is of interest to test whether there are certain uses of these tokens that show social bias.

Furthermore, the study is limited to only focus on two social categories; educational level, where I will operate with one group of people of higher education (i.e. a minimum of 16 years of education) and one group of people of lower education (i.e. less than 16 years of education), and gender, where women's discourse is compared to that of men. Gender was chosen as one social variable because there seems to be an idea that the employment of DMs is predominantly found in women's speech. Due to the varying results found in previous studies, it was of interest to continue this tradition of research to find out which results a new study could support. In the preliminary stages of this project, men and women were the only two social groups that this study was to compare. Later, it was decided to test another social parameter to give the study more sociolinguistic depth. Based on the information about the participants, other social groups could have been based on age, ethnicity, or level of education. However, the corpus contains few participants aged below 20 and participants in their retirement years, so age was rejected as a possibility. There is also little variety in ethnicity; the vast majority of the participants consider themselves white. There is also little heterogeneity with regards to educational level. It was, nevertheless, possible to compare people who have attended college for four years or more to those who have not.

2 Theoretical Background

This chapter will first give an overview of how the term DM has been defined followed by how a series of previous studies have defined functions of *like* and *you know*. Lastly, a more detailed summary of the quantitative studies mentioned in the introduction will be given.

2.1 Definitions of DMs

Jucker & Ziv (1998) claim "there is no generally agreed upon definition of the term 'discourse marker'" and list the following terms which have been used for such elements: pragmatic marker, discourse particle, pragmatic particle, pragmatic expression, and connective. The difficulty of landing on one term for this class of words reflects the fuzziness of DMs (Jucker & Ziv, 1998, p. 1) and the heterogeneity within this field of research (Fischer, 2006, p. 1). Fischer claims "[t]he term *particle* is used (…) [for] clitics, full words, and bound

morphemes" and that particles do not include larger entities (Fischer, 2006, p. 4). A *marker*, on the other hand, has been said to mark off or highlight segments in a discourse (Travis, 2006, p. 220; Aijmer, 2013, p. 6). Fraser (n.d.) claims DMs make up a subcategory of pragmatic markers, and

> [i]n contrast to the other pragmatic markers, discourse markers do not contribute to the representative sentence meaning, but only to the procedural meaning: They provide instructions to the addressee on how the utterance to which the discourse marker is attached is to be interpreted (Fraser, n.d., p. 186).

Defining DMs is not an easy task; no incidence of an uttered DM manifests all the features mentioned in this section; some show many, others few. Jucker & Ziv (1998) elaborate on the issue: "elements demonstrating more of the critical features may be taken to be more prototypical members of the class of discourse markers and those showing fewer characteristic properties may be considered more peripheral" (p. 2).

DMs tend to exploit a "greater syntactic flexibility than Standard English" (Andersen, 1998, p. 147-148), and they are according to Brinton (1996) "semantically empty" (p. 35). However, in the current study they will be, as suggested by Beeching (2016), said to have undergone a long process of semantic bleaching (p. 2), which better explains why DMs have a basic meaning. The process of semantic bleaching makes them "difficult to place within a traditional word class (Brinton, 1996 in Jucker et al., 1998, p. 3).

DMs are predominantly found in oral, rather than written communication, and are more frequent in spontaneous, informal discourse (Fox Tree et al., 2002, p. 727; Brinton, 1996, p. 33; Andersen, 1998, p. 147) likely because they are "strongly constrained by the interactional and situational context of their occurrence" (Pichler, 2010, p. 584 in Aijmer, 2013, p. 3). Moreover, they are frequently used as a way of indicating friendliness and warmth (Beeching, 2016, p. 4), and employing them causes the addressee to be put at ease (Beeching, 2016, p. 18). They operate on the metalinguistic level where the speaker organizes the discourse for example by providing the addressee with some information about the ongoing cognitive processes the speaker undergoes, and they are often inserted in order for the participants of the discourse to "negotiate their common ground" (Jucker & Smith, 1998, p. 172; Fox Tree et al. p. 728; Aijmer, 2013, p. 4, 20 Östman, 1981, p. 4) by "allow[ing] the addressee's opinion to be enjoyed or evoked" (Beeching, 2016, p. 4).

Jucker & Ziv (1998, p. 3) provide a summary of how Brinton (1996, p. 33-35) describes the basic features of DMs. Some of the characteristics regarding their form are listed below:

- They are short and phonologically reduced.
- They form a separate tone group.
- They occur outside the syntactic structure or they are only loosely attached to it.
- They are optional.

In a few cases it is difficult to decide whether a token is a DM or a non-DM with traditional grammatical features although prosodic features, such as those mentioned above, usually resolves the ambiguity. When referring to non-DMs, it is meant as words or phrases that "are long-established and fairly straight-forward" (Romaine & Lange, 1991, p. 244). Although it should be mentioned that Romaine and Lange claim there are scholars who reject some long-established words and phrases as appropriate use of grammar (p. 244).

2. 2 Previous Accounts of Functions

Like and *you know* are highly versatile, and in literature, we find a myriad of different functions DMs are claimed to serve (Müller, 2005, p. 147). Although there is disagreement among scholars with regards to specific functions, there is, nevertheless, much common ground (Fischer, 2006, p 430). A variety of functions attributed to *like* and *you know* by scholars will be presented below.

2.2.1 *Like*

In Schourup's (1985) study on speakers native to Ohio, he claims the DM *like* is an evincive, which he defines as follows:

> a linguistic item that indicates that at the moment at which it is said the speaker is engaged in (…), thinking; the evincive item indicates that this thinking is now occurring (…) but does not completely specify its content.

He further elaborates, claiming an evincive expresses "something about current contents of the private world". Hence such items are a reflection of the speaker's cognitive state. Moreover, he claims, "like is used to express a possible unspecified minor nonequivalence of what is said and what is meant" (underlines in original) and can be considered a hedge.

Upon explaining the more specific functions of the token, he claims *like* can be used for marking imprecision be it a numerical or non-numerical expression. This function is similar, though not equivalent to the adverb "nearly" (p. 38-39). Another function he finds is

when *like* is not used for marking imprecision, but meaning exactly what follows the token, and that synonyms for this function are "as it were" and "so to speak". Schourup was the first to describe *like* as an introducer of direct discourse, and writes it is similar to narrative "say" and "go", but that it can also mark the speaker's reactions, what s/he had in mind, and how s/he "felt at the time" and "[i]t is as if the speaker were saying 'What I am about to report is like what I or someone else had in mind". He found a low number of *like* following questions and *like* is then used for soliciting clarification of what the previous speaker just said. Another function he finds is the "for example" use which indicates an accurate yet selective representation. Aside from the "for example" use, *like* can be an interjection which serves to fill a pause used for holding the floor.

Miller & Weinert (1995) based their study on Scottish English. One part of their corpus is made up by task-related discourse of 18-year-olds. While the other part of the corpus is made up by casual discourse where the youngest speakers are 17 years old, and the oldest are in their fifties. The majority of the speakers, however, are aged 18. In their study, they devote a section to discuss Schourup's (1985) functions, as described above. They argue that they find no indication that *like* + a numerical expression is interpreted as an approximation because approximations are signaled by *about*, and that *like* combines with *about* in their data. Even when *like* occurs in combination with numerical expressions in the absence of *about*, they claim there is no indication that the addressee is to understand the expression as approximative, nor did they find any indications that *like* can approximate non-numerical expressions. They did not find any occurrences of *like* in relation to questions, nor as an introducer to direct speech, but they do not reject the latter as a possible function. They found instances where *like* introduces exemplification. However, they note that the token in these situations also serves to make the marked item more salient. The last function from Schourup which they discuss is when *like* is an interjection. They reject this function because *like* is, in their corpus, not associated with signals for any type of processing problems, and when it is found in an environment of hesitations, *like* is integrated in a construction, i.e. not surrounded by pauses. Once they have discussed and compared their findings to those of Schourup, they propose the function of *like* is to focus an element in the discourse. The speaker may wish to use *like* to gain the addressee's attention, highlight an important entity, highlight particular events or states, counter a misunderstanding, or to contrast one entity with another although the latter is rare.

Underhill (1988), who, like Schourup, used American speech as the basis of his study, centered his study around *like* functioning as a focuser. One type of focus, he found, is marking new concepts or entities that are significant to the discourse— "often (…) the point of the sentence", which is in line with what Miller et al. (1995) concluded. Sometimes *like* appears in questions to mark the focus of the question. Likewise, the token can appear in answers to questions where they are used to focus the "specific information that specifically answers the question". The focus can also be a segment that is not meant to be taken literally, and the speaker can distance himself from it. He makes a digression from *like* marking focus to *like* as a hedge, such as when it is used for approximation. He claims the token is pronounced with rising intonation, and that the hedging function can be found in requests and it appears because it distances the speaker from the request and the speaker is shielded form a possible refusal of that request.

Dailey-O'Cain (2000) conducted a study based on a corpus from 1995 recording 30 (upper) middle class people in southeastern Michigan. She, too, operates with *focuser like*, which she describes as a discourse- or pragmatic marker. It can be used for taking/ holding the floor, repairing, and organizing the discourse, and as Underhill (1988) also found, it marks new information and focus. She also operates with *quotative like*, but she does not go into detail about what can be quoted i.e. whether she includes thoughts, reactions, onomatopoeia, etc.

In Andersen's (2001) study based on London teenagers, he claims *like* signals the meaning of an utterance is to be understood as "a relation of non-identical resemblance" (p. 230). He acknowledges *like* can occur in disfluencies, such as marking a false start, or linking fragmented discourse. However, he rejects that the token functions as a mere pause filler, contrary to Schourup's description of interjection *like*. He agrees with the accounts above that *like* can signal a rough approximation. This function, he claims, is to flag a discrepancy between the proposition and the thought it represents. He, like Schourup, acknowledges this function can be used both for numerical expressions and other clause elements. He also agrees with Schourup that *like* can be used for exemplification. He argues *like* indicates a noun phrase is an "exemplifications of wider categories" (p. 236). He also finds that the token can be used in combination with metaphors and hyperboles where *like* can be glossed as *virtually*. He explains *metalinguistic use* of *like* occurs when the speaker employs an expression which might be less appropriate than an alternative one. Moreover, it can mark an expression which has not been fully incorporated into the speaker's linguistic repertoire. He claims this is the function that has been interpreted as marking focus in previous works, but that the teenagers

he used for his study wish to mark distance from the utterance, which is foreign to their repertoire. *Quotative* BE like, he, like Schourup, claims can mark a thought. It can also signal direct speech. He emphasizes the quotative marker signals a loose rendering of speech or thought; *like* marks the quote is a "representation of another representation which may not have been explicitly uttered" (p. 250).

Hasund (2003) used the same corpus as Andersen for her study on *like*. She claims the discourse use of the token has its origins in preposition and conjunction *like*. She lists pragmatic functions of *like* on three levels. On *the textual level*, the token is located utterance-initially and is used for marking "relations between two units of discourse". Operating on the *textual level*, *like* can also serve as a turn-structuring function and being a verbal filler for holding the floor, and to bring focus to textual elements. The second level in her study is the *subjective level* where the token can be used as a hedge in relation to either the content or form which follows, and *like* can be paraphrased to "approximately" and "for example". On the *subjective level*, *like* could be used as an intensifier emphasizing an "exact rendering of the speaker's mind". This is contrary to what Andersen concluded based on the same corpus material. Her description of *like* as an intensifier is similar to how Underhill (1988) defined focuser *like*. The last level is the *interpersonal level* where the token signals an orientation towards the addressee for him to be involved.

D'Arcy (2007) collected her corpus material from Toronto, Canada. She found that *like* can be a *quotative complementizer*. Her description is similar to that of Schourup and Andersen, in that she acknowledges its use for reporting speech and thought. She also found it introduces nonlexicalized sounds. She named a function *approximative adverb*, which is used in combination with a numeric expression and is the same as that of Schouroup and Andersen. However, it is not clear whether she includes non-numerical expressions. Next, she lists *like* functioning as a *discourse marker*. This is a narrow understanding of the term compared to my study. According to her, DM *like* is a cohesive device operating on the textual level signaling "exemplification, explanation, or the like" and is similar to what Hasund (2003) wrote about the textual level. Her last function of *like* is *discourse particle.* One difference between a marker and particle, she claims, is that only the latter occurs within the clause. This is what Underhill (1988) called focus. They are a plea for cooperativeness in communication and creates a more intimate relationship with the addressee.

2.2.2 *You know*

Lakoff's (2004) study "Language and Woman's Place", which was first published in 1975 discusses a variety of stylistic forms which she associated with women's discourse. The method she used for this study was introspection based on conversations with acquaintances and what she had heard from television shows and commercials. *You know*, she claimed was one of these stylistic forms. She categorized it as a hedge, which could be used for one of three reasons; the speaker is genuinely uncertain about the truthfulness of the utterance, the speaker uses it as a sign of politeness, or when there is no reason to hedge but it, nevertheless, occurs. With regards to the latter type of hedge, she writes:

> the speaker is perfectly certain of the truth of the assertion, and there's no danger of offence, but the tag appears anyway as an apology for making an assertion at all.

The latter type of hedge has gained much attention in subsequent research on DMs. Lakoff hypothesizes this type of hedge is a result of lack of self-confidence.

Holmes (1986) concluded that all instances of *you know* in her New Zealand corpus consisting of approximately 50,000 words uttered by 64 speakers had an overall function as a verbal filler. She used Lakoff's study from 1975 as the basis for her work. Contrary to Lakoff, Holmes finds in her study the DM does not always function as a hedge, but that it can also be used as an intensifier or a booster—opposite of a hedge, a booster is used to reflect speaker-confidence with regards to the assertion, and she created two main categories for the token; one reflecting various kinds of functions which are used for expressing certainty which she calls category I, and another category for reflecting different types of speaker-uncertainty, which she names category II. One subcategory of category I is *conjoint knowledge*, which is used when the speaker introduces something about which the addressee already has knowledge. This function is used for introducing relevant background information. The second subcategory of category I is named *emphatic* and is used for boosting the speech act to reassure the addressee of the validity of the proposition. The last subcategory of category I is called *attributive*, which is used to express not only the validity of an utterance, but also confidence regarding the speaker's knowledge of what the addressee already knows. This function, she claims, can be paraphrased to "I'm confident you know the kind of thing I mean". One subcategory of category II is *appealing*. This function is used when sharing embarrassment or personal information and there is a wish that the addressee validates the speaker's feelings towards what is being shared. *You know* can also be used as saving face when uttering negative or critical comments. The final subcategory of category II us *linguistic*

imprecision which include uncertainty regarding lexical choice, introducing clarification to the previous proposition, and when indicating a false start.

Erman (2001) operates with three main categories. The first one is *textual monitor* where one purpose is to make coherence out of fragmented discourse. These monitors are used for signaling transitions and to highlight certain elements. What she calls *discourse markers* and *editing markers* (including *repair*, and *hesitation markers*) are types of *textual markers*. *Social monitors* are addressee-oriented and are used for negating meaning e.g. for confirmation. They are also used for turn-management. *Interactive markers* (including *turn-regulators*) and *comprehension-securing markers* belong here. *Metalinguistic monitors* deal with modality by signaling the speaker's commitment to the propositional truth, and is, therefore, a face-saving device. Among *metalinguistic markers* she includes *approximators*, *hedges*, and *emphasizers*. She stresses that there is no clear-cut boundary between the three types of monitors.

Searching for basic meanings, Fox Tree & Schrock (2002) find one basic function of *you know* to be *interpersonal* where it can be used for soliciting positive politeness by playing on the addressee's knowledge, invite the addressee to interfere, and a way of conveying negative politeness in face-threatening situations. Another function they found is *turn management* where the speaker might wish to open up for addressee interferences. A different function they found was *repair*, which is used when experiencing expressional trouble, and can be an invitation to the addressee to interfere but is also used strategically to buy time. This function is more similar to Schourup's (1985) definition of *like* as an interjection than Holmes' subfunction as marking linguistic impairment. Fox Tree at al. also find *you know* being used for *monitoring* resulting in backchanneling from the addressee as a form of inference. The speaker wants to assure that the addressee comprehends the implications and relevance of the utterance. The last function they identify is *organization*. In this function *you know* can mark a topic shift, foreshadowing cause and effect, introducing background information, presaging a quote, and highlight a segment. They conclude *you know* always marks a request of addressee inference.

2.3 Sociolinguistics

A rendering of studies testing for sociolinguistic difference with regards to *like* and *you know* will be provided below.

2.3.1 Like

In recent years, a series of sociolinguistic accounts of the DM *like* has emerged, some of which will be mentioned here. In some papers, the quotative function has been the only function of the DM *like* which has been given attention. This might be a result of how different quotative *like* is from the other functions; it is usually combined with the stative verb *BE*, and quotative *like* is not optional in the sense that it cannot be removed without reconstructing the rest of the utterance. There are also a great number of studies which take several functions into account. The rendering of studies focusing exclusively on quotative *like* will be separated from the studies which take several pragmatic functions into account.

2.3.1.1 Studies Focusing on the DM *like* at Large

D'Arcy (2007) used a corpus based on recordings of 48 men and 49 women conducted between 2002 and 2005. She found that quotative *BE like* is used significantly more by women. She also found that the token functioning as an approximation has no significant gender bias. When the token functions as what she refers to as a DM she found that there is a marginal significant difference between men and women, and that women are more frequent users. The last function with which she operated is when the token is used as a discourse particle. The results show men use this function significantly more than women.

Hasund (2003) studied 30 teenage Londoners' use of *like* as a DM in COLT, a corpus containing approximately 500,000 words, and the corpus has been added to the British National Corpus (henceforth written BNC). She found that teenage girls use the token 3.5 times per 1,000 words, while boys employ the token less frequently as they utter it 2.9 times per 1,000 words. Similar results were found by Andersen (2001) in his study on the same corpus; he found that boys use the token 2.78 times per 1,000 words, while girls used it significantly more at 3.24 times per 1,000 words. The small discrepancy of Hasund's and Andersen's results could possibly be explained by how they define DMs and non-DMs. Hasund further investigated the social class of her participants. She compared speakers from the two boroughs Hertfordshire and Hackney. The former primarily contains (upper) middle class speakers, while the latter is mainly made up by working class speakers. The Hertfordshire speakers used the token 3.7 timer per 1,000 words, while the Hackney speakers employed it 3.8 times per 1,000 words, hence showing no particular difference with regards to social class in the overall distribution. No test to detect statistical significance was carried out in her research, however. Andersen, like Hasund, looked into variance between social classes. He divided his speakers into three social classes; high, middle, and low. He found members

from the highest social class to be the most frequent users, employing the token 3.35 times per 1,000 words. The members of the middle social class used it 2.80 times per 1,000 words, while the low social class used it 2.89 times per 1,000 words. However, there was no significant difference between the social classes. He, nevertheless, managed to find a small significant difference when collapsing low and middle class, and comparing it to the high social class (Andersen, 2001, p. 289-290), which in turn correlated to one of his pilot studies (Andersen, 1997d in Andersen, 2001, p. 289-290). He also tested sociolinguistic variation regarding age, ethnicity, and geographical distribution. Based on frequencies of the social groups within these categories he summarizes:

> the prototypical user of the pragmatic marker *like* is a white 17-year-old girl from the highest social class who attends the boarding school in Hertfordshire. Conversely, the least typical *like*-user within the target group is a male ethnic minority member aged 13 from Brent (Andersen, 2001, p. 294).

Beeching (2016) used the entire data from the BNC in her study on various DMs. She used the social classes from the BNC (AB, C1, C2, DE). The social classes are based on occupation and AB denotes the highest level of occupation and DE denotes the lowest class based on occupation (Friginal & Hardy, 2014, p.87). Contrary to Andersen (2001), she found the highest social class to be the least frequent users as they employed the token 43.76 times per 10,000 words compared to 49.23, 51.86, and 50.16 times per 10,000 words respectively. Only the difference between C2 speakers and DE speakers proved to be of an insignificant difference. However, these numbers also include non-DM *like*. Her study also shows women use *like* 51.69 times per 1,000 words, while men use it significantly less at 47.92 times per 10,000 words (Beeching, 2016, p. 140). However, these numbers also contain non-DM uses of the token.

Dailey-O'Cain (2000) conducted both a survey of attitudes towards *like* and an analysis of the sociolinguistic distribution of the token. From the corpus, she found that focus *like* was used more frequently by men than by women, but that the difference is insignificant. The same proved to be true for quotative *like* (Dailey-O'Cain, 2000, p. 66, 68). From the survey, she found that the vast majority of her 40 informants believed women use *like* more frequently (Dailey-O'Cain, 2000, p. 69-70). She concludes there is a direct contradiction between the peoples' perception of the token and the findings from the quantitative study (Dailey-O'Cain, 2000, p. 75).

2.3.1.2 *Studies Focusing Exclusively on Quotation Markers*

Romaine et al. (1991) studied adolescents' and adults' use of quotative *like* from a body of media sources and found nearly 80 instances of it. One of their findings was that adolescents use it more frequently than what is the case for adults. However, there is no mention of how many words were uttered by adolescents compared to adults. This result, nevertheless, seems to support a range of other studies which claims the token is more typical of younger speakers (e.g., Tannen, 1986; Blyth, Recktenwald & Wang, 1990).

Research on the use of quotative *like* according to gender has provided us with a wider range of results than what is the case for age. Romaine et al. (1991) did not only find that this quotative marker is more typically used by younger speakers in their study, they also found that women are by far dominant users of this particular form of reported speech and stood for 83 % of the productions. Just like it in the case of young speakers compared to adults, there is no information on how many words men uttered compared to women.

Ferrara et al.'s (1995) study used corpora recorded in Texas. From a corpus recorded in 1990 containing 115 speakers, they found that young females used quotative *like* in 29 % of the cases where they employ a direct dialogue introducer (e.g. *SAY, GO, BE like*), while the number decreases to 15 % in the case of the young men. The results show that *BE like* is not restricted to women's speech. However, women use it more frequently than what is the case for men. Subsequent studies were conducted on a corpus recorded in 1992 containing 200 different speakers and a 1994 corpus made up by 90 speakers. The results from both these corpora show that men's and women's use of quotative *like* is equally divided by the two genders. They conclude the female bias is becoming neutralized in Texas.

Also Blyth et al. (1990), who based their study on Americans, many of which were associated with Cornell University, conducted research on *SAY, GO* and *BE like* as introducers for direct speech. Of the three forms of direct quotation, they found that only *BE like* had a significant gender bias. Despite publishing the study in 1990, the year in which Ferrara et al. found that there was a female bias towards the token in Texas, they found that their data, primarily, or perhaps exclusively based on people residing in New York, showed men use the token significantly more than women, which was contrary to what they hypothesized. However, it is not certain their results reflect the population given 24 of 30 participants have ties to Cornell University, and they admit to not have controlled for neither, social class, nor education level, nor gender (twenty females, 10 males) (Blyth et al., 1991, p. 216).

2.3.2 *You know*

Lakoff's (2004) study, which as mentioned above, is based on introspection rather than corpus-based data, claimed the third type of hedge is used more frequently my women than by men. She hypothesizes this is a result of women, from an early age, are being taught "to believe that asserting themselves strongly isn't nice or ladylike, or even feminine". Moreover, she claims "the use of these hedges arises out of fear of seeming too masculine by being assertive and saying things directly" (Lakoff, 2004, p. 79). And, as mentioned in chapter 1, she believes women's frequent use is a result of them feeling inferior in society.

Holmes (1986) found no significant difference in men's and women's overall frequency of the employment of the token. When she compared men's and women's frequency of using the token for expressing certainty or uncertainty, she found that women use it 30.4 % of the time for expressing certainty, while men use it for this purpose 20.8 % of the time. For expressing uncertainty, women use it 20.3 % of the time, while men do so 28.4 % of the time. The results proved to be of insignificant difference, however. She also found that *you know* is used significantly more by both men and women in same-gender contexts compared to mixed gender contexts. Based on her comparisons of gender, she concludes her study refutes Lakoff's (2004) claims that women use *you know* more frequently than men for expressing uncertainty, and that a further contradiction to Lakoff is that men used the token more frequently for expressing uncertainty, while it is women who use *you know* most frequently for expressing certainty or confidence despite these results being of insignificant difference.

Stubbe et al. (1995) conducted a quantitative study on a variety of DMs used by New Zealanders. Their corpus contains approximately 75,000 words which are retrieved from interviews and casual conversations. They found that men tended to use *you know* more frequently than women, although there was no significant difference. With regards to social class, which they have defined based on occupation and level of education, they found that although working class-speakers used the token more frequently than middle class speakers in the interviews, it was not of statistical significance. However, they found working class-speakers in casual discourse to use the token significantly more than middle class-speakers. Based on the data from the interviews, they further compare young middle-class men and women, and young working class and men and women. They conclude there is no difference between working-class men and women given working-class women's frequency is 27 and working-class men's frequency is 29. While their results suggest working-class men, whose

frequency of *you know* is 83, use it more than working-class women, whose frequency is 36, and middle-class speakers. It is, however, not known whether these results are of statistical significance.

Koczogh et al. (2011) used American speech retrieved from transcripts of interviews from the talk show *Larry King Live* as a basis for their study on American speech. They operate with one male corpus of 52,000 words which is made up by Larry King and male interviewees, and one female corpus of 53,000 words consisting of various female guest hosts and female interviewees. They operate with raw numbers and percentage of the token, but by calculating the numbers they report, I found men in their study use the token approximately 5.37 times per 1,000 words, while women do so with a frequency of approximately 6.30. Like Holmes (1986), they conclude that there is no difference in the overall frequency of the DM between the two genders. They found the DM to serve a long list of functions which they do not explain in detail. Although most functions were used by both men and women, they found that only women use it in utterances for seeking agreement, and that only men used the token for mitigating disagreements (Koczogh et al., 2011, p. 4-5). They found that there were some functions that were used significantly more by one gender; when the token was said to be a marker for indicating hesitation, a false start or lexical search, men used it significantly more. Men also used the token significantly more when it was used for explanation or elaboration. Interestingly, these categories are very similar to some of the subcategories of Holmes' (1986) category II, which signaled speaker-uncertainty. As written above, she also found men to use the DM more frequently for such purposes although her results were not of significant difference.

Beeching (2016) used, as mentioned above, the full 100-million-word version of the spoken data found in the BNC in her study on various DMs (p. 28-29). She found that women used *you know* 36.35 times per 10,000 words, while men used it 35.8 times per 10,000 words which proved not to be a significant difference (Beeching, 2016, p. 108). Like in her analysis on *like*, Beeching looked at social class as a variable for *you know*. Her results show that C1 speakers, who employ it 36.37 times per 10,000 words, are the most frequent users of the token. The least frequent users are the DE speakers, who employ it 30.91 times per 10,000 words. AB speakers use it 32.50 times per 10,000 words, while C2 speakers use it insignificantly less than C1 speakers, as they use it 36.37 times per 10,000 words. Only the comparison between C1 and C2 speakers was not of significant difference (Beeching, 2016, p. 107).

Macaulay (2002) used two corpora of spoken Scottish English. One corpus is made up by adult speakers in discourse with an interviewer which was recorded in 1978-79, while the other corpus was recorded in 1997. The latter corpus is made up by speech between participants with a preexisting relationship. There are two age groups in this corpus; 13-14-year-olds and people aged 40 or above. The study focuses on the frequency of employment of the token comparing different social groups. His results show that women use the token 4.92 times per 1,000 words, while the males' frequency is 2.41. The results also show the gender bias remains when comparing men and women from the same social class. The only exception is for working-class adolescents where men use it more, although working-class adolescents are infrequent users (Macauley, 2002, p. 753) on the whole. When looking at social class in isolation, he finds that middle-class speakers use the token 2.29 times per 1,000 words, while lower-class speakers use it with a frequency of 3.49 times per 1,000 words. He concludes social class is much less deterministic than gender, age and recoding context, which he also investigated. Although he concludes there are differences between genders, he is careful to make generalization of his results because he finds great individual differences.

2.3.3 Summary

This chapter set out to give an overview of previous accounts of how DMs have been defined, previous accounts of the functions, and sociolinguistic distribution of *like* and *you know*.

DMs have in this chapter been said to be associated with oral discourse and are used on the metalinguistic level as a way of creating coherence. They are notorious for being more syntactically flexible than words of Standard English, and they are semantically bleached. They tend to be phonologically reduced items forming separate tone units, and they can usually be removed without impacting the propositional truth.

The chapter confirms Fischer's (2006) claim that parallels between the functions with which different scholars operate can be found; Schourup (1985), Andersen (2001) and Hasund (2003) claim the DM *like* signals a minor non-equivalence. Schourup writes this always is the case for DM *like*, while Hasund finds it also has its roots in conjunction *like*. Several scholars also acknowledge that *like* marks imprecision of non-numerical and/ or numerical expressions (Schourup, 1985; Underhill, 1988; Andersen, 2001; Hasund, 2003; D'Arcy, 2007) although Miller et al. (1995) argued against this function. The token has also been said to be used for introducing an example, or clarification (Schourup, 1985; Miller et al., 1995; Andersen, 2001; Hasund, 2003; D'Arcy, 2007) focusing essential elements of the discourse (Underhill, 1988; Miller et al, 1995; Dailey-O'Cain, 2000; Hasund, 2003; D'Arcy, 2007), and reporting speech,

thought, sounds, or the like (Schourup 1985; Dailey-O'Cain, 2000; Andersen, 2001; Hasund, 2003; D'Arcy, 2007). Fox Tree et al. (2002) found *you know* can also mark an upcoming quote. Only Schourup (1985) lists *like* as interjection that fills a pause, while only Andersen (2001) addresses that *like* can be used for marking metaphors and hyperboles, and that it can mark expressions which the speaker has not yet fully incorporated in his repertoire.

In literature, *you know* has been described to be used for marking knowledge that the addressee already possesses (Holmes, 1986; Fox Tree, 2002), negating meaning (Holmes, 1986; Erman, 2001), highlighting certain elements (Erman, 2001; Fox Tree et al., 2002), and seeking various kinds of confirmation from the addressee (Holmes, 1986; Erman, 2001, Fox Tree et al., 2002).The two tokens, this chapter suggests, can also be used as a hedge to signal lack of commitment to the truthfulness of the assertion (Schourup, 1985; Underhill, 1988; Hasund, 2003, D'Arcy, 2007, Holmes, 1986, Erman, 2001), a booster, which indicates commitment and certainty (Hasund, 2003; Holmes, 1986; Erman, 2001; Fox Tree et al., 2002), as an organizer of the discourse e.g. for floor management (Dailey-O'Cain, 2000; Hasund, 2003, Erman, 2001; Fox Tree et al., 2002), and to mark discourse disfluencies e.g. filling pauses and repairing (Schourup, 1985; Dailey-O'Cain, 2000; Hasund, 2003, Holmes, 1986, Erman, 2001, Fox Tree et al., 2002).

The previous studies rendered in this chapter show great variance in the results they report. Andersen (2001) and Hasund (2003), who used the same corpus for their studies, reported that teenage girls use *like* more than teenage boys. Beeching found similar trends for adult speakers; women used *like* significantly more than men. D'Arcy (2007) found that gender-based differences could be found based on the purpose they served; men use the token more when it serves as a discourse particle, women use the token more as what she refers to as DM and for introducing a quotation. Romaine et al. (1991) also found women to be more frequent users of quotative *like*, while Ferrara et al. (1995) concluded the initial female bias they found for this quotative marker has been neutralized. Blyth et al. (1990), on the other hand, reported men used quotative *like* significantly more than women. There are also some varying results regarding the distribution of *like* when comparing social background. Andersen (2001), with some difficulties, managed to prove the speakers from the highest social class use the token more frequently when comparing them to a group consisting of speakers from middle and low social class. Beeching (2016) found that the speakers of the highest social class in her data on British language were the least frequent users of *like*.

Lakoff (2004) wrote women were more frequent users of *you know* when she believes the token is not needed, and that it is a result women's inferiority to men. However, her study did not provide any quantitative results to support her claims. Macaulay (2002) found that women are more frequent users of *you know* than men. But neither Koczogh et al. (2012), nor Holmes (1986), nor Stubbe et al. (1995), nor Beeching (2016) concluded men and women use *you know* with different frequencies. Macaulay (2002) concluded there was not much difference in the frequency of middle-class speakers compared to lower-class speakers. Stubbe et al. (1995) found that lower-class speakers use the token significantly more than middle-class speakers in casual discourse, but that the two social classes use it equally often in interviews.

Although the results from the studies mentioned above vary, it becomes clear that neither *like* nor *you know* is exclusive to the discourse style of any of the social groups mentioned above; all the quantitative studies report representatives from all the classes use the tokens, the question is rather whether there are differences in the frequency with which they occur. Not all the studies rendered above have used statistical significance as a basis for the conclusions. Instead, several researchers have made subjective interpretations of the difference between the compared groups. As will be addressed in the following chapter, this current study will base its conclusions on statistical significance.

3 Methodology

This chapter will provide information about the corpus which was used for this study, how pragmatic functions of *like* and *you know* were identified, how the different social classes have been defined, and about the steps taken for conducting statistical analyzes of the data.

3.1 The Corpus

The SBCSAE was selected to form the basis of the current thesis. It contains approximately 249,000 words from naturally occurring spoken interactions across the United States. The corpus was published in four part in the years 2000-2005. This corpus was selected because the material is freely available for everyone, hence making it possible for the readers of my study to listen to the audio files and to read the full transcripts (see appendix II). Moreover, the audio files are of high quality, the different contributions to the data were compiled within a relatively short time span, and the data is relatively new. In comparison, the London-Lund Corpus was compiled between year 1959 and 1990 (Svartvik, 1990). Furthermore, there was no external interviewer present when the data for the SBCSAE was recorded (University of California Santa Barbara, Department of Linguistics, n.d.), which is important to the study at

hand because previous studies suggest DMs are more frequently used between friends than between strangers (Jucker & Smith, 1998, p. 193; Redeker, 1990, p. 375 in Fox Tree et al., 2002, p. 730). Additionally, the absence of an external interviewer lowers the risk of poor data due to the Observer's Paradox i.e. the speaker either consciously or subconsciously alters his linguistic behavior in the presence of someone with whom he does not share a close relationship (Rasinger, 2013, p. 52).

As mentioned above, DMs are said to be predominantly used in spontaneous, naturally occurring discourse, rather than formal discourse (Östman, 1981, p 16; Holmes, 1987, p. 12-13; Fox Tree & Schrock, 2002, p. 729). Furthermore, their functions are closely related to the context in which they are uttered (Aijmer, 2013, p. 6). As a result, utterances produced in classroom lectures, church, town hall meetings, and in the work place where the relationship between the speakers is worker and costumers/ patients were excluded. The corpus has neither been balanced for number of words uttered by each person, nor the different social class (when comparing gender, educational level has not been balanced and *vice versa*), ethnicity, or age although children, teenagers, and people aged 60 or above were excluded from this study. This restriction was done to ensure that the compared groups did not contain an unbalanced number of people representing one of the extreme ends of age.

The subcorpus used for the current study is made up by approximately 86,000 words uttered by 52 speakers although five speakers failed to report their educational level and has only been included when comparing gender. In total, approximately 1,100 tokens of *like* were analyzed. 714 of them were assigned to one of the four functions of *like* (see chapter 4) used in this study. In the case of *you know*, approximately 700 tokens were analyzed, and 560 of them were assigned to one of the four functions of *you know* (see chapter 4).

3.2 Categorizing the DMs

In order to conclude whether there are any differences with regards to how different social groups use *like* and *you know*, the two tokens had to be categorized according to function. To do so, a small pilot corpus was made where a few tokens of *like* and *you know* were analyzed and their functions described. This, and some inspiration from previous works mentioned in chapter 2.1 and the author's BA thesis on the DM *you know* (Jacobsen, 2017) made the framework for analyzing the corpus that was used in this study. While analyzing the corpus, more detailed information about the different functions were added. Based on the corpus analyzes, four functions were assigned to *like*: signaling an approximation, bringing up one of several examples, asking the addressee to focus on a highlighted segment in the discourse,

and signaling an upcoming quote. While it was found that *you know* was used for the following four reasons: introducing information that was previously known to the addressee, reassuring the addressee of the speaker's confidence or conviction of the truthfulness of the marked segment, transitioning from an unclear segment into an elaboration, and to signal an upcoming repair to an utterance. These functions are described in detail in chapter 4.

There were cases when the tokens could not be analyzed properly. On these occasions, the occurrences of the two tokens were placed in the so-called rag bag. The rag bag was usually used because the speaker never uttered what was supposed to be marked by *like* or *you know* either because another participant of the discourse steals the floor, or because the DM is a part of a reparandum. Here, it needs to be mentioned that there is a difference between transitioning from a reparandum to a repair, which is one of the functions of *you know* can serve and being part of a reparandum. As appendix III shows, *like* is frequently found within a reparandum. The reason why *like* does not introduce a repair is because it belongs to the same tone unit as the rest of the reparandum, while *you know* functioning as a transition into a repair belongs to a separate tone unit. Occurrences of the DMs could also be put in the rag bag if they carried a function that the current study did not operate with because there were too few instances of such functions to gain attention here. This only happened two times; one case where *like* appeared to mark transition from a reparandum to a repair, rather than being part of the reparandum and one case where the DM did not fit into any of the described functions, and it did not become evident why the DM was uttered.

Given neither *like* nor *you know* needs to be a DM, a category of non-DMs was also created. The tokens were defined as a non-DM, if they were one of the following:

Verb: (1) `WENDY: ... Do you **like** .. frozen yogurt?`

(SBC013 *Appease the Monster)*

Preposition: (2) `[MILES:] that looks **like** a brother`

(SBC002 *Lambada)*

Conjunction: (3) `[PAMELA:]it's **like** it pulled me under`

(SBC005 *A Book about Death*)

Verb: (4) `[MARIE:] Do **you know** what I mean`

(SBC036 *Judgmental on People*)

There are instances where scholars do not agree to what should, and should not be classified as a DM. In her book *Discourse Markers*, Schiffrin (1992) analyzes several examples of the shortened interrogative form of *do you know* as DMs (Schiffrin, 1992) ((5), (6) and (7) are examples from her work).

(5) [Jack:] **y'know** what Hasidic is?

(p. 269)

(6) [Zelda:] **Y'know** [lunch] enonette? As a waitress.=

[Debby:] [Yeh.]

(p. 269)

(7) Henry: **You know** where Neshaminy is?

(p. 271)

Holmes (1986), however, argues the token "must be distinguished from the superficially similar shortened form of the interrogative *do you know*" (p. 6). In this current thesis the reduced interrogative form will not be regarded a DM for several reasons: (1) despite omitting *do*, the reduced form cannot be seen as separate from the original phrase where *do* is an auxiliary verb which stands to the subject *you* and in combination with the non-finite verb form *know*. Since one of the strongest features of DMs is that they are difficult to place within a traditional word class, it is troublesome to argue in favor of the reduced interrogative being analyzed as a DM. (2) as mentioned above, DMs are only loosely, if at all, attached to the syntactic structure and can be omitted without affecting the remaining utterance. The token in the three examples from Schiffrin above are strongly attached to the rest of the utterances and cannot be omitted. (3) it seems unlikely that either of the three examples displaying the reduced interrogative form *you know* should belong to a separate tone unit.

When assigning the DMs into the different categories, the audio files were played at the same time the transcripts were read. Sometimes they were easy to categorize, but due to the fuzzy nature of DMs (Jucker & Ziv, 1998, p. 2; Hasund, 2003, p. 57) given they are both polysemous and multifunctional (Beeching, 2016, p. 6), it is not always a straight forward job. In cases of doubt, the transcripts were read carefully, and the audio files were replayed. In cases where a given DM serves more than one function on the subordinate level (see chapter

4), it has been categorized according to its most prominent function, see (8). However, choosing to only focus on the most prominent feature does raise an important issue which was addressed by Holmes (1985), who also categorized DMs according to their most prominent feature: the study becomes subjective, and risks leaving traces of the analysist's attitudes towards certain social groups (Holmes, 1986, p. 17). Macaulay elaborates:

> [t]he more the investigator approaches the data with preconceptions about gender differences the greater the risk of biasing the subjective interpretation in one direction or the other, as happened with claims about girls' precocity in language development (Macaulay, 1978 in Macaualy, 2013, p. 224)

Although he only mentions gender, this is equally true for other social constructs, e.g. educational level and social class.

(8)

MARILYN:	% % Cause she said,
	(H) <Q you wouldn't mind if I came back and got a whole ba=g full,
	would you Q>?
PETE:	R=ight.
MARILYN:	I said .. <Q yeah Q>.
	[@@@ (H)
PETE:	[@@@@@@]
MARILYN:	There's **like** one] lemon left on this [2tree that I2] can reach.

(8) shows an example of the token serving two functions at the subordinate level. Although *like* in this example is used as an approximator for *one*, Marilyn is not primarily trying to signal that *one* is supposed to be interpreted as *a few*. The primary function in this case is to focus the point of why the woman referred to as *she* cannot come back and pick lemons from Marilyn's lemon tree. Hence, this example of *like* has only been categorized as a focuser *like* (see chapter 4.1.2.1.4) in this study.

3.2.1 Selecting the Term *Discourse Marker*

The process of selecting the appropriate term for *like* and *you know* for the study at hand has undergone some careful reflection. Some terms were quickly rejected such as the French term *mot de discours* which literally translates into *word of discourse*. This term was rejected because it seems to be too vague given that essentially every word in one's lexicon can be used in discourse. Both *particle* and *marker* are words which have been used in previous literature. As mentioned above, Fischer claims particles do not include larger entities (Fischer, 2006, p. 4). Since *you know* is a two-word, disyllabic phrase, only *like* and not *you know* should, to my understanding, be considered a particle. Employing the word *marker*

seemed a more natural choice given *like* and *you know* mark off, or highlight segments in a discourse (Travis, 2006, p. 220; Aijmer, 2013, p. 6), which are characteristics of markers. The last decision that had to be made in finding a justifiable term was to find a modifier to the head noun. Two of the most common terms used are *pragmatic* and *discourse marker*. As mentioned above, Fraser (n.d.) claimed DMs make up a subcategory of pragmatic markers, and that DMs, in particular, signal how an utterance is to be understood, which seems to be in accordance with how DMs in the current study have been described and analyzed. In addition, Andersen (2000) writes that DMs have come to be closely associated with corpus analyses (p. 3), which lays the foundation of the thesis at hand. Deciding to work with the term DM does not mean that I wholeheartedly reject alternative terminology. The aim was to find a term which is specific, yet wide enough to cover both tokens and to employ a term that is generally known to the field of study.

3.3 Defining Social Groups

For testing gender in this study, women will be compared to men. West & Zimmerman (1987) made a distinction between sex and gender, where sex is ascribed to a person's biological makeup, while gender is determined by a person's psychological, cultural, and social means. Hence, only gender can be taken into account in a study on sociolinguistic differences. In the metadata of the SBCSAE, all participants have been asked to fill out a box where they state their gender. However, they all operate with "male" or "female", which are terms used for sex, whereas "man", "woman" etc. are correct terms for gender. In this study, it has been estimated that the answer "male" in the metadata means that a person identifies as a man, while "female" means the person identifies as a woman.

"Lower education" usually refers to educational level beyond high school. However, very few subjects that could be used for this study had studied for 12 years or less. Several subjects wrote they had taken "some college", and it was decided that a distinction between higher and lower education should be drawn between those who had studied for 16 year or more, and those who had studied less than 16 years. Hence, those who belong to the group of lower education have not acquired a Bachelor of Arts degree, Bachelor of Science degree, or the like.

3.4 Statistics

The SBCSAE does, to my knowledge, not contain any information about how many words each subject utters in a conversation. It was crucial to know how many words were uttered by the different participants in order to compare how frequently people from different

social background employed *you know* and *like*. To say anything about social differences in the employment of these DMs, the speech of each person had to be manually separated from the speech of the other participant(s) in the discourse. The word count was conducted in #LancsBox (Brezina, McEnery & Wattam (2015), a software for corpus linguistics. The problem with #LancsBox is that all signs separated by a space is counted as a word, thus the tags in the corpus such as "(Hx)", which means the speaker exhales, are also counted as words. Therefore, in order to get an accurate number of words uttered, all tags found between two spaces were removed for the data used for word count. The number of both total words uttered, and tokens uttered provided the necessary information needed to compare frequency of DM employment between the different social groups. The frequency was calculated in tokens per 1,000 words (([token]*1000)/[total word]). Afterwards, the results were tested for statistical significance.

Before testing the results for statistical significance, information about whether the results were normally distributed was needed in order to decide which test would be appropriate for testing for statistical significance. The frequencies of the different functions of the two tokens produced by each social group was plotted into SPSS for Windows (v. 25, SPSS Inc, Chicago II, USA) to make histograms for the results. In addition to analyze the histograms, the values of the mean and median were compared. If there is little difference between the values of the mean and the median, the results are normally distributed (Tjønndal, 2018, p. 64).

In cases where the results from both the groups compared are normally distributed, the Student's t-test was employed. When the results from neither group that is to be compared is normally distributed, or when the results from only one of the compared groups are normally distributed, the Mann-Whitney U test, which is a non-parametric test used for small sample sizes (Rasinger. 2016, p. 230), was employed. Given this study seeks to answer whether one social group use it either more or less frequently than the group to which it is compared, two-tailed tests were carried out.

The aim of both the Student's t-test and the Mann-Whitney U test is to decide whether the null hypothesis (H_0), i.e. the hypothesis that there is no difference between the two groups, can be rejected. The decision to reject H_0 is decided by the p-value. For this study, H_0 could be rejected if the p-value was higher than 0.050, i.e. there is a 95.0 % chance of replicating these results by testing another sample form the same population, and the alternative hypothesis (H_1), i.e. there is a difference between the two groups, was accepted.

4 Results

This chapter will first present the findings from the qualitative part of the study, namely which pragmatic functions *like* and *you know* serve. Next, the quantitative part of the study will reveal the frequencies of how the different social groups used the two tokens.

4.1 Functions

As mentioned above, it is generally accepted that DMs simultaneously serve more than one function (Hasund, 2003; Brinton, 1996). By analyzing the SBCSAE, it became apparent that *like* and *you know* have functions serving on two different levels. Some functions are very general and fluid, and do not relate to the basic meanings of the DMs. These have been called superordinate functions in this study and must be combined with the subordinate functions, which have more specific functions, and are related to the basic meaning.

4.1.1 Superordinate Functions

The superordinate functions presented in this section will not be taken into consideration in the quantification of the DMs since they are so fluid on this level, and because the superordinate functions often are intertwined. It is, nevertheless, important to describe these functions in order to understand the full pragmatic force of *like* and *you know*.

4.1.1.1 Hedges and Boosters

In literature, we find that some scholars claim all instances of *like* and *you know* as DMs are hedges. As mentioned in chapter 2.2.2, Lakoff (2004) claimed *you know* and other DMs have three types of hedging functions, but never assigned them any non-hedging functions. Based on what has been said about basic meanings of *like*, e.g. Schoroup's (1985) claim, as mentioned above, that *like* marks a minor non-equivalence (p. 42), and Andersen's (2001) description of *like* as "less-than-literal use of language" (p. 219), we find accounts of *like* being described as a hedge. The data from the SBCSAE suggests these two DMs can be employed for hedging purposes. However, the data shows that *like* and *you know* are not always used for "modify[ing] the degree of membership of a predicate or noun phrase in a set" saying that "that membership (…) is partial, or true only in certain respects" (Brown & Levinson, 1987, p. 145).

There are occasions when the DMs are used deliberately to intensify or boost the marked item, and the speaker exudes certainty and confidence and does not wish to distance himself for the proposition. Holmes (1986), as previously mentioned, found that *you know* cannot always be considered a hedge and claimed that a booster was used for "expressing the speaker's compete confidence in the proposition being asserted (…) and serving to reassure

the addressee of its validity" (p. 16-17). Miller et al. (1995), in their study on *like*, did also go against previous connotations that *like* "has a toning-down effect" (p. 368), and instead focuses on *like* as a focus marker. Underhill (1989) did also emphasize *like* as a focuser, but also acknowledges its function as a hedge. Hasund (2003) calls focuser *like* an intensifier (p. 17-21), which is what is known as a booster in the current thesis.

Like and *you know* in this study can be both hedges and booster depending on the context. Moreover, there are instances where a single instance of the token is used both as a hedge and a booster, which was also found by Hasund (2003), and by Blyth et al. (1990). As has been found in this study, Blyth et al. (1990) write that quotative *like* makes no claims that the reported speech is a syntactically identical representation of the original speech. Hence, it is "approximative in nature" (p. 216), while they also conclude part of the meaning of the quote is found in its semantics (p. 225). Similar discussions are found in Romaine et al. (1991) and Hasund (2003). This is further discussed in chapter 4.1.2.1.5.

4.1.1.2 Turn Management

Many DMs, including *like* and *you know*, can be used strategically for signaling potential speaker shifts in a discourse (Erman, 2001, p. 1345; Beeching, 2016, p.101). This function has been well-attested in literature regarding DMs. Erman (2001), as written above, counts it as a part of social monitors. *Like* and *you know* can be used as subtle signs for taking, and retaining the floor, but only *you know* can be used for yielding the floor.

The DM occurs utterance-initially when it is used for taking the floor. Erman (2001) suggests that if a plea for the floor were spelt out, it could have been worded something like "Can I butt into the conversation here?" or "Excuse me[,] could I say something[?]" (p. 1345). The speaker wishes to gain the other participants' attention in order to weigh in on certain points of the discourse, or to steer the discourse towards a new topic. The data from the SBCSAE shows that devices for taking the floor often occur in high-paced discourse, and it almost becomes a competition to taking the floor, although they are frequent in in slower-paced discourse as well.

In the case of floor retaining DMs, it could be paraphrased to "Hang on, I'm not finished yet" (Erman, 2001, p. 1345), hence it signals the speaker's wish to add something without interruption. This type of floor management is often seen in relation to discourse disfluencies, and the speaker asks for patience while planning the upcoming utterance. In such

cases, it becomes evident that this type of turn-management and verbal filler (see below) are intertwined.

When DMs are used for yielding the floor, they are located utterance-finally and could be paraphrased into something like “What do *you* think” or “Can I get some response, please” (Erman, 2001, p. 1345). Floor-yielding devices allow for a higher degree of interaction, and is, as the paraphrased utterances suggest, often used if the speaker wishes the addressee to take the floor. The SBCSAE shows that there are two types of floor yielding: a fully legitimate turn, and temporary yields. In the case of the former, the speaker signals that s/he has said everything s/he needed to say in one turn and invites one or several addressees to react to what has been said. In the latter case, the speaker does not wish the addressee to take the floor completely. Instead, the speaker wishes to do a comprehension check (Erman, 2001, p. 1345-1346). In these cases, the addressee typically performs some type of backchannelling to signal that the speaker may proceed, while some type of question is usually uttered by the addressee, who has been given the floor temporarily, that is related to the part that prevents comprehension.

4.1.1.3 Verbal Fillers

Verbal fillers are often found in an environment with other DMs, pauses, and false starts. If the processing only provides minor difficulties, a minor pause and a DM can occur. They are used for filling pauses where linguistic planning occurs to signal to the addressee that the speaker undergoes these cognitive processes and wishes to hold the floor and proceed once the obstacle has been overcome. It is largely accepted that when it comes to lexicalization, the lemma (i.e. the meaning of a word) is retrieved prior to the lexeme (i.e. the phonological form) (Harley, 2016, p. 410, 431). Hence, the speaker quite possibly knows the message of what s/he wants to say, but the issue which causes a disfluency is located in the retrieval of the correct word(s).

It is perhaps not that surprising that *like* and *you know* are two candidates for filling pauses. If word retrieval proves to be difficult, the speaker sometimes abandons the process of retrieving the correct word in order to avoid further stalling of time and chooses a word or phrase which is similar in meaning to the word(s) s/he failed to retrieve. Both DMs can be used as an appeal to the addressee to accept less-than-perfect substitutions of word(s). Here the basic meaning *like* and *you know* also become apparent; *like* can be used for saying “this is similar to what I want to say”, while *you know* can be used for saying “you know the kind of thing I really mean”. In the case of *you know*, a substitution does not always happen because it is sometimes expected that the “you” knows what the addressee wishes to convey, and

sometimes the addressee provides an acceptable word. The problem is sometimes also due to difficulties in planning a new stretch of utterance within a turn, which can result in producing false starts (Harley, 2016, p. 433), which is further discussed in chapter 4.1.2.2.5.

4.1.2 Subordinate Functions

This section will first describe the basic meanings of the two DMs because they are related to the subordinate functions. From this point on, subordinate functions will be referred to as functions, and it these functions which lay the foundation for determining whether people from different social groups use *like* and *you know* differently.

4.1.2.1 *Like*

4.1.2.1.1 Basic Meaning

DM *like* seems to be a bleached form of the lexical preposition *like*; just like the preposition, the DM can be used for meaning to be similar to something. Andersen (2001) also found, as mentioned above, DM *like* to mark non-identical resemblance. He claims:

> [*like*] concerns the relation between the encoded propositional content of the utterance and the underlying belief of the speaker, indicating either the speaker's lack of commitment to the literal truth of P or reduced lexical commitment (Andersen, 2000, p. 35).

Hence, the DM explicitly marks such a discrepancy, which is a plea to the addressee to accept the imprecision, and to not take what has been marked off too literally (p. 185; Schourup, 1985, p. 38)

The corpus material from this study suggests that the speaker does not always use *like* to mark something as a non-equivalence between what the s/he has in mind and what s/he says. Instead, sometimes what is marked can be understood as marking resemblance or equivalence. This was also suggested by Romaine et al. (1991, p. 246) and Hasund (2003, p. 12). Therefore, *like* must also originate from conjunction *like*.

4.1.2.1.2 Approximator

When *like* is used as an approximator, it appears to be an adverb. The token, as the name indicates, expresses that the speaker wants the addressee to understand that what follows the DM, which may be a word, phrase, clause or entire utterance, is a rough approximation, and the speaker shows less commitment to a literal interpretation. Consider the two following examples:

(9a) LAJUAN: [And so I was] **like** four years old.

(9b) LAJUAN: [And so I was] **ø** four years old.

(SBC044 *He Knows*)

In (9a), Lajuan expresses that he cannot remember exactly how old he was when an event happened. He modifies "four years old" by inserting the DM to indicate "four years old" is not to be taken too literally. The function of *like* as an approximator becomes even more apparent when it is juxtaposed to (9b) where the DM has been removed. When the subject predicative "four years old" is not modified, the utterance indicates Lajuan is certain the incidence took place when he had entered his fourth year of living. The juxtaposition also makes it evident that this function challenges the notion that DMs have no affection of the truthfulness of an utterance (see chapter 3.1) given (9b) gives the impression that the utterance is held to be true, while (9a) moderates the truthfulness of the utterance. Similar discussions are found in Andersen (2001, p. 260).

As many other studies on *like* have found (e.g. Beeching. 2016; Schourup; 1985; Andersen, 2001), *like* often modifies numerical expressions. This is perhaps not all that surprising given that we normally quantify what we come across in our lives by intuition, which is not particularly accurate. Additionally, an approximation facilitates to keep the flow of the discourse rather than pausing in search of a more precise quantification in cases where accuracy is not needed. But *like* is also used for modifying non-numerical expressions, though used for the same purpose as shown in (10):

(10) PETE: .. Well at [least] they are uh **like**
already breaded --
ROY: [Hm],
PETE: .. I mean crumbed and in the jar,

(SBC003 *Conceptual Pesticides*)

Just like in (9a) where *like* approximates a numerical expression, *like* in this example modifies what comes immediately after the token. Pete struggles to retrieve the term he is searching for and asks his addressees to accept that he means something similar to "breaded". In this case he later succeeds in retrieving the word he was looking for, namely "crumbed".

4.1.2.1.3 Exemplifier

When *like* functions as an exemplifier, the token can either be substituted with "for example" or "such as", as noted by Schourup (1985) or the utterance can be slightly paraphrased to mean such a thing. In many instances *like* introduces an elaboration to support the speaker's

claims. What follows *like* represents only a subset of a pool things. (11) provides an example of exemplifying *like*:

(11)
```
JOANNE: In fact,
        [I eat] .. stuff that he doesn't eat,
KEN:    [<X I don't know X>].
JOANNE: that,
        he wouldn't dare touch.
        you know,
        I eat like a ceviche,
        and,
        and,
        all kinds of salads,
        and,
LENORE: <L2 Gua[ta] L2>.
JOANNE:        [(H)] ... I- I eat all kinds of stuff like that.
```
(SBC015 *Deadly Diseases*)

First Joanne claims she eats things Ken does not eat. To support her claims, she exemplifies some of the foods she eats that he does not. In this example it also becomes obvious that the examples she gives only represent a subset of what she, and not Ken eats by ending the utterance by saying "I eat all kinds of stuff like that". The token can also be used in questions to indicate that the addressee wants the speaker to give examples to the previous statement to get a better understanding.

(12)
```
ROY:            ... Have you heard about these horrific ... uh new..
                genetic ... developments,
                in .. in .. um ... food livestock?
PETE:           .. Y_no.
                .. [Like what].
ROY:               [Like they're],
                .. they're @trying @to bree=d,
                .. <VOX ho ho= VOX>,
                (H) .. they're trying to breed like a forty foot
                long tube chicken?
```
(SBC003 *Conceptual Pesticides*)

In (12), Pete denies having heard about new developments in genetic manipulation and asks Roy if he can give an example of one of these developments, which then follows with a humoristic twist. Examples (11) and (12) show that in essence, what differs between the two examples is that the first example is declarative, while the second is interrogative. *Like* itself holds the same function in both cases.

4.1.2.1.4 Focuser

Focuser *like* is used when the speaker wants to highlight, and to make the addressee to pay close attention to the following item. When used as a focus marker, *like* can be used for both contrastive (14) and information focus (13) (Wise, 2011, p. 1001). Underhill (1988), as

previously mentioned, described focus *like* to mark "the most significant new information in a sentence—often the point of the sentence" (p. 238). This is in accordance with information focus. Underhill's examples and his analyzes of them show the new information can be contrastive.

(13)

[LAJUAN:]	Now=, (H) he calls me and tells me that he wants to be with me. ... And, .. he called -- **Like** he called me last week, and said he wanted to be with me.
CAM:	... You're kidding.

(SBC044 *He Knows*)

(13) shows an example of information focus. Lajuan talks about past affairs he had with a man named Darren who is now married to a woman. In this excerpt, Lajuan first says one time that Darren has said to him that he wants them to be romantically involved, which evokes no reaction from Cam. Lajuan makes a new attempt at conveying the magnitude of the information. He begins by saying "he called", before stopping, which might be related to him being afraid that he will not succeed in evoking a reaction this time either, thus deciding to rephrase himself by adding *like* in order to properly highlight the upcoming information. In this attempt he succeeds, which is shown by Cam's response. In some respects, this function resembles *you know* when it is used for reassuring the addressee (see chapter 4.1.2.2.3); in both cases a DM is used for expressing some degree of confidence with regards to the following.

(14)

[CORINNA:]	... in .. Islam you're cons- .. not considered .. Muslim anymore.
PATRICK:	... Oh[=]?
CORINNA:	[If] you have anal sex. ... But I remember like uh, ... ~Lamar telling me once that, he had a girlfriend? ... Back home? .. (H) A=nd she didn't wanna lose her virginity? ... They were gonna have sex, so it's **like** instead, (H) they always had anal sex?

(SBC045 *The Classic Hooker*)

(14) provides an example of contrastive focus. First, Corinna introduces background information claiming Islam says one can no longer be considered Muslim after participating in anal sex. She then talks about a Muslim woman who did not want to lose her virginity on religious terms and instead decides to have anal sex. By inserting *like* in this case, Corinna

explicitly marks the contrast between where she explains how a girl on one point obeys Muslim doctrine by remaining a virgin and another point where she disobeys the Muslim doctrine. It is worth noting in (14) that the focus is marked by an *it's like* construction rather than the bare *like*. This construction is frequently used when the speaker wishes to focus an element. The pronoun *it* appears to have a vague, rather unspecified reference, which was found by Andersen (2001) and Hasund (2003) as well. The presence of the pronoun does not alter the function of *like* although it increases salience.

4.1.2.1.5 Quotation

When *like* is used for signaling an upcoming quote, it shows a more versatile nature than many other quotation markers found in English as well as other languages. It can quote a previous utterance of the speaker or someone else, a though, a reaction, and gestures. Direct quotations cause the quote to be delivered in a lively, theatrical manner (Buchstaller & Van Alphen, 2012, p. xv; Romaine et al.,1991, p. 266). This becomes obvious when juxtaposed to a marker of indirect speech. See (15) and (16):

(15) [PAMELA:] ... And then she **said that** she dreamt about,
(H) u=m,
.. all of her relatives,
... that had died.
... She wasn't dreaming about anybody
who was living,
but who had die=d.

(SBC005 *A Book about Death*)

(16) [CORINNA:] ... She**'s like**,
oh my God I feel so dirty whenever I do it,
and it hurts so fucking bad.

(SBC045 *The Classic Hooker*)

In (15), Pamela reports what her grandmother had told her right before she died. The manner in which she talks when she introduces the quote does not differ in any way from how she articulates herself outside of the quote, and it does not belong to a separate tone unit. Moreover, she does not draw attention to the quote. Corinna in (16), on the other hand, uses a mocking tone when quoting a girl. She also makes a tiny pause between *like* and the quote, thus marking the upcoming quote to a higher degree than what happens in (15), and she is clearly separating the speech of her friend from that of her own.

(17) [LAJUAN:] and I**'m like**,
... okay,
can I afford to move out here,
get a apartment,

```
until I fi=nd --
```

(SBC044 *He Knows*)

In (17), Lajuan is talking about his time in college when a friend with a stable income asked Lajuan to move closer to him. (17) displays what he was thinking about with regards to the moving proposal. When *like* quotes a thought, it does not necessarily represent inner monologue, but, like in this example, what is quoted is likely a representation of the lines of things that were considered with regards to the given situation. Just like in (16), the quote belongs to a separate tone unit.

As briefly discussed in chapter 4.1.1.1, quotative *like* marks both speaker certainty and uncertainty with regards to the quote. Previous experiments (e.g. Sachs, 1967; Potter & Lombardi, 1998) have shown the *form* of a sentence does not tend to enter the long-term memory and is therefore forgotten within seconds after exposure. The *meaning*, on the other hand, is retained in the long-term memory. The quote is not a verbatim representation of the reported speech, but the speaker is, nevertheless, confidently reporting the meaning conveyed by the original utterance, and quotation *like* is therefore in accordance with the basic meanings although this is also true for other DMs marking direct reported speech. In fact, there are claims that quotative *like* is undergoing a process of grammaticalization and that the basic meaning is fading (Romaine et al., 1991).

4.1.2.2 *You Know*

4.1.2.2.1 Basic Meaning

With the exception of Macaulay (2002), who claims neither *you*, nor *know* of the DM retains any basic meaning, it is commonly accepted among researchers that *you know* is an addressee-oriented DM (e.g. Fox Tree et al., 2002; Stubbe & Holmes, 1995; Beeching 2016) due the second person pronoun it contains. Thus, the speaker explicitly signals a desire to make the addressee an active part of the discourse, both in the sense of contributing to the discourse, and inviting him/her to reflect upon the marked item. The meaning of the token is also rooted in the cognitive verb *know*. The data from the SBCSAE shows a speaker who employs *you know* does not necessarily assume the speaker has previous knowledge of the upcoming information. Instead, what the speaker marks as knowledge is more fluent on the timeline; it may be a piece of information the addressee already possesses, or it may be knowledge the speaker wants the addressee to have for future references. Hence, the knowledge may be close to the time of reference e.g. referring back to information uttered previously in the discourse, or in relation to an adjustment within an utterance such as for

elaborative or repairment use where the speaker pleads for an understanding of the ongoing unclarity of the utterance (Beeching, 2016, p. 98), or the knowledge might be more distant in time of the employment of the token, e.g. bringing up a past event, or wanting the addressee to store the knowledge in their long-term memory. When talking about knowledge in relation to the DM, it is important to note that the speaker does not wish the addressee to understand the utterance as dogmatic knowledge, but rather to let the addressee consider the proposition.

Fig. 4.1 is a modification of Schiffrin's (1992, p. 268) matrix. The matrix is a simple, yet useful illustration of how knowledge is tied to the DM. When *you know* is used for marking previous knowledge, either (a) or (c) is the case; the speaker's assumptions are correct, as exemplified in (18), or they are not, as shown in (19). (b) and (d), on the other hand, are instances of reassuring the addressee. When *you know* is used for elaboration and self-repair, the token does not fit into fig. 4.1 because then it is concerned with how the meaning is conveyed. *You know* in these cases signal that the addressee is about to know what the speaker meant.

		Does the speaker assume the addressee knows of X?	
		Yes	No
Does addressee know of X?	Yes	(a)	(b)
	No	(c)	(d)

Fig. 4.1

Relation between *you know* and knowledge

4.1.2.2.2 Previous Knowledge

When *you know* is used for marking a segment of previous knowledge, the speaker is fairly certain the addressee already knows the information of the marked segment. When the token is used for this purpose, its pragmatic meaning is close to the semantic meaning of the S-V construction *you know*, and Holmes (1986) claims *you know* in this function can be paraphrased into "I'm confident you know the kind of thing I mean" (p. 9). The knowledge on the addressee's behalf may have been acquired from a previous event which the speaker and addressee have mutual knowledge, or the information may be of a kind which is held as common knowledge in the community in which the speaker and addressee are members.

(18)

MILES:	.. And in high school, **you know,** they teach about it.
PETE:	.. Mhm.

In (18), Miles talks about how he does not understand why so many people refuse to protect themselves against HIV, and that they should know about the consequences because this is taught in high school. There is no reason for him not to assume that his addressees do not know this, and Pete even confirms that he knows. (18) is, therefore, an example of scenario (a) described in Fig. 4.1.

(19)

[PAMELA:]	(H) here's Betty [*sic*] Davis. I mean,
DARRYL:	<VOX Sa=rah= VOX>?
PAMELA:	<X<@ z- yeah @>X> -- (H) This incredible ... film legend. ... And we think of her in Jezebe=l, we think of her, (H) **you know,** smoking (H) .. cigarette smoke into the faces of .. William Holden and, (H) and the like.
DARRYL:	I don't, I've never seen those movies.
PAMELA:	... (TSK) You've never seen Betty [*sic*] Davis movie?

(SBC005 *A Book about Death*)

In (19), Pamela has just started to talk about a Bette Davis movie she watched the night before, and her fascination with the actresses' work. Pamela assumes Darryl is familiar with her work. She first lists her act in *Jezebel* before she claims he as well has knowledge about her iconic smoking scenes. When Darryl rejects to have knowledge about the matter, Pamela expresses surprise to learn that her assumption of his familiarity with Davis' works is incorrect. Relating back to Fig. 4.1, (19) is an example of scenario (c). From *you know* in (19), it becomes apparent that when the token is used for marking previous knowledge, it shares some similarities with *like* when used as an exemplifier; Pamela gives an example from a larger set of things people associate with Bette Davis. The difference between *you know* and *like* in such cases is the former's focus on past knowledge the parties of a discourse share, while the latter does not presuppose any prior knowledge. As the two examples of *you know* introducing previous knowledge, the addressee often confirms or denies having knowledge of the marked item.

4.1.2.2.3 Reassuring the Addressee

Fig. 4.1 from chapter 4.1.2.2.2 showed regardless of reality, the speaker assumes the addressee has knowledge of X and the function was fitted with either scenario (a) or (c). However, when *you know* is used for reassuring the addressee, the token fits with scenario (b) and (d); the speaker makes no assumption that the addressee has knowledge about the marked item (Jucker & Smith, 1998, p. 191; Holmes, 1986, p. 8; Erman, 2001). However, it only means the speaker makes no assumption about the addressee's knowledge, not necessarily that the speaker assumes the addressee has no knowledge about the matter. Used for reassuring the addressee, *you know* emphasizes the speaker's knowledge, certainty, or conviction of the highlighted element, which has led previous researchers to name this function of the DM a disjunct marker (Goldberg, 1980 in Schouroup, 1985, p. 108). Used in this function, *you know* can be used ironically where the speaker pretends to be convinced of something but does not mean it. Holmes (1986) suggests that when the token serves this function, it "can be paraphrased as 'let me assure you'" (p. 8).

(20) [MARY:] ... I saw my .. my speedometer just go Brr=.
.. like that just dow=n.
(H) **You know,**
and I knew exactly what it was.

(SBC007 *A Tree's Life)*

In (20), Mary talks to Alice about when she was out driving, and the engine stopped working. She reassures Alice about this because she had no prior knowledge of Mary being able to detect what was wrong with the car. This is also something she takes pride in and wants Alice to incorporate this achievement into her knowledge.

In (21), Alina talks about a friend of a friend who is a cocaine-abusing lawyer. In this excerpt there is less, though some focus, on having the addressee acquire knowledge about the highlighted item than what was seen in (20). Here, Alina focuses on her conviction that the lawyer is no good, and she wants the addressee to know how she feels about him.

(21) [ALINA:] ... **You know,**
he's just,
... (H) no good.

(SBC006 *Cuz*)

4.1.2.2.4 Elaboration

In cases where *you know* is used in connection with elaboration, it is because the speaker knows the meaning of the message was only vaguely encoded (Holmes, 1986, p. 10). When

the token is used for elaboration, the speaker employs it for one of two reasons; the speaker transitions from the vague utterance into a specification of the previous utterance (Erman, 2001, p. 1343), or it can be used for securing that the addressee has comprehended the vaguely delivered message. For the latter case, the SBCSAE shows that *you know* is a chance for the addressee to ask for an elaboration. The speaker has the chance to proceed if the addressee signals that an elaboration is not needed. These signals are usually realized by some type of backchannelling.

(22)
```
[PETE:]    ... the rainy season was all off,
           .. you know,
           it rained during the dry season,
           and was= .. dry during the rainy [season],
```
(SBC003 *Conceptual Pesticides*)

In (22), Pete says that the rainy season did not follow the normal pattern. He the elaborates on what was different for this particular rainy season compared to what the weather is typically like during this season. As the example show, the transition from the vague message into the elaboration is marked by *you know*.

(23)
```
JILL:   .. (H) I think .. there'd be a slight chance,
        .. of it being a false .. negative.
        .. (H) But,
        .. I don't think so,
        cause I'm pretty late?
        And I think I'm late enough,
        (H) where I would have,
        .. like,
        .. enough of .. the hormone that .. the pregnancy test tests
         for?
        (H) I think I would have enough of that in my urine,
        that .. of course it would show up.
        .. if I had any in there?
JEFF:   .. [Yeah=]?
JILL:         [(H)] You know?
```
(SBC028 *Hey Cutie Pie*)

In (23), Jill tells Jeff she just took a pregnancy test which yielded a negative result. First the excerpt shows Jill opens up for the chance of it being a false negative, before claims she believes it is unlikely. What follows next is her argumentation for doubting she could still be pregnant. When *you know* appears in the last line, without an actual elaboration, she gives Jeff the opportunity to ask for clarification of the proposition. When an elaboration does not follow the token, it is used as a comprehension check, but the speaker is fairly certain that the

addressee follows the logic of what is being said, which is different from the cases when an elaboration immediately follows the token.

4.1.2.2.5 Self-repair

Spoken conversations are usually not pre-planned, and as a result of its spontaneous nature, repairs are frequent. (Fox Tree, 1995, p. 710 and cited studies therein). *You know* is employed when the speaker wishes to edit something that was said just prior to when the repair begins. A few different types of repair have been identified in this study. One is recycling where a part of the reparandum (i.e. the error in the discourse) is recycled by repeating a part of the reparandum in the repair (Fox, Maschler & Uhlmann, 2010), as exemplified in (24), and replacement (Fox et al. 2010) where a simple replacement of a word, or another small unit occurs. In the current paper, replacements will include false starts i.e. the speaker makes a syntactic turn when moving from the reparandum to the repair (Holmes, 1986, p. 11). The data from the SBCSAE indicates that when *you know* is used for marking the transition from reparandum to repair, the repair occurs because of the speaker's awareness that an error occurred, and repairs introduced by *you know* are never initiated by the addressee.

(24) [ALINA:] .. They go through this stupid fire science major out at UCLA.
Become firemen.
(H)= And now he- --
%_**You know,**
and now he works for the phone company

(SBC006 *Cuz*)

In (24), Alina recycles "and now he". The first time she utters this, *he* is stretched and is pronounced with rising intonation. She repairs the prosody, and in the repair, the recycle is embedded in the same tone unit as the rest of the utterance in the last line of (24).

(25) [JAMIE:] it's for people who can't move their hips.
... Right?
I mean it's for basic,
.. **you know,**
this is a beginning lambada class,

(SBC002 *Lambada*)

Just prior to Jamie's utterance in (25), her friend, Miles talks about how the lambada class he takes is very different from how Brazilians dance, which causes her to ask whether it is a beginner's class. From the reparandum "it's for basic" and the repair "this is a beginning lambada class" as well as the first line, it is obvious that she tries to convey that one cannot compare Miles' dancing classes taught in America to how Brazilians dance. The major thing

that changes is her wording. Based on the transcript it is impossible to say why this repair occurs, but one possible explanation could be that she is not able to retrieve a suitable NP that "basic" can modify, so instead of stalling for longer than what she already has, she abandons this route in favor of what becomes the repair. Another possible explanation is that she realizes that calling his dance class "basic" could be taken as an offence and changes her wording to a more neutral statement.

4.2 Statistics

In total, 716 instances of *like* (excluding instances put either in the rag bag or analyzed as a non-DM) were analyzed. This token was used with a frequency of 8,23 per 1,000 word by the speakers in my subcorpus of the SBCSAE. *You know* was less used, and 563 instances of the token when serving one of the four pragmatic functions were analyzed. It had a frequency of 6,47 per 1,000 words.

Table 4.1 shows how men and women use *like* with regards to the different functions it serves, the frequency of all four functions combined, the frequency of when *like* is not a DM, and the frequency of the instances which for various reasons had to be put in the rag bag. As Table. 4.1 displays, women are more frequent users of all the four functions of *like*. The p-values, however, show that none of the results are close to being statistically difference. Fig. 4.2 presents clustered bar charts of the mean frequencies of the different categories of *like* according to gender. Several of the standard error bars show great spead from the mean.

	n men	**Men's frequency**	**n women**	**Women's frequency**	**P-value**	**Type of test**
Approximator	25	0.71	79	1.53	0.077	Mann-Whitney U
Exemplifier	29	0.82	57	1.11	0.219	Mann-Whitney U
Focuser	130	3.67	238	4.62	0.660	Mann-Whitney U
Quotation	41	1.16	117	2.27	0.303	Mann-Whitney U
All functions	225	6.35	491	9.53	0.503	Mann-Whitney U
Rag Bag	51	1.44	127	2.46	0.258	Mann-Whitney U
Non-DM	105	2.96	151	2.93	0.757	Mann-Whitney U

Table. 4.1
Distribution of *like* according to gender

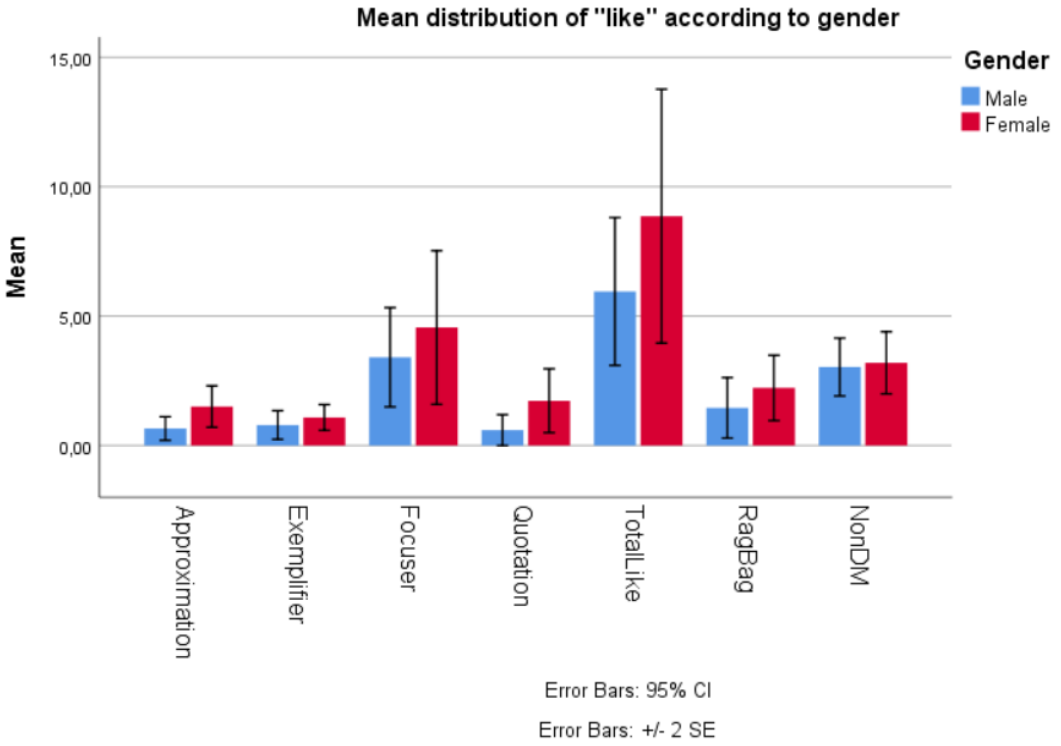

Fig. 4.2.
Mean distribution of *like* according to gender

	n men	**Men's frequency**	**n women**	**Women's frequency**	**P-value**	**Type of test**
Previous Knowledge	22	0.62	27	0.52	0.935	Mann-Whitney U
Reassuring the Addressee	137	3.87	212	4.11	0.613	Mann-Whitney U
Elaboration	35	0.99	39	0.76	0.660	Mann-Whitney U
Self-repair	25	0.71	66	1.28	0.163	Mann-Whitney U
All functions	217	6.18	344	6.68	0.351	Mann-Whitney U
Rag Bag	29	0.82	30	0.58	0.935	Mann-Whitney U
Non-DM	28	0.79	65	1.26	0.386	Mann-Whitney U

Table 4.2
Distribution of *you know* according to gender

Table 4.2 displays the distribution of *you know* according to gender. When comparing the different frequencies displayed in the table, it is evident that men and women use the token for all the purposes with very similar frequencies. As was the case for *like*, there is no statistical difference between the different frequencies between men and women. The standard error bars from fig. 4.3 reveal that there is great spread from the mean frequency of *you know* as well.

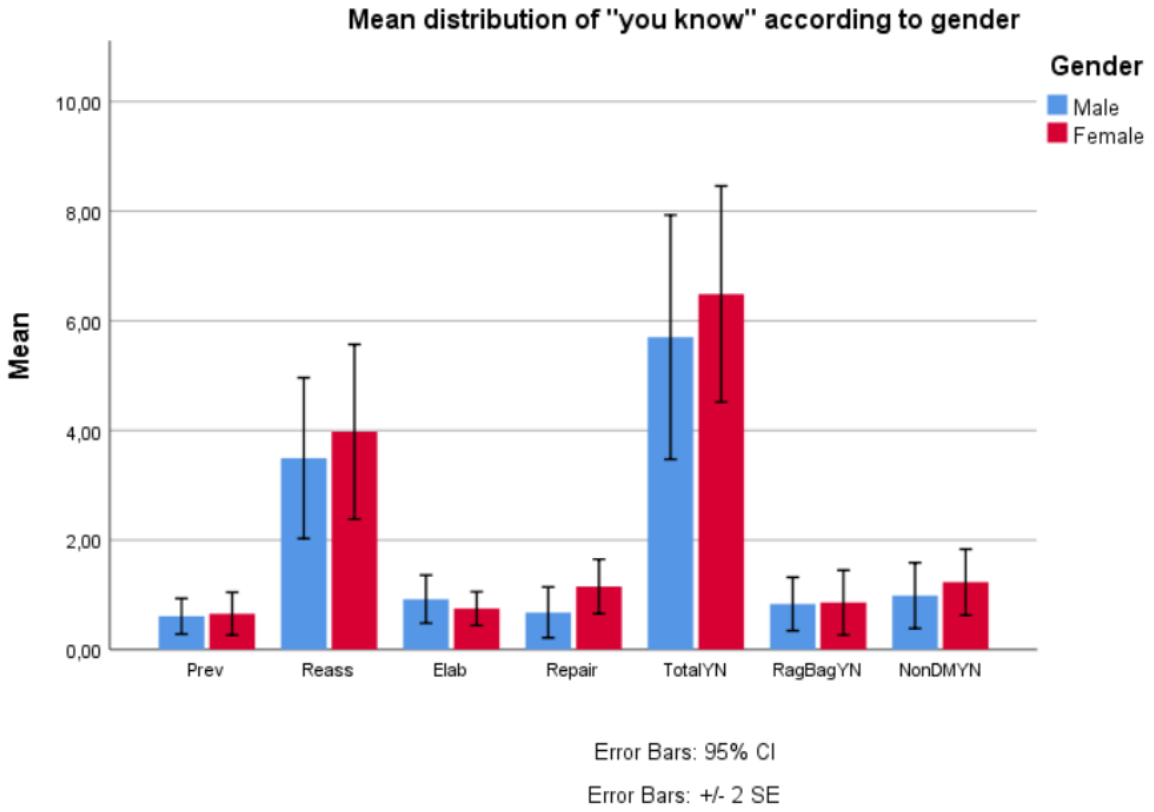

Fig. 4.3.
Mean distribution of *you know* according to gender

The distribution of the different categories of *like* according to educational level is presented in table 4.3. People of lower education appear to be more frequent users of *like* in all the categories. When all four functions of *like* are added together, it is clear that people of lower education are frequent users of DM *like*, and the frequency is more than double the frequency for people of higher education. However, none of these results, as the table reveals, are significantly different. The error bars from the overall frequency of DM *like* referred to as *total like* in fig. 4.4 show there is great overlap in the standard error bars, which is an indication of why the difference between higher and lower education is insignificant.

	n Higher Ed.	Freq. Higher Ed.	n Lower Ed.	Freq. Lower Ed.	P-value	Type of test
Approximator	41	0.76	56	2.16	0.233	Mann-Whitney U
Exemplifier	44	0.82	32	1.24	0.416	Mann-Whitney U
Focuser	150	2.78	191	7.38	0.441	Mann-Whitney U
Quotation	84	1.56	62	2.40	0.819	Mann-Whitney U
All functions	319	5.92	341	13.18	0.546	Mann-Whitney U
Rag Bag	41	0.76	113	4.37	0.233	Mann-Whitney U
Non-DM	160	2.97	83	3.21	0.574	Mann-Whitney U

Table 4.3
Distribution of *like* according to educational level

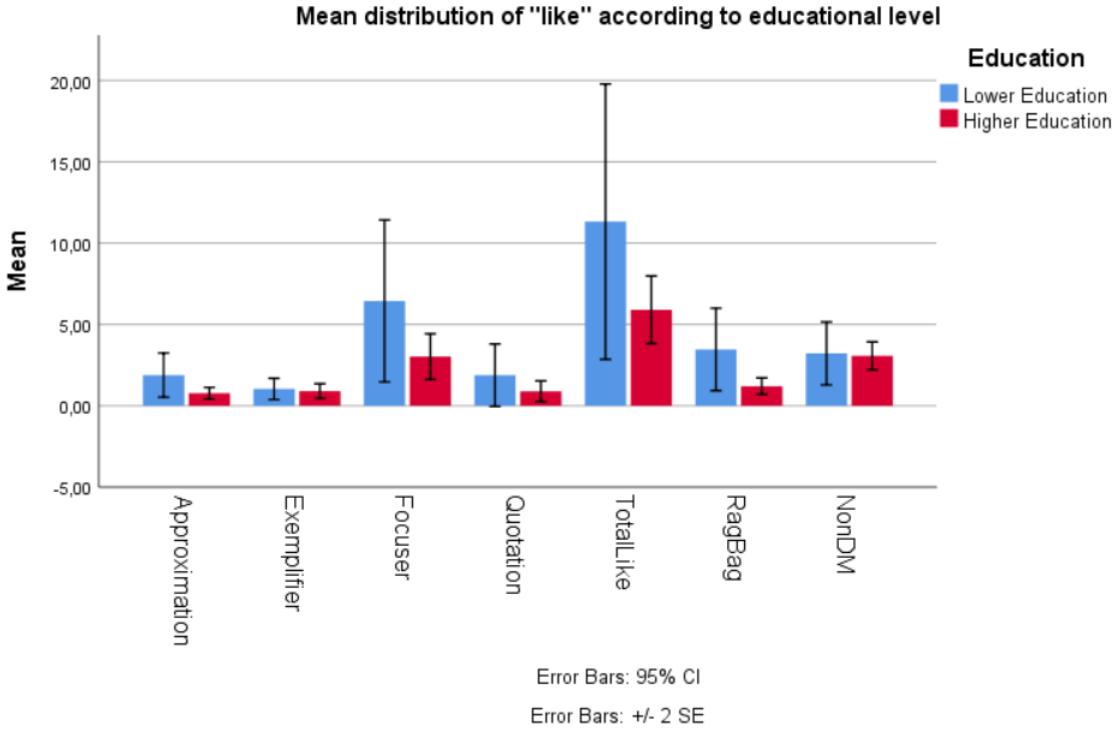

Fig. 4.4
Mean distribution of *like* according to educational level

Table 4.4 reveals that apart from when *you know* is used for marking previous knowledge, people of lower education are more frequent users of the individual functions of

	n Higher Ed.	Freq. Higher Ed.	n Lower Ed.	Freq. Lower Ed.	P-value	Type of test
Previous knowledge	64	1.19	26	1.00	0.250	Mann-Whitney U
Reassuring the Addressee	156	2.90	123	4.75	0.053	Student's t-test
Elaboration	34	0.63	31	1.20	0.233	Mann-Whitney U
Self-repair	41	0.76	38	1.47	0.348	Mann-Whitney U
All functions	295	5.48	218	8.42	0.047 *	Student's t-test
Rag Bag	29	0.54	28	1.08	0.250	Mann-Whitney U
Non-DM	53	0.98	37	4.43	0.983	Mann-Whitney U

Table 4.4
Distribution of *you know* according to educational level

the token. There was no difference of statistical significance for any of the functions although the p-value of reassuring the addressee is very close to the 0.050 margin. As marked by the asterisk, the frequency of the overall use of DM *you know* is significantly different between people of higher and lower education, and people of lower education are more frequent users.

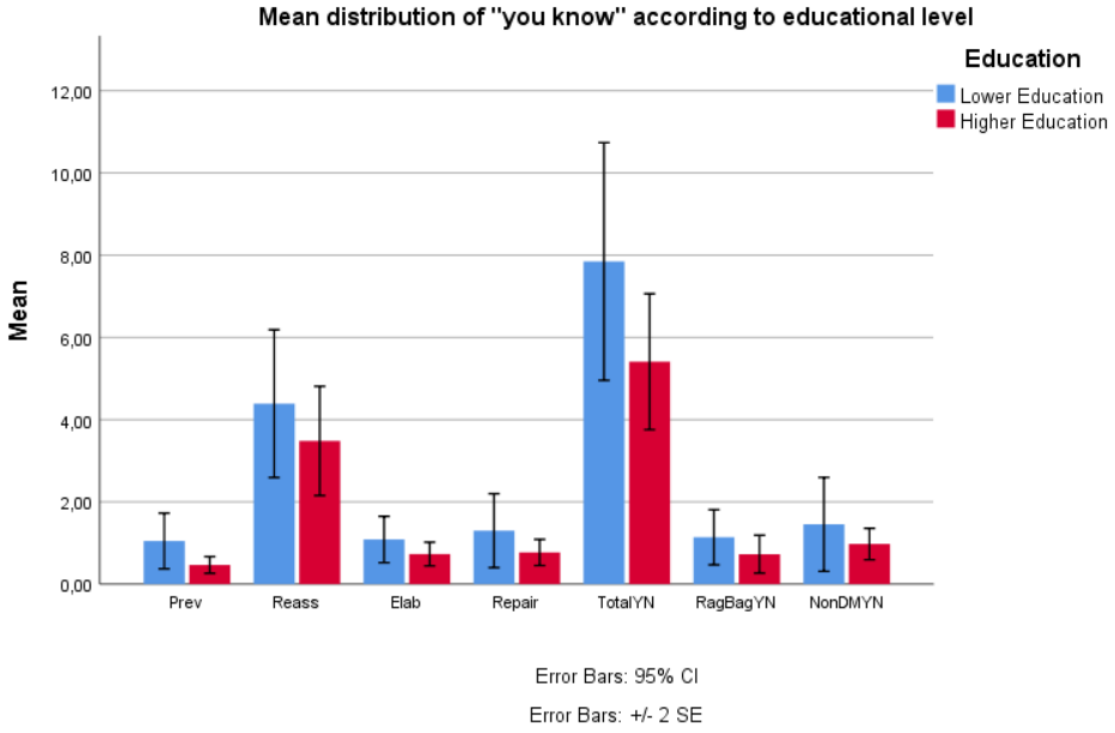

Fig. 4.5

Mean distribution og *you know* according to educational level

The p-value shows this is a small significant difference. The error bars from fig. 4.5 reveal that just like the other results there are great individual differences in the mean frequency of the different categories. They also show that there is some overlap between people of higher and lower education for both *you know* for reassuring the addressee and when the token is used as a DM at large. This is indicative of why these results have a p-value close to 0.050.

The results above only display the distribution when the participants of the study have been categorized according to gender and educational level separately. What follows below is the result of the interaction of gender and educational level. When dividing into four groups, there is very little data for each group. Therefore, there results have not been tested for statistical significance. What these results serve to do then is to provide more information about the social background of the participants that are behind the results reported above. Table 4.5 displays how *like* has been distributed among the four groups. Women of lower education are the most frequent users of *like* in all its categories

	n men higher ed.	**Frequency men higher ed.**	**n men lower ed.**	**Frequency men lower ed.**	**n women higher ed.**	**Frequency women higher ed.**	**n women lower ed.**	**Frequency women lower ed.**
Approx-imation	19	0.82	3	0.30	22	0.71	53	3.34
Exemplifier	20	0.87	5	0.52	24	0.78	27	1.70
Focuser	90	3.90	35	3.50	60	1.95	156	9.82
Quotation	40	1.74	1	0.10	44	1.43	61	3.84
All functions	169	7.33	44	4.40	150	4.87	297	18.69
Rag bag	16	0.69	27	2.70	25	0.81	86	5.41
Non-DM	83	3.60	17	1.70	77	2.50	66	4.15

Table 4.5
Distribution of *like*

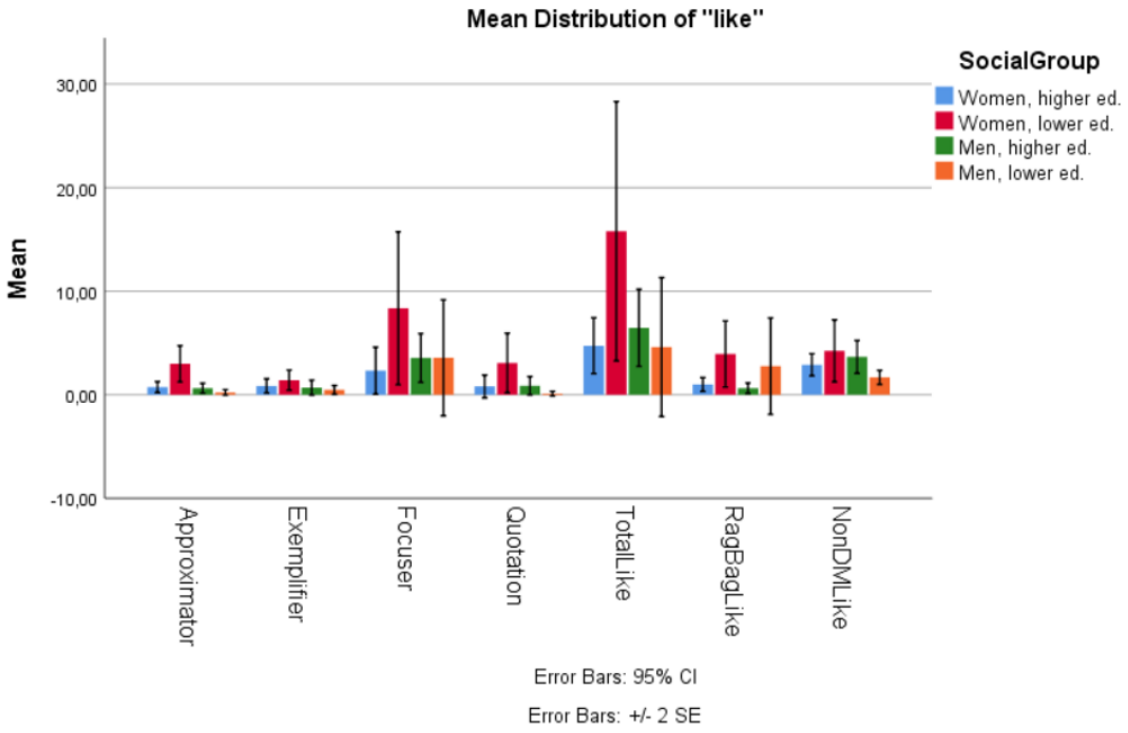

Fig. 4.6

Mean Distribution of *like*

while there is less variation among the other three groups although men of higher education's frequency of *like* as a DM without specifying the function is somewhat higher than the frequencies for men of lower education and women of higher education. The standard error bars from fig 4.6 shows that, like all the previous results, there is great variance within these groups. To demonstrate this, a comparison between two speakers of the same group is helpful. Corinna is a woman of lower education. She utters 2,176 words and uses *like* as a DM 40.44 times per 1,000 words, while Mary, who is also a woman of lower education, utters 1,492 words, but she only uses DM *like* 2,68 times per 1,000 words.

	n men higher ed.	Frequency men higher ed.	n men lower ed.	Frequency men lower ed.	n women higher ed.	Frequency women higher ed.	n women lower ed.	Frequency women lower ed.
Previous knowledge	11	0.48	11	1.10	12	0.39	15	0.94
Reassuring the addressee	52	2.26	70	7.00	104	3.45	53	3.34
Elaboration	17	0.74	17	1.70	17	0.56	14	0.88
Self-repair	6	0.26	17	1.70	35	1.16	21	1.32
All functions	86	3.73	115	11.50	168	5.45	103	6.48
Rag bag	12	0.52	17	1.70	17	0.55	11	0.69
Non-DM	16	0.69	10	1.00	37	1.20	6	1.70

Table 4.6
Distribution of *you know*

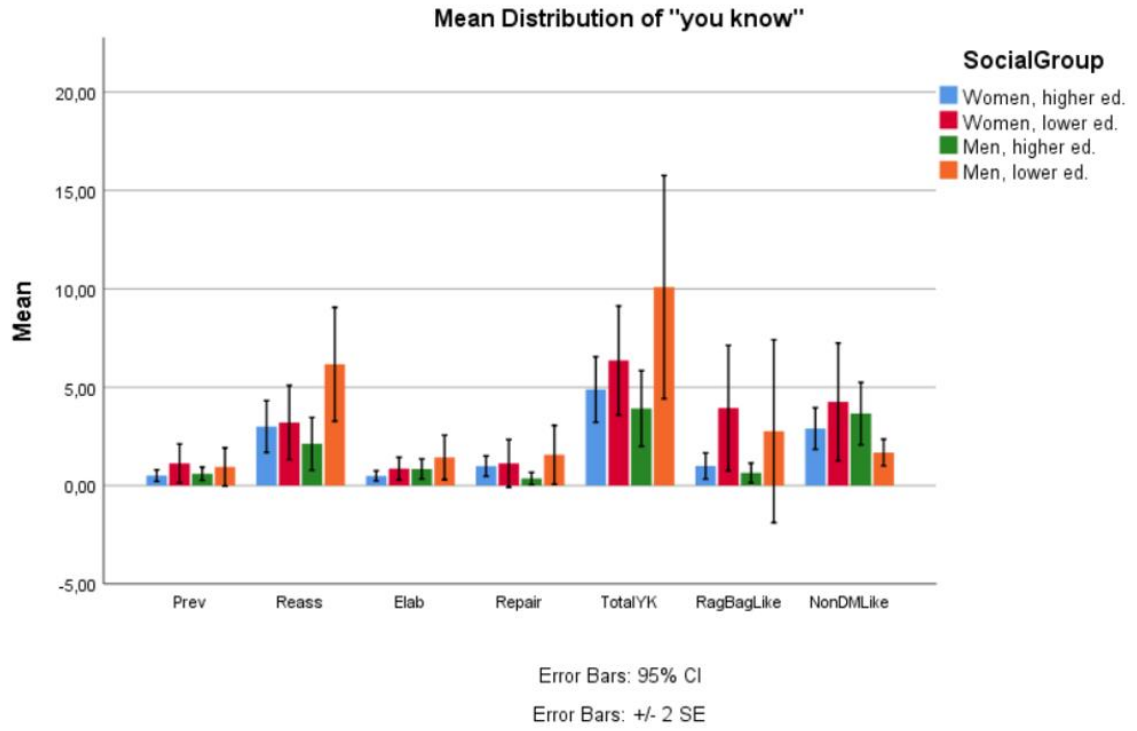

Fig. 4.7

Mean distribution of *you know*

In table 4.6, we find that men of lower education use DM *you know* with a higher frequency than the other groups, which leads frequency of DM *you know* at large to be fairly high compared to the other groups. *You know* used for reassuring the addressee was, as displayed above, close to being used significantly more by people of lower education. Table 4.6 reveals men of lower education was a bigger contributor to the higher frequency than women of lower education. It also shows that men of higher education used it a little less

frequently than women of higher and lower education, who, in turn, did not differ in frequency for this function. The standard error bars from fig. 4.7 reveal there is the is almost no overlap between men of higher and lower education when *you know* is used for reassuring the addressee. Moreover, there is only some overlap between the standard error bars of men of lower education and women of higher and lower education. For the total frequency of the DM *you know*, there is also only some overlap in the standard error bars between men of lower education and the other groups. The standard error bars also reveal it is estimated that there is great spread around the means. As was done with *like* above, two people of the same group will be compared. Michael and Andrew are both men of higher education. In total, Michael uttered 1,479 words and used *you know* as a DM 10.82 times per 1,000 words, while Andrew, who uttered 1,333 words used the token as a DM 2.25 times per 1,000 words. Even though they are of the same group and utter nearly equally many words, Michael uses the DM nearly five times as frequently as Andrew.

4.3 Summary

In this chapter the results from both the qualitative and the quantitative part of the study have been presented. It was first discussed that *like* and *you know* serve functions on two different levels. The superordinate functions are the same for the two DMs. On this level, they can be used as hedges where the DM is inserted so signal that the speaker is distancing himself from an assertion. It can also be used as a booster to signal full commitment and the speaker can vouch for the validity of the assertion. Moreover, in its epistemic nature, a token can simultaneously be used as a hedge and a booster. On this level, the two DMs can also be used for managing the floor. It can be used for holding, taking, or yielding the floor, although the latter is only possible for *you know*. The last function on this level is that of a verbal filler. This is to signal that the speaker is undergoing difficulties in retrieving linguistic elements, and that s/he has not abandoned what s/he has begun to utter but wishes to continue once the retrieval has been successful.

Like seems to originate from preposition and conjunct *like*. Its subordinate functions are closer related to these origins than what is the case for the superordinate functions. On the subordinate level, *like* can be used to signal an approximation of either a numerical expression, or something that is close in meaning to what the speaker wants to covey. It can also be used to signal that what is proposed is only one of more examples, focusing elements of particular importance to make them salient to the speaker, or for introducing reported speech, which causes what is being reported to be delivered more theatrically.

You know finds its basic meaning in the second person pronoun *you*, which signals an orientation towards the addressee and in the cognitive verb *know*. When these two words are put together as a DM, *you know* signals that the addressee already knows, or that the speaker wants the addressee to know something. On the subordinate level, *you know* can be used to mark knowledge which the speaker assumes the addressee possesses. The token can also be used when the speaker does not assume the addressee has knowledge of the marked segment of the discourse. In these cases, the token is used for letting the addressee know about the speaker's knowledge, certainty or conviction of the proposition. *You know* can also be used in relation to vaguely delivered messages; the token is either used for marking a transition into an elaboration or clarification, or the speaker uses *you know* to ask the addressee whether the message was clear enough to proceed or if an elaboration is needed. Lastly, *you know* can be used for editing part of an utterance. *You know* is in these cases used as a transition from a reparandum to a repair.

The comparison of men's and women's use of *like* showed that the women's frequencies were a little higher than those of men for all four functions. However, none were significantly different, and neither was the overall frequency of *like* as a DM. Men and women did not differ noticeably with regards to their employment of *you know* and the small difference was insignificant. The frequency of how *like* was used appeared be of greater different among people of higher education and people of lower education than what was the case when comparing men and women, but neither these results proved to be significantly different. There was no significant difference between people of higher and lower education with regards to *you know* being used in relation to previous knowledge, elaboration, and self-repair. When the token was used for reassuring the addressee, people of lower education used the token more frequently. However, the p-value of 0.053 is minimally higher than what had been decided as the lower line of significant difference. Out of all the compared results of the study, only the total use of *you know* as a DM proved to be statistically different when comparing the discourse of people of higher education and lower education, and that the latter group uses it more frequently although it must be stressed that this is a small significant difference. The error bars reveal that the spread among the people within the different the social groups is large for all the quantitative results.

5 Discussion

5.1 The Qualitative Part of the Study

The findings from this study suggest the speakers of the SBCSAE use up to four different functions of *like* namely approximator, exemplifier, focuser, and quotation. Likewise, four distinct functions of DM *you know* were identified, namely previous knowledge, reassuring the addressee, elaboration, and self-repair. As Fischer (2006) claims, there is little agreement among scholars when assigning DMs with specific functions, which makes it difficult to compare studies. There is, nevertheless, much common ground (p. 430). Although some of the functions have been given different names than what other scholars have given the functions of the DMs in their respective studies. Parts of the definitions in the current study are found in one or more previous studies. The similarities will not be further discussed. Rather, differences between the current study and previous studies will be addressed below.

Andersen (2001), as written above, refused to acknowledge *like* as marking focus despite what other scholars claim (e.g. Underhill, 1988; Miller el al., 1995; D'Arcy, 2007). His argument was that what had been considered a focus marker should instead be interpreted as the speaker creating distance between himself and the assertion. Focuser *like* was the most commonly used function of *like* in this study and the corpus material did usually not suggest distance between the speaker and the assertion when *like* marked focus. In fact, focuser *like* was usually combined with the superordinate function as a booster, meaning the speaker commits to the utterance. It is possible the young speakers he studied sometimes use *like* to signal they have not yet "grown into" certain linguistic expressions. However, Hasund (2003), who used the same corpus as Andersen in her study, concluded that *like* marks focus on the textual level. Andersen wrote *like* can collocate with metaphors and hyperboles to mark a less-than-literal representation. One example he gives of a metaphor is rendered in (26):

(26) `Erm, and, yeah two birds I met in Portugal and and then Kathy just` **`like`** `stormed out.`

(Andersen, 2001, p. 237)

(26) is just one of many examples of collocations between *like* and metaphors/ hyperboles in his study. There is, therefore, no doubt these collocations exist. However, I would argue that it is implicit that metaphors and hyperboles are less-than-literal expressions, and there is no reason why the speaker would need to flag it as such. It is my understanding that *like* in (26) draws focus to the manner in which Kathy left, which indicates she was not happy about the

two Portuguese girls the speaker talks about. This is the main point of the utterance, and the speaker wants to draw focus to this element rather than to mark a metaphor. Hence, the disagreement between Andersen and myself is, then, possibly rooted in the interpretation of *like*.

Miller et al. (1995), as mentioned above, rejected that *like* could be used as an approximator. One of their arguments is that they find no indication that *like* + numerical expression is interpreted as an approximation. The data from the SBCSAE shows several instances where *like* clearly communicates a less-than-literal understanding of the numerical expression although *like* in combination with a numerical expression does not always mean approximation. They argue against this function because *about* is usually associated with approximation, and when *like* is used in the environment of an approximation, they claim *like* is located in front of *about* (p. 370). The data from the current study indeed shows that *about* can introduce an approximation as in (27):

(27) HAROLD: [3For **about** five minutes3], probably.

(SBC002 *Lambada*)

However, proving that *about* is used for introducing approximation does not exclude other words from having the same function. Their claim that *like* precedes *about* when introducing approximations is not supported by the SBCSAE. As displayed above, there are, in this study, several occasions where *like* marks an approximation. The function has also been widely accepted in other studies (e.g. Schourup, 1985; Andersen, 2001, D'Arcy, 2007).

As written in chapter 2, Schourup (1985) claims *like* can be used as a verbal filler to hold the floor. He never comments on whether he views DMs as being able to serve multiple functions simultaneously, but if he is under the impression that *like* can be used as a mere verbal filler, or interjection, as he calls it, he fails to address why *like* and not any other word or sound is used to fill a pause for holding the floor. This is the reason why this study operates with superordinate functions that must be combined with subordinate functions, which in turn are related to the basic meanings of the DMs. Holmes (1986) also touched upon verbal fillers and claimed that all instances of *you know* in her study had an additional function as a verbal filler (p. 16). Although I acknowledge DMs can be used as verbal fillers, it is hard to argue that all the DMs analyzed in this study carry this function, especially if the DM occurs utterance finally with falling intonation, which is usually used for signaling the end of an utterance.

Holmes (1986) also claimed *you know* could be used when the speaker feels embarrassed and wants the addressee to validate the speaker's right to be embarrassed. This function was not found in the current study. It does not mean it is not used by Americans, but the failure to identify this function could simply be explained by the lack of topics regarding embarrassment being present in the SBCSAE. Based on the examples the gives for this function some questions emerge, however. (28) is an example from her study.

(28) `and it was quite well it was it was all very embarrassing` **`you know`**

(Holmes, 1986, p. 10)

The message of the utterance is undeniably about the speaker's embarrassment towards a situation, and it is reasonable to believe the speaker gives yields the floor to the addressee because the token is located utterance-finally. What is missing from both the example rendered in (28) and the other example in her study is a rendering of what is uttered after the token occurs. Hence, there is no evidence in her research paper that embarrassment is validated. Both examples she provides are located utterance-finally and *you know* is pronounced with rising intonation. I noted prosody in the data from the SBCSAE, but due to lack of space, this has not been addressed properly in my study. I found that when *you know* is pronounced with a small rising intonation, the speaker simply wants the addressee to confirm that the s/he pays attention and understands what the speaker talks about, and as mentioned in chapter 4.1.1.2. This is confirmed by a realization of some type of backchanneling. If what follows these examples is backchanneling such as *mhm* or *yeah*, it is reasonable to wonder whether backchanneling to confirm comprehension has been mistaken for validation of embarrassment. Instead, (28) could probably be fitted into the function of reassuring the addressee, which is found in my study and the function which Holmes (1986) called emphatic, which, as mentioned above, also is concerned with reassuring the addressee about the validity of the proposition. This, of course, is only speculations given she does not report what follows *you know* in her study.

5.2 The Quantitative Part of the Study

As the results above suggest, there is little evidence that some social groups are more prone to using neither *like*, nor *you know* as DMs. The only exceptions are that people of lower education use *you know* for reassuring the addressee nearly significantly more than people of higher education, and a small statistical difference which reveal DM *you know* when not specified by function is used more by people of lower education. The SBCSAE does not give

any strong indications to what might be the cause of the higher frequency among people of lower education with regards to DM *you know*, especially when there is no significant difference in why the token is employed in the first place although the near-significant difference with regards to *you know* used for reassuring the addressee is one indication.

The DMs put in the rag bag were also quantified to give an indication to whether the results could be explained by unbalanced numbers of DMs which could not be analyzed. As mentioned above, there was no significant difference in how often the tokens were put in the rag bag. Since previous studies (e.g. Blyth et al., 1990; Dailey-O'Cain, 2000; Buchstaller, 2014) suggest non-linguists ascribe certain social groups to be more frequent users of certain DMs, it was also important to rule out the possibility that these claims were not based on imbalanced observations of non-DMs. The study suggests this is not the case given none of the results from non-DMs showed significant difference.

In general, although none but one comparison of educational level proved to be of significant difference, it appears that the differences in frequencies are higher when comparing educational levels than when comparing genders. As displayed above, women of lower education are the biggest contributors to the non-significant differences in frequencies of *like* between higher and lower education, while men of lower education are the biggest contributors to the differences in the frequency of *you know* when comparing people of higher education to people of lower education. The comparison of men/women of higher/lower education should not be interpreted for anything but a closer look at who the people that contributed to the results when comparing gender and education separately are because, as mentioned above, there are so few speakers in each of these four groups, and these results have not been tested for statistical significance.

Since there is little agreement among scholars how functions of DMs should be defined, I will not attempt to compare the quantitative results of the functions in my study with similar functions found in other studies with the exception quotative *like*, which has a similar definition across literature. The overall frequency of the DMs, however, will be discussed below, although there is also some disagreement to what qualifies as a DM. The comparisons will include a discussion of previous and current findings of both gender and social class/ educational level. The participants in the SBCSAE were not asked to give information about which socioeconomic class they belong to but is reasonable to assume that there is a higher density of people from higher socioeconomic classes among highly educated people in America compared to people of lower education. It is worth noting that the different

studies have different ways of defining social class, and that social class is, nevertheless, not equivalent to educational level as it has been defined in this study.

5.2.1 *Like*

As mentioned above, there is little agreement among scholars how functions of DMs should be defined. Therefore, I will not attempt to compare the quantitative results of the functions with similar functions found in other studies with the exception quotative *like*, which has a similar definition across literature. The overall frequency of the DMs, however, will be discussed below, although there is also some disagreement to what qualifies as a DM. The comparison will include a discussion of previous and current findings of both gender and social class/ educational level. It is worth noting that the different studies have different ways of defining social class, and that social class is not equivalent to educational level as it has been defined in this study.

Both Andersen (2001) and Hasund (2003) used the COLT corpus. They both found that girls use DM *like* more than boys. Andersen found the difference to be statistically significant and concludes the prototypical user of *like* is a girl rather than a boy. Hasund concludes her results support Andersen's conclusion although she did not test for statistical significance. Dailey-O'Cain (2000), on the other hand, found that neither quotative *like* nor other discourse functions of the token showed significant difference between men and women, although men showed a higher frequency, which is contrary to what Andersen and Hasund found. The reason why Andersen found a significant difference which was not replicated in neither my study nor the one by Dailey-O'Cain could be explained by the chance that the discourse pattern in the US is different from that of the UK (Dailey-O'Cain tested all ages). There is evidence of that given Andersen reports the frequency of the token uttered by both boys and girls to be lower than what is the case for men and women in the current study. Dailey-O'Cain found that the frequency is higher for men than women, while the opposite is true for this study. This is probably a random result given neither study found a significant difference, suggesting the gender-based differences are likely to be by chance.

Romaine et al. (1991) found that quotative *like* was produced by women in 83 per cent of the instances they analyzed. There is, to my knowledge, no other study which reports such great difference between men's and women's use of quotative *like*, but also Ferrara et al. (1995) found that women used quotative *like* more than women in 1990, though men's and women's use of quotative *like* did not come near the imbalance reported by Romaine et al. However, no differences were between the two genders were detected in 1992 and 1994, leading Ferrara et

al. to conclude the gender bias has been neutralized. Blyth et al. (1990) found that men used quotative *like* significantly more than women.

There could be a number of reasons why Romaine et al. report much greater varieties than other studies; they fail to mention how many speakers make up their corpus, and how many words are uttered by men and by women. The current study showed normal distribution was rarely found, meaning there is great variation even within a group. Hence, the lower number of participants, the higher the risk is that a selected group is not representative of a population. A second issue is that percentage should only be used if the same amount of words is used by the two groups. Given no such information is given, it is impossible to critique the study properly. Reporting frequency would have solved this issue. Alternatively, one can do as Blyth et al. and only count various markers for direct speech and compare how big a percentage *like* makes up of the total amount of quotative markers and then compare the two groups. Ferrara et al. (1995) never tested for statistical significance. It sounds improbable that a gender bias is neutralized within two years, and the possibility that these results are by chance rather than by what they claim cannot be ruled out. Blyth et al. present an interesting result showing men use quotative *like* significantly more than women. One weakness in the study, however, is that the group of men only contains 10 subjects, and as previously discussed, the results of such small groups may not be representative of a population when there are so great individual differences.

Less work has been conducted to report differences with regards to social class than gender. The three studies which focus on *like* and social class that have been rendered in this study all used corpora based on British speakers. Hasund (2003), who compared speakers from two different boroughs thought to represent two ends to the scale with regards to social class, found the frequencies of the speakers from these two boroughs to be near identical. Andersen (2001) did not find a significant difference between people of high, middle, and low social class, but as mentioned above, in search for results of statistical significance, he collapsed the group of middle and low social class and compared it to the high social class. He then found that people of the high social class were more frequent users. Andersen's findings that speakers of the highest social class were the most frequent users of DM *like* was not supported by Beeching (2016). As reported above, she found that the least frequent users of *like* were the people of the highest class, while the most frequent users were from the two lowest classes. Beeching's work is based on searches on *like* in the BNC and analyzed the number of hits she got, but she never analyzed each occurrence of the token. Hence, her study

also counts non-DM *like* despite what the title of her book, *Pragmatic Markers in British English*, indicates.

In my study, I have concluded there is no significant difference between people of higher and lower education with regards to DM *like*, which is in lines with what Hasund (2003) concluded regarding social class, and the first results reported from Andersen (2001). Although I did not find DM *like* to be used with significantly higher frequencies among people of lower education, their frequency was more than twice as high as the frequency of people from higher education. The lack of a significant difference is probably due to the large individual differences. Despite of a missing significant result on my behalf, it is interesting to see that Andersen found that speakers from the highest social class were more frequent users of the token, while this study finds an (insignificant) higher frequency among people of lower education. Given Beeching's definition with regards to social class is based on occupation, her definition is perhaps closer to educational level as it has been presented in the study at hand than what is the case for Hasund's and Andersen's studies primarily based on teenagers. She, like I, found that the lowest education/ occupation are the most frequent users while the people with highest education/ occupation are the least frequent users. This paragraph, thus, does not only display that the different studies have yielded different results, but also that the results are highly sensitive to how social class has been defined given Andersen and Hasund come to different conclusions based on the same data.

5.2.2 *You Know*

Lakoff's (2004) claims that *you know* can appear without a purpose has been challenged by this study, which shows all instances of *you know* has a specific function. Moreover, a claim that women are the typical users of this style of expressing themselves cannot be supported by the results of the current study given neither the frequency of the functions, nor the overall distribution of this DM proved to be of significant difference. Her method based on introspection rather than a dataset reduces her study to a warning of the dangers of not basing conclusions on evidence from the real word and stresses the importance of corpus linguistics as a science.

Macauay's (2002) study suggests that women's frequency of *you know* is close to twice as high compared to that of men. It would have been interesting to see whether his results were of significant difference. He is, as mentioned above, careful to make any generalizations and stresses that individual differences in frequencies must be taken into consideration. He, nevertheless, concludes women are more likely to use the token. His

conclusions on gender differ from what was concluded by current study, Holmes (1986), Stubbe et al. (1995), Koczogh et al. (2011), and Beeching (2016). These studies concluded there were no differences between the overall use of *you know* between men and women. A shortcoming of Koczogh et al.'s study is that Larry King is the only interviewer in male corpus, while different women conduct the interviews in the female corpus. Both Macauley and I found that there is great individual variety in the employment of DMs, this and should have been accounted for in their study given Larry King is likely to be a big contributor to the word count for the male corpus.

As was the case for *like*, less studies have been conducted to test social classes compared to gender. Beeching's study (2016) revealed that the C1 and C2 speakers are the most frequent users of *you know*. C1 refers people working in junior management, supervisory jobs, or as professionals, while C2 speakers are skilled manual workers. Some of the people who are referred to as C2 speakers would probably be referred to as people of higher education by the definition used in my study, while it is also possible that some of the people categorized as C1 speakers in the BNC would have been categorized as people of lower education by the standards of the current study. It is, therefore, more beneficial to compare my study to Beeching's AB and DE speakers. As written in chapter 2, she found that AB speakers used *you know* significantly more than DE speakers, which is contrary to what the current study found about educational level. As with Beeching's study on *like*, she did not separate DM *you know* from non-DM *you know*. Although my study shows the occurrence of non-DM *you know* is fairly small, it is not known how much the presence of *you know* used in the S-V construction affects her results. It is not known what Macaulay (2002) used as the basis for defining middle-class and lower-class speakers, but he concluded that there is no difference between the frequency of these two social groups. Stubbe et al. (1995) did not find a significant difference between middle-class and lower-class speakers in their data collected from interviews. However, according to their data of New Zealand speakers in casual discourse, lower-class speakers use *you know* significantly more than middle-class speakers. Their definition of social class, is, as mentioned above, educational level and occupation. It is, therefore, therefore interesting that their results from the causal conversations are in accordance with my findings showing people of lower education use *you know* significantly more than people of higher education even though different populations have been tested. Once again one cannot rule out the possibility of the varying results discussed in this

paragraph are due to the difference in how social class/ educational level has been defined. It is also quite possible that these results reflect varieties between nationalities.

5.3 Limitations

As stated above, this study has not been balanced for a range of social parameters such as ethnicity, age, and gender when testing educational level and *vice versa*. The majority of the participants in this study is white. Moreover, few participants reported to attend 12 years of education or less, which means most participants have studied beyond high school level. It is, therefore, not certain that the corpus simulates the adult American society at large. The identity such as age, gender, social status of the addressee has not been taken into consideration either despite Holmes' (1986) claims that the frequency of DM depends on the addressee's gender.

Several scholars have chosen not to do quantitative research on the various functions DMs serve. Hasund (2003) did not look at distributions based on pragmatic functions because DMs can serve a multitude of functions simultaneously, which, she argued, caused many instances of DMs to be fuzzy. Hence, she concluded it was not beneficial to count them for statistical purposes. The current study operated with two levels of functions where only the subordinate functions were quantified. Operating with two levels rather than having all functions on one level decreased the number of fuzzy instances of the DMs. The low amount of instances which could not be properly identified were put in the rag bag, which should increase the reliability and validity of this study. It allows this study to provide more detailed information about social differences with regards why DMs are employed than a study simply measuring frequencies of DMs but the results are based on subjective interpretations from one researcher and it is impossible to ensure that personal bias does not have an impact on the analyses.

6 Conclusions

As the introduction of this thesis displayed, there are indications that the man on the street believes particular DMs are used by certain people. Sociolinguistic works, on the other hand, have reported various findings from around the (English-speaking) world. It was, thus, of interest to learn what could be found when conducting a corpus-based study on adult Americans aged 20-59. The study was restricted to focus on two DMs which are frequent in informal discourse and tend to carry some social stigma namely the DMs *like* and *you know*. The study was also restricted to focus on the two social categories gender and educational

level. Though the use of mixed methods, the following two questions were to be answered: (1) can social differences regarding frequency of employing DMs be detected, and (2) given their versatility, do different social groups use DMs for different purposes.

The results from the study revealed that, despite beliefs that certain DMs are predominantly used by women, there is no significant difference of the frequency with which men and women employ *like* and *you know*. It was of interest to investigate whether other social factors played a role in the way in which people express themselves, and it was decided that the study additionally should focus on linguistic differences with regards to educational level. There results revealed there were tendencies of people from lower education employing *like* more frequently than their counterpart. Despite the apparent difference, the results failed to be significant. *You know*, on the other hand, was used significantly more by people of lower education.

Given the versatile nature of *like* and *you know*, it was of interest to find out, not only how frequently people from different social groups used them, but also why they use them, and whether these groups use them for different purposes. Upon analyzing the corpus material, each DM could be said to be used for one of four different reasons. *Like* is used for signaling an upcoming approximation, introducing one of several examples, drawing focus to particular elements in the discourse that are of importance, and for signaling an upcoming quote. *You know* can be used for introducing previous knowledge as background information to an upcoming main point, reassuring the addressee of the validity or the speaker's conviction of what s/he says. *You know* can also be used when a message has been unclearly delivered where the token either signals the speaker transitions into an elaboration, or where the speaker gives the addressee an opportunity to ask the speaker for an elaboration. Lastly, *you know* could be used when the speaker wishes to go back and repair part of the discourse. The results showed that there was no significant difference with regards to how men and women, or people of lower and higher education use these two DMs although one function came close to being significantly more used by people of lower education compared to people of higher education, namely when *you know* is used for reassuring the addressee.

These results must be taken with caution because the study is of limited scope and one should be careful to make generalizations beyond the SBCSAE since no effort has been made to control for a variety of social factors to make the subcorpus used in this study simulate a true American society.

6.1 Suggestions for Further Research

As is often the case for studies of restricted scope, the current study addresses a few questions but leave many unanswered. This study found tendencies of people from lower education being more frequent users of *like* and *you know* than highly educated people. Further studies need to be conducted to determine whether this is related to social classes in America. Sociolinguistic research on DMs have yielded an array of different results, and more research is needed to determine whether other social factors influence the frequencies with which DMs occur, but also whether other DMs are sensitive to social factors. In the future, research also needs to answer whether the speaker's use of DMs is a result of the social background of the addressee.

7 References

Ajmer. (2013). Understanding Pragmatic Markers: A Variational Pragmatic Approach. Edinburgh: Edinburgh University Press

Andersen. (1998). The Pragmatic Marker Like from a Relevance-theoretic Perspective. In Jucker & Ziv (Eds). Discourse Markers. (147-170). Amsterdam/ Philadelphia: John Benjamins Publishing Company

Andersen. (2000). The Role of the Pragmatic Marker *like* in Utterance Interpretation. Andersen & Fretheim (Eds.) *Pragmatic Markers and Propositional Attitude*. (pp. 17-38). Amsterdam/ Philadelphia: John Benjamins Publishing Company.

Andersen. (2001). *Pragmatic markers and sociolinguistic variation: a relevance-theoretic approach to the language of adolescents*. Amsterdam/ Philadelphia: John Benjamins Publishing Company.

Beeching. (2016). *Pragmatic Markers in British English: Meaning in Social Interaction*. Cambridge: Cambridge University Press.

Blyth, Recktenwald, & Wang. (1990). I'm like, "Say What?!": A New Quotative in American Oral Narrative. *American Speech, 65*(3), 215-227. doi:10.2307/455910

Brizina, McEnery & Wattam. (2015). #LancsBox version 3.0.2 [Computer software]. Retrieved from: corpora.lancs.ac.uk/lancsbox/

Brinton. (1996). *Pragmatic Markers in English*. Berlin/ New York: De Gruyter.

Brown & Levinson. (1987). *Politeness: Some Universals in Language Use*. New York/ Oakleigh: Cambridge University Press.

Buchstaller. (2014). *Quotatives: New Trends and Sociolinguistic Implications*. West Sussex: John Wiley and Sons, Inc.

Buchstaller & Van Alphen. (2012). *Quotatives: Cross-linguistic and Cross disciplinary perspectives*

Dailey-O'Cain. (2000). The Sociolinguistic Distribution of and Attitudes Toward Focuser *like* and quotative *like*. *Journal of Sociolinguistics, 4*(1), 60-80. doi.org/10.1111/1467-9481.00103

D'Arcy. (2007). *Like* and language ideology: Disentangling fact from fiction. *American Speech, 82*(4), 386-419. doi: 10.1215/00031283-2007-025

Du Bois. (1991). Transcription Design Principles for Spoken Discourse Research. *Pragmatics, 1*(1), 71-106. Retrieved from: journals.linguisticsociety.org/elanguage/pragmatics/article/download/464/464-782-1-PB.pdf

Du Bois. (2006) Transciption Convention Updates. Retrieved from: http://transcription.projects.linguistics.ucsb.edu/A05updates.pdf

Du Bois, Chafe, Meyer, Thompson, Englebretson and Martey. (2000-2005) Santa Barbara Corpus of Spoken American English, Parts 1-4. Retrieved from www.linguistics.ucsb.edu/research/santa-barbara-corpus

Erman. (2001). Pragmatic markers revisited with a focus on you know in adult and adolescent talk. *Journal of Pragmatics, 33,* 1337-1359. Doi: 10.1016/S0378-2166(00)00066-7

Ferrara & Bell. (1995). Sociolinguistic Variation and Discourse Function of Constructed Dialogue Introducers: The Case of Be like. *American Speech, 70*(3), 265-290. doi:10.2307/455900

Fischer. (2006). Approaches to Discourse Markers. Amsterdam/ Boston/ Heidelberg/ London/ New York/ Oxford/ Paris/ San Diego/ San Francisco/ Singapore/ Sydney/ Tokyo: Elsevier

Fox, Maschler, Uhlmann, A Cross-linguistic Study of Self-Repair: Evidence from English, German, and Hebrew. Journal of Pragmatics, 42(9), 2487-2505. doi.org/10.1016/j.pragma.2010.02.006

Fox Tree. (1995). The Effects of False Starts and Repetitions on the Processing of Subsequent Words in Spontaneous Speech. *Journal of Memory and Language, 34*(6), 709-738. doi.org/10.1006/jmla.1995.1032

Fox Tree & Schrock. (2002). Basic Meanings of *You Know* and *I Mean*. *Journal of Pragmatics, 33*, 727-474. doi.org/10.1016/S0378-2166(02)00027-9

Fraser (d.u.) Pragmatic Markers. *Pragmatics, 6*(2), 167-190. Retrieved from: journals.linguisticsociety.org/elanguage/pragmatics/article/download/418/418-712-1-PB.pdf

Friginal & Hardy. (2014). *Corpus-based Sociolinguistics: A Guide for Students*. New York: Routledge.

Hashimi & Babaii. (2013). Mixed Methods Research: Toward New Research Designs in Applied Linguistics. *The Modern Language Journal, 37*(4), 828-852. Retrieved from: www.jstor.org/stable/43651724.

Harley. (2016). The Psychology of Language: From Data to Theory. Psychology Press: New York

Hasund. (2003). *The Discourse Markers* like *in English and* liksom *in Norwegian Teenage Language: A Corpus-based, Cross-lingustic Study* (Doctoral dissertation). University of Bergen and Agder University College.

Holmes. (1986). Functions of You Know in Woman's and Men's Speech. *Language in Society, 15*(1), pp. 1-21. Retrieved from: www.jstor.org/stable/4167711

Jacobsen. (2017). *"Or, Like, You Know; Central Iowa and Stuff": A Corpus-based Study of American Men's and Women's Use of the Discourse Marker* You Know *in Informal Speech* (Bachelor's Thesis). University of Agder, Kristiansand

Jucker & Smith. (1998). And People Just You Know Like 'Wow': Discourse Markers as Negotiation Strategies. Jucker and Yael (Eds). *Discourse Markers*. (pp. 171-201). Amsterdam/ Philadelphia: John Benjamins Publishing Company

Jucker & Ziv. (1998). Discourse Markers. Amsterdam/ Philadelphia. John Benjamins Publishing Company

Koczogh and Furkó. (2011). Gender Differences in the Use of the Discourse Markers You Know and I Mean: "It's just like, dude, seriously, it's been a bad week, I mean, kind of thing.". Argumentum (7), 1-18. Retrieved from www.academia.edu/2537039/Gender_differences_in_the_use_of_the_discourse_markers_you_know_and_I_mean

Lakoff. (2004). Language and Women's Place. Bucholtz (Ed.) *Language and Woman's Place: Text and Commentaries*. (pp. 31-102). New York: Oxford University Press.

Livia. (2004). Picking up the Gauntlet. Bucholtz (Ed.). *Language and Woman's Place: Text and Commenaries*. (pp143-149) New York: Oxford University Press

Macauley. (2002). You know, it depends. *Journal of Pragmatics, 34*(6), 749-767. doi.org/10.1016/S0378-2166(01)00005-4

Macaulay. (2013). Discourse Variation. In Chambers & Schilling (Eds). *The Handbook of Language Variation and Change 2nd edition* (220-236). Oxford: John Wiley & Sons, Incorporated

Miller & Weinert. (1995). The Function of LIKE in Dialogue. *Journal of Pragmatics, 23*(4), 365-393. doi.org/10.1016/0378-2166(94)00044-F

Müller. (2005). *Discourse Markers in Native and Non-native English Discourse*. Amsterdam/Philadelphia: John Benjamins Publishing Company

Östman. (1981). *You Know: A Discourse-Functional Approach*. Amsterdam: John Benjamins B. V.

Potter & Lombardi. (1998) Syntactic Priming in Immediate Recall of Sentences. *Journal of Memory and Language, 27*(3), 265-282. doi.org/10.1006/jmla.1997.2546

Rasinger. (2013). *Quantitative Research in Linguistics: An Introduction*. London: Bloomsbury.

Romaine & Lange. (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. Retrieved from:

https://www.jstor.org/stable/pdf/455799.pdf?refreqid=excelsior%3A0b5fda333384b32364ab1a0cc49d9107

Sachs. (1967). Recognititon Memory for Syntactic and Semantic Aspects of Connected Discourse. *Attention Perception & Psycholinguistics, 2*(9), 437-442. doi: 10.3758/BF03208784

Schiffrin. (1992). *Discourse Markers*. Cambridge/ New York/ Victoria: Cambridge University Press.

Schouroup. (1985). *Common Discourse Particles in English Conversations*. New York/ London: Garland Publishing, Inc.

Stubbe & 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 and Communication, 15*(1), 63-88. Retrieved from: www.sciencedirect.com/science/article/pii/0271530994000166

Svartvik. (1990). *The London Corpus of Spoken English: Description and Research.* Lund Studies in English, 82. Lund: Lund University Press. Retrieved from: www.helsinki.fi/varieng/CoRD/corpora/LLC/

Tannen. (1986). Introducing constructed dialogue in Greek and American conversational and literary narrative. Coulmas (Ed.). *Direct and Indirect Speech.* (311-332). Berlin/ New York/ Amsterdam: Mouton de Gruyter. Retrieved from: https://static1.squarespace.com/static/5523ffe4e4b012b2c4ebd8fc/t/56d9c6c759827ead57ecfdd1/1457112776305/introducing+constructed+dialogue.pdf

Travis. (2006). The Natural Semantic Metalanguage Approach to Discourse Markers. Fischer (Ed.). Approaches to Discourse Markers. (pp. 219-241). Amsterdam/ Boston/ Heidelberg/ London/ New York/ Oxford/ Paris/ San Diego/ San Francisco/ Singapore/ Sydney/ Tokyo: Elsevier

Tjønndal. (2018). *Statistisk Analyse i Stata*. Oslo: Cappelen Damm Akademisk.

Underhill. (1988). *Like* is. Like, Focus. A*merican Speech, 63*(3), 234-246. doi:10.2307/454820

Watts. (1989). Taking the pitcher to the 'well': Native speakers' perception of their use of discourse markers in conversations. *Journal of Pragmatics*. *13*, 203-237. doi.org/10.1016/0378-2166(89)90092-1

West & Zimmerman. (1987). Doing Gender. *Gender and Society, 1*(2), 125-151. Retrieved from: www.jstor.org/stable/189945?seq=10#metadata_info_tab_contents

Wiese. (2011). *So* as a Focus Marker in German. *Linguistics: An Interdisciplinary Journal of the Language Sciences, 28*(1), 991-1039. doi.org/10.1515/ling.2011.028

University of California Santa Barbara. (n.d.) Santa Barbara Corpus of Spoken American English. Retrieved from: www.linguistics.ucsb.edu/research/santa-barbara-corpus

Appendix I

Symbols for discourse transcription in the SBCSAE as reported by Du Bois (1991).

UNITS	
Intonation unit	{*carriage return}*
Truncated intonation unit	--
Word	{*space*}
Truncated word	-
SPEAKERS	
Speaker identity/ turn	:
Speaker overlap	[ ]
TRANSITIONAL CONTINUITY	
Final	.
Continuing	,
Appeal	?
TERMINAL PITCH DIRECTION	
Fall	\
Rise	/
Level	—
ACCENT AND LENGHTENING	
Primary accent	^
Secondary accent	‘
Booster	!
Lengthening	=
TONE	
Fall	\
Rise	/
Fall-rise	\/
Rise-fall	/\
Level	—

PAUSE	
Long	...(N)
Medium	...
Short	..
Latching	(0)
VOCAL NOISES	
Vocal noises	( )
Inhalation	(H)
Exhalation	(Hx)
Glottal stop	%
Laughter	@
QUALITY	
Quality	<Y Y>
Laugh quality	<@ @>
Quotation quality	<Q Q>
Multiple quality features	< Y<Z Z> Y>
PHONETICS	
Phonetic/ phonemic transcription	(/ /)
TRANSCRIBERS PERSPECTIVE	
Researcher's comment	(( ))
Uncertain hearing	<X X>
Indescribable syllable	X
SPECIALIZED NOTATIONS	
Duration	(N)
Intonation contour continued	&
Intonation subunit boundary	|
Embedded intonation unit	< | | >
Reset	*{Capital Initial}*
False start	< >
Codeswitching	<L2 L2>

NON-TRANSCRIPTION LINES

Non-transcription line	$
Interlinear gloss line	$G

RESERVED SYMBOLS

Phonemic/ orthographic	’
Morphosyntactic coding	+ * # { }
User-definable	“ ~ ;

During the period in which the SBCSAE was under construction, the transcription convention was updated. Du Bois writes the following: “[s]ome of these changes have already been implemented in the published edition of the Santa Barbara Corpus of Spoken American English (Du Bois, 2000, Du Bois, 2003), while others will be implemented in future editions.” (Du Bois, 2006)

The changes which have been made are the following:

Meaning	**Old notation**	**New Notation**
Unintelligible (syllables)	X	#
Uncertain hearing (words)	<X you’re kidding X>	#you’re #kidding
Pseudograph (fake name, address etc.)	Jill	~Jill
Real name, address, etc.	#Jill	Jill
Long-scope features (various)	<A two words A>	<A> two words </A>
Laughter during speech (1-5 words)	<@ two words @>	@two @words
Laughter during speech (6+ words)	<@ six words @>	<@> six words </@>
Overlap, 3rd instance	[3 word word word 3]	[3word word word]
Overlap, 2nd instance	[[word word word]]	[2word word word]
Vox: voice of another	<Q words Q>	<VOX> words </VOX>

Word truncation/cut-off with no glottal	wor-	wor– (en dash)
Word truncation/cut-off with glottal	wor-	wor%– (en dash)
Intonation unit truncation	--	— (em dash)
Morpheme boundary		- (hyphen)
Extra-long IU	indent	(word wrap)
Pause, timed	...(1.2)	(1.2)
Pause, short (< 150 milliseconds)	..	..
Pause, untimed (> 150 milliseconds)	...	...
Pause location (if at IU boundary)	[line-initial]	[on separate line]
Latching	(0)	=
Speaker label	J:	Jill:
Reset	Capital letter	⊦
Sentence start		Capital letter
Repair/editable	<word>	^{x}word

Appendix II

Links to the transcripts and audio files used in this study:

SBC002 *Lambada*

Transcript: https://www.linguistics.ucsb.edu/sites/secure.lsit.ucsb.edu.ling.d7/files/sitefiles/research/SBC/SBC002.trn

Audio file: http://corpusmedia.linguistics.ucsb.edu/SBC002.wav

SBC003 *Conceptual Pesticides*

Transcript: https://www.linguistics.ucsb.edu/sites/secure.lsit.ucsb.edu.ling.d7/files/sitefiles/research/SBC/SBC003.trn

Audio file: http://corpusmedia.linguistics.ucsb.edu/SBC003.wav

SBC004 *Raging Bureaucracy*

Transcript: https://www.linguistics.ucsb.edu/sites/secure.lsit.ucsb.edu.ling.d7/files/sitefiles/research/SBC/SBC004.trn

Audio file: http://corpusmedia.linguistics.ucsb.edu/SBC004.wav

Additional note: Carolyn has not been considered in this study because of her young age.

SBC005 *A Book About Death*

Transcript: https://www.linguistics.ucsb.edu/sites/secure.lsit.ucsb.edu.ling.d7/files/sitefiles/research/SBC/SBC005.trn

Audio file: http://corpusmedia.linguistics.ucsb.edu/SBC005.wav

SBC006 *Cuz*

Transcript: https://www.linguistics.ucsb.edu/sites/secure.lsit.ucsb.edu.ling.d7/files/sitefiles/research/SBC/SBC006.trn

Audio file: http://corpusmedia.linguistics.ucsb.edu/SBC006.wav

SBC007 *A Tree's Life*

Transcript: https://www.linguistics.ucsb.edu/sites/secure.lsit.ucsb.edu.ling.d7/files/sitefiles/research/SBC/SBC007.trn

Audio file: http://corpusmedia.linguistics.ucsb.edu/SBC007.wav

SBC013 *Appease the Monster*

Transcript: https://www.linguistics.ucsb.edu/sites/secure.lsit.ucsb.edu.ling.d7/files/sitefiles/research/SBC/SBC013.trn

Audio file: http://corpusmedia.linguistics.ucsb.edu/SBC013.wav

SBC017 *Wonderful Abstract Notions*

Transcript: https://www.linguistics.ucsb.edu/sites/secure.lsit.ucsb.edu.ling.d7/files/sitefiles/research/SBC/SBC017.trn

Audio file: http://corpusmedia.linguistics.ucsb.edu/SBC017.wav

SBC028 *Hey Cutie Pie*

Transcript: https://www.linguistics.ucsb.edu/sites/secure.lsit.ucsb.edu.ling.d7/files/sitefiles/research/SBC/SBC028.trn

Audio file: http://corpusmedia.linguistics.ucsb.edu/SBC028.wav

SBC031 *Tastes Very Special*

Transcript: https://www.linguistics.ucsb.edu/sites/secure.lsit.ucsb.edu.ling.d7/files/sitefiles/research/SBC/SBC031.trn

Audio file: http://corpusmedia.linguistics.ucsb.edu/SBC031.wav

Additional note: Jamie is not included in this study because she is a waitress taking the orders of Rosemary, Sherry, and Beth.

SBC036 *Judgemental on People*

Transcript: https://www.linguistics.ucsb.edu/sites/secure.lsit.ucsb.edu.ling.d7/files/sitefiles/research/SBC/SBC036.trn

Audio file: http://corpusmedia.linguistics.ucsb.edu/SBC036.wav

SBC043 *Try a Couple Spoonfuls*

Transcript: https://www.linguistics.ucsb.edu/sites/secure.lsit.ucsb.edu.ling.d7/files/sitefiles/research/SBC/SBC043.trn

Audio file: http://corpusmedia.linguistics.ucsb.edu/SBC043.wav

SBC044 *He Knows*

Transcript: https://www.linguistics.ucsb.edu/sites/secure.lsit.ucsb.edu.ling.d7/files/sitefiles/research/SBC/SBC044.trn

Audio file: http://corpusmedia.linguistics.ucsb.edu/SBC044.wav

SBC045 *The Classic Hooker*

Transcript:
https://www.linguistics.ucsb.edu/sites/secure.lsit.ucsb.edu.ling.d7/files/sitefiles/research/SBC/SBC045.trn

Audio file: http://corpusmedia.linguistics.ucsb.edu/SBC045.wav

SBC047 *On the Lot*

Transcript:
https://www.linguistics.ucsb.edu/sites/secure.lsit.ucsb.edu.ling.d7/files/sitefiles/research/SBC/SBC047.trn

Audio file: http://corpusmedia.linguistics.ucsb.edu/SBC047.wav

SBC050 *Just Wanna Hang*

Transcript:
https://www.linguistics.ucsb.edu/sites/secure.lsit.ucsb.edu.ling.d7/files/sitefiles/research/SBC/SBC050.trn

Audio file: http://corpusmedia.linguistics.ucsb.edu/SBC050.wav

SBC051 *New Yorkers Anonymous*

Transcript:
https://www.linguistics.ucsb.edu/sites/secure.lsit.ucsb.edu.ling.d7/files/sitefiles/research/SBC/SBC051.trn

Audio file: http://corpusmedia.linguistics.ucsb.edu/SBC051.wav

SBC052 *Oh you Need a Breadbox*

Transcript:
https://www.linguistics.ucsb.edu/sites/secure.lsit.ucsb.edu.ling.d7/files/sitefiles/research/SBC/SBC052.trn

Audio file: http://corpusmedia.linguistics.ucsb.edu/SBC052.wav

Appendix III

In total, approximately 1,900 instances of *like* and *you know* have been analyzed. It is too many instances to report. Ten examples from each function of the two DMs will, nevertheless, be presented below.

Approximator (*like*)

(1) we're gonna have **like** .. dietetic ... style ... fish.

(SBC003 *Conceptual Pesticides*)

Note to (1): *Like* is unstressed and pronounced with leveled intonation. *Like* signals lexical approximation i.e. marking there is a better way to describe how the fish is cooked.

(2) .. (H) And I was **like** .. seven years old.

(SBC044 *He Knows*)

Note to (2): *like* is unstressed and pronounced with leveled intonation. *Like* modifies the numerical expression.

(3)
CORINNA: .. Or you could always go down to= that one strip in Chicago.
PATRICK: ... What strip.
CORINNA: ... The one over %,
% .. by Crowbar?
PATRICK: ... No?
CORINNA: ... It's **like** .. by the bridge?

(SBC045 *The Classic Hooker*)

Note to (3): *Like* is unstressed and pronounced with leveled intonation. *Like* modifies the PP.

(4) And so it's **like** .. eight days late?

(SBC028 *Hey Cutie Pie*)

Note to (4): *like* is unstressed and pronounced with leveled intonation. *Like* approximates the numerical expression.

(5)
well he had like um,
a d- --
... (TSK) (H) Uh,
.. **like** a,
.. blue jean ball,
or something like th[at.

(SBC036 *Judgemental on People*)

Note to (5): *like* is unstressed and pronounced with leveled intonation. *Like* signals the speakers is not sure if it was a blue jean ball but is was something the likes of it.

(6) KEVIN: When we were **like**,

.. in middle s- --
When we were in middle school.

(SBC036 *Judgemental on People*)

Note to (6): *like* is unstressed and pronounced with leveled intonation. *Like* modifies the PP.

(7) <X Kinda **like** X> couple years ago.

(SBC050 *Just Wanna Hang*)

Note to (7): *like* is unstressed and pronounced with leveled intonation. *Like* modifies the adverbial of time.

(8) MILES: [He has to double it dow=n to] **like** o=ne f=ifth speed or something,
before they can g=- pick it up-,

(SBC002 *Lambada*)

Note to (8): *like* is unstressed and pronounced with leveled intonation. *Like* modifies the adverbial.

(9) but you can only actually keep them over,
for **like** five minutes.
[Cause they have like buses to catch and stuff],

(SBC004 *Raging Bureaucracy*)

Note to (9): *like* is unstressed and pronounced with leveled intonation. *Like* modifies the adverbial of time.

(10) KELLY: Your grandma called.
ARIANNA: ... She did.
KELLY: .. Yeah **like**,
.. fifteen minutes ago.

(SBC050 *Just Wanna Hang*)

Note to (10): *like* is unstressed and pronounced with leveled intonation. *Like* modifies the adverbial of time.

Exemplifier (*like*)

(1) ALICE: (SWALLOW) ... We should all get some money together and,
... is there any way he could [**like**,
MARY: [(H)=]
ALICE: meet us in Great Falls] or something?

(SBC007 *A Tree's Life*)

Note to (1): *like* is unstressed and pronounced with leveled intonation. *Like* signals Great Falls is only one example of where they could go.

(2) MARILYN: ... Should I make **like** tartar sauce,

(SBC003 *Conceptual Pesticides*)

Note to (2): *like* is unstressed and pronounced with leveled intonation. *Like* signals she is happy to make something else as well.

(3)
```
things like,
.. making sure that the script gets to the people
doing the sound effects,
so that they can follow alo=ng,
and make certain that their= sound effects are going in the right
spots
```

(SBC006 *Cuz*)

Note to (3): *like* is unstressed and pronounced with leveled intonation. An exemplification of what an assistant needs to do at a movie set.

(4)
```
.. these guys would like tickle me=,
and hug me,
```

(SBC044 *He Knows*)

Note to (4): *like* is unstressed and pronounced with leveled intonation. *Like* introduces an exemplification of what the speaker was aroused by as a little child.

(5)
```
JEFF:   [(H)] from Nixon,
        to like Rush Limbaugh,
        to abortion,
to capital punishment,
```

(SBC028 *Hey Cutie Pie*)

Note to (5): *like* is unstressed and pronounced with leveled intonation. *Like* introduces an exemplification of the topics the speaker and a stranger talked about.

(6)
```
Like they'll have Footprints on it,
.. (H) or um,
I can't remember.
... Just little thing --
Little prayers,
and stuff like that,
```

(SBC036 *Judgemental on People*)

Note to (6): *like* is unstressed and pronounced with leveled intonation. *Like* signals an exemplification of things that were printed on prayer cards.

(7)
```
everybody's making noises around here.
<X Make [me kinda irritated X>].
ARIANNA:        [Be XX].
DANA:   (H) No.
.. The like banging.
        [And sawing.
```

(SBC050 *Just Wanna Hang*)

Note to (7): *like* is unstressed and pronounced with leveled intonation. *Like* introduces two examples of the type of noice Dana talks about.

(8)
```
(H) and then painted ... garish colors,
like aqua,
and yellow,
and everything.
```
(SBC051 *New Yorkers Anonymous*)

Note to (8): *like* is unstressed and pronounced with leveled intonation. *Like* introduces two examples of the "garish colors" he mentions.

(9)
```
SHARON: cause I4] made the whole class learn,
        like,
        (H)= good morni=ng,
        good bye=,
        (H) can I go to the bathroo=m,
        can I stand u=p,
```
(SBC004 *Raging Bureaucracy*)

Note to (9): *like* is unstressed and pronounced with leveled intonation. *Like* introduces a series of things Sharon taught her students in language class to say.

(10)
```
... (H) I don't like it,
like and when I was ho=me,
.. just went home to Indiana.
(H) I went to hug my sister,
.. and I still feel that she finds a coldness in hugging.
```

Note to (10): *like* is unstressed and pronounced with leveled intonation. The speaker talks about feeling distanced from his family and gives and *like* gives an example of a specific occasion when he felt this distance between him and his sister.

Focuser (*like*)

(1)
```
(H) .. It's like sometimes you go through things,
... and you come out the other side of them,
<WH you WH> .. come out so much better.
```
(SBC005 *A Book About Death*)

Note to (1): *like* is unstressed and pronounced with leveled intonation. The speaker has talked about a bad marriage she was in. What is rendered in (1) is the point of the conclusion of the utterance that, and she want the addressee to notice this.

(2)
```
MARY:          I bet he could do it.
               ... When though.
ALICE:         ... I don't know=,
MARY:          ... He goes [back] to school like the secon=d.
```

(SBC007 *A Tree's Life*)

Note to (2): *like* is unstressed and pronounced with leveled intonation. Alice and Mary are planning an event. Mary highlights this has to happen before the second.

(3)
```
          (H) you know like how we have always talked about life being
           out there,
          .. [you know,
JILL:        [Unhunh].
JEFF:   like there] has to be.
          Cause there's planets evolving around,
[(H) um],
JILL:   [Unhunh].
JEFF:   (SNIFF) you know various stars and stuff.
```

(SBC028 *Hey Cutie Pie*)

Note to (3): *like* is unstressed and pronounced with leveled intonation. Jeff focuses on the main point of the utterance because he will elaborate on his conviction of the existence of extraterrestrial life.

(4)
```
          ... They just built a .. a great big gray water
          processing center,
          .. at the laundromat,
          .. in the .. complex where I live.
MARILYN:.. Oh.
PETE:   It's like right outside our back door.
```

(SBC003 *Conceptual Pesticides*)

Note to (4): *like* is unstressed and pronounced with leveled intonation. Pete highlights that the grey water processing center is located on the property he lives.

(5)
```
(H) So I'm driving up to the house,
... and there's a car in front of me,
and the guy is just like sitting there,
<VOX in the middle of the roa=d,
and he's not moving,
and,
.. you know I wanna park the [car] VOX>.
```

(SBC006 *Cuz*)

Note to (5): *like* is unstressed and pronounced with leveled intonation. *Like* marks important information which is related to the speaker not being able to park.

(6)
```
we joined a book club.
And I got it for hardly any money,
you know,
they'll send you all these free books,
all you have to do is do postage,
(H) and then supposedly,
you don't have to u=m (Hx),
.. (TSK) (H) buy any more books,
but they want you to (Hx).
[So I got like f=]ive of em.
```

(SBC052 *Oh you Need a Breadbox*)

Note to (6): *like* is unstressed and pronounced with leveled intonation but *five* carries additional stress. The speaker wants to draw focus to the fact that despite hardly paying anything, she received five books.

(7)
```
it was just like,
everybody was real ... friendly and every[thing],
```
(SBC047 *On the Lot*)

Note to (7): *like* is unstressed and pronounced with leveled intonation. The speaker recently terminated a relationship with a woman. The addressee has asked whether her family were heartbroken over the break up. The speaker recently met her family and as (7) shows, he highlights the information about them being good to him.

(8)
```
First she hires me n- like,
.. the Friday before school starts.
(H) ... And expects me to get my room ready,
... (H) and then=,
... and then I find out on Thursday,
in the first week of school,
that I might lose my jo=b.
```
(SBC004 *Raging Bureaucracy*)

Note to (8): *like* is unstressed and pronounced with leveled intonation. The speaker highlights *Friday* as important information to make sure the addressees notes this because it is important so they can understand how little time passed from when she was hired to when she learned she might lose her job.

(9)
```
        there was a pr- --
        a Russian prostitute,
        in Victor's,
        on Wednesday night.
PATRICK: ... Really,
        I didn't see her.
CORINNA:        ... She was j- --
        She was like standing right by us.
```
(SBC045 *The Classic Hooker*)

Note to (9): *like* is unstressed and pronounced with leveled intonation. Upon learning that Patrick did not see the Russian prostitute, Corinna highlights he should have seen her because they were so close to her.

(10)
```
CORINNA: It's like,
         if you go in the summertime,
         I m- mean it's just l=ined with em.
```
(SBC045 *The Classic Hooker*)

Note to (10): *like* is unstressed and pronounced with leveled intonation. In (10), Corinna has just talked about one place where there are many prostitutes. As (10) shows, *like* is used for highlighting when there are particularly many prostitutes in that particular area.

Quotative (*like*)

(1) WENDY: [I was gonna ask the doctor,
I'm **like**,
.. wh]=at is wr=ong with me,
that I am sleeping so much.

(SBC013 *Appease the Monster*)

Note to (1): *like* is unstressed and pronounced with leveled intonation. The quotation belongs to a separate tone unit. Wendy quotes something on the lines of what she had planned to ask the doctor but it never happened.

(2) KEVIN: And he's **like**,
no=,
I] [2just want you to smell it2].

(SBC013 *Appease the Monster*)

Note to (2): *like* is unstressed and pronounced with leveled intonation. The quotation belongs to a separate tone unit. Kevin quotes what someone else said to him.

(3) CAM: I'll look at some of the houses,
I'm **like** ooh,
is it this one,
is this one it,
.. and then I just- --
Oh no,
that's the one.

(SBC044 *He Knows*)

Note to (3): *like* is unstressed and pronounced with leveled intonation. Cam quotes what his thoughts.

(4) LISA: ... I'd be **like** Mom,
uh,
what should we do with this child.

(SBC036 *Judgemental on People*)

Note to (4): *like* is unstressed and pronounced with leveled intonation. Lisa quotes what she would have said in a hypothetical situation.

(5) Her mom's like an alcoholic,
and stuff like that,
and her dad's like a <VOX biker VOX> now,
and I'm **like**,
God,

(SBC036 *Judgemental on People*)

Note to (5): *like* is unstressed and pronounced with leveled intonation. The speaker quotes her feeling of hopelessness towards the parental situation of someone.

(6)
```
and I'm like,
okay=,
this is kinda cool,
```

(SBC045 *The Classic Hooker*)

Note to (6): *like* is unstressed and pronounced with leveled intonation. Quotation about what the speaker thinks about in a specific situation.

(7)
```
... He was like that's fine.
And I'm like fine,
there just won't be our night then.
<P And he was like P>,
... okay,
if that's what you want.
```

(SBC043 *Try a Couple Spoonfuls*)

Note to (7): *like* is unstressed and pronounced with leveled intonation. The speaker renders a conversation between her and someone else. The speaker switches between *BE* in the present and past tense.

(8)

```
ANNETTE: .. Dad there's a little left for you and Lou.
         .. (SNIFF) ... But,
         it's like I'm not gonna be here on your birthday,
         is that <@ okay?
         I'm like I'm not @> gonna [be here] on my birthday.
```

(SBC043 *Try a Couple Spoonfuls*)

Note to (8): *like* is unstressed and pronounced with leveled intonation. Annette uses the pronoun *it* in relation to the *like* in bold. The quote is meant to represent her dad. The reason why she uses *it* instead of *he* could be because he has never actually uttered the quote and *it* distances her father from the quote.

(9)
```
         (H) Something she just said,
         <VOX well Jesus is in your heart,
         not on your forehead VOX>.
         Or something.
MARIE:   @ (Hx)
         [And I was just like] --
KEVIN:   [@@@@@]@@@@[2@@2]@
```

LISA: [2@2]

(SBC036 *Judgemental on People*)

Note to (9): *like* is unstressed and pronounced with leveled intonation. Marie uses her body to express her reaction to what someone told her. She is not cut off because the addressees do not start to laugh until after she has said *like*.

(10) So we were **like** no,
seniors,

(SBC050 *Just Wanna Hang*)

Note to (10): *like* is unstressed and pronounced with leveled intonation. The speaker quotes what she and some friends said even though it is possible that only one person said the quote but answering behalf of the others.

Rag bag (*like*)

(1) KEVIN: [4That's why,
cause that's4] --
WENDY: [5You have to **like**5] --
KENDRA: at all5]=.

(SBC013 *Appease the Monster*)

Note to (1): Kevin, Wendy, and Kendra all fight for the floor. Wendy gives up and stops talking before she said what she meant to mark by *like*. It is, therefore, impossible to categorize the DM.

(2) ROY: [**Like** they're],
.. they're @trying @to bree=d,
.. <VOX ho ho= VOX>,
(H) .. they're trying to breed like a forty foot
long tube chicken?

(SBC003 *Conceptual Pesticides*)

Note to (2): There are so many false starts after the *like* in bold and the token is probably a part of the reparandum given the second *like* which is not in bold in the same syntactic position as the first one.

(3) ROY: [2this is **like**,
MARILYN: UC Davis2].
ROY: H) this2] is like some kind of horrific nightmare.

(SBC003 *Conceptual Pesticides*)

Note to (3): Roy recycles the first line in the repair. Although it is possible to determine the function of the *like* in bold based on the repair, recycles have not been counted in this study.

(4) ... but they'll say it really,
.. **like** pretty clear,

(SBC036 *Judgemental on People*)

Note to (4): *like* is used for transitioning from a reparandum to a repair. This is the only time *like* is used for such a purpose in this corpus. Therefore, this study does not address the use of *like* for marking a transition into a repair.

(5) JEFF: **Like** could you get –
Is it possible] that you could still be posi- .. positive?

(SBC028 *Hey Cutie Pie*)

Note to (5): *like* is a part of a false start.

(6) .. [It's **like**] --
JEFF: [It is]?
JILL: Unhunh.
Like seventy degrees,

(SBC028 *Hey Cutie Pie*)

Note to (6): the *like* in bold is uttered simultaneously as Jeff talks. Jill continues to talk and it becomes apparent that the first *like* was supposed to be used for an approximation but, as mentioned above, the reparandum of a recycle are not counted in this study.

(7) JEFF: .. You know **like** in,
like --
.. (H) you know **like** you're ch-,
.. u=m,
.. just- % --
even with people.

(SBC028 *Hey Cutie Pie*)

Note to (7): all the three occurrences of *like* are parts of false starts.

(8) ... (H) And then carjack em.
... And then **like like**,
@ (H)
and then so they got up on

(SBC036 *Judgemental on People*)

Note to (8) *like* occurs within a reparandum.

(9) FRED: Take a shower.
Why not.
Because,
[**like**],
RICHARD:[Get] there at seven and,
play from seven to quarter to nine

(SBC047 *On the Lot*)

Note to (9) Richard takes the floor before Fred utters what was related to *like*.

(10) b- you know the **like** s- fall- --

(SBC043 *Try a Couple Spoonfuls*)

Note to (10): the speaker speaks disfluently and never utters anything related to *like*.

Non-DM (*like*)

(1) .. I] didn't like the book

(SBC005 *A Book About Death*)

(2) ... And I f=eel like I'm in a spaceship.

(SBC005 *A Book About Death*)

(3) You know if you ... put a situation like that to ~Tim or ~Mandy,

(SBC007 *A Tree's Life*)

(4) ... [2what it2] would be **like** to ... train her

(SBC007 *A Tree's Life*)

(5) KENDRA: [3What is this3].
... s- Feels **like** a pair and a spare,

(SBC013 *Appease the Monster*)

(6) MARILYN:[Just **like** Dickens].

(SBC003 *Conceptual Pesticides*)

(7) (H) <VOX I always **like** it when,
(H) both .. of you .. are married to each o[ther] VOX>.

(SBC006 *Cuz*)

(8) KEVIN: [3How many3] .. people did they [4get **like** that.

(SBC036 *Judgemental on People*)

(9) ... To me it was **like** a game,

(SBC051 *New Yorkers Anonymous*)

(10) me and about five other people,
drove around to all these houses,
to see what they look **like**.

(SBC052 *Oh you Need a Breadbox*)

Previous knowledge (*you know*)

(1) (H) you know like how we have always talked about life being
Out there,
.. [**you know**,
JILL: [Unhunh]
JEFF: like there] has to be.

(SBC028 *Hey Cutie Pie*)

Note to (1): *you know* does not carry any stress. Small rising intonation on *you know* to signal he is not finished talking. He introduces familiar background information to the main point in the last line.

(2)
```
Cause there's planets evolving around,
[(H) um],
JILL:   [Unhunh].
JEFF:   (SNIFF) you know various stars and stuff.
Just like ours.
```

(SBC028 *Hey Cutie Pie*)

Note to (2): *you know* does not carry any stress. Falling intonation on *you know* to signal end of turn. (2) is a continuation of the example in (1) directly above. *You know* introduces familiar background information to support his belief that there has be to life on other planets.

(3)
```
        ... (TSK) (H) I just knew I didn't like the other guy.
        .. Who was --
        .. Dave Cargo.
KEVIN:  ... [Ye=],
LISA:   [It was like],
KEVIN:  oh my,
        <X he gets a X> --
        You know.
LISA:   .. Yeah.
```

(SBC036 *Judgemental on People*)

Note to (3): *you know* does not carry any stress. *You know* is pronounced with falling intonation. Dave Cargo is a politician. Kevin knows he has Lisa share the same knowledge about and he does not need to finish his utterance. Lisa confirms she knows what he means.

(4)
```
        We don't have ca]ble.
        That's the only thing that sucks.
DANA:   ... (TSK)
NANCY:  ... [So we don't have anything] --
DANA:   [We have a ster]eo.
NANCY:  .. Yeah.
DANA:   [You know],
ARIANNA:        [That's true].
NANCY:  [You're right].
```

(SBC050 *Just Wanna Hang*)

Note to (4): *you know* does not carry any stress. The DM is pronounced with leveled intonation. The girls are planning a party. Despite not having cable for entertainment, Dana reminds them they have a stereo.

(5)
```
... I put down on the card,
you know,
no cases.
Because it was lost time.
```

(SBC047 *On the Lot*)

Note to (5): *you know* does not carry any stress. The DM is pronounced with leveled intonation. He mentioned this earlier and marks the repetition.

(6) .. **you know**,
once it comes along with experience,
and the more people I work with,
the easier it will,
it'll be,

(SBC047 *On the Lot*)

Note to (6): *you know* does not carry any stress. The DM is pronounced with falling intonation. It is common knowledge that things become easier with experience.

(7) she al[4ways was,
FRED: [4Yeah,
RICHARD: **you know**4],
FRED: .. exac4]tly.
RICHARD: ... (H)= pretty much uh,
... able to do anything that I wanted to do.
She was never negative or anything and uh,
it was basically me=, you know going out.
The problem going out.

(SBC047 *On the Lot*)

Note to (7): *you know* does not carry any stress. The DM is pronounced with leveled intonation. *You know* introduces background information that the addressee already has to support a claim that Richard is to blame for them terminating their romantic relationship.

(8) ... He's only paid for it.
You know.

(SBC045 *The Classic Hooker*)

Note to (8): *you know* does not carry any stress. The DM is pronounced with falling intonation to signal end of turn. In (8) the speaker refers to a guy who has not been with other women but prostitutes. He has mentioned this earlier in the discourse and marks that he repeats something the addressee should remember from earlier.

(9) MICHAEL: And gangs are coming back,
now,
you know?

Note to (9): *you know* does not carry any stress. The DM is pronounced with rising intonation to signal he has more to say. Prior to what is rendered in (9), Michael talks about how history repeats itself and that some people always have the power, but that the people in power

changes through history. To support his statement, he plays on the addressee's knowledge of gangs gaining more power. They are in southern California in the late 1990's or early 2000's. At this time, Los Angeles was gang ridden.

(10)
```
                .. this darn dog keeps ... breathing,
                <X and like X> ... dreaming,
                <X you know X> I wonder if we should wake her up?
MARY:           ... No,
                .. she'll get scared and want to go outside.
                ... Kinda nervous,
                you know.
```
(SBC007 *A Tree's Life*)

Note to (10): *you know* does not carry any stress. The DM is pronounced with falling intonation. Mary brings up the addressee's knowledge that the dog is nervous as background information to support her claim that it is not a good idea to wake up the dog.

Reassuring the addressee (*you know*)

(1)
```
PAMELA:         You know,
                (H) and I h- --
                I bit my tongue the other day,
```
(SBC005 *A Book About Death*)

Note to (1): *you know* does not carry any stress. The DM is pronounced with falling intonation. Pamela tells the addressee something he did not know

(2)
```
MARILYN:        .. You know,
                they eat it,
                ... when they're up there,
```
(SBC003 *Conceptual Pesticides*)

Note to (2): *you know* does not carry any stress. The DM is pronounced with a small rising intonation. Marilyn lets her addressees know that her parents eat salmon they catch when they are on fishing trips.

(3)
```
                we went to see Oba Oba.
LENORE:         .. @[@@]
ALINA:              [You know],
                !Ruben loved it.
```
Note to (3): *you know* does not carry any stress. The DM is pronounced with falling intonation. Alina introduces unfamiliar knowledge to Lenore.

(4)
```
                Uh .. !Cathy,
                don't you understand that,
                .. you know,
                things are different now,
```
(SBC006 *Cuz*)

Note to (4): *you know* does not carry any stress. The DM is pronounced with falling intonation. The speaker talks as if she addressed Cathy, who is not present. The speaker believes Cathy does not understand that things have changed, and in (4) she speaks as if she reassures her that things are different.

(5)
.. (H) They would see how= .. things were done and,
he- --
... He like took them around .. (TSK) the world.
.. (H) **You know** doing this.

(SBC044 *He Knows*)

Note to (5): *you know* does not carry any stress. The DM is pronounced with falling intonation. The speaker knows the addressee has no knowledge about this.

(6)
I took that as a pretty good (H) sign.
JEFF: ... Oh [my god].
JILL: [**You know**].

(SBC028 *Hey Cutie Pie*)

Note to (6): *you know* does not carry any stress. The DM is pronounced with falling intonation. Jill reassures Jeff about how she interpreted the given situation.

(7)
he just wants to be up in the mountains,
he wants to write,
and just,
(H) .. you know,
... just (H) (Hx),
you know,
get in tune with him- .. with himself.

(SBC028 *Hey Cutie Pie*)

Note to (7): *you know* does not carry any stress. The DM is pronounced with falling intonation. The speaker lets the addressee know what a friend wants to do.

(8)
... (TSK) like Cargo'd just kinda be like,
.. (TSK) (H) <VOX all these low-lifes VOX>.
You know

(SBC036 *Judgemental on People*)

Note to (8): *you know* does not carry any stress. The DM is pronounced with falling intonation. The speaker reassures the addressee that she believes Cargo, a politician, would look down at other people if he were elected.

(9)

it's not that they weren't Christian at all.
.. **You know,**
(H) it was just,
it annoyed them.

(SBC036 *Judgemental on People*)

Note to (9): *you know* does not carry any stress. The DM is pronounced with falling intonation. The speaker reassures the addressees that the reason some costumers did not like certain religious things is not because of their religious view.

(10)

(H) You say,
you know,
how's it going.
He'll te[ll you]

(SBC051 *New Yorkers Anonymous*)

Note to (10): *you know* does not carry any stress. The DM is pronounced with falling intonation. The speaker reassures the addressees that you will get an honest question when talking to a New York taxi driver.

Elaboration (*you know*)

(1)

I want it to be homemade.
You know,
something special.

(SBC003 *Conceptual Pesticides*)

Note to (1): *you know* does not carry any stress. The DM is pronounced with a small rising intonation. The speaker elaborates on why she wants something to me homemade.

(2)

... and it's just really a @scary @book.
You know.
(H) There's (H) no natural world left.

(SBC003 *Conceptual Pesticides*)

Note to (2): *you know* does not carry any stress. The DM is pronounced with falling intonation. The speaker elaborates on what makes the book scary.

(3)

they said they'd come over here.
NANCY: ... Mm.
DANA: **You know.**
... If we wanna just hang out and drink.

(SBC050 *Just Wanna Hang*)

Note to (3): *you know* does not carry any stress. The DM is pronounced with falling intonation. Dana elaborates on why someone would like to visit them.

(4)
```
PETE:    [Sort of like] Tahitian dancers or some[2thing.
HAROLD:  [2Yeah=2].
PETE:          <X You know X>2],
               close.
```

(SBC002 *Lambada*)

Note to (4): Prosody has not been marked in this example because Pete and Harold talk simultaneously, which makes it hard to hear how *you know* was uttered. Pete asks if lambada dancers dance like Tahitian dancers. It is unclear what he means by it, so he elaborates to say what he wonders is whether lambada dancers dance close together.

(5)
```
... twenty minutes later,
they were kinda like .. all over each other.
You know.
... kissing,
et cetera,
et cetera.
```

(SBC002 *Lambada*)

Note to (5): *you know* does not carry any stress. The DM is pronounced with falling intonation. The speaker elaborates on what he means by *all over each other*.

(6)
```
ALICE:  ... Cause Beth.
        ... You know,
        .. [my friend,
```

(SBC051 *New Yorkers Anonymous*)

Note to (6): *you know* does not carry any stress. The DM is pronounced with falling intonation. Alice realizes that it might not be apparent who Beth is and adds that she is her friend.

(7)
```
        ... MD Anderson,
        .. you know,
        the children's,
        (H) .. uh,
        MD Anderson cancer ... [thing,
CINDY:  [Oh].
DARLENE:       in] Houston?
```

(SBC052 *Oh you Need a Breadbox*)

Note to (7): The prosody of this DM has not been analyzed because the quality of Darlene's microphone is poor. Darlene realizes Cindy may not know who MD Anderson is, so she gives additional information.

(8)
```
and that she would .. give me that chance,
```

you know,
FRED: ... [To- go- --
RICHARD: [to go back with her,
FRED: .. To go back],
RICHARD: and try to m]ake our life work together.

(SBC047 *On the Lot*)

Note to (8): *you know* does not carry any stress. The DM is pronounced with a small rising intonation. The speaker elaborates on what chance a woman gave him.

(9)
... that just teaches the third-grader,
with the lesser intelligence,
that,
(H) .. that he's worthless,
you know,
that he can't learn [stuff on his own].

(SBC004 *Raging Bureaucracy*)

Note to (9): *you know* does not carry any stress. The DM is pronounced with falling intonation. The speaker elaborates on what the feeling of worthlessness is rooted in.

(10)
(H) me and Janine and all of us were walking dow=n the hill.
... **You know,**
going dow=n,
... where the pier is?

(SBC045 *The Classic Hooker*)

Note to (10): *you know* does not carry any stress. The DM is pronounced with falling intonation. The speaker elaborates on which hill they walked down.

Self-repair

(1)
... Now if i- --
You know if you ... put a situation like that to ~Tim or ~Mandy,

(SBC007 *A Tree's Life*)

Note to (1): *you know* does not carry any stress. The DM is pronounced with falling intonation. The speaker changes the pronoun from first to second person.

(2)
... well I called,
... **you know,**
this thing was going on with !Buck and everything,
... so I called um,
... ~Mandy's,

(SBC007 *A Tree's Life*)

Note to (2): *you know* does not carry any stress. The DM is pronounced with falling intonation. The speaker decides to add some background information before letting the addressee know she called Mandy.

(3)
```
It was,
.. you know.
... (H) It was cel[ebration cake,
```
(SBC013 *Appease the Monster*)

Note to (3): *you know* does not carry any stress. The DM is pronounced with falling intonation. *You know* is used as the transition into a the recycle.

(4)
```
.. so I didn't have to haggle about what kind of tires,
or where to k_ __
you know,
.. put em,
front or back,
```
(SBC013 *Appease the Monster*)

Note to (3): *you know* does not carry any stress. The DM is pronounced with falling intonation. Only *k* _ __ is a part of the reparandum.

(5)
```
and then it gets down to s- --
you know,
I mean that brings into play,
other people's remarks
```
(SBC003 *Conceptual Pesticides*)

Note to (5): *you know* does not carry any stress. The DM is pronounced with leveled intonation. The first line in (5) is a false start and *you know* signals this should be replaced with what follows the token.

(6)
```
.. and I- --
You know,
of course it's this long drive,
so I --
I probably look like ... total hell,
```
(SBC003 *Conceptual Pesticides*)

Note to (6): *you know* does not carry any stress. The DM is pronounced with a small rising intonation. The speaker begins to say something before realizing additional information is needed to understand why what caused the bad-looking appearance.

(7)
```
well I thought you'd,
.. you know=,
m- I didn't think you were coming ~Alina
```
(SBC006 *Cuz*)

Note to (7): *you know* does not carry any stress. The DM is pronounced with falling intonation. *You know* marks the transition from reparandum to repair.

(8)
```
ALINA:  .. (TSK) A- --
        .. You know,
        at least that's not my shtick.
```
(SBC006 *Cuz*)

Note to (8): *you know* does not carry any stress. The DM is pronounced with falling intonation. Alina hesitates when she pronounces *A- --*. *You know* marks a transition into a repair where the utterance is fluidly pronounced.

(9)
```
        ... But,
        % you know,
        cause nobody's up there now?
```
(SBC028 *Hey Cutie Pie*)

Note to (9): *you know* does not carry any stress. The DM is pronounced with falling intonation. *You know* marks that the speaker substitutes one conjunction for another in the repair.

(10)
```
(H) It's such a great,
... you know.
.. (H) That would be so great if people could get that.
```
(SBC028 *Hey Cutie Pie*)

Note to (10): *you know* does not carry any stress. The DM is pronounced with falling intonation. *You know* marks what was previously said was a false start and the speaker wants to repair it.

Rag bag (*you know*)

(1)
```
PAMELA:     the illusions of this life,
            .. (H) you know,
            I --
            (H) .. % I,
            % I,
DARRYL:     X X [X],
PAMELA:         [<VOX my] favorite word when I was twelve VOX>,
            ... was paradox.
```

Note to (1): Pamela fumbles too much after uttering *you know*, which makes it difficult to say why she uttered the token.

(2)
```
MARY:       ... you know and,
```

.. especially the way um,
... I mean ~Tim gets .. in- .. himself into a=,
uncomfortable situation or whatever,
.. (H) and his first reaction is to blow up about
it.

(SBC007 *A Tree's Life*)

Note to (2): *you know* is inside the reparandum which is marked by *I mean*.

(3)

PETE: [8Yeah8].
.. You know,
.. the early man probably said ... the same thing
about the first domestic chicken.
MARILYN: ... @
ROY: .. <@ **you know** @>.
MARILYN: .. (H) .. <VOX Can you imagine,

(SBC003 *Conceptual Pesticides*)

Note to (3): Roy never gets the floor.

(4)

MARILYN: .. This is typical.
PETE: .. Yeah.
.. You know,
ROY: ... And --
and then it gets down to s- --
you know,
I mean that brings into play,
other people's remarks like,
(H) ... you know,
the --
... w- it's all fine [and good for us in this]
ge[2neration2],

(SBC003 *Conceptual Pesticides*)

Note to (4): Pete does not get the floor, which Roy takes.

(5)

MARILYN: [2Well not if you were care2]ful.
... I mean,
... **you know,**
if you --
if you --
... if you didn't have,
... if you were really careful about your [dishes],

(SBC003 *Conceptual Pesticides*)

Note to (5): Marilyn fumbles too much.

(6)

(Hx) you **you know,**
oh[= you] --
LENORE: [you space out] on it.

(SBC006 *Cuz*)

Note to (6): Lenore takes the floor before the Alina utters what is marked by *you know*.

(7)

that it's not a choice.
.. **You know,**
[because] --
LAJUAN: [They do] believe it now?

(SBC044 *He Knows*)

Note to (7): Lajuan cuts Cam off before he utters what he planned to mark with the DM.

(8)
```
(H) you know,
like I nailed him on the contradiction <X you know X>,
```

Note to (8): based on the audio file, it does not sound like the *you know* in bold is uttered at all.

(9)
```
... he doesn't let us put anything up.
Like as far as,
you know like um=,
KEVIN: ... Can't put anything Catholic up?
[Or nothing secul]ar.
```

(SBC028 *Hey Cutie Pie*)

Note to (9): Kevin takes the floor before what would have been marked by *you know* is uttered.

(10)
```
SHERRY: I thought they were l]ower.
ROSEMARY:      %r=ight],
... you know,
BETH:  .. Yeah that's what we [always thought too,
```

(SBC031 *Tastes Very Special*)

Note to (10): Rosemary does not succeed in taking the floor.

Non-DM (*you know*)

(1)
```
[2how do you know that2].
```

(SBC005 *A Book About Death*)

(2)
```
You know what it would be,
a real good lesson for them,
```

(SBC007 *A Tree's Life*)

(3)
```
... You know what I was thinking of doing?
```

(SBC007 *A Tree's Life*)

(4)
```
Oh and you know another thing that ~Tim had the audacity to bitch
about?
```

(SBC007 *A Tree's Life*)

(5)
```
you know= I need to get sleep over the weeke=nd.
```

(SBC013 *Appease the Monster*)

(6)
```
WENDY:  I'll let you know,
```

(SBC013 *Appease the Monster*)

(7) Did **you know** that?

(SBC013 *Appease the Monster*)

(8) Do **you know** what I'm talking a[bout],

(SBC013 *Appease the Monster*)

(9) [2**You know** what I mean2],

(SBC013 *Appease the Monster*)

(10) ARIANNA: [<X **You know** her X>].

(SBC050 *Just Wanna Hang*)