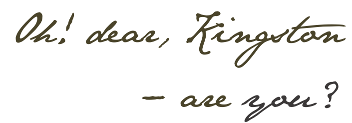
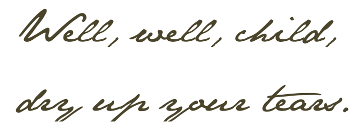
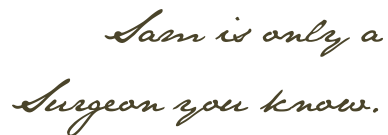
Oh, well, you know, do not you?:

Socio-historical discourse analysis of Jane Austen's fictional works







Name: Diane-Eugenie Jonker

Student number: 0907642

Date: 04-08-2014

Supervisor: Prof. I.M. Tieken-Boon van Ostade

Second reader: Dr. F.K. Ameka

The cursive font based on Jane Austen's handwriting is made available by the artist Pia Frauss.
URL: http://www.pia-frauss.de/fonts/ja.htm. Downloaded: 15-07-2014.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 4
1.1 Overview 4
1.2 Inspiration for and Relevance of the Study 4
1.3 Discourse Markers and Tag Questions: Past and Present 7
1.4 Research Variables, Questions and Hypotheses 9
1.5 Thesis Overview 12
CHAPTER 2: BACKGROUND LITERATURE 13
2.1 Introduction 13
2.2 The Language of Jane Austen 13
2.3 Research on Discourse Markers 15
2.3.1 Interjections: *Oh* and *Ah* 15
2.3.2 General Discourse Markers: *Well* and *Why* 18
2.3.3 Phrasal Discourse Marker: *You Know* 20
2.4 Research on Tag Questions 23
2.5 Concluding Remarks and Hypotheses 25
CHAPTER 3: METHODOLOGY 27
3.1 Introduction 27
3.2 Corpus 27
3.2.1 Explanation for the Corpus Selection 27
3.2.2 Preparation of the Corpus and Subcorpora 29
3.2.3 Sociolinguistic Categorization of the Utterances 32
3.3 Corpora Analysis 37
CHAPTER 4: RESULTS 41
4.1 Introduction 41
4.2 Interjections: *Oh* and *Ah* 41
4.2.1 Gender 41
4.2.2 Intimacy 43
4.2.3 Social Class 44
4.2.4 Setting 47
4.3 General Discourse Markers: *Well* and *Why* 48
4.3.1 Gender 48
4.3.2 Intimacy 51
4.3.3 Social Class 52

4.3.4 Setting....55
4.4 Phrasal Discourse Marker: *You Know*....56
4.4.1 Gender....56
4.4.2 Intimacy....58
4.4.3 Social Class....59
4.4.4 Setting....61
4.5 Tag Questions....62
4.5.1 Gender....62
4.5.2 Intimacy....64
4.5.3 Social Class....65
4.5.4 Setting....67
CHAPTER 5: CONCLUSION....68
5.1 Introduction....68
5.2 Main Findings....68
5.2.1 Interjections: *Oh* and *Ah*....68
5.2.2 General Discourse Markers: *Well* and *Why*....69
5.2.3 Phrasal Discourse Marker: *You Know*....69
5.2.4 Tag Questions....70
5.3 Discussion....71
5.4 Conclusion....73
BIBLIOGRAPHY....74

LIST OF ABBREVIATIONS

Cath	*Catharine, or the Bower.*
EModE	Early Modern English
LModE	Late Modern English
MP	*Mansfield Park*
NA	*Northanger Abbey*
P&P	*Pride and Prejudice*
PDE	Present-Day English
S	*Sanditon*
TW	*The Watsons*

CHAPTER 1: INTRODUCTION

[I]t was so shocking you know to come away without a change of Cloathes—
Why indeed said Kitty, it seems to have been a very shocking affair from beginning to end.

(*Catharine, or the Bower*, in Bree et al. 2013:190)

1.1 Overview

The present study will report on a socio-historical analysis of discourse markers (*oh*, *ah*, *well*, *why*, and *you know*) and tag questions by looking at the variables of gender, intimacy, social class, and setting in order to gain insight into the way in which these variables occur in Late Modern English speech as represented by the well-known author, Jane Austen (1775–1817), in her fiction. This chapter will place the topic into its wider context by explaining its relevance to current research on discourse markers and tag questions, and by formulating the chosen research variables, the research questions, and a preliminary hypothesis; it will also provide a brief outline of the structure of the paper.

1.2 Inspiration for and Relevance of the Study

Earlier quantitative research I did on the use of interjections by Jane Austen in her fiction, both the early writings (short stories and plays), referred to as the *Juvenilia* (Bree et al. 2013:11), and her later novels, showed that certain interjections were gender-dependent (Jonker 2014). *La* and *psha*, for instance, were only used by female characters, which confirmed Stokes's observation, based on a qualitative analysis of Austen's vocabulary in her published works, that *la* in particular characterized the speech of "affected/vulgar/ignorant and usually young women" (Stokes 1991:19). Examples may be found in the following sentences:

1. **La**! my dear […] it is not Lady Catherine. (Maria Lucas, *P&P*)[1]
2. **Psha**! my dear creature […] do not think me such a simpleton as to be always wanting to confine him to my elbow. (Catherine Morland, *NA*)

Hey, on the other hand, which Columbus (2009:404) describes as an "invariant tag [or] interjection-based discourse marker" proved a specifically male interjection:

3. A pretty good thought of mine, **hey**? (John Thorpe, *NA*)

[1] For my analysis of Jane Austen's published novels, I drew on digitized versions of the texts found in Project Gutenberg.

La, which is still used in Singapore English today, is according to Columbus, an interjection-based discourse marker with different pragmatic functions: it can be a marker of emphasis, contrast, persuasion, and approval, to name a few (Columbus 2009:405).

Studies on gendered language, as McElhinny (2003:22) argues, have been conducted by analyzing cross-sex and same-sex interaction as these interactions were argued to be the most salient in language and gender analysis (Brown and Levinson 1987:53). Furthermore, the language of the speaker may vary depending on the addressee(s), as Giles proposes in his "speech [or communication] accommodation theory" (Swann 2009:150). In addition, speakers may vary their speech to adapt to the speech style of the addressee or to emphasize their own accent, dialect, style, or even social status (Swann 2009:150–151). Social status plays an important role in a person's speech style, as appeared from studies in interactional sociolinguistics and linguistic ethnography (Swann 2009:183). Not merely social status, but also the intimacy between speakers and the context, or setting, plays a part in the way in which people communicate. Brown and Levinson's (1987:17–22, 91) politeness model applies to studies in language as speakers manage their 'face' (i.e. their public self-image) according to their relation with the addressee. Thus, the speaker-addressee relationship is of importance when analyzing speech (Swann 2009:188).

Jane Austen has been celebrated for her language use (Phillipps 1970:11), specifically the conversations in her novels (Page 1972:116–117; Burrows 1987:2). Wyld's *A history of modern colloquial English* (1921) describes Austen's language as:

> [T]he representation of actual life and dialogue as the author knew it. There can be no doubt that this is the real thing, and that people really spoke like this in the closing years of the eighteenth century […] It is not Miss Austen who is speaking, it is the men and women of her day (1921:185).

Discourse analysis, formerly primarily studied through analyzing spoken interactions, has increasingly been applied to written texts during the last twenty years in order to understand interactions of earlier periods for which no spoken data was available (Taavitsainen and Fitzmaurice 2007:18; Jucker 2008:895; Lutzky 2012:1–2). Spoken language as recorded in writing is merely an approximation of what speech was like in the past, but it is the only gateway into the past for historical discourse analysis (Rissanen 2008:60; Lutzky 2012:45–48). Wyld praised Austen's writing skills for the accuracy of the dialogues. Additionally, Burrows's *Computation into Criticism* (1987) demonstrates by analyzing high-frequency words such as personal pronouns (e.g. *I*, *you*, *we* and *us*), adverbs (*quite* and *very*), prepositions (e.g. *in* and *of*), (in)definite articles (*the* and *an*), auxiliaries (e.g. *have* and *be*), and conjunctions (e.g. *and*) in several statistical analyses that

Austen's characters are undeniably different in their use of these common words: "the differences of incidence from character to character are too pronounced to be ignored", Burrows writes, adding that "the patterns [of the prepositions *of* and *in*] indicate that […] the idiolects of many of Jane Austen's major characters are firmly and appropriately differentiated" (Burrows 1987:57, 69). Burrows's findings show that the (major) characters of Jane Austen's novels have their own idiolects, which provides proof for Wyld's claim that Austen's dialogues are like those she heard around her.

The present study, focusing on discourse markers and tag questions, relates to pragmaphilology, a strand of historical pragmatics which deals with the "contextual aspects of historical texts" such as the speaker and addressee, their relationship both social and personal, the setting of text production and reception, and the specific goal of the text (Jacobs and Jucker 1995:11; Lutzky 2012:3). In addition, the study is placed within the realm of variationist sociolinguistics. Pichler (2013) notes that studies "which have systematically correlated their use with contextual predictors" are modest in number and this observation led to the conclusion that discourse variation as a study is still at an early stage (2013:3). In order to conduct a pragmaphilological and sociolinguistic variationist study on late eighteenth- and early nineteenth-century discourse, I will study Austen's praised dialogues and the idiolects of the statistically distinctive characters as these allow for an interesting as well as important source of conversations during that period. As Lutzky (2012:47) stresses, speech-based genres, such as drama, are those which allow us to see spoken language as it was centuries past, since these "record or mimic spoken language [and] show conversational and colloquial everyday language features [here, discourse markers and tag questions] and therefore mirror the oral register of earlier language varieties more closely than others". Lutzky (2012), which focuses on Early Modern English (EModE), studies the discourse markers *marry*, *well*, and *why* in terms of their collocation and functions both structurally and interactionally by analyzing several subcorpora. The present study will take a snapshot of Late Modern English (LModE) in this respect through a selection of Austen's works during a twenty-five year period (1792–1817) (see further Chapter 3). Moreover, it will correlate the findings with certain selected sociolinguistic variables (e.g. gender) in order to see how these variables influence the use of discourse markers and tag questions. In some ways, it will be a further step into the world of discourse markers and tag questions, and their socio-historical history in the English language.

1.3 Discourse Markers and Tag Questions: Past and Present

The study of discourse markers is a relatively young area of research, as pragmatics is in general. In the past, discourse markers were referred to by many different terms, such as "pragmatic expressions" (Erman 1987), "discourse connectives" (Blakemore 1987; 1992), "pragmatic markers" (Andersen 1998), and "discourse particles" (Aijmer 2002). Currently, the term 'discourse marker' has increasingly been used to refer to items such as *well* and *you know*. The precise definition of the term, however, is still unclear. Discourse markers could refer to items having solely a structural function in announcing a change in topic or as having an interactional function to show hesitation or as a stall for a speaker when formulating sentences. Blakemore (2002:1) argues that the *discourse* part of a discourse marker pertains to how the discourse marker should be interpreted, i.e. at the level of the conversation and not merely the propositional sentence, whereas the word *marker* refers to the meaning of a discourse marker which pertains to what it indicates (e.g. hesitation or conviction). The function of discourse markers varies from context to context, but as Crystal (1988:48) puts it, discourse markers allow a speaker to smoothly and efficiently "perform the complex task of spontaneous speech production and interaction", which is about as general as the definition of a discourse marker's function gets.

Discourse markers, as proposed by Blakemore (2002:94–95), should be classified with regard to their cognitive effect (e.g. strengthening an assumption or contradicting/eliminating an assumption). Furthermore, Blakemore suggests that discourse markers are categorized in coherence-based approaches, i.e. how they relate to the preceding and succeeding phrase: for example, *so* and *therefore* are regarded as indicating causal relations and *but* and *however* as expressing contrastive relations (Blakemore 2002:161). However, whether certain words or phrases can be classified as discourse markers might differ from linguist to linguist depending on the preferred categorization and definition. Apart from this, there are certain characteristics that are generally accepted, such as optionality (syntactically and semantically optional), flexibility (position in a phrase), connectivity (relationship between current and previous discourse) and multi-grammaticality (as a discourse marker's grammatical category can range from adverbials to interjections) (Huang 2011:23–25).

The study of discourse markers in older stages of the language is relatively young. Defour (2007) and Lutzky (2012) are two of the most comprehensive studies I have been able to find that deal with language from the LModE and EModE periods, respectively. Most research has been conducted on actual spoken interactions or on collections of transcribed speech as opposed to fictional speech (see, for example, Östman 1981; Erman 1987; Schriffin 1987; Crystal 1988; Aijmer 2002, 2009; Müller 2005; Columbus 2009; Huang 2011; Koczogh and Furkó 2011;

Popescu-Belis and Zufferey 2011; and Pichler 2013). Recently, research has been done on historical discourse (see Taavitsainen and Fitzmaurice 2007; Jucker 2008; Rissanen 2008; Person 2009; Lewis 2011; Lutzky and Demmen 2013).

Similar to discourse markers, tag questions are a relatively new topic of study. The origin of research into tag questions lies in the study of the differences between male and female speech. Lakoff (1973) argued that the "marginality and powerlessness of women is reflected in both the ways women are expected to speak [e.g. by using tag questions], and the ways in which women are spoken of" (1973:45). Today, the powerlessness of the use of tag questions is still maintained in that these type of questions have a negative impact on the sociability, credibility and even trustworthiness of the speaker (Blankenship and Craig 2007:112). However, tag questions have been shown to occur in a context involving power relations, such as when doctors, lawyers, and judges interact with patients, clients, and witnesses in court, respectively (2007:113). This shows that Lakoff's assumptions about women particularly using tag questions was not entirely correct. In fact, Blankenship and Craig's (2007) research shows that credible speakers who provided strong arguments and used tag questions came across as more favorable than credible speakers that did not use tag questions (2007:115–116).

Just like discourse markers, tag questions have been analyzed with regard to their function. Especially their prosody (i.e. intonation and stress) is important for interpreting their discourse function: "[the] relationship between the final intonational contour of the tag and discourse function [and the] dependency between intonational phrasing and interpretation" (Reese and Asher 2007:448). Syntactically, tag questions are the reduced inversions of main sentences:

4. You are pregnant, **aren't you**?

In this example, there is a change in polarity, from 'positive' to 'negative', e.g. the main sentence is positive and the tag question is negative because of *n't*, which can either indicate uncertainty or confidence depending on the intonation (Collins and Mees 1984:234). However, it is also possible that the polarity does not change, i.e. the main sentence and the tag question are positive, which is either used to "respond to recently acquired information or [to voice hostility]" (Collins and Mees 1984:234):

5. She did it, **did she**?

Syntax may play a role in determining whether the tag question is an assertion rather than a question, but the main way to interpret a tag question correctly is by evaluating its intonational phonology, referred to above.

Tag questions such as *aren't you?* in (4) and *did she?* in (5) are canonical tag questions. There are also what Columbus (2009) describes as invariant tags, like John Thorpe's *hey?* in example (3) above, which have the same function as the canonical ones. While invariant tags are regarded as discourse markers, canonical tag questions are not, but there is some evidence that these too are discourse markers. Stenström (1994:96) argues that tag questions may function as discourse markers, as does Komar (2007:51). Similarly, Andersen (2001) believes that certain tag questions are "non-turn-yielding" tags and are therefore more like discourse markers than actual propositional tag questions, which canonical tag questions generally are (2001:135). Furthermore, Pichler (2013:5–6) argues that tag questions should be added to the analysis of discourse markers in general as "the exclusion of comment clauses and clause-final tags from the category of discourse-pragmatic features is theoretically unjustified and counter-productive to developing a comprehensive theory of linguistic elements that function primarily in the non-referential domains of language use".

1.4 Research Variables, Questions and Hypotheses

As Austen's fiction is of considerable size, for the purpose of this study a selection was made of her short stories and novels. The first two markers I decided to analyze are the interjections *oh* and *ah*:

6. **Oh**! I do not mind it. (Charles Blake, *TW*, in Bree et al. 2013:287)

7. Irish! **Ah**! I remember—and she is gone to settle in Ireland.
   (Mr. Edwards, *TW*, in Bree et al. 2013:282)

In my earlier paper on the topic, I found that these interjections were the most frequent in both Austen's *Juvenilia* and her later writings. Therefore, analyzing these interjections would seem most fruitful as they provide enough data for a more elaborate analysis. *Oh* has been extensively studied in real spoken speech (Aijmer 1987; Schiffrin 1987; Fox Tree and Schrock 1999; Müller 2005; Huang 2011); for written discourse, however, there has not been much research on the topic. *Oh*s have been analyzed during the EModE period (Person 2009; Taavitsainen 1995); *ah*, however, has not been studied to the same extent as *oh* (Aijmer 1987).

Two other discourse markers I selected for the analysis are *well* and *why*, and I will use Lutzky's (2012) study on LModE as a model for this. *Well* and *why* are referred to as general discourse markers, since *oh* and *ah* are interjection-based discourse markers:

8. **Well**, Fanny, and how do you like Miss Crawford now? (Edmund Bertram, *MP*)

9. **Why**, child, I have but this moment escaped from his horrible mother.
(Julia Bertram, *MP*)

Well and *why* are not discourse markers in every instance in which they occur. It is important to note these different uses of *well* and *why*, which is what makes their position and relation to the sentence very important, as they do not always function as discourse markers and should therefore be excluded from the analysis when we look at the distribution and frequency of markers. In example 10, *well* functions as an adverb and in example 11, *why* functions as an interrogative adverb:

10. That's **well** thought of. (Fanny Price, *MP*)

11. **Why** should you dare say that? (Edmund Bertram, *MP*)

In British English, *why* is almost exclusively an interrogative adverb, whereas in American English, *why* as a discourse marker is relatively common (Jucker 2002:221). Culpeper and Kytö (2010) studied the occurrence of pragmatic *why*s across genders and within the EModE period and found that overall, Drama and Prose fiction had the most occurrences with Drama favoring *why* in particular (2010:385–386). *Well* has been studied in Middle English, in EModE, and in LModE, but it has only recently been touched upon as a part of a diachronic study from Old English to Present Day English (PDE) (Defour 2007; 2009).

The next discourse marker that will be studied is the phrasal discourse marker *you know*.

12. She has had the advantage, **you know**, of practising on me.
(Emma Woodhouse, *Emma*)

You know has been extensively studied in spoken data in Östman's *You Know: A Discourse-Functional Approach* (1981). In addition, just as with other discourse markers, it has most often been studied in transcriptions of real spoken speech (Erman 1987; Crystal 1988; Aijmer 2009; Koczogh and Furkó 2011; Huang 2011).*You know* is one of the discourse markers that speakers use unconsciously (Lutzky 2012:25). Generally, the use of *you know* is perceived as a marker of

insecurity, irrationality, and even immaturity (Östman 1981:70). *You know* in fictional speech has not been studied, as the focus has predominantly been on one-word items.

In section 1.3, I argued that tag questions can function as discourse markers, particularly in the case of invariant tags. However, since seeing canonical tag questions as discourse markers is still quite controversial, I will only refer to them as tag questions. Tag questions have mostly been analyzed in terms of the main clauses they are attached to or their pragmatic function (Wichmann 2007; Reese and Asher 2007; Columbus 2009). Apart from this, the main interest is in what the use of a tag question conveys about the speaker (Lakoff 1973; Blankenship and Craig 2007). Due to the fact that tag questions have mostly been examined in terms of their function by way of analyzing intonational phonology, analyzing them in a historical context has not been an area of research. However, it could prove to be interesting to see what the use of tag questions can say about a person's social class, for instance.

In this study, the discourse markers *oh*, *ah*, *well*, *why*, and *you know*, and tag questions will be analyzed by correlating them with different sociolinguistic variables which have briefly been referred to in section 1.1. The four main variables in the present study are (1) the gender of the speaker and the addressee, (2) the intimacy between the speaker and addressee, (3) the social class membership of the speaker and addressee, and (4) the setting in which the conversation takes place (see further Chapter 3).

The present study aims to gain further insight into Late Modern English dialogue from a socio-historical perspective to identify which sociolinguistic variables correlate with the use of certain aspects of speech. The main research question therefore is:

> In what way do the sociolinguistic variables of (1) gender, (2) intimacy, (3) social class, and (4) setting correlate with the use of discourse markers and tag questions in LModE?

As mentioned in section 1.2, the period studied covers twenty-five years (1792–1817), during which the published and unpublished novels and stories by Jane Austen were written.

What I expect to find are results similar to what has been discovered in studies of PDE. Previous research has shown that in terms of function and expression not much has changed over the years regarding the use discourse markers. Tag questions have been regarded as more characteristic of female than male usage, and I expect to find similar results in the studied period. Furthermore, women tend to use more emotive language (e.g. insertions of *Oh* and *Ah*) and I expect to find this in Austen's works as well (Newman et al. 2008:221). For the variable intimacy, I hypothesize that the closer the intimacy between the speaker and the addressee, the more likely it is for discourse markers and tag questions to occur in their speech. In terms of social class, I

expect that members of the lower-middle classes in Austen's works, such as Miss Bates in *Emma*, use more discourse markers than middle-class and upper-class speakers, as it might be the case that negative connotations are attached to the use of the analyzed items just as there are in PDE. For the variable setting, I would expect for settings in which strangers are present that society would dictate proper speech (i.e. without emotive elements or optional additions): "in ordinary eighteenth-century usage, character was firmly situated in a social context and manifested itself through speech and action" (Michaelson 2002:185). In Chapter 2, these hypotheses will be more formally stated and related to previous research.

1.5 Thesis Overview

The outline of this study is as follows: the next chapter, Chapter 2: Background literature, will provide a discussion of the variables gender, intimacy, social class, and setting in relation to the discourse markers and tag questions (2.3), but before this, (2.2) will deal with previous studies on Jane Austen's language in her novels and stories. Chapter 3 covers the subcorpora which I compiled in order to answer the research questions and the approach I used to analyze the discourse markers and tag questions. Chapter 4 presents the findings and discusses them by correlating them with the variables and sub-variables I selected. In the final chapter, Chapter 5, I will summarize my findings relevant to the research questions and will answer the main research question as well as critically reflect upon the study I have undertaken.

CHAPTER 2: BACKGROUND LITERATURE

2.1 Introduction

The use of discourse markers (*oh*, *ah*, *well*, *why*, *you know*) and tag questions that are the subject of this study will be correlated with the variables gender, intimacy, social class, and setting. Some of the literature referred to in Chapter 1 will be referred to here, but most of the literature relating to studies on the language of Jane Austen, whose stories and published novels make up the main corpus for the study, discourse markers, and tag questions were not mentioned before. Before turning to the selected discourse markers and tag questions, I will turn to previous studies on the language of Jane Austen's novels.

2.2 The Language of Jane Austen

> [T]he sheer wit that sparkles in so many of the verbal exchanges is the most memorable of her conversations. But they are also distinguished by their lifelike flow. [….] [T]here is a wonderfully easy movement within the single speech which combines with a natural progress from one speech to the next to give the dialogues an air of artlessness, of truth of life. […] [N]one of her contemporaries or immediate predecessors among the novelists commands a dialogue at once so fluent and so brilliant as hers. (Babb 1962:242)

Austen's discourse was extensively studied by Babb (1962), who performed a stylistic analysis of both the *Juvenilia* and the later (un)published works. Babb describes the growth of Austen's dialogues by taking *Catharine, or the Bower* (*Cath*) as the starting point, which, despite inconsistencies of character as represented in speech, shows conversations similar to those in Austen's published novels (1962:34–37). Babb points out that Austen "dramatizes her characters through their linguistic habits" and that the dialogues constantly display the relationship between the speakers and their conventions and the society to which they belong (1962:242–243). Similar to Phillipps's (1970) analysis of the vocabulary in Austen's narrative and dialogue, sentence structure, words and phrases of address and Page's (1972) analysis of Austen's language in general, Babb's study is a qualitative one which provides proof to the genuineness of Austen's characters by analyzing their language, "for that is where the characters define themselves" (1962:28). Stokes's (1991) study of Austen's language further qualitatively analyzes the language of both the characters and the narrative with respect to the time in which Austen wrote the novels compared to present day English. The word *country*, for instance, is used to contrast London or the more metropolitan area with the provinces; in fact, in *Mansfield Park*, *country* indicates goodness of character and the opposing *London* represents the vices of the

'sophisticated' world (Stokes 1991:10). Furthermore, Stokes pays attention to the verbal expressions of characters which stand out. Page (1972:96) noted similar characteristics and summarized them as the exclamatory style in Austen which he regards as more poetic than rational. Stokes finds that the expressions show information about the character itself, as the exclamatory style of characters are a distinguishing mark (1991:16), as in the use of *la*. Austen's characters, Stokes argues (1991:34), manifest themselves "in social and personal relationships; for it is one capable of identifying and distinguishing personal and social strengths and weaknesses, intrinsic worth (head and heart) and social charm (address and spirits)" that can move beyond their social environment. *Mansfield Park*'s Fanny is, for instance, defined by her sweetness of character which highlights her good nature, kind-heartedness and obligingness (Stokes 1991:35). Stokes's focus, however, seems to be more on the language describing the characters' strengths and weaknesses than on the language the characters use, in contrast to Babb (1962).

Burrows's (1987) quantitative study, on the other hand, centers on the dialogue of Austen's characters. Burrows provides statistical proof of Babb's stylistically motivated claims. His analysis of thirty high-frequency grammatical and lexical words and an additional thirty words demonstrates the individuality of Austen's characters based on their dialogue. An example is the analysis of *her* in the speech of Lydia Bennet and Mr. Collins, characters in *Pride and Prejudice*: "Collins uses *her* [referring mostly to his patroness, Lady Catherine de Bourgh] at almost eight times as high a rate as Lydia, who has little attention to spare for the remainder of the female population" (Burrows 1987:84). This analysis alone shows that Lydia is more focused on herself than any other female or male and that for Mr. Collins, the predominant topic of his conversations is a female rather than himself. Furthermore, the analysis of *know* and *think* provide information about the assertiveness of characters, which is demonstrated by *Mansfield Park*'s Mrs. Norris who uses these verbs frequently as well as the "habitual *I am sure*" which reflects her "uneasy self-assertiveness" (Burrows 1987:151). Not only did Burrows concern himself with spoken dialogue, he also analyzed the narrative and thoughts of the characters. Even though there are differences between the thirty most common words in the dialogue and narrative of the characters, the dialogue itself with relation to the other characters remains highly similar (Burrows:1987:168).

A similar quantitative study of Austen's language is Fischer-Starcke's *Corpus Linguistics in Literary Analysis: Jane Austen and her Contemporaries* (2010). Her study demonstrates new techniques for interpreting and gaining literary insight into a text. Fischer-Starke's (2010:27–29) focus is on *Northanger Abbey* and how this novel compares to the remaining five novels, to contemporary literature as well as to roughly contemporary Gothic novels, as *NA* parodies the Gothic novel.

She studies the language (both dialogue and narrative) by means of several programs: *kfNgram*, which analyses the frequency of strings of words; *Vocabulary Management Profiles*, which pinpoints the moment new lexis is introduced and by which it is thus possible to locate a change in topic; *Word-Distribution*, which demonstrates where a particular word occurs and identifies its location numerically; and *WordSmith Tools*, used in the present study (see Chapter 3), which generates data pertaining to the context of a word, the frequency of words, and the 'keyness' of a word compared to other texts (Fischer-Starke 2010:31). The phraseology analysis showed that the most frequent phrases or clusters are delexicalized and function as discourse markers such as *I do not know* and *I am sure I* (2010:135, 143). Fischer-Starke also discovered that while literary critics claimed differences between *NA* and the other novels by Austen, there is no linguistic reason for there to be one (2010:194). Earlier critics had claimed that the narrative of *NA* was found lacking, even incomplete, and seemed more of a narrative experiment rather than an actual Austen novel (2010:184). Furthermore, Fischer-Starke also found that by analyzing keywords and their context, the protagonists of *NA* can be characterized by their preferred literature and the perceptions of family members and other social relations, compared to Austen's other novels (2010:106). Fischer-Starke's study shows that not only does corpus linguistics provide important interpretative insights into texts, but it also gives structural insights into language such as grammatical negations, which are more prevalent in *NA* (2010:199). Page's request for the analysis of the "notion of *realism* in [Austen's] dialogue […], especially since it has often been evoked in relation to Jane Austen's art" (1972:115), has been demonstrated by both Burrows (1987), who focused on individuals, and by Fischer-Starke's (2010) research, which studied *Northanger Abbey*.

The focus of this study will be on Austen's dialogue, but the language will be studied by looking at specific items, discourse markers and tag questions, a bottom-up approach similar to Burrows (1987), to uncover whether the use of these items sheds light on their distribution across gender, intimacy, social class, and setting.

2.3 Research on Discourse Markers

2.3.1 Interjections: *Oh* and *Ah*

Oh and *ah* are regarded as a word class by themselves in both past and present times, according to Ameka (1992:101). *Oh* and *ah* belong to the primary interjection subclass, as they are not used outside their interjection domain (Ameka 1992:105). Primary interjections, however, may be taken as "interjection-based discourse markers", as argued by Columbus (2009:404), which will similarly be done in this study where they are regarded as a subclass of discourse markers. The

precise nature of what an interjection truly is, is captured by an utterance of *Emma*'s Harriet in which *oh* occurs:

13. **Oh** yes! – that is, no – I do not know – (Harriet Smith, *Emma*)

Oh is privileged in being an area of study in both present-day language as well as in that of the past. Huang (2011) analyzes the occurrence of *oh* in PDE with reference to the nature of the multitude of pragmatic uses, whereas Taavitsainen (1995), Jucker (2002), and Person (2009) focus on the use of *oh* in EModE. Person's analysis of *oh* in Shakespeare's works reveals that *oh* expresses similar sentiments and is used in similar ways to present day *oh* (2009:102). Taavitsainen (1995:463) argues that the interjections found in EModE are "far removed from purely emotive cries" and can function as a form of address or a reactionary response. Taavitsainen (1995:453–457) found instances in which *oh* was merely prefixed to an exclamatory sentence and/or swear word(s) (see example 14), was a vocative optionally followed by a noun of address, was emotionally colored by anger or feelings of doubt, was an indication of a mental process, and was a marker of a change in topic.

14. **Oh! D— it**, when one has the means of doing a kind thing by a friend, I hate to be pitiful. (John Thorpe, *NA*)

All these characteristics clearly indicate that *oh* is much more than an expression of emotion. Jucker (2002:219), on the other hand, remarks that *oh* in particular has an exclamatory function but also a text-structuring one and that therefore EModE *oh* is an interjection and not the discourse marker it is in PDE. Person (2009:104), however, supports the claim that *oh* has the same characteristics of pragmatic markers and therefore ranks *oh* as a discourse marker. The classification of *oh*, similar to PDE, seems to be undetermined and open to interpretation.

Ah, which has not been studied to the same extent as *oh*, has nevertheless been studied in similar ways. Aijmer (1987) analyzed the functions and context of *oh* and *ah* in PDE conversation. Similar to Aijmer's findings for *oh*, *ah* marks a previous utterance as pleasant or crucial as well as noting the significance of something which just came to mind (see example 7 in Chapter 1) (1987:65), while Taavitsainen's study on interjections in EModE shows that *ah* expresses consent or appreciation, but also a pain, sympathy, or that it introduces a clause conveying regret or sorrow (1995:446). An example of *ah* introducing a clause expressing sorrow sarcastically occurs, for example, when *Emma*'s Mr. Elton walks in while Harriet and Emma are conversing:

15. **Ah**! Harriet, here comes a very sudden trial of our stability in good thoughts. […] Oh! dear. (Emma Woodhouse and Harriet Smith, *Emma*)

Furthermore, Taavitsainen found that the collocation of *ah ha* caused the function of the interjection to shift towards *ha*, expressing a mental process such as, having formulated an idea (1995:446). Furthermore, *ah* can also indicate disbelief over what has been said:

16. **Ah**! – so you say; but I cannot believe it. (Harriet Smith, *Emma*)

Harriet expresses disbelief after Emma relates that she is not going to marry nor that she has any inclination to get married in the future.

As the study of discourse markers is more centered around their pragmatic function, there has not been much research on the correlation of discourse markers and sociolinguistic variables. Aijmer (2009) is one of the studies that looks at *oh* and *ah* in relation to gender. She found that women had a tendency to use *oh* more often than men and that, contrastively, men tended to use *ah* more than women (2009:13). Similar to Aijmer (2009), in Jonker (2014) I found that there was a tendency, though not confirmed statistically, that *ah* was more of a male interjection than *oh* in *Emma*. The present study will look more into the use of other discourse markers and will be focused on spoken dialogue only, whereas in my earlier paper I included thoughts and inner monologues.

How were *oh* and *ah* regarded in Austen's time? Johnson's (1755) comprehensive dictionary of the English language provides us with a relatively contemporary interpretation. *Oh* has two entries in Johnson: *o* and *oh*. Interestingly, they were not defined identically: *o* is used when a speaker wishes or exclaims something, whereas *oh* is specified as an exclamation indicating pain, sorrow or surprise (*JohnsonOnline*, s.v. *o* 1. and *oh*).

17. **O** that we, who have resisted all the designs of his love, would now try to defeat that of his anger! (*JohnsonOnline*, s.v. *o* 1.)

18. **Oh** me! all the horse have got over the river, what shall we do? (*JohnsonOnline*, s.v. *oh*)

Ah, on the other hand, has only one entry and has a predominantly negative undertone as it shows dislike, censure, contempt, and complaint.

19. In youth alone, unhappy mortals live; But, **ah**! the mighty bliss is fugitive. (*JohnsonOnline*, s.v. *ah*)

At times, *ah* can also be used to express compassion or joy. In addition, a special note is given when *ah* is followed by *that* which expresses "vehement desire" (*JohnsonOnline*, s.v. *ah*).

2.3.2 General Discourse Markers: *Well* and *Why*

Well, a well-known discourse marker, has been studied quite extensively (Schiffrin 1987; Crystal 1988; Jucker 1993, 2002; Defour 2007, 2009; Popescu-Belis and Zufferey 2011; Huang 2011; Lutzky 2012). Crystal (1988) investigated why *you know* (see 2.3.3) and *well* were considered as "markers of unclear thinking, lack of confidence, inadequate social skills, an a range of other undesirable characteristics" (1988:47). He argues that the reason for these prejudiced views on discourse markers are a result of their status of "stylistic excesses of certain groups of speakers in the past" (1988:48). Defour (2007)'s analysis of *well* showed that, as opposed to the other discourse marker studied (*now*), the word has undergone a delexicalization process to a greater degree. Defour discovered that the grammaticalization of *well* and *now* was due to the subjectivity attached to them by the speakers who "guide the addressee in [the] interpretation of the utterance" using these markers (2007:295). Just as Crystal (1988:49) argued, discourse markers "facilitate the often thorny task of making communication between speakers successful".

Well can have a multitude of functions. Lutzky (2012) explains that *well* can have a structural function (e.g. indicating a change in topic); a quotative function which "introduces direct reported speech […] in a similar way to quotation marks in writing"; a coherence function; an acknowledging and continuative function (e.g. acknowledging a statement and elaborating on it); a filler function (e.g. marker of hesitation), a self-repair function (e.g. introducing self-corrections); a non-acceptance function (e.g. indicating that the previous statement failed to be sufficient in content); an answer or question function (e.g. answering a question with *well* as an act of non-compliance, e.g. the second *well* in example 20); a face-threat mitigator function (see example 21); and an emotional function (e.g. conveying impatience or disapproval) (2012:77–84). An example of *well* functioning both as a question and an answer of non-compliance is the following:

20. My brother's heart, as you term it, on the present occasion, I assume you can I can only guess at

    Well?

    Well! Nay, if it is to be guesswork, let us all guess for ourselves.

    (Henry Tilney and Catherine Morland, *NA*)

In this case, *well* can serve as a one-word answer or utterance, just as *oh* and *ah* can. Another example of one of the functions described by Lutzky is the face-threat mitigator function in a conversation between Isabella and Catherine:

21. Are they? **Well**, I never observed that. They always behave very well to me. (Catherine Morland, *NA*)

Lutzky (2012:268) found that *well* occurred primarily at the beginning of an utterance to signal a shift in turn or initiating a conversation in the EModE period. Furthermore, her analysis of a tagged sociopragmatic corpus uncovered that *well* was a marker that occurred in the speech of the upper social class and that the use of *well* decreased steadily when stepping down the social ladder. However, Lutzky found that the lowest group, the servant class, almost equaled in the number of *well*s compared to the upper social class, which may have been due to servants imitating the speech of their masters. As for gender differences, she found that *well* was used most often by female characters and occurred most often in same-gender dialogues.

Why is to *well* as *ah* is to *oh* to a certain extent. As a discourse marker, *why* has not garnered as much attention as *well* in both historical and current studies. Similar to *oh*, *why* has been analyzed in Shakespeare's works and has been found to introduce a conclusion deduced from a previous utterance with an undertone of superiority and belittlement (qtd. in Brinton 2001:142). Jucker (2002:219) claims that *why* as a discourse marker indicates that the previous utterance has not been understood. Furthermore, an analysis of *why* in trials showed that the marker "signal[ed] a break-down in the question-answer sequence [and] express[ed] disbelief in a witness's evidence, [while] the witnesses use it to present an answer as if it were a self-evident truth" but it also occurred as a marker of surprise at "the lack of understanding" (qtd. in Jucker 2002:219–220). The following example is taken from a scene at a ball in *Mansfield Park* where *why* expresses disbelief at an observation:

22. Poor Fanny! […] how soon she is knocked up! **Why**, the sport [i.e. dancing] is but just begun. (William Price, *MP*)

Similar to *well, why* can have several functions. *Why* can have a coherence function in the case it precedes "a logical conclusion to what has been [said] before and [thus precedes] some kind of definitive view" (Blake 1996:127). Alternatively, it may have an emotional function when it "signal[s] surprise at a question because the interlocutor ought to know the answer themselves or because the speaker cannot be expected to know the answer" (Lutzky 2012:87). *Why* can also have a contrastive function as it can precede a statement in which a counter-argument is introduced (Blake 1996:129). Culpeper and Kytö (2010:396) found that *why* has two prevalent functions: the first expressing "a challenge to the previous speaker's irrelevancy […] and an expression of a negative attitude towards [the utterance]" and the second indicating "a cue to

optimise relevancy when a change in speaker is signalled" (i.e. the coherence function mentioned earlier).

Lutzky's (2012) study on the discourse markers *marry*, *well*, and *why* showed that *why* was the most prominent discourse marker. At the beginning of the EModE period, *why* was not very frequent but it increased enormously over the years. Lutzky found that *why* was often used in a turn-initial position, as a marker of surprise, to indicate factual or new information, and to draw a conclusion. In relation to sociolinguistic variables, Lutzky (2012:269) pointed out that *why*, just as *well*, was a marker of upper- and lower-class speech. She argued that people belonging to the lower social classes use *why* as an imitation of speech of the upper social class, but also that *why* is used to address people that are of a higher social class. As for gender, Lutzky (2012:269) found evidence that suggested that *why* was a predominantly male speech feature and that it was used most often in male-to-male conversations.

Why and *well* also have entries in Johnson's dictionary (1755), which help us in understanding what they meant around Austen's time. *Well* has many different definitions but the closest one that resembles a discourse marker is: "It is a word by which something is admitted as the ground for a conclusion" (*JohnsonOnline*, s.v. *well* adv. 9.), which is similar to Lutzky's findings.

23. **Well**, let's away, and say how much is done. (*JohnsonOnline*, s.v. *well* adv. 9.)

Why also has an entry which resembles a discourse function: "It is sometimes used emphatically" (*JohnsonOnline*, s.v. *why* adv. 4.). This sole entry relating to an emotional function is quite surprising as *why* was used very often in the eighteenth century, as Lutzky pointed out, with similar functions as *well*.

24. You have not been a-bed then?
Why, no; the day had broke before we parted.
(*JohnsonOnline*, s.v. *why* adv. 4.).

2.3.3 Phrasal Discourse Marker: *You Know*

You know, according to Crystal (1988:47), garnered much attention after his publication of *Who cares about English usage?* (1984). Crystal discusses three positions where *you know* is permitted in everyday conversations: at the beginning, in the middle, and at the end of an utterance. In sentence-initial position, *you know* has a softening or sympathetic effect (Crystal 1988:47).

25. **You know**, Mrs. Weston, you and I must be cautious how we express ourselves. (Mrs. Elton, *Emma*)

In example 25, Mrs. Elton gently puts Mrs. Weston on her guard by using *you know*, but also by using it in combination with the addressee's name, which removes some of the directness the sentence would otherwise have. In sentence-medial position, *you know* clarifies or amplifies the first part of the utterance and stresses the importance of the second part (Crystal 1988:47). The following example illustrates the use of *you know* in sentence-medial position where *Sanditon*'s Arthur Parker explains to Charlotte Heywood how he is not 'bilious' (i.e. suffering from a disorder to the liver or gallbladder).

26. If I were Bilious, he continued, **you know** Wine would disagree with me, but it always does me good. (Arthur Parker, *S*, in Bree et al. 2013:373)

Previously, Arthur stated that his sisters considered him to be bilious and therefore the second part of the utterance is stressed as it is a counterargument to the opinion of his sisters. The first part of the utterance repeats that statement hypothetically, adding *you know* to amplify the previous statement, while the second part of the utterance provides information that contradicts the hypothetical statement, which is important information for the purpose of Arthur's argumentation. In sentence-final position, Crystal (1988:47) claims, *you know* is a type of tag question to inquire after the addressee's understanding of the utterance, often with an incomplete preceding utterance.

27. They are at Brighton now, **you know**. (Mrs. Norris, *MP*)

Mrs. Norris is relating a tale to Sir Thomas, and Dr. and Mrs. Grant about some of her friends and acquaintances. By using *you know* in final position, she both stresses the preceding question and inquires after whether they absorbed that information.

You know has many other functions than softening, stressing, or inquiring. Östman (1981) and Erman (1987; 1992) both analyzed *you know* as a discourse marker, while Erman (1987) also looked at two other phrasal discourse markers, or "pragmatic expressions" as she prefers to call them, *you see* and *I mean*. Östman (1981:16) found that *you know* is most frequently used in narrations that are part of conversations, similar to example 27. It is a type of hedge used to transfer implicit information (i.e. feelings and attitudes). Furthermore, *you know* can both refer to a single lexical item (*bilious* in example 26), but it may also influence the entire utterance. Östman (1981:17) formulates the core meaning of *you know* as follows:

> The speaker strives towards getting the addressee to cooperate and/or to accept the propositional content [i.e. the information the utterance provides without the emotional and/or attitudinal implication] of his utterance as mutual background knowledge.

Furthermore, Östman (1981:21–23) distinguishes between the declarative and interrogative uses of *you know*. Declarative *you know* implies that the speaker is certain that the addressee has knowledge of the information being presented and that no type of argument is expected, also found as *as you know*, which is a hedge. Interrogative *you know*, on the other hand, implies that the speaker is not sure of whether the information he or she gives is accurate or whether the information is known or regarded as truthful by the addressee, similar to *don't you know*, which is a kind of tag question. In relation to gender, Östman (1981:70) found that women use *you know* more often than men. By contrast, Erman (1992)'s study showed that men use *you know* more frequently than women. In addition, she found that the discourse markers *you know*, *you see*, and *I mean* were more often used in same-sex conversations than in mixed-sex ones. Furthermore, Erman's (1992:228) results indicated that men and women use *you know* for different purposes. Men use it more for rhetorical purposes or to repair phrases, whereas women use it to organize the discourse or as a hesitation marker. Korzogh and Furkó's (2011) analysis of *you know* and *I mean* in transcribed conversations of an American talk show, *Larry King Live*, indicated that there was no significant difference between the frequencies of *you know* in the male and female subcorpora. In their analysis of the function of *you know* in the two corpora, male speakers used *you know* as a topic changer least, while female speakers, on the other hand, used it least with the function of seeking agreement (Korzogh and Furkó 2011:7). Additionally, male speakers use it more frequently as a hesitation marker than female speakers, which contradicts Erman's (1992) findings.

Fox-Tree and Schrock (2002:735–740) provide a useful overview of the functions of *you know*. *You know* could have an interpersonal function, where it is a marker of membership to a community or of friendship based on the premise of mutual knowledge. Furthermore, it could also be viewed as an invitation of inferences in the mind of the addressee. *You know* can both be a marker of confidence as well as one of marking uncertainty, similar to Östman (1981)'s definition. In addition, *you know* may also function as a politeness marker. By using *you know*, the utterance seems less definitive and speakers use them to retain their face and inquire after the addressee's interpretations. But then again, *you know* can also be used as a definitive marker that rebukes retort. *You know* may also be a signal to the addressee to give them a turn to speak, to provide them with some time to think about what was said, and/or to point out their views after the speaker's turn has ended. The phrasal marker can also indicate repair of the speaker's previous part of the utterance or as a stall tactic to formulate a coherent sentence. There are many more functions that *you know* can have, but the ones discussed here provide a clear enough picture that *you know* most definitely is a multi-functional marker, similar to *oh* and *well*.

Even though the relevant research on *you know* is based on present-day use of the marker, the functions and meaning of *you know* do not seem to have changed, as we can conclude from the examples provided (12 and 25–27).

2.4 Research on Tag Questions

Tag questions entered linguistic research by way of Lakoff (1973) who undertook a study to analyze the differences between male and female speech. They can be both a marker of power and of powerlessness, and a marker of certainty and uncertainty at the same time (Lakoff 1973; Blankenship and Craig 2007). Even though tag questions do no always belong to the same category as discourse markers, their function is perceived as highly similar (Stenström 1994; Andersen 2001; Komar 2007) and, at times, they are regarded as a fully-fledged discourse marker (Pichler 2013). Schiffrin's (1987) work on discourse markers briefly mentions tag questions, since her focus is on the more stereotypical discourse markers, like *oh*, *well*, and *you know*. Schiffrin defines tag questions as "declarative statements with postposed tags through which questioners seek agreement with the content of the statement" (Schiffrin 1987:88). In addition, she argues that tag questions are not generally facilitative, but that they can also restrict the response a hearer can give since a tag question can transform an open question into a Yes/No-question or it can be used to direct the addressee towards an answer that satisfies the speaker (1987:23); for example,

28. How came she to think of asking Fanny? Fanny never dines there, you know, in this sort of way. I cannot spare her, and I am sure she does not want to go. Fanny, you do not want to go, **do you?**

    If you put such a question to her […] Fanny will immediately say No; but I am sure, my dear mother, she would like to go; and I can see no reason why she should not. (Lady Bertram and Edmund Bertram, *MP*)

Meyerhoff (2011:235) explains that there are three predominant functions of tag questions: "the speech act of asking questions, a stance of attentiveness, and the discourse activity of eliciting the contributions of others". The main function a tag question has is to keep the conversation going. Meyerhoff (2011:235) further argues that the use of tag questions are not distributed equally across a speech community, for which she gives the example of the teacher:

> Teachers, for instance use tag questions quite often and this can be explained by the fact that in order to fulfil the role of a teacher a person is expected to do all these things: ask questions, be attentive, and elicit contributions from students.

Even though there does not seem to be a difference in the number of tag questions used by male and female speakers, there is a difference in their function. Supportive tags (e.g. to elicit a response or to mitigate criticism) are used more often by women than by men (Meyerhoff 2011:235–236). LaFrance (2001:247) discusses both gender roles and power of the speaker and addressee in relation to tag questions. She finds that tag questions used by women are more facilitative, whereas tag questions used my men are used to seek an answer. In terms of power, when a speaker has a powerful role, tag questions serve to continue a conversation, whereas speakers with a powerless role use it to seek reassurance about what has just been said. An example of a 'powerless' tag question is the following:

29. Mr. Knightley was standing just here, **was not he?** I have an idea he was standing just here? (Emma Woodhouse, *Emma*)

Emma is clearly not sure of her own observation that Mr. Knightly "*was standing just here*". A 'powerful' tag question also occurs in *Emma* when Frank Churchill and Emma are discussing Jane Fairfax and her musical talents:

30. She plays charmingly.
    You think so, **do you?** – I wanted the opinion of someone who could really judge. She appeared to me to play well, that is, with considerable taste, but I know nothing of the matter myself. (Emma Woodhouse and Frank Churchill, *Emma*)

In this case, the tag question is not used to ask reassurance of what Emma has just said, but rather as a starting point for the speaker himself to continue talking about the subject.

Pichler (2013) analyzed negative polarity tag questions, such as example 29 (*was not he?*) in a corpus of transcribed speech collected in Berwick upon Tweed, near the Scottish border. Her analysis of tag questions also looks at the function of within-subject and between-subject tag questions:

> Subjectively, they function to signal speakers' degree of commitment and attitudinal stance towards their propositions. Intersubjectively, [on the other hand], they serve to mitigate potential face-threats, to draw listeners into the discourse and maintain their active involvement, and to signal speakers' alignment with prior talk and active involvement in the interaction. (Pichler 2013:192).

Pichler's data indicate two within-subject and three between-subject functions (2013:269–270): the within-subject functions are tag questions as 'epistemic markers', which signal uncertainty in

the speaker, and 'attitudinal stance markers', which emphasize preceding utterances and, for example, underline the obviousness of the statement. Between-subject, or 'intersubjectively', the tag question can function as a 'mitigation device', which lessens dismissals and disagreements, as an 'involvement inducer', which involves addressees in discourse, and as an 'alignment signal', which shows agreement and cooperation with previous speakers.

Apart from the functions of tag questions, their grammaticalization has also been a topic of interest among researchers. Wichmann (2007) studied tag questions in order to discover whether tag questions (specifically those in the English language) may have undergone grammaticalization or whether they are on their way to being grammaticalized. She found evidence that tag questions display loss of prominence in that they elicit no response from the hearer, or are not meant to elicit a response, which is actually the case in example 30 (Wichmann 2007:357).

Not only has the tag question been proven to have lost some of its propositional content in conversations, it may also be replaced with an invariant tag. An example of such an invariant tag is John Thorpe's *hey?* in 1.2:

31. A pretty good thought of mine, **hey**? (John Thorpe, *NA*)

Invariant tags are similar to tag questions but they differ in that a change of polarity is impossible as is the inversion of subject and verb (Columbus 2009:401). Columbus (2009:403) defines invariant tags as response elicitors in which the speaker enquires after the addressee's acceptance and understanding of the preceding statement. Invariant tags occur in many forms across different varieties of a language. Columbus (2009) looked at utterance-final invariant tags and the invariant tags that occurred more than fifty times were analyzed further regarding their frequency in the five varieties of English. The tags that were relevant were "*eh*, *yeah*, *la*, *right*, *OK/okay*, *you see*, *no*, *na* and *you know*" (Columbus 2009:407). Across the five varieties, there were only four tags that occurred in all of them: "*okay/OK*, *right*, *you know*, and *you see*" (Columbus 2009:407). Even though his research showed differences in the frequencies of tags and in their functions, he did not look at the social implications of the occurrences of tag questions, which has been demonstrated in earlier studies to be important aspect of the tag question. The social implications of tag questions will be of interest in this study. Furthermore, *you know* is regarded as a tag question in Columbus' study, but in this paper it will only be regarded and analyzed as a phrasal discourse marker.

2.5 Concluding Remarks and Hypotheses

Most of the earlier studies on discourse markers looked at their multifunctional properties in PDE and therefore not much is known about the sociolinguistic properties of discourse markers

apart from gender, let alone those in historical discourse. I will briefly recapitulate the sociolinguistic findings per discourse marker before forming my definitive hypotheses: *oh* is a predominantly female marker, whereas *ah* is a male marker (Aijmer 2009; Jonker 2014); *well* is a marker of both upper-class and lower-class speech, of female speech, and of same-gender conversations, in which it is similar to *why* except for the fact that *why* is a predominantly male marker; but it occurs most in same-gender conversations, specifically those of the male-to-male variety (Lutzky 2012); *you know* has been regarded as both a female marker (Östman 1981) and a male marker (Erman 1992), but the function of the phrasal discourse marker has been proven to be different for men and women (Erman 1992; Korzogh and Furkó 2011). Tag questions, on the other hand, have been examined in a sociolinguistic context. They are found to be distributed equally amongst men and women, though women use tag questions for different purposes (e.g. to offer support) than men do (e.g. to seek an answer) (LaFrance 2001; Meyerhoff 2011).

Based on these findings, I will expect to find in a corpus of a selection of Jane Austen's fictional works that female speakers prefer to use *oh* and *well* and male speakers *ah* and *why*. As for *you know* and tag questions, I expect to find an equal distribution across gender. In addition, I will expect to find most instances of *well* in same-gender conversations and *why* in male-to-male directed speech. Furthermore, in relation to social class, my hypothesis is that speakers from a high social class will use discourse markers most often and that there will be a decrease in use in comparison with the lower social classes. Though none of the earlier research covered intimacy and setting, I still expect to find a difference when the relationship between speaker and addressee is more intimate or less intimate: the more intimate the relationship, the more discourse markers and tag question are used. For setting, I expect to find more discourse markers and tag questions in more informal situations, which coincides with the intimacy of the relationship between speakers. However, due to certain discourse markers being part of upper-class speech, I will also expect to find discourse markers in formal situations as a marker of distinction.

CHAPTER 3: METHODOLOGY

3.1 Introduction

The chapter introduces the selected stories and novels by Jane Austen that make up the main corpus for analysis as well as the character selection for the published novels. Furthermore, it will demonstrate how the corpus was prepared and how it has been divided into subcorpora, which lexical tool was used to analyze the subcorpora, and which statistical analysis was used in order to determine the correlation between the selected discourse markers and tag questions in general, and the sociolinguistic variables (gender, intimacy, social class, and setting).

3.2 Corpus

The corpus for the study comprises six novels by Jane Austen written during the selected twenty-five year period (1792–1817): three unpublished novels, *Catharine, or the Bower* (1792), *The Watsons* (1804), and *Sanditon* (1817), and three published novels, *Northanger Abbey* (1798/99–1816), *Mansfield Park* (1811–1813), and *Emma* (1814–1815). The latter were retrieved from Project Gutenberg, through which texts are made available, among others, in .txt format. The former were retrieved from a pdf version of Bree et al.'s (2013) *Jane Austen's Manuscript Works*, and were subsequently converted into .txt files.

3.2.1 Explanation for the Corpus Selection

To account for an equal number of unpublished and published stories or novels, three unpublished works and three published ones were selected. The unpublished novels and stories were, most probably, only altered by Jane Austen herself, while in the published novels, editors could have changed phrases, words or punctuation marks from the original text. Furthermore, including the unpublished novels or stories provides some evidence of the language at the beginning of the time period and at the end.

Catharine, or the Bower is the first of the *Juvenilia* which "contain[s] conversations that seem nearer in tone to the finished novels" (Babb 1962:34) and thus demonstrates the first glimpses of what conversations were like at that point in time. *The Watsons* is an unfinished novel written during the period in which the novels of *Sense and Sensibility* (1795–1810), *Pride and Prejudice* (1796–1812), and *Northanger Abbey* (1798/99–1816) were being written. *Sanditon* is Austen's final novel on which she worked until her death on 18 July, 1817. *The Watsons* and *Sanditon* were the only other unpublished novels written after *Catharine, or the Bower* that were made available in Bree et al. (2013) that were not epistolary novels (i.e. novels existing of only documents such as diary pages, or letters). Furthermore, *The Watsons* and *Sanditon* demonstrated more interjection use in

male speech than in female speech (Jonker 2014) and therefore deserve more attention in relation to gender and might prove to be interesting regarding the other sociolinguistic variables.

Northanger Abbey is the first of the published novels that was selected because of the untrustworthy and trustworthy characters and their conversations with one another. John Thorpe, one of the main antagonists, proved to be quite interesting regarding his speech and it could be that more characters have interesting speech characteristics in this novel. In *Mansfield Park* and *Emma*, there are clear differences between male and female speech regarding the use of interjections (Jonker 2014) and it would be relevant to see whether discourse markers and tag questions also have a clear distribution across male and female speech or in relation to other variables. *Emma* has much more direct speech and many dialogues. It stands to reason that in relation to discourse markers and tag questions, if there is a correlation with their use and the sociolinguistic variables, *Emma* would most probably display this.

In the unpublished novels, direct speech of all the characters will be used as there is less direct speech in general, compared to the published novels. For the published novels, I made a selection of important characters to lessen the amount of work involved, but also because of the fact that the important characters have more direct speech to begin with. The *Northanger Abbey* characters that I have selected are Catherine Morland, Isabella Thorpe, Henry Tilney, and John Thorpe, the four main characters with the most lines of speech (see Burrows 1987:xiv). Catherine Morland is a naïve character of lower-middle-class origin who is introduced to higher social circles by friends of the family, the Allens, when they take her to Bath. In Bath, Catherine meets Isabella and John Thorpe who are the children of an upper-middle-class childhood friend of Mrs. Allen. Catherine and Isabella quickly become close friends and John develops feelings for Catherine. At a ball, Catherine meets Henry Tilney, an upper-class character, who has an instructive role in the novel and who eventually helps Catherine to overcome her naivety. The selected *Mansfield Park* characters are the main protagonist Fanny Price, Sir Thomas Bertram, Mrs. Norris and Edmund. Even though they are not the characters with the most lines, this selection provides for both upper-class and lower-middle-class speech, family intimacy, and a fair distribution of gendered speech (see Burrows 1987:xiv). Fanny Price is a lower-middle-class character that comes to live with her aunt and uncle, Sir Thomas Bertram and Lady Bertram who are portrayed as members of upper-class characters. Mrs. Norris is Fanny's other aunt who is the caretaker at the Bertram estate. Due to her profession I have categorized her as a lower-middle-class character. Edmund Bertram is the second son and one of the four children of Sir Thomas and Lady Maria Bertram. Similar to Henry Tilney in *Northanger Abbey*, Edmund has an instructive role to Fanny and helps her become a more level-headed person. As for *Emma*, the main

protagonist, Emma Woodhouse, was chosen as were Frank Churchill, George Knightley, Harriet Smith, and Miss Bates. Emma, Frank and George have the most lines of speech, but Harriet and Miss Bates have about half to a quarter of the amount (Burrows 1987:xiv). Emma Woodhouse is an upper-middle-class character that lives with her father. George Knightley is Emma's brother-in-law, an upper-class character, but they have been friends since before his brother, John, married her sister, Isabella. George has a similar role to Henry Tilney and Edmund Bertram. Frank Churchill is an upper-middle-class character who comes to town to visit his father. Frank is portrayed by Emma as a possible match for herself, but his interest lies in Jane Fairfax to whom he has been engaged in secret. Harriet Smith is the protégé of Emma and Emma teaches Harriet about the ways of the higher social classes. As Harriet's origins are unknown, I have classified her as lower-middle class since she is lower than Emma. Miss Bates is a neighbor of Emma and the aunt of Jane Fairfax. She is a very talkative and friendly person of lower-middle-class origin.

For a more detailed overview of the intimacy level between characters and the social classes of the characters, consult Table 3.1 and 3.2 in Section 3.2.3 below.

3.2.2 Preparation of the Corpus and Subcorpora

The Project Gutenberg texts *Northanger Abbey*, *Mansfield Park*, and *Emma* were cleaned up so that only the text of the novels themselves were part of the corpus. All the chapter titles were deleted as well, because the titles would otherwise add to the total number of words as would the numbers, which would influence the normalized data. In addition, all the single and double quotations marks were altered to regular quotations marks that are recognized as such by the lexical analysis software (see 3.3 below). Furthermore, due to my earlier interest in Austen's use of the dash, all dashes were replaced by hyphens which made it easier to do a frequency analysis as the dash was not recognized by the software. Similarly, Bree et al.'s (2013) transcriptions of Austen's *Catharine, or the Bower*, *The Watsons*, and *Sanditon* were cleaned up by deleting the chapter titles and footnotes, and replacing the quotation marks and dashes. Moreover, as I am specifically interested in discourse markers and tag questions in direct speech, the narrative text, indirect speech (e.g. letters), the inner- and outer-monologues, and thoughts were removed from the texts. Looking at quotation marks alone was not enough as some of the thoughts and monologues occurred within double quotation marks. Some examples of thoughts, inner- and outer-monologues, and indirect speech may be found below:

32. Well done, Mrs. Martin! **thought** Emma. You know what you are about.
    (Emma Woodhouse, *Emma*)

33. She **hoped** no one could have said to her, "How could you be so unfeeling to your father? – I must, I will tell you truths while I can." (Emma Woodhouse, *Emma*).

34. Can Camilla's explanation be true? **said** Catherine **to herself**, when her freind had left the room. And after all my doubts and Uncertainties, can Stanley really be averse to leaving England for my sake only?
(Catharine Percival, *Cath*, in Bree et al. 2013:203).

35. Sir Thomas's look **implied**, "On your judgment, Edmund, I depended; what have you been about" She [Fanny] knelt in spirit to her uncle, and her bosom swelled **to utter**, "Oh, not to *him*! Look so to all the others, but not to *him*!" (Sir Thomas Bertram and Fanny Price, *MP*).

In order to be able to run the analysis software successfully, the discourse marker *you know* and the tag questions were altered so that they did not consist of multiple words, as the software only counts separate words and not phrases. The relevant *you know*s, as *you know* can also occur as the subject and verb of a sentence, were altered to *youknow* and the tag questions were altered to *tagquestion*.

36. **You know** how uncomfortable I feel with her. (Fanny Price, *MP*)

37a. She will have a companion in Fanny Price, **you know**, so it will all do very well.
(Mrs. Norris, *MP*)

37b. She will have a companion in Fanny Price, **youknow**, so it will all do very well.

38a. Fanny has been cutting roses, **has she**? (Edmund, *MP*)

38b. Fanny has been cutting roses, **tagquestion**?

Example 36 above demonstrates that *you know* does not always occur as a discourse marker, but that it may also be a subject and a verb followed by a phrase. When you compare this example to the following one, 37a, *you know* in this case clearly stands apart from the sentence as you can delete it without losing the sentence structure or the propositional meaning of the sentence, which is not the case in example 36. The examples 37b and 38b demonstrate how the discourse markers and tag questions are altered in the corpus for the analysis software. Tag questions were not hard to distinguish from other questions but *you know*s at times were difficult to label as discourse marker or subject-verb sequence. An example of a *you know* that is difficult to categorize is the following:

39a. **You know** we have settled all that already. (Emma, *Emma*)

If I were to go by what I proposed earlier: if you can delete it, it is a discourse marker principle; *you know* in example 39a is a discourse marker as the sentence can stand alone without it:

39b. We have settled all that already.

However, when we look at the context, it is clear that *you know* is a subject-verb sequence:

39c. They are to be put into Mr. Weston's stable, papa. **You know** we have settled all that already. We talked it all over with Mr. Weston last night.
(Emma Woodhouse, *Emma*)

Emma is clearly stressing to her father that he knows of the talk about putting the horses in Mr. Weston's stable. *You* has a demonstrative function and is not a generic *you know*, as was the case in 37a above. Furthermore, in the published novels, the discourse marker *you know*s are generally followed by a comma, preceded by a comma or both. However, in the unpublished novels, this is not the case (see example 40 below), and thus looking at the context as well as applying the deletion principle is important to determine the status of the phrasal discourse marker *you know*.

40. Those who tell their own Story **you know** must be listened to with Caution.
(Mr. Parker, *S*, in Bree et al. 2013:334)

In the case of an ambiguous *you know*, it was not regarded as a discourse marker but a regular subject-verb sequence.

For tag questions, the different types of tag questions (e.g. negative *don't you?* or positive *do you?*) are not regarded as different items for frequency, due to all of them being replaced by *tagquestion*, which is not necessary for the study as it is merely the use of tag questions in general that is of interest regarding frequency. It is important to note that the *youknow* and *tagquestion* alterations are only applied in separate subcorpora for the software analysis. For the in-text analysis, I will look at the unaltered subcorpora to determine the different types and functions of the tag questions. These alterations would affect the results, but as I only applied the frequency analyses to the altered subcorpora the alterations will have no noticeable effect, which would be the case with the unaltered subcorpora.

3.2.3 Sociolinguistic Categorization of the Utterances

I have decided to analyze the correlation between the selected discourse markers and tag questions and the four variables ((1) gender, (2) intimacy, (3) social class, and (4) setting). In order to do so, I have divided the main variables into several sub-variables.

Firstly, the main variable (1) gender is only regarded from the biological point of view (i.e. sex) and therefore will not be taken to reflect upon the femininity or masculinity of a speaker but purely on whether the speaker is a man or a woman. The terms sex and gender will be used interchangeably throughout the study. The sub-variables of gender are male and female speech as well as same-gendered (male-to-male or female-to-female) and mixed-gendered (male-to-female or female-to-male) dialogue. The character Miss Bates has the tendency to walk into a room and start up conversations with all of the characters that happen to be present:

41. As the door opened she [Miss Bates] was heard,

> So very obliging of you! – No rain at all. […] Jane, Jane look! – did you ever see any thing? Oh! Mr, Weston, you must really have had Aladdin's lamp. […] Ah! dear Mrs. Elton, so obliged to you for the carriage! – excellent time. […] Oh! And I am sure our thanks are due to you, Mrs. Weston, on that score. […] Oh! Mr. Frank Churchill, I must tell you my mother's spectacles have never been in fault since; […] Dear Miss Woodhouse, how do you do? […] (Miss Bates, *Emma*)

Example 41 is an excerpt of a one-sided conversation, as we do not get to see the responses to Miss Bates' observations and enquiries. The entire monologue comprises 710 words. Miss Bates has several of these 'conversations' which have been added to the female speaker sub-variable but not to same-gendered or mixed-gendered variables, since it is at times unclear to whom the conversation is directed.

Secondly, the main variable (2), intimacy, has been divided into four sub-variables: close family (e.g. parent – child), distant family (e.g. cousins), friends (both close and acquaintances), and other (e.g. strangers or other relationships, such as mistress – servant, that do not fall within the first three categories). Furthermore, the intimacy level will be looked at from the speaker's point of view, so even if the addressee perceives the relationship in a different way the utterance will be judged from the speaker's standpoint. In Table 3.1 below, the relationship between characters (speaker to addressee) per story and novel are provided. The novels are ordered by way of date of completion or publication.

	Close Family	Distant Family	Friends	Other
Catharine, or the Bower	1. The Stanleys: *Mr.* and *Mrs. Stanley, Camilla* and *Edward* 2. The Percivals: *Mrs. Percival* and *Catharine*	1. *Catharine – Mr.* and *Mrs. Stanley* (distant relatives) 2. *Mrs. Percival –* The Stanleys (distant relatives)	1. *Catharine – Camilla* 2. *Catharine – Edward* (later on)	1. *Catharine – Anne* (mistress – servant) 2. *Catharine – Edward* (strangers)
The Watsons	1. The Watsons: *Mr. Watson, Elizabeth, Margaret, Jane, Robert* 2. The Edwards(es) 3. The Osbornes 4. The Blakes	1. *Mr. Blake – Charles Blake* (uncle – cousin)	1. The Watsons – *Tom Musgrave* 2. The Osbornes – *Tom Musgrave* 3. *Emma Watson – Elizabeth Watson* 4. The Edwards(es) – The Watsons	1. The Blakes – *Emma Watson* 2. The Watsons – *Nanny*
Northanger Abbey	1. The Thorpes: *Mrs. Thorpe, Isabella* and *John* 2. The Morlands: *Catherine* and *James* 3. The Tilneys: *General Tilney, Henry, Eleanor, Captain (Frederick)*		1. The Thorpes – The Morlands 2. The Thorpes – The Tilneys 3. The Thorpes – The Allens 4. The Tilneys – The Allens (later on) 5. The Tilneys – The Morlands (later on) 6. The Morlands – The Allens	1. The Tilneys – The Morlands (strangers) 2. The Tilneys – The Allens (strangers)
Mansfield Park	1. The Bertrams: *Sir Thomas* and *Lady (Maria) Bertram, Tom, Edmund, Maria, Julia* 2. The Bertrams – *Mrs. Norris* (sister and sister-in-law) 3. The Prices: *Mr.* and *Mrs Price, Fanny, Susan, William* 4. The Crawfords: *Mary and Henry*	1. The Bertrams – The Prices 2. *Mrs. Norris – Fanny Price*	1. The Bertrams – The Crawfords 2. *Fanny Price –* The Crawfords 3. *Mrs. Norris –* The Crawfords	

Emma	1. The Woodhouses: *Mr. Woodhouse, Emma, Isabella* 2. The Knightleys: *John* and *George* 3. The Westons: *Mr.* and *Mrs. Weston, Frank Churchill*	1. The Woodhouses – *John Knightley* 2. *Miss Bates – Jane Fairfax*	1. *Emma Woodhouse – Mr. George K.* 2. The Woodhouses – *Harriet Smith* 3. The Woodhouses – The Westons 4. The Woodhouses – The Eltons 5. *Miss Bates* – All (except Jane) 6. *Mr. George K. – Harriet Smith* 7. *Harriet Smith* – The Westons	
Sanditon	1. The Heywoods: *Mr. Heywood* and *Charlotte* 2. The Parkers: *Tom* and *Mary, Diana, Susan, Sidney, Arthur* 3. The Denhams: *Lady Denham* and *Sir Edward*		1. The Parkers – The Denhams 2. The Parkers – The Heywoods (later on) 3. The Denhams – *Charlotte Heywood*	1. *Tom and Mary Parker – Mr. Heywood* (strangers)

Table 3.1 Categorization of the (selected) characters by intimacy level (close or distant family, friends, or other) per novel or story. NB. The italicized words/names are separate characters, whereas those non-italicized are the family names.

The intimacy categorization deserves some explanation as, for example, in *Catharine, or the Bower*, Edward comes to Catharine's house to take her to a ball. Edward is aware of who Catharine is to him, but Catharine is not aware of who he is. At the beginning, they are strangers from Catharine's point of view but not from Edward's. Afterwards, when Edward reveals that he is the son of the Stanleys and the brother of her friend Camilla they are no longer strangers to each other. Furthermore, the reason why I categorized Emma and Elizabeth Watson as friends rather than close family is due to their more intimate relationship than the rest of the siblings in the story. Emma has lived with a rich aunt for most of her life and has hardly been in contact with her father or any of her siblings. The only one Emma has stayed in contact with, by way of correspondence, is Elizabeth. In this case, the relationships are chosen relative to Emma's position as the main character of the story, who regards Elizabeth as her friend but her other sisters, sister-in-law, and brother as something else. Relationships can become quite tricky in

Austen's fictional works at times, which was the case in *The Watsons*, but in her published novels the intimacy between characters is more intricate but also more developed.

Thirdly, the main variable (3), social class, has been divided into three sub-variables: lower-middle class, upper-middle class, and upper class. The social class of both the speaker and the addressee will be taken into account; since it could be that the social class of the speaker as well as the social class of the addressee correlates with the use of discourse markers and tag questions. The lower-class characters are barely represented in Jane Austen's novels and therefore I cannot study their language on its own. There are only two lower-class characters whose speech is added to the corpus, Anne in *Cath* who is Catherine's maid and Nanny in *TW* who is a servant to the Watson family. Table 3.2 provides an overview of which characters are classified as belonging to which class per novel:

	Lower(-middle) class (LMC)	Upper-middle class (UMC)	Upper class (UC)
Catharine, or the Bower	1. *Anne* (maid)	1. The Percivals	1. The Stanleys
The Watsons	1. *Nanny* (servant)	1. The Watsons 2. The Blakes 3. Tom Musgrave	1. The Edwards(es) 2. The Osbornes
Northanger Abbey	1. The Morlands	1. The Thorpes 2. The Allens	1. The Tilneys
Mansfield Park	1. Fanny Price 2. Mrs. Norris	1. The Crawfords	1. The Bertrams
Emma	1. Harriet Smith 2. Miss Bates 3. Jane Fairfax 4. The Eltons	1. The Woodhouses, 2. The Westons	1. The Knightleys
Sanditon		1. The Heywoods 2. The Parkers	1. The Denhams

Table 3.2 Categorization of the (selected) characters by social class (LMC, UMC, or UC) per novel or story.

Similar to the intimacy categorization, the social class of the characters is occasionally hard to determine, especially in the unpublished works. Characters that are being spoken to but are not mentioned in terms of their status or profession occur in the published novels as well. When the social class was left unspecified, I categorized the characters regarding their position in the community by way of the relationship to other characters or by what was mentioned about them. Anne is the maid in *Catharine, and or the Bower* and even though she is not lower-middle class, I classified her as such since she is below Catharine on the social scale. For *Emma*, not much is known about Miss Bates but what is known is that she is regarded as lower than the Woodhouses, especially so by the protagonist Emma Woodhouse. The Crawfords in *Mansfield Park* are not analyzed but they are being spoken to by the other characters. Mary and Henry

Crawford are described as lower in social class than the Bertrams but higher than Fanny Price, which is why I ranked them as upper-middle-class characters.

Lastly, the main variable (4), setting, refers to the pace where the discourse occurs. Setting has three sub-variables: festivity/social outing, discourse occurring during a ball in, for example *The Watsons* where Emma Watson goes to her first ball courtesy of the Edwards family:

> 42. The party passed on – Mrs. Edwards's sattin gown swept along the clean floor of the Ball-room, to the fire place at the upper end, where one party only were formally seated, while three or four Officers were lounging together, passing in and out from the adjoining cardroom.

The domestic setting refers to discourse occurring in someone's house, for example, the first dialogue in *Emma* where Emma Woodhouse and her father are having tea in their home:

> 43. Emma smiled and chatted as cheerfully as she could, to keep him [Mr. Woodhouse] from such thoughts; but when tea came, it was impossible for him not to say exactly as he had said at dinner[.]

The outdoors setting indicates discourse occurring while walking outside or meeting someone outside, for example, when Isabella and Catherine are walking outside in *Northanger Abbey* and come across their brothers John and James, respectively:

> 44. An inquiry now took place into the intended movements of the young ladies; and, on finding whither they were going, it was decided that the gentlemen should accompany them to Edgar's Buildings, and pay their respects to Mrs. Thorpe.

These variables and their corresponding sub-variables will help to understand the use of discourse markers and tag questions in the sociolinguistic light of the LModE period.

The main variables and sub-variables are used in order to split the selected texts into several subcorpora. The variable gender comprises six subcorpora: female speech, female-to-female speech, female-to-male speech, male speech, male-to-male speech, and male-to-female speech. The variable intimacy has four subcorpora: close family, distant family, friends, and other. The variable social class has been divided into nine subcorpora: upper class to upper class, upper class to upper-middle class, upper class to lower-middle class, upper-middle class to upper class, upper-middle class to upper-middle class, upper-middle class to lower-middle class, lower-middle class to upper class, lower-middle class to upper-middle class, and lower-middle class to lower-middle class. The variable setting comprises three subcorpora: festivity/social outing (festive),

domestic, and outdoors. In total, there are forty-four sets but only half of them, twenty-two, have been used for the lexical analysis software, namely, the altered subcorpora with *youknow* and *tagquestion* in order to count the separate instances. Table 3.3 gives an overview of the twenty-two altered subcorpora and the number of tokens per subcorpus:

Variable	Subcorpus	Tokens
Gender	SpeakerFemale	73,580
	SpeakerFtoF	44,390
	SpeakerFtoM	27,717
	SpeakerMale	49,544
	SpeakerMtoF	44,564
	SpeakerMtoM	4,808
Intimacy	IntimacyCloseFamily	15,876
	IntimacyDistantFamily	18,858
	IntimacyFriends	85,390
	IntimacyOther	3,321
Social Class	SocialHigh-High	2,804
	SocialHigh-Middle	17,409
	SocialHigh-Low	19,462
	SocialMiddle-High	7,214
	SocialMiddle-Middle	24,029
	SocialMiddle-Low	19,930
	SocialLow-High	12,473
	SocialLow-Middle	17,271
	SocialLow-Low	2,655
Setting	SettingDomestic	69,825
	SettingFestive	26,387
	SettingOutdoors	28,685

Table 3.3 The number of tokens per subcorpus per main variable. NB. For gender, the F stands for female and the M stands for male. For social class, High stands for the upper class, Middle for the upper-middle class, and Low for the lower(-middle) class.

3.3 Corpora Analysis

For the analysis of the corpora I will use the lexical analysis software called WordSmith Tools, version 5.0. WordSmith was designed by Mike Scott and the late Tim Johns and subsequently published by a collaboration of the Oxford University Press and Lexical Analysis Software Ltd. in

1996. Version 5.0 came out in 2010. Even though this is not the most current version, it is the only one to which I have full access. WordSmith consists of several sub-programs that allow the user to get a deeper understanding of a text by analyzing its words, or more eloquently put, by analyzing how words behave in the text. There are three main sub-programs: Concord, WordList, and KeyWords.

Concord allows you to search for a word or a phrase in the context of the sentence in which it occurs and allows you to see where in the text it is located and which words are in its immediate proximity. Concord also provides information on how often a word occurs per 1,000 words. An edited example may be found below in which the concordance of *oh* is given for the first four instances in *Northanger Abbey*:

45.

N	Concordance		
1	what are you reading, Miss - ?" "	**Oh**!	It is only a novel!"
2	you have not been here long?" "	**Oh**!	These ten ages at least.
3	"Are you, indeed? How delightful!	**Oh**!	I would not tell you
4	Are not you wild to know?" "	**Oh**!	Yes, quite; what can

WordList generates a frequency list to see how many times a word occurs in a text as well as an alphabetical list to see the different inflections and derivations of a headword. Example 46 below demonstrates a frequency list of *Emma* with the four most used words as it occurs in WordSmith and example 47 shows the inflections and derivations of *accept* for *Emma*.

46.

N	Word	Freq.	%	Texts	%
1	TO	5,242	3.26	1	1000.00
2	THE	5,204	3.23	1	1000.00
3	AND	4,897	3.04	1	1000.00
4	OF	4,293	2.67	1	1000.00

47.

N	Word	Freq.	%	Texts	%
40	ACCEPT	11		1	1000.00
41	ACCEPTABLE	8		1	1000.00
42	ACCEPTANCE	2		1	1000.00
43	ACCEPTED	11		1	1000.00
44	ACCEPTING	6		1	1000.00
45	ACCEPTS	1		1	1000.00

WordList also provides information on the number of types (i.e. the total number of different words) and tokens (i.e. the total number of words). Furthermore, creating Wordlists is an integral part to the following function, KeyWords, which only accepts wordlist files.

KeyWords is a specific type of analysis with which you can analyze a text, or a smaller corpus, in relation to a reference corpus, a larger one. The analysis allows you to find so-called *key words* in a text, "whose frequency is unusually high in comparison to some norm [i.e. the reference corpus]" (Scott 2010). I will not go further into the KeyWords function as I will not use it for my own analysis due to the variability in the number of tokens per subcorpus, see Table 3.3, which make it hard to analyze one subcorpus within a variable to the other subcorpora belonging to that variable as the reference corpus needs to be substantially larger.

For discourse marker and tag question analysis, I have used both Concord and WordList. Concord is used to show patterns that occur with the discourse markers *oh*, *ah*, *well*, *why*, and *you know* (see example 44 above), their position in the sentence, and also the occurrence of an item per 1,000 words. Concord is especially important when analyzing the status of *well* and *why* in an utterance due to their function of both a discourse marker and an (interrogative) adverb. The irrelevant instances of *well* and *why* are deleted and removed from the concord list by *zapping* those away. The data can be *zapped* by way of a button which can be found under *Edit* or by pressing the Ctrl and Z key simultaneously. This allows for a reanalysis of the original data without the irrelevant entries. WordList calculates the frequency of the discourse markers and tag questions per subcorpus as well as generates wordlists. Since I cannot use KeyWords due to the differing subcorpus sizes, I have decided to use the chi-squared test (cf. Burrows 1987:34–45) using the statistical analysis software SPSS.

The chi-squared test analyzes whether the differences of the frequencies with which items occur differ significantly, with a significance level of $p<0.0001$ as recommended by Rayson et al. (2004) for corpus studies (referred to in Huang 2011:49). In other words, I will be able to determine whether the use of the discourse markers and tag questions in one subcorpus differ significantly from that of the other subcorpora.

This means that for variable (1), gender, I will analyze whether the use of the items are significantly different when comparing the female speaker subcorpus to the male speaker subcorpus. In addition, I will look at whether there is a correlation between usage when a female speaker addresses a female speaker or a male speaker, and when a male speaker addresses a female speaker or a male speaker, i.e. same-gendered and mixed-gendered discourse. Similarly, I will compare whether there is an overall difference when looking at the addressee, i.e. comparing the SpeakerFtoM and SpeakerMtoM subcorpus to the SpeakerFtoF and SpeakerMtoF subcorpus.

The SpeakerMtoM subcorpus consists of only 4,808 tokens, which is almost ten times as small as that of the male-to-female subcorpus (see Table 3.3). However, as I will normalize the data, the size of the subcorpus should not matter, but relative to the number of discourse markers or tag questions this might become an issue which will be addressed when analyzing the data.

For (2) intimacy, I will compare the close family subcorpus, the distant family subcorpus, the friends subcorpus, and the 'other' subcorpus to one another. I will also look at the family subcorpora combined and see how speech to the family differs from that to friends or other relationships. Just as the male-to-male subcorpus, the subcorpus IntimacyOther is quite small compared to the other subcorpora of the intimacy variable; in fact, the friends subcorpus is almost twenty times as big.

The variable (3) social class is divided into higher classes, middle classes, and lower classes. I will compare the subcorpora belonging to the upper classes, the upper-middle classes, and the lower(-middle) classes to one another to see whether there are differences between speakers belonging to different social classes. Additionally, I will look at the speech of speakers belonging to a specific social class to addressees of the same social class and those of different social classes; for example, comparing upper-class speech directed towards the upper-class speakers, the upper-middle class, and the lower(-middle) class. Moreover, I will also look at the addressees specifically to determine the use of discourse markers and tag questions when they are being spoken to by members of the upper class, of the upper-middle class, and the lower(-middle) class.

In relation to the variable (4) setting, I will contrast the subcorpora belonging to the different settings to one another: domestic to festive, festive to outdoors, and outdoors to domestic. Furthermore, I will use the chi-squared test to determine whether there is a correlation between specific discourse markers and the variables using only the number of discourse markers that were found per subcorpus. The previous comparisons are based solely on the normalized data regarding the size of the subcorpora. The analysis when looking specifically at the discourse markers will determine whether they are more frequent as opposed to other discourse markers in a single subcorpus. For example, if there are a total of 150 discourse markers in a subcorpus and 60 of those are the interjection-based discourse marker *oh*, I will try to determine whether the use of *oh* is significantly different from the remaining discourse markers in that subcorpus.

CHAPTER 4: RESULTS

4.1 Introduction

The present chapter reports on the findings generated by my analysis described in the previous chapter. The structure of the chapter will be as follows: firstly, I will discuss the findings relating to *oh* and *ah* in section 4.2, both quantitatively and qualitatively, and similarly to the other discourse markers of which *well* and *why* in section 4.3, are also grouped together, just as was the case in Chapter 2. The discourse marker *you know* has a separate section, section 4.4, as do the tag questions, located in section 4.5. Furthermore, each section has been structured by the four variables that were analyzed: (1) gender, (2) intimacy, (3) social class, and (4) setting. In relation to the tables, the numbers in the tables are based on normalization per 10,000 words so that the numbers become more understandable. Some subcorpora do not consist of 10,000 words and therefore the table shows an estimate of how many items there would have been of the subcorpus comprised 10,000 words. The statistical analyses, however, are based on the real numbers which are provided in brackets after the normalized results.

4.2 Interjections: *Oh* and *Ah*

4.2.1 Gender

In the corpus, 288 instances of *oh* were found and only 32 of *ah*. There is a clear and significant difference between the use of *oh* and *ah*: *oh* is used significantly more (p<.0001) than *ah* in general. In the female subcorpus, as represented in Table 4.1, women were more prone to use *oh* than *ah* both in same-gendered and mixed-gendered conversations (p<.0001).

	Oh	*Ah*
Female	33/10000 (242)	3/10000 (22)
Female to Female	32/10000 (141)	2/10000 (11)
Female to Male	33/10000 (91)	3/10000 (8)

Table 4.1 Distribution of interjection-based discourse markers *oh* and *ah* in the female subcorpora normalized per 10,000 words.

In certain instances, the use of *oh* and *ah* was used in mixed company or as was the case in example 42 of one of Miss Bates' monologues, the addressee of the speaker was not clear. For this reason, the number of instances of the female subcorpus are not attained by adding the female-to-female and female-to-male subcorpus. Table 4.1 shows that women barely adjusted their speech in relation to their addressee regarding *oh* or *ah* use. Women had a tendency to use *oh* and *ah* slightly more in mixed-gendered conversations. For the male speakers, there also was a difference in *oh* and *ah* use, as they used *oh* more, but it was not a significant difference, see Table 4.2.

	Oh	*Ah*
Male	9/10000 (46)	2/10000 (10)
Male to Female	10/10000 (43)	2/10000 (9)
Male to Male	6/10000 (3)	2/10000 (1)

Table 4.2 Distribution of interjection-based discourse markers *oh* and *ah* in the male subcorpora.

Regarding *oh* there was a difference between same-gendered and mixed-gendered conversations, as men tended to use *oh* more in female-directed speech. In *Ah* use, there was no difference between male- and female-directed speech.

As for the differences between male and female speech, there is a significant correlation, $p<.0001$ in the use of the interjections *oh* and *ah*. Women use more interjections than men in general and specifically so regarding the use of *oh* in both same-gendered as mixed-gendered conversations. In addition, there is a difference between speech directed to women and men irrespective of the gender of the speaker:

	Oh	*Ah*
to Female	21/10000 (184)	2/10000 (20)
to Male	29/10000 (94)	3/10000 (9)

Table 4.3 Distribution of interjection-based discourse markers *oh* and *ah* in relation to the sex of the addressee.

The findings in Table 4.3 demonstrate that male-directed speech has both more *oh*s and *ah*s than female-directed speech. There is a significant difference, $p<.0001$, between the use of *oh* and *ah* in speech directed at men, as the speech has more *oh*s than *ah*s.

The position of *oh* in female speech is predominantly turn-initial, ($p<.0001$). There are only four exceptions: one by Fanny Price in *Mansfield Park* and three in *Emma* by Emma Woodhouse, Harriet Smith, and Miss Bates whose example can be found below:

48. Mr. Churchill, **oh!** you are too obliging. (Miss Bates, *Emma*)

Furthermore, *oh* clusters tend to be followed by either a negation (i.e. *Oh no*) or affirmation (i.e. *Oh yes*): thirty-five and thirty-three times, respectively. *Oh* is also frequently followed, twenty-five times, by *dear* which either indicated whether it was used as a marker of surprise, sympathy, or whether *oh* was used to add to the negation of the statement, as in example 49, or as part of an address, see example 50.

49. I do not think it will.
Oh! dear, no. (Emma Woodhouse and Harriet Smith, *Emma*)

50. **Oh! dear** Mrs. Percival, you must not beleive every thing that my lively Camilla says. (Mrs. Stanley, *Cath*, in Bree et al. 2013:193)

Similar to the female *oh* position, the position of *oh* in the male subcorpus is predominantly turn-initial, p<.0001. There are two occurrences where *oh* is in turn-medial position, one of them is the following example where *oh* reflects an outburst of sentimentality:

51. [It] was the detection, in short – **oh**, Fanny! it was the detection, not the offence, which she reprobated. (Edmund Bertram, *MP*)

Furthermore, the only *oh* cluster highlighted in the male subcorpus is the affirmation cluster, *oh yes*, which occurs five times. Women seem to be more constant in what succeeds *oh*.

Ah is only used in turn-initial position in both male and female speech. There are no recurrent patterns for either of the male or female subcorpus, which is probably due to the number of *ah*s: ten in the male subcorpus and twenty-two in the female one.

4.2.2 Intimacy

The distribution of *oh* in relation to the variable of intimacy did not have any significant results at the p<.0001 level, but there were certain tendencies that could be perceived.

	Oh	*Ah*
Close Family	16/10000 (25)	2/10000 (3)
Distant Family	13/10000 (25)	1/10000 (2)
Friends	27/10000 (232)	3/10000 (27)
Other	12/10000 (6)	0
Total	288	32

Table 4.4 Distribution of interjection-based discourse markers *oh* and *ah* in terms of the intimacy level of the speaker and addressee.

Table 4.4 shows that the level of family, irrespective of closeness or distance, did not show a lot of difference. In close family interaction, *oh*s were used more frequently than in discourse between distant family members. At the level of friends, *oh* was used quite often in comparison to the family and 'other' subcorpus, twenty-seven times per 10,000 words. Furthermore, regarding the distribution of *oh* and *ah* in the friends subcorpus, *oh* is used significantly more frequent than *ah*. *Ah* only shows a slight difference in use. *Ah* is used most often in the friend subcorpus, similar to *oh*, and least often in the 'other' subcorpus. At the family level, *ah* is used more in close family interaction than in distant family interaction.

It was mentioned above that there were six *oh*s that did not occur turn-initially. Four of these occur in the friends subcorpus one of which was shown above in example 47, and two in the distant family subcorpus, see example 50 for Edmund's example. Furthermore, the clusters, as obtained through concordance, of *oh* in the friends subcorpus are similar to the female subcorpus as thirty-three instances are a negation, *oh no*, and twenty-six an affirmation, *oh yes*. The

third most occurring cluster is also *oh dear* for which *dear* can either be an interjection or a mode of address. In the combined family subcorpus, the *oh* clusters are predominantly used in affirmation: *oh yes* occurs ten times and *oh no* five times. A mode of address occurs as well in the form of *oh cousin*, which occurs five times. The close family subcorpus in itself does not have any recurrent patterns regarding *oh*, but the distant family subcorpus does. *Oh yes* occurs six times at the distant family intimacy level and *oh cousin* occurs five times. *Ah*, on the other hand, does not show any recurrent patterns at any of the intimacy levels.

4.2.3 Social Class

Though the upper-class speakers use both the interjections *oh* and *ah*, *ah* is again used least often, but not significantly so. Table 4.5 shows that *Oh* is used thirteen times and *ah* is only used once per 10,000 words.

	Oh	*Ah*
High-High	7/10000 (2)	0
High-Middle	25/10000 (44)	2/10000 (4)
High-Low	3/10000 (6)	0
Total	13/10000 (52)	1/10000 (4)

Table 4.5 Distribution of interjection-based discourse markers *oh* and *ah* in the upper-class subcorpora.

Oh is used in all of the upper-class subcorpora: towards fellow members of the upper class, towards the upper-middle class, and towards the lower(-middle) class. *Ah*, on the other hand, is only used in upper-class-to-upper-middle-class speech and it is used significantly less so compared to the *oh* use in the upper-middle-class directed speech (p<.0001). In addition, there is also a correlation between upper class to upper-middle and upper class to lower(-middle) class. *Oh* occurs significantly more frequently in upper-class-to-upper-middle-class directed speech, i.e. twenty-five times per 10,000 words (p<.0001), compared to speech directed at the lower(-middle) class, i.e. three per 10,000 words. Furthermore, there is also a substantial difference between upper-to-upper-class and upper-to-upper-middle-class speech but not a significant one.

Oh was found only once in turn-medial position, for which see example (50) from Edmund, which is directed to a lower(-middle)-class character, his cousin Fanny in *Mansfield Park*. The clusters of *oh* in general are *oh dear* and *oh yes*, which occur six times, and *oh I*, which occurs five times.

52. [F]or you know he is Lord Amyatt's Brother.
Oh! **I** know all that very well, but it is no reason for their being so horrid. (Catharine Percival and Camilla Stanley, *Cath*, in Bree et al. 2013:175)

Example 52 shows an example of *oh* which is immediately followed by *I*. In this case, *Oh* functions as an affirmative of the fact that Camilla is aware of the relationship of that the *he* in question is Lord Amyatt's Brother. Camilla uses the *Oh I* cluster four of the five times and the only other time, her brother, Edward uses it:

53. **Oh**! **I** knew you would be delighted with me for making so much haste –. (Edward Stanley, *Cath*, in Bree et al. 2013:187)

All instances are directed towards their friend, Catharine, who is an upper-middle-class character. In the High-Middle subcorpus specifically, the *oh dear* cluster occurs six times and the *oh I* cluster occurs five times. Interestingly, all instances of *oh dear* and *oh I* occur in the first part of the subcorpus, which corresponds with the first story, *Catharine, or the Bower*.

Table 4.6 shows the distribution of *oh* and *ah* in the speech of the upper-middle-class speakers.

	Oh	*Ah*
Middle-High	24/10000 (17)	6/10000 (4)
Middle-Middle	15/10000 (36)	3/10000 (7)
Middle-Low	25/10000 (50)	3/10000 (6)
Total	20/10000 (102)	3/10000 (17)

Table 4.6 Distribution of interjection-based discourse markers *oh* and *ah* in the upper-middle-class subcorpora.

The upper-middle-class characters use *oh* and *ah* regardless of the social class of the addressee. *Oh* is used least often in upper-middle-to-upper-middle-class discourse, which contrasts with upper to upper-middle-class speech where *oh* was used the most. The distribution of *oh* in upper-middle to upper and upper-middle to lower(-middle) is similar and therefore it is in the speech directed to their fellow class members that there appears to be a decrease in *oh*. Furthermore, there is a significant difference between the use of *oh* and *ah* in upper-middle to lower(-middle) class, as *oh* is used more often than *ah*, ($p<.0001$). *Ah* is used most often in upper-middle-to-upper-class discourse, six occurrences per 10,000 words. The distribution of *ah* amongst upper-middle and lower(-middle)-class directed speech is the same, three per 10,000 words. In addition, the upper-middle-class speech directed to the lower(-middle) class has significantly more occurrences of *oh* than the upper class speech directed to the lower(-middle class.

Oh is predominantly turn-initial; there are two instances where *oh* is preceded by either a coordinator or an affirmation (and also followed by one), both of which occur in upper-middle-to upper-middle-class conversations in *Emma*:

54. [F]or, **oh!** Mrs. Weston, if there were an account drawn up of the evil and the good I have done Miss Fairfax! (Emma Woodhouse, *Emma*)

55. Yes, **oh!** yes; I was just going to mention it. (Frank Churchill, *Emma*)

The first *oh* expresses the shame that Emma feels about her treatment of Jane Fairfax and the second *oh* interjects a thought that just came up in Frank Churchill's mind. The most predominant *oh* clusters in the upper-middle speaker subcorpora are *oh no*, which has fifteen occurrences, *oh yes*, which has fourteen, and *oh I*, which has eleven. In upper-middle-to-upper-middle-class speech, *oh yes* and *oh no* were found seven times which are more or less equally distributed amongst the stories and novels as apparent from the concordance plot. Similarly, upper-middle speech to lower(-middle) speech predominantly has the clusters *oh yes*, which occurs five times, and *oh I*, which also occurs five times. *Ah*, on the other hand, is solely turn-initial and has no recurrent patterns with a minimum of five.

The lower(-middle) classes use both *oh* and *ah*, but they use *oh* significantly more than *ah* (p<.0001) no matter at whom the speech is directed: forty-one instances per 10,000 words for *oh* and three instances for *ah*, as can be seen in Table 4.7.

	Oh	*Ah*
Low-High	36/10000 (46)	3/10000 (4)
Low-Middle	46/10000 (79)	3/10000 (6)
Low-Low	30/10000 (8)	4/10000 (1)
Total	41/10000 (132)	3/10000 (11)

Table 4.7 Distribution of interjection-based discourse markers *oh* and *ah* in the lower(-middle)-class subcorpora.

Similar to the upper class, the lower(-middle) class uses *oh* the most in speech directed to the upper-middle class, the second most to the upper class, and the least to members of their own class. The lower(-middle) class, in comparison with the upper class, use *oh* significantly more: forty-one times per 10,000 words to the upper class's seven. Furthermore, the lower(-middle) class compared to the upper-middle class uses significantly more *oh* when conversing to the upper-middle-class characters. The same holds true for the lower(-middle)-class speech directed to the upper class and the upper-class speech directed to the members of their social class. Just as was the case with the upper-middle to the lower(-middle) speech, there is a significant difference between the occurrences of *oh* in lower(-middle)-to-lower(-middle) class and upper-to-lower(-middle)-class speech. *Oh* is used predominantly more in lower class speech in general.

The main *oh* clusters in lower(-middle) class speech are similar to the other social classes. The main cluster is *oh no*, which occurs twenty times. The other main clusters are *oh yes*, with seventeen occurrences, and *oh dear*, with fifteen. Lower speech to higher speech has both *oh no* and *oh yes*, but also *oh cousin*, referred to in 4.2.2. In lower to upper-middle speech, *oh dear* is the prevalent cluster with *oh no* being a close second. The third most frequent cluster is *oh miss* which occurs eleven times of which ten times are *oh miss woodhouse* used by both Miss Bates and Harriet

Smith in *Emma*. Discourse amongst lower-class characters rarely took place and therefore resulted in a smaller subcorpus, which had only eight occurrences of *oh* and no recurrent patterns.

Table 4.8 displays the findings regarding the distribution of *oh* and *ah* in the combination of the subcorpora based on the addressee, i.e. regardless of the social class of the speaker.

	Oh	*Ah*
to High	28/10000 (65)	4/10000 (8)
to Middle	27/10000 (159)	3/10000 (17)
to Low	15/10000 (64)	2/10000 (7)

Table 4.8 Distribution of interjection-based discourse markers *oh* and *ah* in relation to the social class of the addressee.

Speech directed towards the upper-class characters has both the most *oh*s and *ah*s, twenty-eight and four occurrences per 10,000 words. Lower(-middle)-class directed speech has the least interjections. There is a significant difference regarding the use of *oh* and *ah* in speech directed to upper- and upper-middle-class characters, which has been the case in general with *oh* as the more predominant interjection.

The three main *oh* clusters in upper-class directed speech are *oh no*, eleven times, *oh yes*, ten times, and *oh cousin*, five times. *Oh yes* also occurs as a main cluster in upper-middle class, nineteen times, and lower(-middle)-class directed speech, eight times. *Oh cousin* has been referred to earlier in 4.2.2 in relation to the distant family subcorpus. There are two clusters that occur most often in speech directed at upper-middle-class characters: *oh no* and *oh dear* each occurring twenty-three times. *Oh dear* appears to be a cluster that is a marker of speech directed towards the upper-middle class as it does not occur in any of the other subcorpora combinations as a frequent cluster. Furthermore, there were also frequent three-word clusters with *oh dear*: *oh dear I*, six times, and *oh dear yes*, five times. In lower(-middle)-class directed speech, the clusters *oh my* and *oh I* occur five times, with *oh yes* being the most frequent.

4.2.4 Setting

Oh and *ah* occur in each of the settings. Table 4.9 shows that o*h* is used more in all settings. However, only the domestic and festive setting have significantly more *oh*s than *ah*s.

	Oh	*Ah*
Domestic	24/10000 (166)	3/10000 (20)
Festive	29/10000 (77)	3/10000 (8)
Outdoors	15/10000 (45)	1/10000 (4)

Table 4.9 Distribution of interjection-based discourse markers *oh* and *ah* in relation the setting in which the discourse takes place.

Oh occurs most often in a festive setting and least often in an outdoors setting: twenty-nine and fifteen per 10,000 words, respectively. *Ah* is used just as frequently in a festive setting as in a

domestic setting. There is no significant difference regarding the distribution of *oh* or *ah* where p<.0001.

Oh no and *oh yes* are the most frequent clusters in each of the settings. *Oh dear* is a frequent cluster in the domestic setting which occurs fifteen times. In a festive setting, *oh mr* occurs six times.

56. **Oh! Mr.** Knightley, one moment more; something of consequence – so shocked! (Miss Bates, *Emma*)

It is very likely that festive settings are where formal address towards a male character is deemed appropriate. Phillipps (1970:208) explains that "modes of address are a good deal more formal in the novels than they are today" and therefore addressing George by Mr. Knightley was common and an expected form of address. However, Phillipps further mentions that calling people solely by their first name or by their last name is highly informal and only appropriate depending on the relationship between the characters (1970:208–209). Therefore, it is expected when out in society to use *mister* when addressing the person or when referring to the person. The use of *oh* may not be stigmatized or at least, not in the way that not using *mister* would have been.

4.3 General Discourse Markers: *Well* and *Why*

4.3.1 Gender

In general, there are altogether 149 instances of well and 22 of *why*. As I have referred to *well* and *why* as general discourse markers, I will compare them to one another just as I have done with *oh* and *ah* in the previous section. *Well* is used significantly more than *why* in the corpus (p<.0001). There are fourteen instances of *well* per 10,000 words in the main female subcorpus, see Table 4.10.

	Well	*Why*
Female	14/10000 (103)	2/10000 (14)
Female to Female	14/10000 (62)	2/10000 (8)
Female to Male	12/10000 (34)	1/10000 (2)

Table 4.10 Distribution of the discourse markers *well* and *why* in the female subcorpora.

Female-to-female discourse uses the same number of *well*s as female discourse in general. Women use *well* less often when talking to men. *Why* is similar to *well* in that it is used more in same-gendered discourse, two per 10,000 words, than in mixed-gendered discourse, one per 10,000 words.

Well occurs predominantly in turn-initial or sentence-initial position which usually functions as a response to what was said before, be it a disagreement or an agreement. The

following example is from a scene in which Catharine is being shown around the grounds of Northanger Abbey by Henry Tilney and his father, the general:

57. [I]t waits only for a lady's taste!
Well, if it was my house, I should never sit anywhere else.
(General Tilney and Catherine Morland, *NA*)

Well could be taken to function as a marker of acknowledgement of the fact that Catherine is a lady, but also that she wants to give her input and that she uses *well* as a turn-taking mechanism to give her opinion of the room she is being shown. The most frequently occurring cluster of *well* is *well I*, which shows that *well* is used most often in an initial position. *Well and* is also a frequent cluster, which occurs predominantly in a turn-medial position to add a phrase to the previous one of the current speaker or in a turn-initial position as a response to a statement of the previous speaker.

58. Oh! yes, that has vexed me more than you can imagine–. **Well, and** so Edward got to Brampton just as the poor Thing was dead […].
(Camilla Stanley, *Cath*, in Bree et al. 2013:190)

59. Six years hence! Dear Miss Woodhouse, he would be thirty years old!
Well, and that is as early as most men can afford to marry, who are not born to an independence. (Harriet Smith and Emma Woodhouse, *Emma*)

Well also occurs as *well well*, signaling impatience and enthusiasm on behalf of the speaker:

60. Good God! **Well! Well!** now tell me every thing; make this intelligible to me.
(Emma Woodhouse, *Emma*)

In the female-to-female subcorpus, *well miss* and specifically *well miss woodhouse* were found as a frequent cluster similar to *oh*. There are no recurrent patterns in conversations with men.

Similarly *why* is a predominantly turn-initial marker which has a vast number of functions even though there were only fourteen instances in the female subcorpus. *Why* is used to agree or disagree with a previous statement, to express astonishment over it, or to state that it is considered as being obvious to the speaker. The example below shows Catherine commenting acquiescently to Camilla's story about her brother Edward.

61. **Why** indeed, it seems to have been a very shocking affair from beginning to end.
(Catharine Percival, *Cath*, in Bree et al. 2013:190)

There are no recurrent patterns with *why* in the female subcorpus, but there are but few instances of *why* in the corpus in general.

In the male subcorpus, there are less *well*s than in the female subcorpus, ten per 10,000 words, and an equal number of *why*s, two per 10,000, see Table 4.11.

	Well	*Why*
Male	10/10000 (46)	2/10000 (8)
Male to Female	9/10000 (42)	2/10000 (7)
Male to Male	6/10000 (3)	2/10000 (1)

Table 4.11 Distribution of the discourse markers *well* and *why* in the male subcorpora.

Male speakers tend to use more *well*s in conversations with women than with men: nine and six instances per 10,000 words, respectively. Regarding *why*, there are no differences between same-gendered and mixed-gendered conversations in the male subcorpora, two per 10,000 words.

There are two prevalent clusters in the male general subcorpus: *well I* and *well well*, similar to the female subcorpus. In the male subcorpus, specifically the male-to-male subcorpus, *well* has a concluding function to mark the end of the discourse:

62. **Well** Sir, we will settle this point some other time […].
    (Edward Stanley, *Cath*, in Bree et al. 2013:193)

63. **Well**, there is nothing more to be said. (Edmund Bertram, *MP*)

Well well also has different functions or expresses different sentiments than those in the female subcorpus. *Well well* can be used as a type of 'there, there' with a comforting undertone: for example,

64. **Well**, **well**, child, dry up your tears. (Sir Thomas Bertram, *MP*)

Why is used in similar ways in the male subcorpus as in the female subcorpus expressing feelings of shock or surprise over a question just asked:

65. Going to? **Why**, you have not forgot our engagement! (John Thorpe, *NA*)

Why is also frequently turn-initial. The example above is one of the two, out of the eight occurrences, that is not turn-initial. The other turn-medial *why* is also by John Thorpe and both directed at his love-interest, Catherine Morland.

Table 4.12 reveals that there is barely any difference between the use of *well* directed towards female or male speakers, nor is there for *why*. However, proportionally speaking, men use more *well*s, eleven to one, in comparison to *why* than women do, six to one.

	Well	*Why*
to Female	12/10000 (104)	2/10000 (15)
to Male	11/10000 (37)	1/10000 (3)

Table 4.12 Distribution of the discourse markers *well* and *why* in relation to the sex of the addressee.

4.3.2 Intimacy

There is no significant difference between the use of *well* and *why* regarding the relationship between the speaker and addressee, be it family, friends or some other connection. In fact, there is little difference when we look at the discourse markers individually. The friends subcorpus carries most instances of *well*, thirteen per 10,000 words. For *why*, it is the 'other' subcorpus that carries most *why*s, nine per 10,000 words, but bear in mind that this subcorpus has the least occurrences when looking at the actual number and not the normalized one. Fewest *well*s and *why*s occur in the family subcorpora.

	Well	*Why*
Close Family	11/10000 (17)	1/10000 (1)
Distant Family	10/10000 (19)	1/10000 (2)
Friends	13/10000 (109)	2/10000 (16)
Other	12/10000 (4)	9/10000 (3)
Total	149	22

Table 4.13 Distribution of the discourse markers *well* and *why* in terms of the intimacy level of the speaker and addressee.

The family subcorpus has a couple of frequent clusters such as *well my*, *well I*, and *well Fanny*. *Well Fanny* occurs predominantly in *Mansfield Park* and is used by Fanny's uncle, aunt, and cousin, i.e. present in the distant family subcorpus. *Well Fanny* is used as a conversation starter, see example 66, or when there are more individuals present and the utterance is solely directed at one.

66. **Well**, **Fanny**, I have seen Mr. Crawford again, and learn from him exactly how matters stand between you. (Sir Thomas Bertram, *MP*)

Well my is used in the same way as it is often followed by *dear* and thus has an air of familial love or affection. In the friends subcorpus, there are no instances of *well my* or *well* followed by a first name as a common cluster. *Well miss*, in particular *well miss woodhouse*, is a common cluster. Similar to the male and female subcorpus, *well I* and *well well* are common clusters. In the 'other' subcorpus, there are only four instances and therefore no recurrent pattern.

There are only three occurrences of *why* in the family subcorpus, each followed by another discourse marker: *why you know* and *why indeed*, twice, followed by the name of the addressee. Regarding the friends subcorpus, there is a phrasal construction which occurs three

times: *why*(,) *to own the truth*. In the 'other' subcorpus, *why* is used by Henry Tilney when he is not acquainted with Catherine yet in *Northanger Abbey* and by the servant of Catharine, Anne, in *Catharine, or the Bower*:

67. What livery has his Servants?
Why that is the most wonderful thing about him Ma'am, for he has not a single servant with him […].
(Catharine Percival and Anne, *Cath*, in Bree et al. 2013:183).

4.3.3 Social Class

There were no significant differences between the use of *well* and *why* individually regarding the social class of the speaker, the social class of the addressee, or the interaction of the social classes. As represented in Table 4.14, the upper class uses *well* more than *why* in general.

	Well	*Why*
High-High	18/10000 (5)	0
High-Middle	10/10000 (18)	3/10000 (5)
High-Low	8/10000 (15)	2/10000 (4)
Total	10/10000 (38)	2/10000 (9)

Table 4.14 Distribution of the discourse markers *well* and *why* in the upper-class subcorpora.

Well is used the most in upper-to-upper-class speech, eighteen times per 10,000 words; contrastively, *why* is used the least in this context as it does not occur. *Well* occurs second most frequently in the upper to upper-middle-class context, where *why* is most frequent. *Well* is used least often in speech directed to the lower classes. In general, the upper-class speakers use *well* ten times per 10,000 words and *why* two times per 10,000 words.

Upper-class speakers as a whole, i.e. regardless of the social class of the addressee, have two recurrent patterns *well well* and *well I*. *Well well* occurs in both the upper-to-upper-middle as the upper-to-lower(-middle)-class speech. In the latter, the cluster is used as a comforting gesture as well as a concluding one and in the former, *well well* is used to indicate subdued surprise or disbelief:

68. **Well**, **well**, means to make her an offer then. (Mr. Knightley, *Emma*).

Another frequent cluster is *well I*, which occurs most of the time in upper-to-upper-middle class discourse. The *well Fanny* that was mentioned above in section 4.3.2 in relation to the distant family subcorpus occurs four times in the higher class subcorpus as her relations are mainly upper class.

Why has one recurrent pattern in relation to the upper class, namely, *why indeed* which occurs three times. The combination is used to express agreement with the previous statement. In the following example, Mrs. Stanley is conversing with Mrs. Percival about Catharine's sentimental behavior:

69. **Why indeed** Mrs. Percival, I must think that Catherine's affection for the Bower is the effect of a Sensibility that does her Credit.
(Mrs. Stanley, *Cath*, in Bree et al. 2013:180)

The upper-middle class uses *well* slightly more than the upper class, twelve times per 10,000 words, whereas *why* is used just as much by the two classes. Table 4.15 shows that the upper-middle class uses *well* most often when speaking to the lower classes and that *why* is used most often when addressing the upper class.

	Well	*Why*
Middle-High	8/10000 (6)	6/10000 (4)
Middle-Middle	10/10000 (24)	1/10000 (3)
Middle-Low	15/10000 (30)	2/10000(3)
Total	12/10000 (60)	2/10000 (10)

Table 4.15 Distribution of the discourse markers *well* and *why* in the upper-middle-class subcorpora.

Well is used least in upper-middle-to-upper-class discourse, for *why* this is the case in upper-middle-to-upper-middle-class speech.

The clusters *well I* and *well well* are also prevalent in the upper-middle-class subcorpora. *Well this is* and *well I have* are common three-word clusters, each occurring five times.

70. **Well**, **this is** very extraordinary! Very extraordinary indeed!
(Diana Parker, *S*, in Bree et al. 2013:376)

71. **Well**, **I have** so little confidence in my own judgment, that whenever I marry, I hope some body will chuse my wife for me. (Frank Churchill, *Emma*)

Well this is occurs as a cluster in which *well* most often expresses shock or astonishment as in example 70. The *well* in *well I have* has a predominantly concluding undertone and is either followed by a concluding statement or is followed by a change in topic. A topic change can be found in example 71 where the preceding conversation covers the making of new acquaintances. In the upper-middle-to-upper-class subcorpus, there are no recurrent patterns as there are only six instances. The upper-middle-to-upper-middle subcorpus does have one recurring pattern which is *well I*, but this cluster is more than twice as much represented in the upper-middle-to-lower(-middle)-class subcorpus, namely eleven times to four.

Why only occurs ten times in the upper-middle-class subcorpora and therefore there are no real patterns to discover. However, when regarding the distribution in relation to the class of the addressee, it seems that when addressing someone of the lower class, the speech is more direct (i.e. followed by *you*). John Thorpe's example, 65 above, shows this, whereas in speech towards the upper-middle class *why* is followed by phrases such as *in truth Sir* or *to own the truth*; and in speech directed towards the upper class *why* is followed by *indeed*, *yes*, or *really*.

As can be seen in Table 4.16, the lower(-middle) class uses *well* most often in comparison to the other classes, sixteen per 10,000 words, and *why* the least, one per 10,000 words.

	Well	*Why*
Low-High	14/10000 (18)	1/10000 (1)
Low-Middle	15/10000 (26)	1/10000 (2)
Low-Low	26/10000 (7)	0
Total	16/10000 (51)	1/10000 (3)

Table 4.16 Distribution of the discourse markers *well* and *why* in the lower(-middle)-class subcorpora.

The normalized figures for *well* show that it is used the most in the same class subcorpus, whereas *why* is not used at all, which is a significant difference. This may indicate that *why* is a marker used most often by the comparatively higher classes and in speech directed towards higher classes. *Well* is used second most towards the upper-middle class and least towards the upper class. Regarding *why*, it is used in equal amounts in upper- and upper-middle-class directed speech.

Similar to the upper- and upper-middle-class speakers, the lower(-middle)-class speakers use the clusters *well I* and *well well* most frequently. Another frequent cluster is *well miss (woodhouse)* which occurs in both the friends subcorpus and the female-to-female subcorpus.

72. **Well**, **Miss Woodhouse**, is he like what you expected? (Harriet Smith, *Emma*)

Well I and *well miss woodhouse* are prevalent clusters in speech directed towards the upper-middle class. The speech directed at the upper-class characters did not have any notable clusters. Lower(-middle)-class speech, on the other hand, does seem to have a pattern. *Well* is mainly followed by a first name or a term of endearment, such as *Fanny*, *my dear Jane*, or *my dear*.

73. **Well**, **my dear**, I shall say you are laid down upon the bed, and I am sure you are ill enough. (Miss Bates, *Emma*)

As *why* only occurs three times, the lower(-middle)-class speaker subcorpora do not have any specific reoccurring patterns.

Even though there are differences regarding the social class of the speaker, there are hardly any when analyzing the speech regarding the social class of the addressee.

	Well	*Why*
to High	13/10000 (29)	2/10000 (5)
to Middle	12/10000 (68)	2/10000 (10)
to Low	12/10000 (52)	2/10000 (7)

Table 4.17 Distribution of the discourse markers *well* and *why* in relation to the social class of the addressee.

The only difference in the distribution of *well* and *why* is that *well* is used one per 10,000 words more towards the upper class than towards any of the others. *Why* is used equally as much to the three social classes.

The upper-class addressee subcorpora do not have any clusters. However, similar to *why* followed by *you* in the upper-middle-speech subcorpora, the address following *well* directed at upper-class characters is predominantly formal; for example, *well Ma'am, well Mr. Knightley, well sir,* and *well Lady Bertram*. There is one *well* which is followed by *you* and one which is followed by *Mary*, which are both used by close friends and close family. *Why* only occurs five times though without any striking patterns. The upper-middle-class addressee subcorpora, on the other hand, comprise several clusters both two-word and three-word ones. The most frequent patterns are *well I*, *well well*, *well miss (woodhouse)*, *well and*, and *well this (is)*. In relation to modes of address after *well*, formal words such as *miss*, *sir*, *ma'am*, *Mr. Parker* but also informal words such as *my dear Mary*, *papa*, and *you* occur. In the case of *why*, phrases such as *to own the truth* and *in truth* are frequent when addressing the upper-middle class as well as being followed by a relative phrase starting with *what.*

74. **Why to own the truth**, I do think the Miss Parkers carry it too far sometimes and so do you my Love, you know. (Mrs. Parker, *S*, in Bree et al. 2013:346)

Speech directed to the lower-middle class also has the frequent clusters *well I* and *well well*, but also *well Fanny* and *well my*, followed by *dear* most often, which have a more informal undertone. *Why* is followed by either *indeed*, another discourse marker, or *you* as in example 75 below:

75. **Why**, **you** do not suppose a man is overset by a bottle? (John Thorpe, *NA*).

4.3.4 Setting

Well and *why* occur in each of the settings but their distribution differs across the various settings. *Well* occurs most often in a festive setting, fifteen times per 10,000 words, whereas it occurs the least in an outdoors setting, ten per 10,000 words. *Why*, on the other hand, occurs the most in a domestic setting, two per 10,000 words, but equally less frequently in a festive or outdoors setting.

	Well	*Why*
Domestic	11/10000 (80)	2/10000 (15)
Festive	15/10000 (40)	1/10000 (3)
Outdoors	10/10000 (29)	1/10000 (4)

Table 4.18 Distribution of the discourse markers *well* and *why* in relation to the setting in which the discourse takes place.

The clusters occurring with *well* are *well I* and *well well* and they are found in the domestic and festive subcorpus. The outdoors subcorpus also has the cluster *well I*, but not *well well.* It does have the cluster *well miss (woodhouse)* to which has been referred above. In addition, the domestic subcorpus also has *well my (dear)* as a cluster and *well and.* The outdoors setting has *miss* following which may be taken to be a more formal setting as *well my* only occurs in a domestic setting and is thus more informal. However, *why* occurs with *my dear cousin* in the outdoors subcorpus:

76. **Why**, **my dear Cousin**, this will be a most agreable surprise to every body to see you enter the room with such a smart Young Fellow as I am […].
(Edward Stanley, *Cath*, in Bree et al. 2013:187)

Since the speaker addresses a cousin, informal language may still be appropriate even in an outdoor setting. *Why, in truth sir* is also a cluster which occurs in the outdoor setting which is clearly more formal, but it only occurs once. In the domestic setting, *why* occurs with *to own the truth* and *indeed* which are both discourse markers, but there is little direct address in this setting immediately following *why*. *Indeed* also occurs after *well* in the festive subcorpus.

4.4 Phrasal Discourse Marker: *You Know*

4.4.1 Gender

There are a total of 117 instances of *you know* in the corpus. *You know*, in the corpus, does not occur in sentence-initial position. It occurs predominantly in sentence-medial position, see example 77, in almost 80 per cent of the instances for the female subcorpus; the rest of the instances are sentence-final, as in example 78:

77. Mr. Elton, **you know**, is out of the question. (Emma Woodhouse, *Emma*)

78. My mother does not hear; she is a little deaf **you know**. (Miss Bates, *Emma*)

The proportion also holds true with respect to the sex of the addressee of the female speakers, in both the female and the male subcorpora. In relation to gender, *you know* is especially prevalent in the female subcorpus, with thirteen per 10,000 words.

	You Know
Female	13/10000 (98)
Female to Female	16/10000 (71)
Female to Male	9/10000 (24)

Table 4.19 Distribution of the phrasal discourse marker *you know* in the female subcorpora.

Furthermore, *you know* is used almost twice as often when the addressee of the female speaker is another female, sixteen per 10,000 words, than when the addressee is a male, nine per 10,000 words. The female subcorpus contains four clusters: *you know I*, *you know it (is)*, *you know is*, and *you know and*. The cluster *you know I* is most often an inter-sentential cluster (i.e. *you know* is the end of one clause and *I* the beginning of the other). The other clusters are usually found within an utterance, especially so with *you know it is* and *you know is*, as in example 77. In the female-to-female subcorpus, *you know I* and *you know it* are prevalent clusters. The female-to-male subcorpus, on the other hand, does not have any striking clusters.

The phrasal marker *you know* occurs almost three times as less in the male subcorpus. Less than 75 per cent of the instances are in sentence-medial person.

	You Know
Male	4/10000 (19)
Male to Female	4/10000 (18)
Male to Male	2/10000 (1)

Table 4.20 Distribution of the phrasal discourse marker *you know* in the male subcorpora.

You know is used four times as often when the addressee of the female speaker is a female than when the addressee of the male speaker is a female, see Table 4.20. Regarding the male addressee subcorpora, *you know* only occurs once in the male-to-male subcorpus and is thus almost used five times less compared to the female-to-male subcorpus.

79. One is never able to complete anything in the way of Business **you know** till the Carriage is at the door. (Mr. Parker, *S*, in Bree et al. 2013:324)

The male-to-female subcorpus has one prevalent cluster, *then you know*. An example in which it is used twice in quick succession is when John Thorpe is hinting at his intentions towards Catherine Morland to Catherine:

80. And **then you know** I say, **then you know**, we may try the truth of this same old song. (John Thorpe, *NA*)

Then you know and its repetition are both signs of anxiety and insecurity in this example in relation to the ‘same old song’ which refers to the song with the lyrics where one wedding leads to

another. However, the cluster does not necessarily carry these emotions as in the example by Henry Tilney where it adds to the flow of the sentence:

81. But **then you know**, madam, muslin always turns to some account or other. (Henry Tilney, *NA*)

Regardless of the speaker, *you know* is used most often in speech directed to women, ten per 10,000 words, than to men, eight per 10,000 words.

	You Know
to Female	10/10000 (89)
to Male	8/10000 (25)

Table 4.21 Distribution of the phrasal discourse marker *you know* in relation to the sex of the addressee.

There are no patterns in the male addressee subcorpus except for the position of *you know*, which is predominantly sentence-medial (80 per cent). In the female addressee subcorpus, all of the previous patterns occur: *then you know*, *you know I*, *you know it*, *you know is*, but also *you know you*. Similarly, 75 per cent of the instances of *you know* are also in sentence-medial position. And therefore, it appears that *you know* is used most often to clarify what has been said previous to the discourse marker by what is said after it.

4.4.2 Intimacy

In relation to intimacy, *you know* occurs at each of the examined relationship levels.

	You Know
Close Family	11/10000 (18)
Distant Family	2/10000 (4)
Friends	11/10000 (92)
Other	9/10000 (3)
Total	117

Table 4.22 Distribution of the phrasal discourse marker *you know* in terms of the intimacy level of the speaker and addressee.

You know is most often present in the close family and friends subcorpora. In both, the discourse marker occurs eleven times per 10,000 words. *You know* occurs least in the distant family subcorpus. As the 'other' subcorpus can contain both distant and close relationships, *you know* may be regarded as a marker used at the close intimacy level rather than the distant one.

In both the distant family subcorpus and the 'other' subcorpus, there are exclusively sentence-medial instances of *you know*. *You know* occurs in sentence-final position just over a quarter of the instances in the close family subcorpus, four out of eighteen, and a bit below a quarter of the instances in the friends subcorpus, i.e. twenty-four out of ninety-two. As for

clusters or recurring patterns, there are none in the close family subcorpus, nor are there in the distant family or 'other' subcorpus as there are only four and three occurrences of *you know*, respectively. The patterns that do occur are only found in the friends subcorpus: *you know I*, *then you know*, *you know you*, *you know it*, and *you know is*. This is due to the fact that 79 per cent of the instances occur in the friends subcorpus. Similar to *you know I*, *you know you* also occurs predominantly inter-sententially, and in the case of the following example, intra-subjectively:

82. Ten miles there, and ten back, **you know**. **You** must excuse my sister on this occasion […]. (Mrs. Norris, *MP*).

4.4.3 Social Class

You know occurs in each of the social classes but least in the upper-class subcorpora. In fact, Table 4.23 shows that *you know* does not occur at all in upper-to-upper-class discourse.

	You Know
High-High	0
High-Middle	6/10000 (11)
High-Low	1/10000 (2)
Total	3/10000 (13)

Table 4.23 Distribution of the phrasal discourse marker *you know* in the upper-class subcorpora.

You know occurs most frequently in upper-to-upper-middle class discourse, six per 10,000 words. Of the eleven instances, four are in sentence-final position. The cluster *then you know* occurs four times, which is either preceded by *but*, *for*, or *and*. In upper-to-lower(-middle)-class discourse, there are only two instances of *you know* which are both sentence-medial and both used by Henry Tilney. Henry has an instructive role in relation to Catherine, see section 3.2., and the use of *you know* demonstrates this quality, as sentence-medial *you know* is used to explain the relevance of what was said or to amplify it (Crystal 1988:47):

83. He loved her, I am persuaded, as well as it was possible for him to – we have not all, **you know**, the same tenderness of disposition. (Henry Tilney, *NA*).

Furthermore, the upper-class speakers have a total of three *you know*s per 10,000 words.

In the upper-middle-class subcorpora, there are nine instances of *you know* per 10,000 words, which altogether points to *you know* being more of an upper-middle-class speech feature than an upper-class one.

	You Know
Middle-High	8/10000 (6)
Middle-Middle	11/10000 (26)
Middle-Low	6/10000 (16)
Total	9/10000 (48)

Table 4.24 Distribution of the phrasal discourse marker *you know* in the upper-middle-class subcorpora.

You know occurs most frequently in the subcorpus directed towards the middle class, eleven times per 10,000 words, which is similar to the upper-class subcorpora. The discourse marker is used least often when addressing the lower(-middle) class.

Sentence-final *you know* occurs most often in upper-middle-to-upper-middle discourse, in less than one third of the instances. It is used least often in the upper-middle-to-upper-class subcorpus, one of the six instances. To lower(-middle)-class characters, sentence-final *you know* is used in one-fifth of the cases. As for clusters with *you know*, there are none in either of the upper-middle-class subcorpora, nor are there any in the upper-middle-class subcorpora when combined.

The lower(-middle)-class speech seems to possess the most *you know*s, seventeen per 10,000 words.

	You Know
Low-High	9/10000 (11)
Low-Middle	25/10000 (44)
Low-Low	4/10000 (1)
Total	17/10000 (56)

Table 4.25 Distribution of the phrasal discourse marker *you know* in the lower(-middle)-class subcorpora.

You know is least often used in conversation with lower(-middle)-class characters; there is only one instance:

84. Oh! yes, my dear, I dare say you might, because **you know**, the door was open, and the window was open, and Mr. Knightley spoke aloud. (Miss Bates, *Emma*).

Similarly in upper-class discourse, *you know* is not frequent either and only one instance is sentence-final. *You know* is most represented in the lower(-middle)-to-upper-middle subcorpus. In fact, *you know* is significantly more present in speech directed towards the upper-middle class than towards the lower-middle class ($p<.0001$). Just as with the other upper-middle-class directed discourse, *you know* in sentence-final position is more prevalent here.

Furthermore, the only frequent cluster is *you know I* which is most often found as an inter-sentential cluster:

85. [B]ut I can go into my mother's room **you know**. **I** dare say he will come in when he knows who is here. (Miss Bates, *Emma*)

Another pattern with *you know* is that it is frequently preceded by *and*, be it *and yet* or *and then* or followed by *and*. This could refer to *you know* being part of an utterance in which a summation is present or where additions to a statement previous to *you know* are given.

The distribution of *you know* in terms of the social class of the addressee is distinctive.

	You Know
to High	8/10000 (17)
to Middle	14/10000 (81)
to Low	5/10000 (19)

Table 4.26 Distribution of the phrasal discourse marker *you know* in relation to the social class of the addressee.

The discourse marker occurs most often when in conversation with the upper-middle class, fourteen instances per 10,000 words. *You know*, though most prevalent in lower(-middle)-class speech, is used least in speech directed towards that social class, five instances per 10,000 words.

Upper-class and lower(-middle)-class directed speech do not have prevalent patterns. They are similar in that around 85 per cent of the instances of *you know* are sentence-medial and thus meant for exemplification or clarification. The subcorpus relating to upper-middle-class discourse, on the other hand, has close to 70 per cent sentence-medial *you know*s. Furthermore, this subcorpus is the only one with prevalent clusters, due to their being over four times as much *you know*s in the subcorpus than in the others. Common clusters are *you know it (is)*, *you know I*, which is predominantly inter-sentential, and *then you know*.

4.4.4 Setting

You know occurs in each of the settings and is used almost equally frequently regardless of the setting, though somewhat more frequently in a festive setting but that is merely a difference of one instance per 10,000 words compared to the domestic and outdoors setting.

	You Know
Domestic	9/10000 (66)
Festive	10/10000 (26)
Outdoors	9/10000 (25)

Table 4.27 Distribution of the phrasal discourse markers *you know* in relation to the setting in which the discourse takes place.

The proportion of the *you know*s in sentence-medial and sentence-final position are three to one in the festive and outdoors subcorpus. *You know* is used most often in sentence-medial position in comparison to the other two subcorpora, close to six to one proportionally. Furthermore, the festive and outdoors subcorpora do not have any common clusters, but the domestic subcorpus does, i.e. *then you know* and *you know I* which are both predominantly sentence-medial. Sentence-final *you know* is most often used to ask for agreement with the

previous statement, but in relation to the setting in which it occurs there may be a difference. In the domestic setting, *you know* is used to ask for agreement of the capacities of the speaker, as in example 86:

86. I could not have misunderstood a thing of that kind, **you know**!
(Catherine Morland, *NA*)

In an outdoors setting, however, it is used by the speaker to ask whether the addressee understands the implications of the phrase (e.g. that being a surgeon is a lowly profession as in example 87):

87. Her Father and Mother would never consent to it. Sam is only a Surgeon **you know**. (Elizabeth Watson, *TW*, in Bree et al. 2013:277–278)

4.5 Tag Questions

4.5.1 Gender

In the corpus, there are a total of 58 tag questions, both polarity changing and unchanging ones. There are slightly more tag questions in the female subcorpus than in the male subcorpus, one per 10,000 words difference.

	Tag Question
Female	5/10000 (37)
Female to Female	5/10000 (22)
Female to Male	4/10000 (12)

Table 4.28 Distribution of tag questions in the female subcorpora.

Female-to-female discourse similarly has one tag question less than female-to-male discourse. The female subcorpus has polarity changing tag questions from positive to negative, twenty-five, and from negative to positive, six.

88. We are very proud of the children, **are not we**, papa?
(Emma Woodhouse, *Emma*)

Example 88 is one of the seven instances in which a tag question is not the end of the sentence but the person to whom the tag question is addressed. The addressee's name following the tag question is predominantly used in the female-to-female subcorpus. In addition, there are six unchanging polarity tag questions of which two can be found below:

89. [T]herefore you know I could not go with them, **could I**?
(Catherine Morland, *NA*)

90. Proof indeed! Mr. Dixon is very musical, **is he**? (Emma Woodhouse, *Emma*)

The tag questions in examples 88 and 89 both seek confirmation of or agreement with what is said, whereas the one in example 90 is mainly to give feedback on information acquired in the conversation that Emma had with Frank Churchill previous to the example.

The male subcorpus relevant to tag questions only consists of the male-to-female data as tag questions do not occur in male-to-male conversation.

	Tag Question
Male	4/10000 (21)
Male to Female	5/10000 (21)
Male to Male	0

Table 4.29 Distribution of tag questions in the male subcorpora.

Tag questions thus occur most often in same-gendered and mixed-gendered discourse as long as there is at least one female speaker present. Furthermore, ten of the instances, nearly 50 per cent, are polarity unchanging tags, with an accusatory or disapproving undertone:

91. So, you ended with Captain Hunter Mary, **did you**?
(Mr. Edwards, *TW*, in Bree et al. 2013:293)

92. Fanny has been cutting roses, **has she**? (Edmund Bertram, *MP*)

This may also be because due to the teacher role of the male characters. Mr. Edwards wishes to instruct Mary, his daughter, in the right ways to act in a ballroom and Edmund teaches Fanny to act in a more upper(-middle)-class manner.

Furthermore, the utterances to which tag questions are adjoined are also frequently combined with jokes or vulgar language, such as example 93 and 94, respectively:

93. [T]he old Coachman will look as black as his Horses – **won't he** Miss Edwards?
(Tom Musgrave, *TW*, in Bree et al. 2013:296)

94. Old Allen is as rich as a Jew – **is not he**? (John Thorpe, *NA*)

Tag questions in female language seem to either be intended to be helpful to the speaker, to ask for input or confirmation of an utterance, or to the addressee, to show interest into what is said or to include the addressee in the conversation.

Similar to the sex of the speakers, there is barely any difference between the distribution of tag questions regarding the sex of the addressee.

	Tag Question
to Female	5/10000 (43)
to Male	4/10000 (12)

Table 4.30 Distribution of tag questions in relation to the sex of the addressee.

Tag questions addressed to female speakers range from supportive to confrontational as the previous examples have shown, whereas tag questions directed towards a male addressee seek for agreement or are markers of insecurity such as when Fanny asks her cousin Edmund about his opinion of the Miss Owens:

95. The Miss Owens – you liked them, **did not you**? (Fanny Price, *MP*)

Fanny seems anxious about the topic of the conversation which revolves around Mr. Crawford's intentions with her and she is worried that Edmund judges her based on her lack of interest in Mr. Crawford's proposal.

4.5.2 Intimacy

The family subcorpora have the fewest tag questions, three per 10,000 words, whereas the friends and 'other' subcorpus have an equal number of tag questions, six per 10,000 words.

	Tag Question
Close Family	4/10000 (6)
Distant Family	2/10000 (3)
Friends	6/10000 (47)
Other	6/10000 (2)
Total	58

Table 4.31 Distribution of tag questions in terms of the intimacy level of the speaker and addressee.

In the family subcorpora, two of the nine tag questions are polarity unchanging ones. These tag questions are used by Mr. Edwards towards his daughter, as in example 91, and Edmund towards his mother, example 92. Both of them are part of the close family subcorpus and both are disapproving tag questions, whereas the other polarity changing tag questions are supportive tags. The friends subcorpus comprises supportive tags, but also answer-restricting tags and confrontational tags.

The two instances in the 'other' subcorpus are both by John Thorpe in *Northanger Abbey* and are both preceded by incomplete sentences:

95. A neat one, **is not it**? (John, *NA*)

96. No sure, **was it**? (John, *NA*)

In example 95, *it* refers to John's gig, a carriage for a horse. The underlying *it is* has been elided. The root sentence in the second example has also been elided (possibly from *no to be sure it wasn't*). *It* in this case refers to the fact that *Udolpho* was written by Mrs. Radcliffe which seems to have escaped John's mind.

4.5.3 Social Class

Tag questions occur in each of the subcorpora regardless of speaker or addressee. In the upper-class speaker subcorpora, tag questions are found most often, when dealing with normalized results, in upper-to-upper-class discourse. Of the three instances, two are the confrontational tags mentioned earlier, see example 91 and 92, but one is a facilitative one by Camilla's mother in *Catharine, or the Bower*.

	Tag Question
High-High	11/10000 (3)
High-Middle	3/10000 (5)
High-Low	1/10000 (2)
Total	3/10000 (10)

Table 4.32 Distribution of tag questions in the upper-class subcorpora.

Speech addressed to the upper-middle class, the upper-class speakers tend to use polarity unchanging tags as well with a disparaging or accusatory undertone. This is also the case for one of the tags in the speech directed towards the lower(-middle) class.

97. Oh! you would rather talk of her person than her mind, **would you**?
(Mr. Knightley, *Emma*)

In the example above, Mr. Knightley judges Harriet for her superficial ways that are focused on the looks of a person, Emma in this case, rather than the person's intelligence or character.

The upper-middle-class speech contains most of the tag questions, seven per 10,000 words. Tag questions occur most often towards the lower classes.

	Tag Question
Middle-High	6/10000 (4)
Middle-Middle	6/10000 (14)
Middle-Low	8/10000 (16)
Total	7/10000 (34)

Table 4.33 Distribution of tag questions in the upper-middle-class subcorpora.

In conversations with the upper class, the polarity unchanging tag question is either facilitative in the sense that it keeps the conversation going or emotionally loaded and expressing excitement. These occurrences are both by the same character, the somewhat unpolished Catharine:

98. Said I was a nice girl, **did he**?
(Catharine Percival, *Cath*, in Bree et al. 2013:202)

The other tag questions are either used to check a statement or to ask for further explanation. Regarding same-class discourse, the tag questions range from expressing excitement similar to example 98 to uncertainty and indicating facilitativeness to functioning as mainly fact checking tags.

Speech directed to the lower classes contains more tag questions and also in many different forms. There are both canonical tag questions, polarity changing and unchanging ones, and invariant tags (e.g. *hey* and *surely*):

99. [T]here can be no danger, **surely**? (Emma Woodhouse, *Emma*).

In this subcorpus, the examples of elided root sentences by John Thorpe as in examples 95 and 96 are also a part of the speech towards the lower(-middle) class.

The lower(-middle)-class speech has a similar number of tag questions as the upper class does.

	Tag Question
Low-High	3/10000 (4)
Low-Middle	5/10000 (8)
Low-Low	8/10000 (2)
Total	4/10000 (14)

Table 4.34 Distribution of tag questions in the lower(-middle)-class subcorpora.

Speech directed towards the upper class has an undertone of insecurity or shock attached to it, as in example 95 above. The normalized data shows that tag questions are used most in same-class speech. Both of the instances are accompanied by the name of the addressee, *Jane* and *Mrs. Ford.* These instances occur in Miss Bates' speech and as she has the propensity to strike up conversation with many characters at (almost) the same time, it is necessary for her to clarify to whom the tag question is addressed. The tag questions directed towards the upper-middle class are all asking for support. Catherine uses tag questions in a specific setting where she is trying to rectify what happened earlier in the story when she had arranged an outing with the Tilneys but ended up going away with the Thorpes and her brother. Miss Bates, on the other hand, uses tag questions at several points in the story which usually have the function of seeking confirmation on her statements from characters around her.

The distribution of tag questions in relation to the social class of the speaker has some subtle differences but regarding the social class of the addressee in general, there are no difference as all subcorpus combinations have five tag questions per 10,000 words.

	Tag Question
to High	5/10000 (11)
to Middle	5/10000 (27)
to Low	5/10000 (20)

Table 4.35 Distribution of tag questions in relation to the social class of the addressee

Speech towards the upper class seems to have either a confrontational or highly emotive undertone, whereas upper-middle and lower(-middle)-class directed speech is overall facilitative or tentative.

4.5.4 Setting

Tag questions occur more frequently in a festive setting than in a domestic or an outdoors one.

	Tag Question
Domestic	5/10000 (33)
Festive	6/10000 (17)
Outdoors	3/10000 (8)

Table 4.36 Distribution of tag questions in relation to the setting in which the discourse takes place.

Tag questions are usually utterance final, but in the case of the festive setting, the name of the addressee (e.g. *Mrs. Allen*) at times comes last. There is one instance in the corpus where the tag question comes sentence-medial which occurs in a festive setting:

100. A long time, **is not it**, for a cold to hang upon her? (Miss Bates, *Emma*)

The example is similar to the incomplete sentences that John Thorpe uses in *Northanger Abbey*. These incomplete utterances are a frequent occurrence with Miss Bates as later on she also says the following:

101. I am going to Kingston. Can I do any thing for you?
Oh! dear, Kingston – **are you**? (Mr. Knightley and Miss Bates, *Emma*).

John Thorpe's examples, see 95 and 96, belong to the outdoors subcorpus (e.g. *no sure, was it?*). Apart from these, the tag questions in the outdoors subcorpus are either inquisitive or confrontational. The tag questions in the domestic subcorpus are predominantly followed by the name of the addressee (e.g. *Miss Edwards*) or by a term of endearment (e.g. *my love*). The nature of the tag questions can be facilitative on behalf of the speaker or addressee, but it can also be accusatory as previous examples have illustrated.

CHAPTER 5: CONCLUSION

5.1 Introduction

The present study focused on a specific slot in time covering twenty-five years in the Late Modern English period in which the well-known author Jane Austen was active. The corpus comprises six of Austen's novels. The study covers socio-historical discourse analysis by way of analyzing four single-word discourse markers (*oh*, *ah*, *well*, and *why*), one phrasal discourse marker (*you know*), and tag questions in general both canonical (e.g. *do not you?*) and invariant ones (e.g. *hey?*). The main variables that were under investigation were gender, intimacy, social class, and setting. The discourse markers and tag questions in combination with the variables led to the following research question:

In what way do the sociolinguistic variables of (1) gender, (2) intimacy, (3) social class, and (4) setting correlate with the use of discourse markers and tag questions in LModE?

Before answering the main question, I will first summarize the findings from the previous chapter following the same structure: firstly, the interjections *oh* and *ah*; secondly, the general discourse markers *well* and *why*; thirdly, the phrasal discourse marker *you know*; and lastly, the tag questions.

5.2 Main Findings

5.2.1 Interjections: *Oh* and *Ah*

The interjection *oh* occurs most of all the discourse markers and tag questions with 288 instances, twenty-three per 10,000 words, in the entire corpus. It is used significantly more than the other interjection *ah*. *Ah* occurs 32 times, three per 10,000 words.

In relation to gender, *oh* is used significantly more by female speakers than by male speakers. Furthermore, *oh* occurs most often when the characters are friends, as does *ah*. In addition, *oh* is used significantly more in upper- to upper-middle-class speech compared to upper- to lower(-middle)-class speech. Compared to usage in upper-class speech, *oh* is used significantly more in lower(-middle)-class speech. Regarding the setting, *oh* and *ah* occur most often in a domestic and festive setting. *Oh* and *ah* occur most often in sentence-initial or utterance-initial position.

The literature indicated that *oh* was used most often by women, and that *ah* is generally a feature of male speech (Aijmer 2009; Jonker 2014). This was continued for *oh*, but *ah* was found to be neither a male or female marker, although it is used slightly more in speech directed towards men.

5.2.2 General Discourse Markers: *Well* and *Why*

The general discourse marker *well* occurs second most frequently of the discourse markers and tag questions with twelve instances per 10,000 words. *Why*, on the other hand, occurs least of all with a mere two instances per 10,000 words.

Usage of *well* has certain tendencies regarding the distribution across the variables. *Well*, similar to *oh*, shows to be used more often by female speakers. For *why*, there are hardly any differences in any of the variables studied as there were only few instances overall. Furthermore, *well* is most often used when the speaker and addressee are friends. *Well* corresponds to *oh* in this way but also in that it is used least often in upper-class speech but most often in lower(-middle)-class speech. Moreover, *well* is most often used in a festive setting, whereas *why* appears to be a marker of domestic discourse. As for the position in the sentence, *well* and *why* are both predominantly sentence- or utterance-initial.

Lutzky (2012) found that *well* was used most often by female speakers, which has similarly been the case for the present study. She also found that *well* occurred most often in same-gender dialogues, but this study's findings show a difference in favor of mixed-gender discourse, five instances per 10,000 words. Lutzky further found that *well* was used equally often in upper-class speech as in lower-class speech, which is not supported by my findings as *well* occurs most often in lower(-middle)-class speech and is actually used least often in upper-class speech. *Why* was described as the more prominent marker nearing the end of the EModE period as compared to *well* and *marry* (Lutzky 2012). By contrast, *why* is used least of all the discourse markers in the present corpus, which could indicate that just as *why* was rarely in use at the start of the EmodE period it went out off use in the LModE period. However, this decrease in the LModE period could also suggest that *why* is not a discourse marker Austen would frequently use, which resulted in the appearance of a decrease. There is not enough evidence to suggest that *why* is more of a male speech feature than a female one as they both occur two times per 10,000 words, nor is there evidence to suggest that *why* is equally used as much in higher-class speech as in lower(-middle)-class speech; there is only one *why* per 10,000 words difference with the upper class being more prone to use *why*.

5.2.3 Phrasal Discourse Marker: *You Know*

You know is the third most frequent discourse marker in the corpus with 117 hits, nine instances per 10,000 words.

It is used more by female speaker than male speakers, just as *oh* and *well*. Furthermore, the phrasal discourse marker is used most often in same-gendered discourse than mixed-gendered,

especially regarding female-to-female discourse compared to female-to-male discourse. Regarding intimacy, *you know* is predominantly used when the relationship between the speaker and addressee is of a close nature, i.e. close family and friends. For the social class the following is true: the higher the social class, the less *you know* is used, similar to *oh* and *well*. Apart from that, *you know* is most often used in speech directed towards the upper-middle class. Previous discourse markers seem to be used most often in a festive setting which is also the case for *you know*, but only by a small margin.

The literature was inconclusive on whether *you know* was a male feature or female feature in relation to earlier corpus study. Östman (1981) found that *you know* was a marker of female speech, whereas Erman (1992) found that it was a marker of male speech. A more recent study on *you know* determined that there was no difference between male and female speakers in relation to tag question use (Korzogh and Furkó 2011). The present findings indicate that female speakers use *you know* more often but not significantly so, nine instances per 10,000 words more than male speakers.

5.2.4 Tag Questions

Tag questions are used more often than *why* and *ah*, but less often than *oh*, *well*, and *you know*. There are a total of 58 occurrences which count for five instances per 10,000 words.

In relation to the variables, there does not seem to be much difference. There is no correlation between male or female use. In the friends subcorpus, tag questions seem more prevalent, as was the case with the discourse markers, but not by a substantial number. Furthermore, tag questions occur least often in the upper-class subcorpora but occurs, regarding normalized numbers, most often in upper-to-upper-class speech. In general, the upper-middle-class speaker uses tag questions most often, but only slightly so. For setting, tag questions are more prevalent in festive settings, which has generally been the case, and least in an outdoors setting.

Earlier research has not found any substantial correlation with tag questions and gender in terms of frequency. LaFrance (2001) and Meyerhoff (2011) did find that there were differences regarding the use of a tag question in relation to gender. Women tended to be more supportive, whereas men used tag questions to seek for an answer. The present findings indicate that female speakers are indeed facilitative with their tag questions, but also tend to ask for input or confirmation. The tag questions belonging to male speech have the tendency to be more accusatory and disapproving rather than answer-seeking.

5.3 Discussion

Earlier extensive historical socio-pragmatic research on discourse pointed to the higher class speakers being more prone to use discourse markers. Furthermore, it showed that the lower classes, specifically servants, adopted the speech of the higher-class speakers, thus resulting in the frequent use of discourse markers in their speech. My findings indicate that it is the lower(-middle)-class speakers who use discourse markers most often, which possibly indicates a change from above as tag questions became a feature of the lower classes, as indicated by Lutzky (2012), and further influenced the speech of lower(-middle)-class speakers. Regarding correlations with gender, I expected *oh* and *well* to be used more by female speakers, which was indeed the case. Furthermore, earlier research suggested that *ah* and *why* were male markers; however, my study found no differences in this respect. As for *why*, male-to-male discourse rarely occurred in Austen's works, presumably because she heard that type of conversation least often and could not portray it as realistically as the type of discourse where women took part. I expected that the intimacy level of relationships and setting would influence the use in that the more intimate the relationship or setting, the more discourse markers and tag questions are present. The findings indicate that this is the case, as the speech used among friends and relatives contains more discourse markers and tag questions, as is the case for discourse in a domestic setting, in which either friends and family converse, and a festive setting, where characters with a close intimacy tend to converse; at least, this was the case in the corpus. As for the relevance of the results, they may be used to gain further insight into the artistic abilities of Jane Austen and understanding the place of characters in her novels and short stories. Furthermore, the results shed some light on the correlation which seems to be present regarding the use of discourse markers and the variable social class as well as discourse markers and gender.

The study was limited by focusing on one author in the LModE period and also limited further by using the speech of the main characters in the published novels, but all the characters were used in the unpublished ones. Regardless of these limitations, significant correlations were found, especially pertaining to social class. *Oh* was used significantly more in lower(-middle)-class speech than in upper-class speech but amongst upper-class speech, *oh* was used significantly more often when conversing with upper-middle-class characters than with lower(-middle)-class characters. For the phrasal discourse marker *you know*, there is one significant correlation which relates to lower character's speech and to the social class of the addressee. *You know* is used significantly more in conversation with upper-middle and lower(-middle)-class characters for the lower(-middle)-class characters. The significance in relation to the social class of the speaker as well as the interaction between the speaker and the addressee with the same or a different social

class could prove to be interesting in relation to discourse markers or even speech features in general regarding socio-historical research. Additionally, the EmodE period, for fiction, has been extensively studied and it would prove to be interesting to do a diachronic study of the LModE period too as there already seem to be differences in the corpus for this study regarding the use compared to Lutzky (2012)'s one.

Furthermore, it could prove to be interesting to look into the language of morally bad, manipulative, or silly characters, but also antagonists, such as the Thorpes in *Northanger Abbey* and the Crawfords in *Mansfield Park* (not analyzed in this paper), as these characters seem to have the most discourse markers and tag questions and could prove to be relevant in order to study which speech features or grammatical constructions are deemed as normal speech or 'proper' speech and which are not. As discourse markers have proven to be relevant regarding the variable social class, it might be pertinent to look into it further by looking at phrases such as *I am sure*, which seem prevalent throughout Austen's works as well, or *I suppose*, but also other single-word discourse markers like *indeed* or *then*, which occurred quite often during the concordance analysis. Future study into socio-historical discourse could be helpful to understand more about discourse of the past and how certain features of language can tell something about a speaker whether it be their gender or their position in society. Historical discourse analysis will help to pinpoint when changes in language occurred, such as the increase of *why* in the EmodE period to decrease in the LModE period, as well as the direction of the change in use of a discourse marker regarding the social class of the speaker, as *well* was a marker of the upper-class speech but it moved from the upper class to the lower class and decreased in use in upper-class speech or increased in lower-class speech, or both. The rise and fall of *why* could prove to be interesting to see whether there is an actual decrease or whether it is the preference of the author. *Well* has 'won' in present-day conversation, but *why* was clearly more favorable in EModE, so when was the balance tipped in favor of well? Another point of study which could prove to be interesting could be the function of tag questions by male speakers and female speakers in historical discourse as well as the interaction with the polarity of the tag question in relation to the root sentence. Studying tag questions from a historical perspective is still a relatively new area of study and seeing that with the number of tag questions studied certain trends became visible, it may be prudent to look into the function of tag questions, their emotional charge and what tag question use may say about the speaker.

As there are hardly any socio-pragmatic corpora it could prove interesting to create them which would allow for a more complete corpora than the handmade ones used for the present study. Furthermore, the corpus had some inconsistencies as I had forgotten to add utterances to

certain subcorpora, which I had to add after I discovered the mistakes during my in-text analysis. Using corpus linguistics software would result in a more consistent corpus and thus analysis, but as tagging a corpus would have taken up a quite substantial amount of time, this would not have been possible. Moreover, some of the subcorpora had fewer than five instances which influenced the normalized results as was the case with certain subcorpora being considerably smaller than others belonging to a specific sociolinguistic variable. This might have influenced some of the results especially regarding the social classes as there were nine different subcorpora and the main corpus was subdivided into these parts, but also in relation to the subcorpora which were considerably smaller to begin with such as the 'other' subcorpus for the intimacy variable and the male-to-male subcorpus for the gender variable.

5.4 Conclusion

To conclude and fully answer the research question on the correlation of the variables gender, intimacy, social class, and setting with the use of discourse markers and tag questions in LModE, the study has shown that the use of discourse markers in Austen's fiction correlates with gender in that female speakers are more prone to use them. Regarding social class, the lower(-middle) class seem more likely to use certain discourse markers than the upper classes. Furthermore, there were certain trends, not statistically significant, when it came to the variables intimacy and setting, the closer the intimacy of the setting or people, the more discourse markers and tag questions were used. In addition, regarding tag questions there seemed to be a correlation with the gender of the speaker and the tone of the tag, which was predominantly facilitative with women and more aggressive with men.

Austen's talent for creating idiolects, as was shown in Burrows 1987, is not her only talent regarding the speech of her characters. The use of discourse markers as well as the function of tag questions demonstrate patterns in relation to the sociolinguistic variables. Austen appears to be aware of by whom discourse markers were used most and also which discourse markers were preferred, but also when and where they occurred. She used this knowledge and applied it to the speech of her characters, which further demonstrates her artistic genius. In her fiction, Austen actually mentioned the qualities of women's language which shows her awareness of gender and speech. I find it fitting to end the study on Austen's note as expressed by Mr. Knightly in *Emma*:

> *This is all that I can relate of the how, where, and when. Your friend Harriet will make a much longer history when you see her. – She will give you all the minute particulars, which only woman's language can make interesting.*

BIBLIOGRAPHY

Aijmer, K. (1987). *Oh* and *ah* in English conversation. In W. Meijs (ed.), *Corpus linguistics and beyond*. Amsterdam: Rodopi. 61–86.

Aijmer, K. (2002). *English discourse particles: Evidence from a corpus*. Amsterdam: John Benjamins.

Aijmer, K. (2009). Interjections in the COLT Corpus. In S. Slembrouck, M. Tarverniers, and M. van Herreweghe (eds.), *From* Will *to* Well: *Studies in linguistics offered to Anne-Marie Simon-Vandenbergen*. Gent: Academia Press. 11–20.

Ameka, F. K. (1992). Introduction – Interjections: The universal yet neglected part of speech. *Journal of pragmatics*, 18(2–3), 101–118.

Andersen, G. (1998). The pragmatic marker *like* from a relevance-theoretic perspective. In Jucker, A.H. and Y. Ziv (eds.), *Discourse markers: Description and theory*. Amsterdam: John Benjamins. 147–170.

Andersen, G. (2001). *Pragmatic markers and sociolinguistic variation: A relevance-theoretic approach to the language of adolescents*. Amsterdam: John Benjamins.

Babb, H.S. (1962). *Jane Austen's novels: The fabric of dialogue*. Columbus, Ohio: Ohio State University Press.

Blake, N.F. (1996). *Why* and *what* in Shakespeare. In N.F. Blake (ed.), *Essays on Shakespeare's Language* (1st series). Misterton: The Language Press. 121–137.

Blakemore, D. (1987). *Semantic constraints on relevance*. Oxford: Blackwell.

Blakemore, D. (1992). *Understanding utterances: An introduction to pragmatics*. Oxford: Blackwell.

Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge: Cambridge University Press.

Blankenship, K.L. and T. Y. Craig (2007). Language and persuasion: Tag questions as powerless speech or as interpreted in context. *Journal of experimental social psychology*, 43, 112–118.

Brinton, L.J. (2001). Historical discourse analysis. In D. Schriffin, D. Tannen, and H.E. Hamilton (eds.), *The handbook of discourse analysis*. Malden, MA/Oxford: Blackwell Publishers. 138 160.

Bree, L., P. Sabor and J. Todd (eds.) (2013). *Jane Austen's manuscript works*. Claremont: Broadview Editions.

Brown, P. and S. Levinson (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Burrows, J.F. (1987). *Computation into criticism. A study of Jane Austen's novels and an experiment in method*. Oxford: Clarendon Press.

Coates, J. (2004). *Women, men and language: A sociolinguistic account of gender differences in language* (3rd ed.). London: Longman.

Collins, B. and I. Mees (1984). *The sounds of English and Dutch*. Leiden: Leiden University Press.

Columbus, G. (2009). A corpus-based analysis of invariant tags in five varieties of English. In A. Renouf and A. Kehoe (eds.), *Corpus Linguistics: Refinements and Reassessments*. Amsterdam – New York, NY: Rodopi B.V. 401–414.

Crystal, D. (1988). Another look at *well, you know*. *English today*, 13, 47–49.

Culpeper, J. and M. Kytö (2010). *Early Modern English dialogues: Spoken interaction as writing*. Cambridge: Cambridge University Press.

Defour, T. (2007). *A diachronic study of the pragmatic markers* well *and* now: *Fundamental research into semantic development and grammaticalisation of a corpus study*. PhD thesis for Ghent University. URL: http://lib.ugent.be/fulltxt/RUG01/001/272/289/RUG01-001272289_2010_0001_AC.pdf. Accessed: 27-05-2014.

Defour, T. (2009). *Well, well! What a surprise!* A diachronic look at the relation between *well*, *well* and the pragmatic marker *well*. In S. Slembrouck, M. Tarverniers, and M. van Herreweghe (eds.). *From* Will *to* Well: *Studies in linguistics offered to Anne-Marie Simon-Vandenbergen*. Gent: Academia Press. 161-172.

Erman, B. (1987). *Pragmatic expressions in English. A study of* you know*,* you see *and* I mean *in face-to-face conversation*. Stockholm: Almqvist & Wiksell.
Erman, B. (1992). Female and male usage of pragmatic expressions in same-sex and mixed-sex interaction". *Language Variation and Change*, 4, 217–234.
Fischer-Starke, B. (2010). *Corpus Linguistics in literary analysis: Jane Austen and her contemporaries*. New York, NY/London: Continuum International Publishing Group.
Fox-Tree, J.E. and J.C. Schrock (1999). Discourse markers in spontaneous speech: Oh what a difference an *oh* makes. *Journal of memory and language*, 40, 280–295.
Fox-Tree, J.E. and J.C. Schrock (2002). Basic meanings of *you know* and *I mean*. *Journal of Pragmatics*, 34, 727–747.
Huang, L-F. (2011). *Discourse markers in spoken English: A corpus study of native speakers and Chinese non-native speakers*. PhD thesis for the University of Birmingham. URL: http://etheses.bham.ac.uk/2969/. Accessed: 10-03-2014
Jacobs, A. and A.H. Jucker. (1995). The historical perspective in pragmatics. In A.H. Jucker (ed.), *Historical pragmatics: Pragmatic developments in the history of English*. Amsterdam: John Benjamins. 3–33.
JohnsonOnline. (1755). *A dictionary of the English language: A digital edition of the 1755 classic by Samuel Johnson.* URL: http://johnsonsdictionaryonline.com/. Accessed: 29-05-2014.
Jucker, A.H. (1993). The discourse marker *well*: A relevance-theoretical account. *Journal of Pragmatics*, 19(5), 435–452.
Jucker, A.H. (2002). Discourse markers in Early Modern English. In R. Watts and P. Trudgill (eds.), *Alternative histories of English*. London: Routledge. 210–230.
Jucker, A.H. (2008). Historical pragmatics. *Language and linguistics compass*, 2(5), 894–906.
Jucker, A.H. and I. Taavitsainen (2013). *English historical pragmatics.* Edinburgh: Edinburgh University Press.
Koczogh, H.V. and P. B. Furkó (2011). Gender differences in the use of the discourse markers 'you know' and 'I mean'. *Argumentum*, 7, 1–18. URL: http://epa.oszk.hu/00700 /00791/ 00008/pdf/furko-koczogh.pdf. Accessed: 24-04-2014.
Komar, S. (2007). The interface between intonation and function of discourse markers in English. In S. Komar and U. Mozetič, *Studies in the English language and literature in Slovenia* (Volume IV/1-2). Ljubljana: Trajanus d.o.o., Kranj. 43–55.
LaFrance, M. 2001. Gender and Social Interaction. In R.K. Unger, *Handbook of the psychology of women and gender*. Hoboken, NJ: John Whiley & Sons. 245–255.
Lakoff, R. (1973). Language and Woman's place. *Language in society*, 2(1), 45–80.
Lewis, D. M. (2011). A discourse-constructional approach to the emergence of discourse markers in English. *Linguistics*, 49(2), 415–443.
Lutzky, U. (2012). *Discourse markers in Early Modern English*. Amsterdam/Philadelphia, PA: John Benjamins.
Lutzky, U. and J. Demmen. (2013). *Pray* in Early Modern English drama. *Journal of Historical Pragmatics*, 14(2), 263–284.
McElhinny, B. (2003). Theorizing gender in Sociolinguistics and Linguistic Anthropology. In J. Holmes and M. Meyerhoff (eds.), *The handbook of language and gender*. London: Blackwell. 21–42.
Mesthrie, R., J. Swann, A. Deumert, and W. L. Leap (2009). *Introducing Sociolinguistics* (2nd ed.). Edinburgh: Edinburgh University Press.
Meyerhoff, M. (2011) *Introducing Sociolinguistics*. New York, NY/Oxon: Routledge.
Müller, S. (2005). *Discourse markers in native and non-native English discourse.* Amsterdam/Philadelphia, PA: John Benjamins.
Norrick, N. R. (2009). Interjections as pragmatic markers. *Journal of pragmatics*, 41, 866–891.
Östman, J-O. (1981). You know: *A discourse-functional approach*. Amsterdam/Philadelphia, PA: John Benjamins.

Page, N. (1972). *The language of Jane Austen*. Oxford: Blackwell Publishing.
Person, R. F. (2009). *Oh* in Shakespeare: A conversation analytic approach. *Journal of historical pragmatics*, 10(1), 84–107.
Phillipps, K.C. (1970). *Jane Austen's English*. London: André Deutsch.
Pichler, H. (2013). *The structure of discourse-pragmatic variation*. Amsterdam/Philadelphia, PA: John Benjamins.
Popescu-Belis, A. and S. Zufferey (2011). Automatic identification of discourse markers in dialogues: An in-depth study of *like* and *well*. *Computer speech and language*, 25, 499–518.
Reese, B. and N. Asher (2007). Prosody and the interpretation of tag questions. *Proceedings of sinn and bedeutung*, 11, 448–462.
Rissanen, M. (2008). Corpus linguistics and historical linguistics. In A. Lüdeling and M. Kytö (eds.), *Corpus linguistics: An international handbook. Volume 1*. Berlin: de Gruyter. 53–68.
Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.
Scott, M. (2010). *WordSmith Tools manual for version 5.0*. URL: http://www.lexically.net/downloads /version5/HTML/index.html?wshell.htm. Accessed: 10-06-2014.
Scott, M. and T. Johns (2010). *WordSmith Tools*. Version 5.0. Oxford: Oxford University Press and Lexical Analysis Software Ltd.
Stenström, A-B. (1994). *An introduction to spoken interaction*. London: Longman.
Stokes, M. (1991). *The language of Jane Austen*. London: The Macmillan Press.
Swann, J. (2009). Language choice and code-switching; Language in Interaction. In R. Mesthrie, J. Swann, A. Deumert, and W.L. Leap (eds.), *Introducing Sociolinguistics* (2nd ed.). Edinburgh: Edinburgh University Press. 146–212.
Taavitsainen, I. (1995). Interjections in Early Modern English. In A.H. Jucker (ed.), *Historical pragmatics: pragmatic developments in the history of English*. Amsterdam: John Benjamins. 439–465.
Taavitsainen, I. and S. Fitzmaurice. (2007). Historical pragmatics: what it is and how to do it. In S. Fitzmaurice and I. Taavitsainen (eds.), *Methods in historical pragmatics*. Berlin: de Gruyter. 11–36.
Traugott, E.C. and R.B. Dasher (2002). *Regularity in semantic change*. Cambridge: Cambridge University Press.
Wichmann, A. (2007). Can English tag questions grammaticalise? *Nouveaux cahiers de linguistique française*, 28, 349–357.
Wyld, H.C. (1921). *A history of modern colloquial English* (2nd ed.). London: John Murray.