# Gender Difference in the Use of Thought Representation—A Corpus-Based Study

Anne Riissanen[1] & Greg Watson[1]

[1] University of Eastern Finland, Joensuu, Finland

Correspondence: Greg Watson, School of Humanities, University of Eastern Finland, Joensuu Campus, Joensuu, Finland. Tel: 358-294-452-160. E-mail: greg.watson@uef.fi

## Abstract

This study (Note 1) investigates potential differences in language use between genders, by applying a modified model of thought representation. Our hypothesis is that women use more direct forms of thought representation than men in modern spoken British English.

Women are said to favour "private speech" that creates intimacy and nearness through discourse, which involves direct forms of speech, and thought, representation. Men are said to prefer a more distancing "public speech" style, in order to maintain independence and to hold and negotiate status, which often involves the display of skill and knowledge. In order to investigate this hypothesis, we examine a slightly modified form of the Lancaster SW & TP Spoken corpus, which has been tagged for the full spectrum of the primary categories of thought representation.

The results of this study prove our thesis to be correct, there are statistically highly significant differences between the genders' use of direct forms of thought representation. British women use the direct forms more than their male counterparts. This greater use of direct thought categories in their daily discourse depicts a lucidity and consciousness, supposedly faithfully repeating the actual thoughts of the speaker, and it often occurs in a moment of heightened emotional or cognitive state. Therefore, because women seem to be able to express their emotions more lucidly then men, and are more inclined to express their thoughts in more detail, they tend to use more direct forms of thought representation than men in daily discourse.

**Keywords:** gendered thought differences, thought representation, corpus linguistics

## 1. Introduction

The primary aim of this paper is to investigate if women use more direct forms of thought representation than men in actual spoken language. In order to do this, we use a model for the representation of thought originally developed by Geoffrey Leech and Mick Short (1981) that was later revised by Short (1996), Watson (1997) and Semino and Short (2004). A secondary aim is to determine the usefulness of this model for language and gender research. The database consists of contemporary British English obtained from the spoken component of the Speech, Writing and Thought Presentation (SW & TP) corpus, which was compiled in 1995-1999 by Dan McIntyre, Carol Bellard-Thompson, John Heywood, Tony McEnery, Elena Semino and Mick Short at Lancaster University. The size of the original spoken corpus is approximately 260,000 words, its texts having been taken from the spoken section of the British National Corpus and from the archives housed in the Centre for North West Regional Studies (CNWRS) at Lancaster University.

The topic of language and gender has always been associated with folkloristic beliefs, but during the last few decades many of these beliefs have been proven false by studies based on empirical data. Although language and gender has not been examined from the viewpoint of thought presentation before, our thesis, 'women use more direct forms of thought in spoken language than men', finds support from other theories found within the field of gender and language. Most models of thought representation describe direct forms of thought to be utterances that depict faithfulness to an original statement or express personal involvement by the speaker. Direct thought appears to be more immediate and intimate than indirect thought. Research suggests that women use language to express solidarity and support, while men use it to express power (Coates, 2004, p. 126). However, the issue of power and solidarity is not that straightforward. Tannen (1994, pp. 22-25) points out that although solidarity and

power appear to be opposites, each still entails the other. Therefore, the aims of interaction of men and women are often the same, but their way of expression differs. Women's way of talking is seen as cooperative, because they use hedges, minimal responses and questions in order to acknowledge and build on other's utterances (Coates, 2004, p. 127). Men's way of talking, on the other hand, appears to be more competitive, characterized by monologues that are often used to play the expert and in verbal sparring (Coates, 2004, p. 133). Additionally, men and women seem to also talk differently about their problems, women often being more personal than men (Coates, 2004, p. 127). These findings suggest that women tend to use more direct forms of thought than men, because a personal and cooperative way of expression is hard to achieve by using distancing indirect forms of thought. This paper will further discuss this issue and offer a more extensive account of previous research conducted on gender and language.

Representation of thought is an area of stylistics that was originally used when focusing on researching fiction and has been applied to non-fiction only quite recently. To date, there has been only limited research on the representation of thought in actual spoken language, and virtually no investigation of language and gender using this model, hence with this in mind we have adopted a model for the representation of thought originally proposed by McIntyre et al. (2004) and that had been used to tag the SW & TP corpus. We first review previous research on language and gender before presenting and explaining a model for the representation of thought. Subsequently, we present the quantitative and qualitative analysis of our results, before offering our conclusions and suggestions for possible future studies.

## 2. Gender and Language

Coates (2004, p. 9) points out that observation on the gender differences in language have always been a topic of interest among humans. She states that the early views on gender differences in language, recorded in novels, poems, letters and other writings, have always been echoes of the ideas of the viewers' contemporary time. According to Litosseliti (2006, p. 2), as a term, language and gender research accounts for the cross-disciplinary discussions of how language is used by women and men and also how language is used to convey things about women and men.

Litosseliti (2006, p. 1) states that the feminist movement in the 1960s had an impact on the social sciences and humanities, including linguistic research, which led to increased interest in gender and language and especially in gender difference. She points out that the development of gender and language research reflects the development of feminism and debates on gender during the last thirty years.

According to Coates (2004, pp. 5-6), Robin Lakoff's Language and Woman's Place is the most well known piece of work that represents the deficit approach, the earliest line of investigation in the field of gender and language. Lakoff's work formed the idea of women's language, which was regarded implicitly as lacking and weak. Litosseliti (2004, p. 28) points out that Lakoff's views about women's language being over-polite, lacking in vocabulary and use of weaker expletives and tag questions have been heavily criticised, because it is not backed by empirical data and does not take linguistic differentiation into consideration. In addition, Coates (2004, p. 6) explains that Lakoff's work was criticised because it implicates that there is something wrong with women's language and that women should adopt the way men speak in order to be taken seriously. Nevertheless, Litosseliti (2006, p. 32) admits that Lakoff's work is important despite the criticism, because it was the starting point for research on "actual speech behaviour in context" and on "asking more critical, social questions about language".

Furthermore, Litosseliti (2006, pp. 2-3) explains that more recent approaches during the last two decades have concentrated on how men and women are established through language, instead of how differently they use language. These approaches are more complex in nature, shifting interest from male and female language use differences to discourse and its social context. According to Litosseliti (2006, p. 2), the earlier approaches on gender and language viewed language as a 'closed system with internal rules, and not as a dynamic entity influenced by external social factors and used variably by real speakers and writers.'

Litosseliti (2006, p. 27) points out that the gendered language discussion has focused on two primary directions; in the 1970s, on theories of dominance, and, in the 1980s, on theories of difference. Litosseliti (2006, pp. 27-32) explains that theories of dominance consider the differences in the way that women and men use language as an indication of men's dominance over women in interaction, and that this position was influenced significantly by the political status of women at the time and also by the existing "deficit" theories of women's language. Research concentrated on interaction within single-sex and mixed-sex groups, aiming to expose prejudice in language in general, as well as in language use.

According to Coates (2004, p. 6), the dominance approach regards linguistic practise as a tool, which enables men to dominate women, and that in discourse both genders contribute towards sustaining female oppression and male dominance. Coates (2004, p. 113) lists three ways in which a speaker can break the fundamental conventions of turn-taking in order to dominate other speakers. Firstly, a dominating speaker can interrupt the current speaker by "grabbing the floor". A second way of dominating is to talk too much, participating in behaviour called "hogging the floor". Lastly, a speaker may talk too little, or be completely silent, which is seen as non-cooperative behaviour and which frequently ends the conversation.

All of these discourse strategies have been studied widely. Coates (2004, p. 124) states that sociolinguistic research into mixed-sex talk indicates that women and men do not have equal rights in conversation. She mentions a piece of research conducted by West and Zimmerman (1998) on similarities between interchanges between men and women and parent and child. They found that women and children in modern American society have limited rights to speak and that interruptions are used by men as a reflection of dominance.

However, Tannen (1991, pp. 189-190) critiques the way in which interruptions are identified and interpreted, pointing out that recording conversations and counting instances of interruption does not take into consideration the substance of the conversations. She believes that in order to discover if a speaker is violating other speakers' rights, one has to know more about the speakers and the situation in which the interaction occurs. Additionally, she points out that different speakers have different conversational styles, which can influence the effects of linguistic strategies such as interruptions. Hence, the reason why men and women might feel interrupted by each other are the differences in what they want to achieve with talk.

According to Coates (2004, p. 116), talking too much is a conversational strategy that needs to be examined with consideration to context. In mixed-sex talk, the consensus is that women talk more than men, although studies consistently prove otherwise. There are studies that indicate that in many situations men talk significantly more than women (e.g. B. Eakins & G. Eakins, 1978; Swacker, 1979; Edelsky & Adams, 1990).

Furthermore, the third discourse dominance strategy mentioned by Coates (2004, p. 120), non-cooperation, appears to be used in informal talk in private. She continues that non-cooperation is a strategy in which one participant in interaction does not want to commit to having a conversation. For example, Sattel (1983) found that inexpressiveness is used by men in order to dominate and achieve control, in all-male and mixed-sex discussions.

According to Litosseliti (2006, p. 37), the difference approach to gender and language views differences in male and female language as products of different socializations of men and women. Coates (2004, p. 6) states that this approach started gaining interest at the beginning of the 1980s and was a reaction to women's resistance to being considered as a subordinate gender. Litosseliti (2006, p. 37) explains that unlike the deficit and dominance views, the difference approach views women's language as positively appreciated.

Litosseliti (2006, p. 38) states that cultural differences can materialise in girls and boys learning different styles and choices of interaction. She argues that girls are pressured to be nice and boys strong and that, linguistically, this often tends to work against girls and women, as it might be seen as unfeminine or bossy for females to use direct language. Tannen (1991, p. 244) argues that gender style differences are 'symmetrically misleading'; they learn language in different worlds and understand the other genders' way of interacting in relation to their own way. According to Tannen (1991, p. 244), in mixed-sex interaction women and men tend to speak in a way that is closer to men's style of talking than women's and, additionally, both ways of interacting are usually evaluated according to the rules of men's speech, which is considered the norm.

Tannen (1991) argues that women and men use language differently in relation to directedness. Women are often more indirect, preferring to not make demands or give orders, whereas men use more direct language. Tannen (1991, pp. 225-226) suggests that women being indirect is not because women are powerless or feel that they do not have a right to speak directly, but because they seek connection, wanting to achieve something without demanding it or being impolite. She states that indirectness itself does not indicate powerlessness, but the belief about the position of women in society influences the way in which women's speech is perceived.

Coates (2004, p. 110) concentrates on "gender differences" in conversational practise and suggests that men and women have different interactive styles. She presents evidence that women use more hedges and compliments, whereas men are more talkative, swear more and use directives in order to get what they want. Coates (2004, p. 110) calls linguistic characteristics that men and women use "men's style" and "women's style", and argues that linguistics modes should not be labelled in a simplistic manner, such as 'powerless' or 'powerful', because, for example, calling women's language powerless language supports the myth that women's language is weak.

Men's language has not received nearly as much attention in the field of language and gender as women's language. Coates (2004, p. 5) states that the issues of men and masculinity were left unexamined so long mainly because the terms man and person were typically considered as synonyms. However, during the 1990s researchers started to take greater interest in men and masculinity and there has also been a change in how men perceive themselves, considering themselves more than "unmarked representatives of human race".

The first book that concentrated on men, masculinity and language was Language and Masculinity (Johnson & Meinhof, 1997). The articles in this book focus mainly on relationships between males and females and male dominance in those relationships. According to Johnson and Meinhof (1997, pp. 2-7), the articles aim to show 'a range of positions regarding the conceptualization of masculinities' and to encourage further debate on the issue. They explain that a book that concentrates on language and masculinity and relies on spoken or written data complements other studies in the field of gender and language.

Johnson and Meinhof (1997, pp. 12-13) argue that since women are seen as objects of problematization in the field of gender and language, men are then considered to represent the normative status, which stems from a lack of sufficient research on men and masculinity. They believe that it is important to explore men linguistically, as constructed individuals, not only as ungendered representatives of the human race. In addition, they point out that focusing solely on women and femininity is insufficient, and that in order to understand all the aspects of gender and language, linguists should consider the input that the study of men and masculinity can offer to the field.

Coates (2003, pp. 1-2) studies men's language in her book *Men Talk* by exploring stories that men tell each other in everyday conversation. She concentrates on narratives that occur in informal, all-male conversations and are set in various contexts, in order to discover the cultural principles that "lie behind men's lives and masculine identities at the turn of the century in Britain". Coates (2003) tries to demonstrate how masculinity is built into discourse, and how men's talk maintains the accepted forms of being male.

Furthermore, according to Litosseliti (2006, p. 63), the current trends in examining gender and language from the point of view of feminist linguistics are more complex than before and include the re-evaluation of the issue of differences in genders. Litosseliti (2006, pp. 63-68) explains that there has been a shift in gender and language study towards a more complicated inquiry on discourse, gender, the position of discourse in constructing identity, and language in general. She continues by stating new thinking in the field of gender and language focuses on the dynamics of the situations and societies, where enactment of gender occurs and that the whole field of study has become more interdisciplinary and diverse.

Coates (2004, pp. 215-218) agrees with the notion that the emphasis in the field of gender and language has shifted from looking at language to looking at discourse. She explains that the concept, discourse, shows the "value-laden nature of language" and that our formation of ourselves as feminine and masculine is influenced by 'the discourses on gender current at any given time'. In addition, she states that a new sociolinguistic approach has urged researchers to examine the speech patterns of both genders in a range of different cultures, because gender is formed locally and it is influenced by age, class, race and sexuality.

As has been shown, there are many different areas of research on gender and language, which indicates that this field is versatile and has many different aspects that interest researchers. Age is also a factor that can influence how men and women use language, which we originally wanted to include in our study. However, the corpus used in this study is not very suitable for this purpose, hence this variable needs to be further pursued in future, complimentary research. We shall now present the model for thought representation that we adopted for this study.

## 3. Thought Representation

Simpson (2004, p. 30) points out that a special interest of modern stylistics has been the way in which thought is represented in texts. Therefore, researchers in the field of stylistics are interested in explaining writers' methods for transcribing the thoughts of imaginary or real people. This area of stylistics, thought representation, contains a variety of methods for reporting thought. These methods ease identification of the modes used in texts and help us to evaluate their effects.

### 3.1 A Model for the Representation of Thought

Leech and Short (1981, pp. 336-338) (Note 2) state that the categories of representation of thought are the same as those of representation of speech, although it should be remembered that, like speech, the representation of thoughts of characters is fundamentally an artifice, in spite of how direct the form of representation would be.

Naturally, as Leech and Short (1981, pp. 336-338) point out, it is impossible to see directly into the thoughts of other people, but it is still necessary for novelists to try to present their characters' flow of thought.

Short (1996, p. 311) points out that although the categories of representation of speech and thought are formally very similar, the effects of some of the categories are different, namely those of Direct Thought (DT) and Free Indirect Thought (FIT). He continues by stating that Indirect Thought (IT) is a relatively rare category, because it is so indirect and therefore does not suit well for presenting and foregrounding a speaker's exact thoughts.

We have omitted the categories that are irrelevant to our study and replaced them with the ones that McIntyre et al. (2004) have adopted for their study. Watson (1997, p. 143) presents a revised model of speech and thought representation, which also clarifies the correspondence of the categories, which has been modified to suit this study:
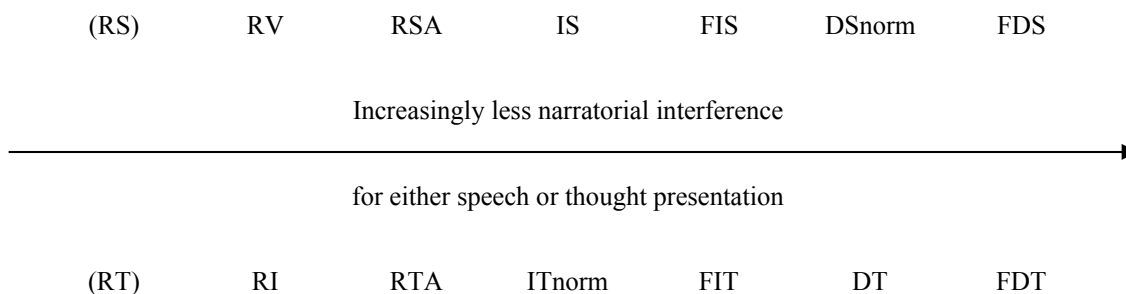
| (RS) | RV | RSA | IS | FIS | DSnorm | FDS |

Increasingly less narratorial interference

for either speech or thought presentation

| (RT) | RI | RTA | ITnorm | FIT | DT | FDT |

Figure 1. Modified correspondence of speech and thought representation categories and "interference" in report

The category of Direct Speech is assumed to be the norm of speech representation, but as we can see from the diagram above the norm of thought representation is Indirect Thought instead of Direct Thought. Short (1996, p. 315) states that this is because the thoughts of other people can never be examined directly and that we can merely infer what people think, for instance from their actions and speech. Therefore, it is reasonable to regard IT as the norm of thought presentation instead of DT.

In addition, according to Short (1996, p. 315) the differences between DT and DS or between FIT and FIS are partly due to IT being the norm of thought presentation, because when we move on the scale of thought presentation from the norm category IT towards FIT the narratorial influence decreases, whereas when we move from the norm category of speech presentation, DS, towards FIS the influence of the narrator increases.

## 4. Methodology

The Lancaster Speech, Writing and Thought Presentation Spoken Corpus was compiled by McIntyre et al. (2004, p. 49) in order to study "the ways in which speakers present speech, thought and writing in contemporary spoken British English". They state that one aim was to also compare the results to a previous corpus study of SW&T presentation in written texts (see Semino et al., 2004). The composite texts were taken from two different archives, namely the British National Corpus (BNC) and archives housed in the Centre for North West Regional Studies (CNWRS). According to McIntyre et al. (2004, p. 51), the previous research on SW & TP in speech has concentrated on direct speech or has been based on quantitatively small amounts of data that has been acquired from a specific context. The Lancaster SW & TP spoken corpus is an attempt to construct a balanced corpus of contemporary spoken British English in order to analyse the presentation of speech, thought and writing in a systematic way.

McIntyre et al. (2004, p. 53) used texts from the BNC that were from the spoken demographic part of the corpus because it allowed them to compare spontaneous dialogue with elicited monologues of the CNWRS archives. They explain that the texts drawn from the BNC cover all age ranges and there is equal representation of male and female respondents and the texts are face-to-face conversations, which constitute spontaneous and unscripted material. See Appendix 1 for further detail.

The CNWRS material was collected from two archives, namely the "Family and Social Life" archive and the "Childhood and Schooling" archive. According to McIntyre et al. (2004, p. 52), the previous archive was collected in the 1970s and 1980s by Elizabeth Roberts and Lucinda Beier and it includes 250 hours of interviews, taped on reel to reel tapes and audiocassettes, and transcripts of those interviews, while the "Childhood and Schooling" archive was compiled by Penny Summerfields in the 1980s and it consists of approximately 200 hours of interviews on audiocassette and it also includes transcripts. In this dataset, McIntyre et al. (2004, p.

52-53) aimed to acquire equal amount of texts from male and female interviewees. See appendix 2 for further detail.

McIntyre et al. (2004, p. 56) annotated the Spoken Corpus for SW & TP (approx. 260 000 words) using tags that allowed them to compare their results with findings of the Written Corpus project. They explain that the system of annotation in the Spoken Corpus is adopted from Wynne et al. (1998), though due to the differences between the written and spoken data they made some modifications to the system. Appendix 3 shows the acronyms used to indicate instances of SW & TP in the Written Corpus project and their equivalents in the Spoken Corpus.

The Lancaster SW & TP corpus is very useful for researching the thought representation model, and gender differences, since the amount of female and male speakers is already approximately the same and the instances of thought categories are tagged in the text. However, some modifications to the corpus and to the thought model had to be made for this study. As we started to work on the Lancaster SW & TP corpus, we had to make many decisions, since McIntyre *et al.* used the corpus generally, without considering the gender or age of the speakers. We counted the instances of thought categories manually, using the "search" function on the text files. We tried to use several search engines on the corpus but because they were incompatible we decided on a manual approach. As always, when working with large quantities of data manually, there is a chance of human error. Nevertheless, the annotation of the corpus was reasonably straightforward and clear, which made our work easier.

Furthermore, when we decided to use the Lancaster SW & TP corpus, we also decided to adopt their thought model. McIntyre et al. (2004) also used the written presentation categories in their research, but we decided to exclude them from our study because we did not think that including them would add very much to this study and also because our own preference and interest lies in the thought categories. As mentioned before, McIntyre et al. (2004) included additional features in their categories that are marked by lower-case letters, for example, according to McIntyre et al. (2004, p. 64), the instances tagged with the suffix e indicates occurrences of "discoursal embedding where one SW & TP category is embedded discoursally, but not necessarily syntactically, in another". We did not count these additional features separately, but included them in their main categories, for example if an instance of RTA was annotated as RTAp (RTA with topic), we counted it as simply RTA.

McIntyre et al. (2004) included the information about the speakers in file headers, from which we could discover their gender and age. However, this issue was not straightforward, since some speakers' information was incomplete, e.g. missing information on gender. Therefore, we decided to omit the files that had speakers whose information did not state the speakers' gender and the files whose header information did not correspond to the text in the file. This modification resulted in a total number of 109 files. In addition, we did not include the interviewers in the CNWRS part of the corpus as speakers, because almost all of them were female, and it would have resulted in there being significantly more female speakers than male speakers in the data. However, after these changes there were still an approximately equal number of files between the CNWRS part and the BNC part of the corpus and an equal number of male and female speakers in the corpus. The statistical procedure employed was chi-squared test, which enabled us to obtain objective statistical results and to use them to support our thesis.

However, the issue of the age groups was not as straightforward as the division of male and female speakers. Many of the corpus file headers did not necessary include information about the speakers' ages and therefore, the amount of male and female speakers and age groups was not equal. The corpus files had to be rearranged according to the amount of words that the speakers use. This data rearrangement was conducted by Dr. Jukka Mäkisalo from the University of Eastern Finland. The age groups, in which we divided all the speakers, are obtained from the BNC part of the corpus, because the CNWRS part did not have any clear age group division. However, after examining the rearrangement of the data, it became clear that there were still some problems with the header information in the corpus, so, as a result, research on the age groups had to be excluded from this study. In the following chapter, we present our results on the representation of thought categories and gender, and discuss our findings.

## 5. Results and Discussion

The aim of this chapter is to offer a quantitative and qualitative analysis of the data obtained from the SW & TP corpus (Note 3). We begin with the general results for the representation of thought categories, then move on to our findings on gender differences for these categories.

*5.1 Thought Representation Categories*



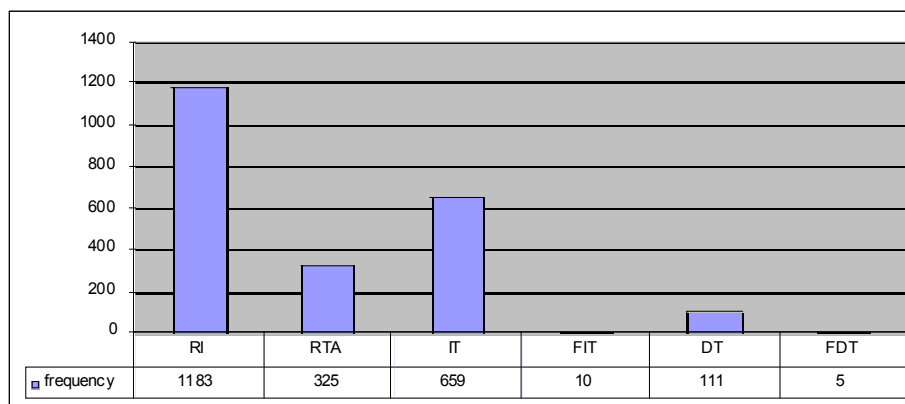| frequency | RI | RTA | IT | FIT | DT | FDT |
|---|---|---|---|---|---|---|
| | 1183 | 325 | 659 | 10 | 111 | 5 |

Figure 2. The number of occurrences of thought representation categories in the SW & TP spoken corpus

The overall frequency and division of instances of thought representation categories in the SW & TP corpus are shown in figure 5.1. The difference between the most frequent and the least frequent categories is clear. Thought representation categories are quite different in function, use and form than speech categories and according to McIntyre et al. (2004, p. 71) this is mainly because "speech and writing are both modes of ostensible communication leading to the physical production of 'discourse', while thought is a private and often non-verbal phenomenon". The difference between the frequency of the categories is statistically very highly significant ($x2$ = 2814, 30; df = 5; p ≤ 0.001) (Note 4). The high frequency of Indirect Thought (IT) was expected, and since it is considered the norm of thought representation it is also understandable. The issue of even higher number of occurrences of Representation of Internal State (RI), on the other hand, may not be as straightforward. In addition, although the number of instances for Free Indirect Thought (FIT) and Free Direct Thought are low, it was expected based on McIntyre et al.'s (2004, p. 67) results for the same categories.

As mentioned earlier, RI was not in the original Leech and Short (1981) thought representation model. The category was formed by Semino and Short (2004) as Internal Narration and later modified to Representation of Internal State by McIntyre *et al.* in order to be able to apply it to spoken language. According to McIntyre et al. (2004, p. 62), RI basically catches references to mind states that can be emotional or cognitive, which are not really distinct thoughts. On the thought representation cline, RI is located at the most restricted end. As with most of the categories, McIntyre *et al.* do not offer any extensive quantitative or qualitative insights into the high frequency of RI in their spoken corpus.

Furthermore, Semino and Short (2004, pp. 133-134) state that in their written corpus study the category NI (which is basically the same category as RI in effect and function, only applied to written language) was also the most frequent form of thought representation, which corresponds to McIntyre et al.'s findings, as well as to the results in this study. Semino and Short (2004, pp. 133-134) explain that the pure cases of NI are especially present in the fiction section of their corpus, and also in the (auto) biography section and that the reason for this could be that in fiction there is often a narrator, who describes the internal processes of a character, or a first person character, who gives accounts of his or her own internal state. According to Semino and Short (2004, pp. 133-134), NI is frequent in the (auto) biography section, because there is always a narrator in autobiographies, telling about "their own past cognitive states and changes". In this study, the elicited monologues of CNWRS contained most of the instances of RI. With 823 instances, it has remarkably more occurrences of RI than the spontaneous speech of the BNC, which only showed 360 instances. An explanation for this might be the same as for the (auto)biography section of Semino and Short's (2004) study, namely that as in autobiographies, in the elicited monologues of CNWRS the speakers often recall their past, telling the interviewer about their states of mind at certain times and situations in their childhood or youth. The following examples are from the CNWRS component of the SW & TP corpus:

(1) The room seemed to be full of flowers [RI] *and oh I I hated that smell*.

(2) As I say [RI] *I was unhappy at that school* because somehow I didn't fit in, in that way.

RI conveys a wide range of mental states or processes and emotional experiences. The instance of RI in example (1) involves an emotional impact of a scent, a reaction to the smell of flowers. A man is telling a story about the

time his brother died and the coffin was at his house, in a room that smelled like lilies because of all the funeral flowers. In example (2) the RI conveys an emotional state, which seems to occur over a relatively long time.

Representation of Thought Act or RTA is the third most frequent thought representation category in the SW & TP corpus with 325 instances. RTA can be seen as the formal counterpart of RSA in the speech representation cline, but it actually is not the same in effect. Semino and Short (2004, pp. 130-131) explain that because thought acts are not communicative, RTAs cannot, in essence, present illocutionary force, like RSAs. According to Semino and Short (2004, pp. 130-131), RTAs often refer to a particular individual thought and do not give any account of the specific 'wording' of the thought, and RTAs do not include a reported clause. As in the case of RI, there were more instances of RTA in the CNWRS section of the SW & TP corpus, 234 occurrences, while there were only 91 occurrences in the BNC section. Since RTA does not usually have the same kind of summarizing effect as RSA, it is possible that RTA is more frequent in the monologues than in the spontaneous speech for similar reasons as RI, since Semino and Short (2004, pp. 131-132) also found that RTA was most frequent in fiction and (auto) biography sections of their written corpus.

(3) the man who took me for Latin was a marvellous teacher, he was an Oxford man too, er and I remember we read erm the second book of Virgil's Aeneid which is the story of the Trojan Horse, [RTA] *which fired my imagination*, and its language, i in my opinion is absolutely marvellous!

(4) My my favourite subject was geography, and it still is. [RTA] *And I've tried to learn geography er a and, and ignore history* and I found that I can't.

In example (3) the stretch of RTA conveys a mental act that was impacted upon by something the speaker read. This is very typical RTA, since there is no indication as to what the speaker imagined; only that he did so. In example (4) the instance of RTA shows a mental process, learning, and although the subject of learning is indicated, it is not specific enough to be considered a topic.

The category of Indirect Thought (IT) is next on the thought representation cline, and it is the second most frequent thought category, with 659 occurrences, and it is significantly more frequent than Free Indirect Thought, which had only 10 instances. McIntyre et al. (2004, p. 70) suggest that if we would consider the scale of thought representation without the new category RI, IT would be the 'quantitative norm' of the cline, which would support Leech and Short's stand that IT is the norm of thought representation (since RI was not in Leech and Short's original model). Semino and Short (2004, pp. 127-128) state that IT does not have a summarizing effect like IS, but in contrast it can have the impact of giving an account of the actual wording of some specific thought, without actually claiming to do so. In the SW & TP corpus, the frequency of IT was higher in the CNWRS files (398 instances) than in the BNC files (261 instances). The following examples are from the CNWRS part and BNC part of the corpus, respectively:

(5) Er and er I remember being kissed, and I mean really kissed and [RT] I thought [IT] *I was going to have a baby* and I would be thirteen or fourteen I suppose then. I kissed boys at parties, but when this boy kissed me like that I worried myself to death.

(6) Did you have a word with Angela about 45s?

No cos [RT] I thought [IT] *you were going in Friday*.

In both examples above, the instances of IT are quite typical, since there is a reporting clause with the verb "think". In example (5) IT presents the fear in the speaker's mind about having a baby after being kissed by a boy and in example (6) the occurrence of IT realizes the specific thought of the speaker in a situation, where she is asked if she has talked to a woman called Angela on behalf of the person questioning her.

As mentioned earlier, Free Indirect Thought is quite infrequent category in the whole SW & TP corpus, since it occurs only ten times. Furthermore, eight of these are in the CNWRS section, but with so few occurrences it is impossible to make any meaningful conclusions about it. This is in stark contrast with the findings of Semino and Short (2004, pp. 123-124), who found that in their written corpus study FIT was overall more frequent than IT and especially frequent in the fiction section. Based on this, Semino and Short (2004, p. 123) state that "FIT is primarily, but not exclusively, a fictional phenomenon". Since IT is the norm of the thought representation scale, FIT is one of the freer forms in the scale and therefore, cannot be compared to FIS, which is one of the indirect categories in the speech representation cline. According to Semino and Short (2004, p. 124), FIT creates the effect of closeness and empathy and it gives a more constant access to characters' consciousness than the other thought representation forms. Due to the fact that this type of access to a person's thoughts cannot really be achieved in real life and that FIT seems to be typically associated with fiction, it is not difficult to understand why it might not be frequent in spoken language.

(7) He would go in the Police Force, [FITgi] *not to be a policeman* he didn't think he wanted to be [FITi] *but in the police force somehow*.

In example (7) there are two stretches of FIT, both are inferred (the suffix i) which means that the speaker does not have direct access to the thoughts, in essence, the conveyed thoughts are someone else's and in this case the brother's of the speaker. Also, the suffix g in the first instance means grammatical negative. In the example, the occurrences of FIT are actually accentuated by the IT in between the FIT stretches, highlighting their freer form.

The most direct categories on the thought representation cline are Direct Thought (DT) and Free Direct Thought (FDT), with 111 and 5 instances respectively. Because of the low frequency of FDT in the SW & TP corpus it is not sensible to form any specific conclusions regarding its use, other than it seems that it is clearly not a category that is used much in spoken English. Leech and Short (1981, pp. 342-343) explain that the use of DT and FDT results in a monologue, a stream of consciousness in which the character (or in this case the speaker) talks to him or herself. So why would there not be clearly more instances in the CNWRS part of the corpus, since it contains recorded monologues? The reason could be that the monologues are elicited, so there is always an interviewer involved, which means that there is always someone to address the stories to and this might somewhat restrict a "speaking to oneself" behaviour.

Although Direct Thought clearly shows more instances than Free Direct Thought in the corpus, it is only the fourth most frequent category of all the thought presentation categories. Furthermore, according to Semino and Short (2004, pp. 118), DT occurs often "at moments of heightened emotion or of sudden and momentous realization". Therefore, it is not surprising that there are more instances of DT in the BNC (71 instances) section, than in the CNWRS (40 instances) section, since the BNC includes spontaneous speech, moments of realization or heightened emotion might perhaps occur in such situations more easily than during the elicited monologues of CNWRS.

(8) He said, Oh well I'm Raymond Winder from next door, and I've just come to tell I want no troub no trouble from him, erm Keep out of my way. Something of that sort and th I was furious, thought he was only about fourteen something like that, and [RT] I thought, [DT] *That's a good start!* you know, I mean we were only just moving in.

(9) and er anyway he answered this and in the train, he'd never been to this part of England he heard somebody saying J. P. Smith [RT]and he thought [DT] *oh that's the man who wrote to me* and he listened a bit and the man said er oh he's a very staunch Roman Catholic so dadda leaned back [RT]and thought [DT]*I'll be alright then with him*.

In example (8) the stretch of DT is an emotional reaction to the audacity of the boy next door coming to the speaker's door and dictating how she and her husband should act with him. There are two occurrences of DT in example (9), where a man is telling a story about his father going to a job interview and overhearing a conversation in the train about the man who wrote to him concerning the job. The effect of DT in both of these instances is the same; the speaker seems to convey the thoughts of his father directly, like the hearer could be actually listening in on what he is thinking to himself.
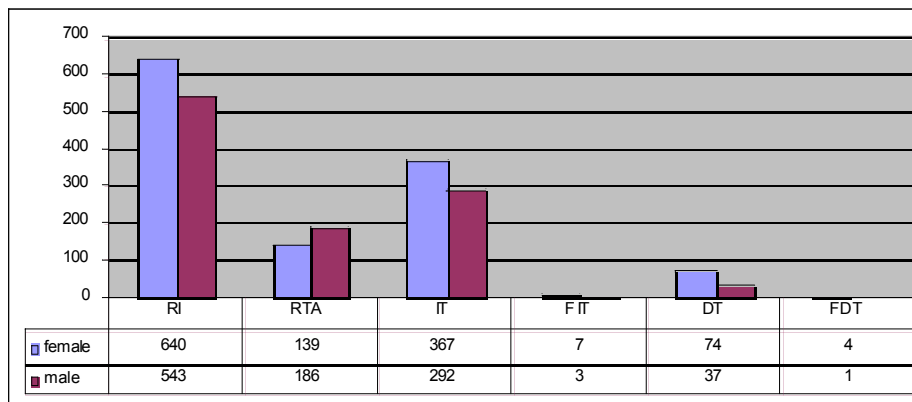
*5.2 Thought Representation, Male/Female*



Figure 3. The frequency of thought representation categories by gender

The distribution and frequency of the thought representation categories is quite distinct. The three most frequent categories are at the most indirect end of the scale, and the categories of Free Indirect Thought and Free Direct Thought have such very low number of instances that it is virtually impossible to form any valid conclusions about them. Nevertheless, there are thought representation categories that did show statistically significant differences and what makes this interesting is that a few of those categories are at the indirect end of the cline and only one at the direct end, which indicates that there are some issues to examine, despite the low frequency of FIT and FDT. However, since not much research on thought representation and gender in spoken language has been conducted, it is not a very simple matter to present evidence or reasons for differences in the use of thought categories between genders. The issue is slightly more straightforward regarding speech representation, since discourse and language use in context of gender has been researched more than the expression of thoughts in language by men and women.

Representation of Internal State is the most frequent thought representation category in the whole corpus and there is a statistically highly significant difference in the use of RI between male and female speakers ($x2 = 7,953$; $df = 1$; $p \leq 0.01$). As mentioned before, RI has a broad scope, since it can convey many different kinds of mental states. According to Semino and Short (2004, p. 132), the category "captures the mental states and changes which involve cognitive and affective phenomena but which do not amount to specific thoughts". Furthermore, one reason for the difference in use between male and female speakers could be the notion that men do not express thoughts and emotion as easily as women. Tannen (1991, pp. 83-84) studies the expression of feelings and thoughts, especially in and about relationships, and suggests that women speak about their mental states, emotions and thoughts as they come, while men are accustomed to dismissing their fleeting thoughts.

The Representation of Thought Act is the third most frequent thought category in the SW & TP corpus. Furthermore, it is the only category that shows a statistically highly significant difference between the genders ($x2 = 6,797$; $df = 1$; $p \leq 0.01$), where men use the category more than women. The instances of RTAp, or RTAs with topic were counted as RTAs and not as a separate category. In addition to men using RTA significantly more than women, men used the category more in the CNWRS part of the corpus than in the BNC part, with 142 and 44 instances respectively. However, due to the nature of the category and results from other studies, it is not surprising that it is more frequent in the monologues and storytelling of CNWRS. The reason for male dominance in the use of RTA could be that it allows the expression of individual thoughts without needing to relay the actual wording or specific topic of the thought, thus avoiding getting into too much detail. As mentioned before, relaying details is more common in female speech and according to Tannen (1991, p. 115) "men … often find women's involvement in details irritating".

Indirect Thought is the second most frequent thought representation category in the SW & TP spoken corpus, and it is considered to be the norm of the thought representation cline. IT shows a statistically highly significant difference in use between male and female speakers ($x2 = 8,536$; $df = 1$; $p \leq 0.01$). According to Semino and Short (2004, pp. 127-128) IT is the most typical mode for thought representation, it conveys the propositional content of a thought without actually claiming to repeat any words in the character's or speaker's mind and in effect it is more "understated and less dramatic" than FIT, DT and FDT, which create the impact of immediacy and lucidity. The use between the genders shows a statistically highly significant difference, with female speakers using the category more than male speakers, but it seems that they use IT similarly, because both male and female speakers tend to use IT more in the elicited monologues of the CNWRS part of the corpus that involve storytelling than in the spontaneous speech, or non-story discourse, of the BNC part.

The fourth most frequent thought representation category is Direct Thought, and it is the only category at the direct end of the thought representation cline that shows difference in use between male and female speakers at a statistically very highly significant level ($x2 = 12,333$; $df = 1$; $p \leq 0.001$). This supports our thesis that women use more direct forms of the representation of thought (and speech) than men in spoken English. The overall high number of instances of Direct Speech in the spoken corpus was expected and so was the low frequency of Direct Thought. With only 111 instances, it is quite low compared to RI, RTA and IT. This is due to the nature of the DT category, namely its function and effect. Compared to Indirect Thought, Direct Thought comes with a claim to faithfully present the thoughts of a speaker, conveying the actual wording of the thoughts, and one reason why women might use DT more than men could be their tendency to communicate their thoughts in more detail than men (Tannen 1991: 83-84). According to Leech and Short (1981, p. 345), DT (as well as FDT) has a conscious quality that is not present in the other thought categories. Semino and Short (2004, pp. 118-120) discuss DT and FDT as one category, and state that (F)DT requires the translation of thoughts into words and that the result is a "conscious and deliberate thought", the effect of the speaker talking to him or herself. Semino and Short (2004, p. 119) explain that the (F)DT occurs often at a moment of heightened mental state, emotional

or cognitive. Furthermore, according to Johnson and Meinhof (1997, p. 17) men "are unable to express their emotions with the same lucidity as women due to the pressure of a patriarchal society which demands that they appear rational and unemotional". Tannen (1991, pp. 76-77) supports this view by explaining that women tend to use dialogue and talk in general in establishing and maintaining relationships, and in conveying and creating closeness and connection, while men prefer "public speaking" in order to maintain independence and social status, and to show knowledge. These theories give weight to the suggestion that women use more direct forms of thought representation than men.

## 6. Conclusion

This study has approached the issue of potential differences in language use between male and female speakers by utilizing the thought representation model originally formed by Leech and Short (1981) for written language, initially literature. The model used in this study is a modification of that original model, with added categories and a new representation of a writing cline plus variants that highlight different features of the categories. This modified model was presented by McIntyre et al. (2004). The source data of this study is the Lancaster Speech, Thought and Written Presentation spoken corpus, also compiled and annotated by McIntyre et al., but which was also modified in order to meet the purposes of this study.

The application of this model to spoken language is a relatively new direction for research on thought representation, and it seems that the model is very applicable to spoken language, with some minor changes. The categories of thought representation have been applied to spoken language in only a few studies, but they have not been used to examine any potential difference in the language use of men and women, even though there has been considerable research on gender differences and language, conducted from multiple viewpoints and using a variety of methods. Earlier research focused on women's language from a feminist point of view, but later the focus shifted to examining how men use language, discourse between men and women and discourse between same gender speakers, from, for example, a linguistic or sociolinguistic point of view. This study draws possible explanations from different gender studies, in order to explain why certain patterns of use of language between male and female speakers have emerged and how the motivation for men and women to use language in certain ways might be the reason for these patterns.

According to the quantitative results, our thesis proved correct, female speakers do use direct forms of thought representation more than men. Four of the six thought representation categories showed a statistically significant difference in use between the two genders, and three of those categories were at the indirect end of the cline and only one in the direct end. However, in the scope of our study, this still proves our assumption right, i.e. women use more direct forms of thought than men. The most frequent thought representation categories across the two genders were at the indirect end of the thought representation cline and the direct thought categories were the least frequent. Nevertheless, since expressing thought in spoken language is always an artifice, as we can never truly know the exact words of the thoughts that pass through a speaker's head, the more indirect categories are a way to avoid trying to repeat thoughts word for word, since they convey thoughts without claiming to be faithful to the wording of those thoughts. The reason why female speakers seem to use more direct forms of thought representation might be related to the notion that women seem to be more inclined to express their specific thoughts and emotions in general and in more detail, than men.

This study shows that it is clear that a thought representation model applied to corpora can offer many insights. Further studies on the use of the different forms of thought in spoken language should prove fruitful. Since corpora are, in general, versatile sources of data, there are many focuses available; perhaps applying the model in order to examine dialects or different kinds of discourse situations would provide interesting results. In addition, the study of different variables, such as age or occupation should prove useful in future research on thought representation in spoken language.

## References

Coates, J. (2003). *Men Talk: Stories in the Making of Masculinities*. Oxford, UK: Blackwell Publishing.

Coates, J. (2004). *Women, Men and Language: A Sociolinguistic Account of Gender Differences in Language* (3rd ed.). United Kingdom: Pearson education limited.

Eakins, B., & Eakins, G. (1978). *Sex Differences in Human Communication*. Boston: Houghton Mifflin.

Edelsky, C., & Adams, K. (1990). Creating Inequality: Breaking the Rules in Debates. *Journal of Language and Social Psychology*, *9*(3), 171-190. http://dx.doi.org/10.1177/0261927X9093001

Johnson, S., & Meinhof, U. H. (Eds.). (1997). *Language and Masculinity.* UK and USA: Blackwell Publishers Ltd.

Lakoff, R. (1975). *Language and Woman's Place*. New York: Harper & Row.

Leech, G. N., & Short, M. H. (1981). *Style in Fiction: a Linguistic Introduction to English Fictional Prose*. London and New York: Longman Group Limited.

Leech, G. N., & Short, M. H. (2007). *Style in Fiction: a Linguistic Introduction to English Fictional Prose*. London: Pearson Longman.

Litosseliti, L. (2006). *Gender & Language: Theory and Practise*. London: Hodder Education.

McIntyre, D., Bellard-Thompson, C., Heywood, J., McEnenry, T., Semino, E., & Short, M. (2004). Investigating the Presentation of Speech, Writing and Thought in spoken British English: A Corpus-based Approach. *International Computer Archive of Modern and Medieval English Journal*, *28*, 49-76.

Semino, E., & Short, M. (2004). *Corpus Stylistics: Speech, Writing and Thought Presentation in a Corpus of English Writing*. London: Routledge.

Sattel, J. (1983). Men, Inexpressiveness and Power. In B. Thorne, C. Kramarae, & N. Henley (Eds.), *Language, Gender and Society* (pp. 119-124). Cambridge, Mass.: Newbury House.

Short, M. (1996). *Exploring the Language of Poems, Plays and Prose*. London: Longman.

Simpson, P. (2004). *Stylistics: A Resource book for Students*. London: Routledge.

Swacker, M. (1979). Women's Verbal Behaviour at Learned and Professional Conferences. In B. Dubois, & I. Crouch (Eds.), *The Sociology of the Languages of American Women* (pp. 155-160). Texas: Trinity University.

Tannen, D. (1991). *You Just Don't Understand: Women and Men in Conversation*. London: Virago Press Limited.

Tannen, D. (1994). *Gender and Discourse*. New York: Oxford University Press.

Watson, G. (1997). *Doin' Mudroroo: Elements of Style and Involvement in the Early Prose Fiction of Mudroroo*. Joensuu: University of Joensuu.

West, C., & Zimmerman, D. (1975). "Sex Roles, Interruptions and Silences in Conversations". In B. Thorne, & N. Henley (Eds.), *Language and Sex: Difference and Dominance* (pp. 105-129). Rowley, Mass.: Newbury House.

**Notes**

Note 1. This paper is based upon Anne Riissanen's original, unpublished MA research, University of Eastern Finland.
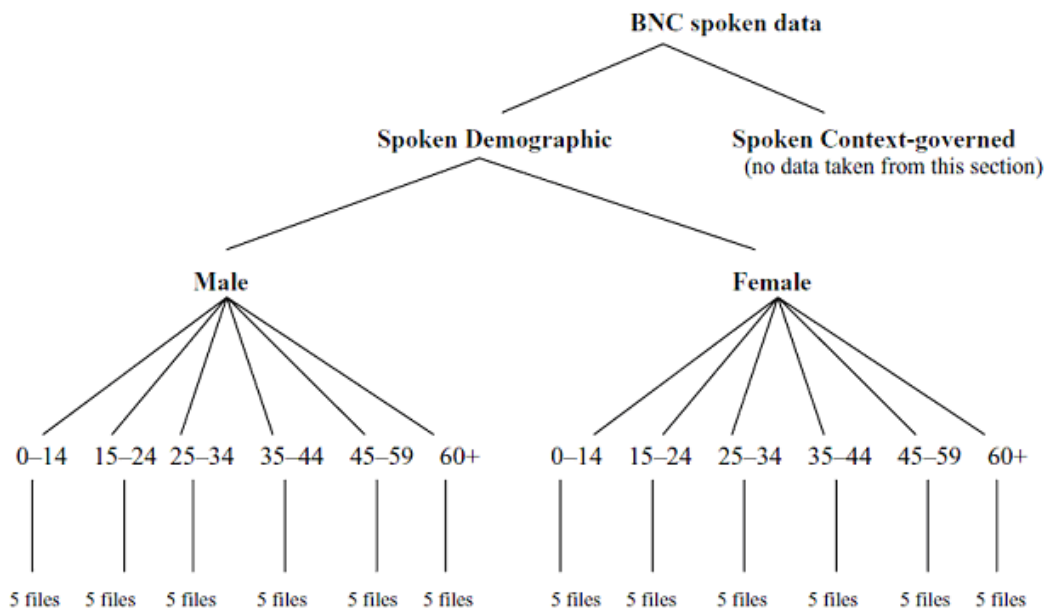
Note 2. Although Leech and Short published their second edition in 2007, we refer to the original edition, as their model has not changed.

Note 3. Unless otherwise stated, SW & TP spoken corpus refers to the corpus with the modifications made for this study (see Methodology).
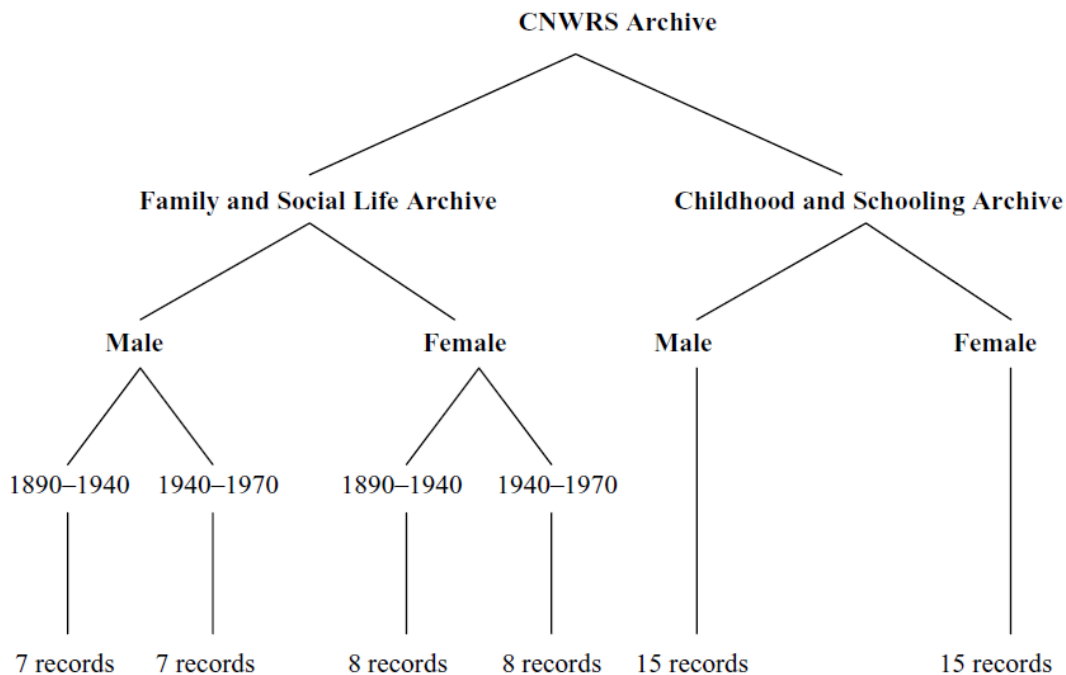
Note 4. All p values refer to Chi-squared analysis, unless otherwise stated. The results are statistically significant when p = 0.05; very significant when p = 0.01; and very highly significant when p = 0.001.

**Appendices**

Appendix 1. Number and distribution of BNC files in the Lancaster SW & TP spoken corpus (McIntyre et al., 2004, p. 54)

**BNC spoken data**

Spoken Demographic       Spoken Context-governed
(no data taken from this section)

Male                                          Female

0–14  15–24  25–34  35–44  45–59  60+        0–14  15–24  25–34  35–44  45–59  60+

5 files  5 files  5 files  5 files  5 files  5 files       5 files  5 files  5 files  5 files  5 files  5 files

Appendix 2. Number and distribution of CNWRS texts in the Lancaster SW & TP spoken corpus (McIntyre et al., 2004, p. 53)

**CNWRS Archive**

Family and Social Life Archive            Childhood and Schooling Archive

Male                Female                Male                Female

1890–1940   1940–1970   1890–1940   1940–1970

7 records   7 records   8 records   8 records   15 records          15 records

Appendix 3. Acronyms used to mark instances of SW & TP in the written corpus and their counterparts in the spoken corpus (McIntyre et al., 2004, p. 57)

| Categories outside the discourse presentation clines | | | |
|---|---|---|---|
| **Written Corpus** | | **Spoken Corpus** | |
| **Category** | **Definition** | **Category** | **Definition** |
| N | Narration | A | Anything other than SW&TP (narrative and non-narrative) |
| | | RU | Report of Language Use |
| NRS | Narrator's Report of Speech | RS | Report of Speech |
| NRT | Narrator's Report of Thought | RT | Report of Thought |
| NRW | Narrator's Report of Writing | RW | Report of Writing |
| **Discourse Presentation Categories** | | | |
| **Written Corpus** | | **Spoken Corpus** | |
| **Category** | **Definition** | **Category** | **Definition** |
| NV | Narrator's Representation of Voice | RV | Representation of Voice |
| NI | Narrator's Representation of Internal States | RI | Representation of Internal State |
| NW | Narrator's Representation of Writing | RN | Representation of Writing |
| NRSA | Narrator's Representation of Speech Act | RSA | Representation of Speech Act |
| NRTA | Narrator's Representation of Thought Act | RTA | Representation of Thought Act |
| NRWA | Narrator's Representation of Writing Act | RWA | Representation of Writing Act |
| NRSAp | Narrator's Representation of Speech Act with Topic | RSAp | Representation of Speech Act with Topic |
| NRTAp | Narrator's Representation of Thought Act with Topic | RTAp | Representation of Thought Act with Topic |
| NRWAp | Narrator's Representation of Writing Act with Topic | RWAp | Representation of Writing Act with Topic |
| IS | Indirect Speech | IS | Indirect Speech |
| IT | Indirect Thought | IT | Indirect Thought |
| IW | Indirect Writing | IW | Indirect Writing |
| FIS | Free Indirect Speech | FIS | Free Indirect Speech |
| FIT | Free Indirect Thought | FIT | Free Indirect Thought |
| FIW | Free Indirect Writing | FIW | Free Indirect Writing |
| DS | Direct Speech | DS | Direct Speech |
| DT | Direct Thought | DT | Direct Thought |
| DW | Direct Writing | DW | Direct Writing |
| FDS | Free Direct Thought | FDS | Free Direct Thought |
| FDT | Free Direct Thought | FDT | Free Direct Thought |
| FDW | Free Direct Writing | FDW | Free Direct Writing |

**Copyrights**