On the search for metadiscourse units

Annelie Ädel
annelie.adel@eng.gu.se
English Department
Göteborg University, Sweden

1. Introduction

To what extent do writers anchor their discourse in the current discourse situation and make the presence of the writer and/or reader overt? What types of situations are at hand when writers refer to their text as text, to themselves or to their readers? To what extent are texts monologic or dialogic? These types of questions have recently attracted a lot of attention within research on metadiscourse. Metadiscourse refers to discourse about on-going discourse and is interesting to study from the perspective of how the writer's or reader's presence in the text are made explicit. It has been studied by various branches of linguistics, for example text linguistics (Mauranen 1993, Markkanen et al 1993) – also in a historical perspective (Taavitsainen 2000), pragmatics (Hyland 1998) and genre studies (Bäcklund 1998, Bondi 1999) and is rapidly becoming a dynamic field of research.

One particularly interesting perspective is the study of cultural differences in the use of metadiscourse. Several researchers have shown that metadiscourse typically differs across cultures (e g Markkanen et al 1993, Mauranen 1993, Vassileva 1998, Bäcklund 1998). Cultural differences in writing have been studied within the field of contrastive rhetoric, on the basis of the primary hypothesis that there are culture-specific patterns of writing, and that these cause interference in L2 writing (see Connor 1996:90). In this paper, I will report ongoing research into the use of metadiscourse in written argumentative texts by native and non-native speakers of English. The study is corpus-based and comparative, contrasting the use of metadiscourse by Swedish advanced learners' writing in English to the writing of native speakers of British and American English.[1] All writers are university students. The argumentative essays are full-length, and are available within the framework of the International Corpus of Learner English (Granger 1993). In case there may be cultural differences in the use of metadiscourse, the British English and American English parts of the control corpus are kept separate.

One of the aims of my thesis is to investigate metadiscoursal patterns with explicit reference to the writer or reader. Some examples of metadiscourse specify discourse acts that the writer intends to perform are, for instance:

> (1a) to introduce a topic or state an aim: *In this essay I will discuss some of the problem...; We must now consider the pros and cons of Britain joining "The Single Market".*
> (1b) to sum up a discussion: *To make a short summary of what I have been trying to say in this essay...; I have presented some of the most important benefits of drug legalization…*
> (1c) to close the topic: *…one may conclude that the American people does not approve of political leaders with low moral standards; The conclusion one might draw is rather depressing…*

I will first discuss the definition of metadiscourse and how I delimit the area. This is not unproblematic and metadiscourse has been defined differently by different researchers (see e g Markkanen et al 1993 and Mauranen 1993). The focus of the paper is on a subcategory within the broader field of metadiscourse which will be referred to as 'metatext'. Metatext is text about the current text, and includes references to the evolving text itself rather than its subject matter. Since I am concerned with personalised examples here, this means linguistic elements through which the writer comments on his or her own discourse actions (see examples (1a-c) above). The method used in this work will only capture personalised types of metadiscourse, where the writer and/or reader have been explicitly mentioned in the texts, including expressions such as *as I showed above* and *I will give an example,* but not *as shown above, to exemplify* or *this essay will show*, etc. First person pronouns *I* and *we*, which are

---

[1] The native speaker corpus, called LOCNESS also has one British English part consisting of A-level essays, which has been excluded from the investigation. The Swedish subcorpus will be referred to as SWICLE in the following. Examples will be marked either (Swicle), (BrE) or (AmE).

potential explicit references to the writer and/or the reader, have been retrieved from the corpus material and will be considered in the present paper.

Further questions that will be posed within the span of the paper are: Who do the pronouns *I* and *we* include? How frequent are metatext units across corpora? What do they look like? and How are the metatext units distributed in the texts? The comparative perspective involves investigating the differences and similarities in the use of metadiscourse between learners and native speakers. It should be stressed that this is work in progress, so no definite analyses will be given.

2. The definition and delimitation of metatext

All uses of *I* or *we* are not metatextual. *I* also occurs when the writer appears in the text to express his or her feelings, or to talk about personal experiences. I call this category involvement, and do not include it in the concept of metadiscourse. The way the term is used here is somewhat different from Chafe's (1982) definition. Chafe refers to involvement with the audience as typical for a speaker, and 'detachment' from the audience as typical for a writer, and the concept of involvement includes a range of linguistic features that are more prevalent in speaking[2]. Involvement in the sense it is used in this paper adds a narrative quality, or the writer as a person in the discourse-external world adds some personal experience to the discourse, or expresses his or her personal feelings or attitudes towards phenomena in the 'real world'. Involvement including *we* stresses shared personal experiences in discourse-external situations. One would not expect a great deal of involvement in argumentative essays, but this type is extremely frequent in the learner material.

I make a distinction between two subcategories within metadiscourse, which are metatext and commentary. Commentary has to do with writer-reader relations. It refers to features used to address readers directly, and draw them into an implicit dialogue (Vande Kopple 1988). I plan to carry out a future study, looking at the following commentary features: second person pronoun *you* (*your*), questions, exclamations, and imperative forms, and possibly also lexical items such as *dear reader* (vocatives). Speaker attitude is not dealt with here, although it may have to do with the writer-reader relationship and especially with persuasion. In integrative approaches to metadiscourse (e g Markkanen et al 1993), however, the writer's attitude to what is said is included.

In order to distinguish metatext and commentary from other uses of the pronouns, I also make a distinction between (a) whether the action takes place within the world of discourse, or (b) within object language, dealing with the 'real world'. This can be compared to the distinction between metalanguage and object language in linguistics, an idea which originates with Roman Jakobson (see for instance 1980:81-92). Jakobson described metalanguage as a language in which we speak about the verbal code itself. It was contrasted to object language, in which we deal with objects or items that are external to language itself. Similarly, in a text perspective, we can speak of discourse-internal versus discourse-external phenomena. Thus, one basic question I examine when analysing the data is: Is the ongoing discourse in focus or other, 'worldly', activities that are external to the text?

The notion of *current* text (cf Mauranen 1993) is also important to metadiscourse, meaning that we are interested in how texts refer to themselves, and not to other texts. Texts about other texts (alluding to or commenting on other texts) are described by the concept of intertextuality, which is not our concern here. This is also the reason why quote markers, or introductions to reported speech, which are often included in models of metadiscourse, will be disregarded in the present study. Reporting the speech of others or quoting from other sources than the current text are intertextual rather than metatextual activities.

Criteria or parameters for metadiscourse are generally not specified by researchers; it is usually simply referred to as 'text about text'. However, I have extracted features, which work well for the analysis of the type of personalised, explicit types of metadiscourse present in my collection of corpus data. When looking at an occurrence of *I*, I decide whether it refers to the writer of the current text, or to an experiencing subject outside the realm of discourse. When the examples contain *we*, the question to ask is whether the pronoun includes [+writer persona] and [+reader persona], and not some other group that does not directly participate in the ongoing discourse. These parameters are summarised in figure 1 below. To a great extent, the parameters concern rhetorical roles taken on by the writer and offered to the reader.

---

[2] These features are for example first and second person reference; the presence of the speaker's mental processes; monitoring by the speaker of the communication channel; emphatic particles that express enthusiastic involvement in what is being said, like *just* and *really*; vagueness and hedges; and direct quotes (see Chafe 1982).

Fig. 1. Parameters for metatext.

Metatext
[+current discourse]
[+world of discourse]
[+writer persona] (*I, we*)
[+reader persona] (*we*)

In order to analyse an element as metatext, we need to ask whether it refers to the current discourse, whether it deals with the world of discourse, and, more specifically, concerning first person singular, whether *I* refers to the writer presenting him- or herself as a writer. Concerning first person plural *we*, it should include the writer persona and/or the reader persona, and no other persons that do not belong in the world of discourse.

3. Explicit references to the writer and reader: What do the pronouns refer to?
The first and second pronouns are most typically used to refer to specific individuals identified in the situation of communication (Biber et al 1999:328). They may point to 'the one who is speaking' and 'the one(s) who is/are being addressed'" (Wales 1996:51). These pronouns help us to capture the current writer and the current reader of the current text. In my search for metadiscourse, I look for what Quirk et al (1985:347) call 'specific reference', i e when the first and second person pronouns are used to refer to the writer(s) and the reader(s), those directly involved in the discourse situation.

The first person singular pronoun *I* is "unambiguous in referring to the speaker/writer" (Biber et al 1999:329), in contrast to *we*, which often involves a "fluidity and ambivalence of meaning" (Wales 1996:58). In my data, first person singular *I* points to the writer of the current text in the majority of cases, except when it occurs in quoted material or reported speech, or for example when the pronoun *it* has been misprinted and lost its last letter.

In order to classify an example as metadiscourse, not only has the *I* to point to the writer of the current text, but the action that the *I* performs has to be discourse-internal. The verb *discuss*, for example, involves a speech act which is quite frequently performed in argumentative essays. In the first example, however, the action does not take place within the realm of the current discourse:

> (2) One does not need to read the papers to notice how the antagonism towards immigrants has increased. Lately *I have discussed* the increasing hostility towards immigrants with my friends, relatives and fellow workers. Almost everyone I spoke to wants tougher immigration rules. (Swicle)

The discussion in example (2) is held between the writer as a person in 'real life' and text-external individuals and it counts as involvement. Another example including the verbs *discuss* and *analyse*, which does concern the writer and the reader is the following, classified as metatext:

> (3) [] and secondly that the executive had been strengthened vis-à-vis the parliament. *I will briefly discuss* the Prime Ministers role and then elaborate on the Presidents functions. Then *I will analyse* each presidency showing how the presidents role has evolved. (BrE)

Examples (10)-(16) in section 5 give further examples of metatext having *I* as subject.

3.1 *We* and metatext
Let us consider some examples of the more complex first person pronoun *we*. Most occurrences in my material are not metadiscoursal, but relate to discourse-external phenomena. There is a fairly small number of examples that refers to the writer (but not as a writer) and other persons (who are not the readers of the current text), as in the following example:

> (4) [], but 2 a.m. struck and he bad [sic] to go. We talked some more in the lobby but *we had to keep* our voices down, out of respect. I wanted him to spend the night because [] (AmE)

Examples of this kind have been left out. What I am primarily interested in is the type that has been called 'inclusive authorial *we*' (Quirk et al 1985:350). A small part of the overall material is of this kind (see section 4), for example:

(5) students ask themselves the question; is it worth the cost and the great effort it takes to study at the university? Therefore *we are now going to look at* some advantages of higher education, both in a short-term perspective and also over a longer period (Swicle)

The effect of the fact that both the writer and the reader are addressed in this example is that co-operation is emphasised. The writer is showing his or her helpfulness and will to guide the reader in the discourse, thereby bonding with the reader, as it were.

One subgroup of the so-called 'exclusive *we*', according to Wales (1996) is the collective *we*, indicating several writers. This type does not occur in the selected material, since there are only single writers of the essays. A second subgroup is 'editorial *we*', which is restricted to very few occurrences in the present material, for example "In the course of this essay, *we shall attempt to analyse* whether this is a belief founded in reality and, if it is, why it should cause such fear." (BrE). Here, the single author uses the plural form for his own discourse actions. Quirk et al (1985:350) explain the motivation for using this type as a "desire to avoid *I*, which may be felt to be somewhat egotistical".

Another special usage of *we* is found in several of the literary essays of the British English material, where *we* is equivalent to 'we the audience', as in the following extract:

(6) Through Caligula's own dialogue, and through the opinions of Cherea, *we are shown that* there is a logic to Caligula's approach. Caligula insists on following this logic through to the completion (BrE)

This reference is made explicit in another instance, through apposition: "*We the audience can identify with* her view" (BrE). These types have been omitted. Since the reader is excluded, it could be argued that these are not cases of metatext. Before leaving this category, I will bring up one more example, which is intertextual (*throughout the play*) rather than metatextual, discussing how readers of another text might interpret that other text:

(7) In the play 'Caligula' Camus is dealing with the themes of death and the absurd, and throughout the play *we can see* different characters reacting in different ways. (BrE)

The criterion that metatext instances should refer to the current text is not fulfilled here, but the writer is referring to another text, which belongs to a different genre even. Also, unless the writer assumes that the reader of his essay has read (or seen) the play, the *we* does not include the reader of the current text. Several examples of 'we the audience' type can be found in the British English material (as many as 75 out of a total of 262, which is nearly 30 per cent), but none at all in the other corpora. In this case, the use of first person plural *we* is probably different in the British English material due to the presence of a large number of literary argumentative essays, illustrating how different topics may influence linguistic features.

A number of examples are ambiguous, for example: "[] However this system is once again proved wrong as Cacambo, Candide's manservant remains faithful disproving Martin's theory that Cacambo would run off with the money collected from Elderado. *We can therefore assume* that Voltaire is attacking Optimism in its context of a system, just as he criticises other systems in the book, such as the Church system, the military system and the caste system." (BrE). Here, does *we* refer to the readers of *Candide* or to the writer and reader of the current text? Probably the latter, since it is possible to insert [from what I have said so far] in connection to the instance.

As in the *we the audience* example, it is fairly often the case in the material that the pronoun is followed by a specification, which makes it easier to identify what it is intended to refer to. The 'generic *we*', which is by far the most frequent type in the SWICLE and American English material, is sometimes modified by prepositional phrases and apposition: *in Sweden/in the US/all/people/women*. In some cases, a subsection of the whole is referred to by the addition of *as*-phrases, e g *as individuals/as borrowers*. This type has also been called 'rhetorical *we*' (used in a collective sense: e g nation, party, see Quirk et al (ibid.)), and is particularly frequent in the SWICLE essays, presupposing solidarity with the reader.

In example (8) below, the persons the writer is talking about are 'people in general' (also specified by *most of us*). This includes the writer him- or herself, but not in the role of a writer but rather as an experiencing person in the world. Moreover, other people than the reader are included.

(8) Most of us are rather selfish and *if we are frank* we have to confess that we can not compare ourselves with Mother Theresa when unselfishness is considered. (Swicle)

It can be argued that this example is not metadiscoursal. It is still fairly persuasive, and it seems as if the writer very much would like to include the reader in the group of selfish mortals, but the *most of* still leaves some freedom to the reader to decide for him- or herself whether the description fits or not.

This type of 'freedom' given to the reader, however, is rare. As Clark and Ivanic (1997:165) have pointed out, "[i]n building the dialogue with readers, writers in all genres often take for granted that readers are going to share their beliefs and values [], for example, [] by using the pronoun 'we'. In this way they position their readers as consenting, part of an 'usness' that is hard to resist []". This is particularly true in the case of rhetorical *we*. The 'usness' is particularly hard to resist in a phrase such as *as we all know,* which includes everyone, also the reader. These examples have been classified as commentary.

Interestingly, one example in the material seems to contain some awareness of the fact that *we* can be quite powerful in its inclusiveness:

(9) So, in the immediate future eastern Europe is an enormous market for consumers. At the same time *"we" receive* alarm reports from all over the world we have millions of people longing for things we have taken for granted. (Swicle)

The pronoun is put within quotation marks, which seems to have the effect of hedging the statement made by the writer. Thus, the reader is also made aware that the all-inclusiveness of *we* can be questioned.

4. Frequencies of pronouns and metatext units

There are huge differences in the overall frequencies between the three groups, in particular concerning *I*. What is most striking is that the Swedish learner material has overwhelmingly higher frequencies of first person pronouns.

Table 1. The overall frequency of *I* across corpora

Corpus	Approx. number of words in corpus	Raw frequency	Frequency per 10,000 words
SWICLE	204,630	1,851	90
LOCNESS BrE	95,508	83	9
LOCNESS AmE	149,767	649	43

Table 2. The overall frequency of *we* across corpora

Corpus	Approx. number of words in corpus	Raw frequency	Frequency per 10,000 words
SWICLE	204,630	1,893	93
LOCNESS BrE	95,508	262	27
LOCNESS AmE	149,767	426	28

We is the most preferred first person pronoun, except in the American English corpus, which uses the more individualistic *I* to a greater extent.[3] The preference in the British English data for *we* over *I* (27 versus 9) is not in accordance with the results of a study by Vassileva (1998:167) on academic writing, who found that "the 'I' perspective clearly dominates in English" in comparison to the 'we' perspective. The main part of her English corpus consists of research articles written by speakers of the British English variety. The reason why academic writing and the British English argumentative essays differ remains to be solved.

The overuse of *I* by Swedish learners in contrast to native speakers has been pointed out by Altenberg (1997) and by Ringbom (1996, as reported in Altenberg 1997:127). Previous research on learner writing has also found that learner writers within the ICLE project generally are much more overtly present within the discourse than native speaker writers (Altenberg 1997 and Petch-Tyson 1998), suggesting that this may be a general learner strategy.

---

[3] However, first person plural forms are most frequent in the American English data in the accusative and possessive forms.

The table below reveals the great difference in the number of metatext units per 100,000 words including *I* across corpora, particularly in the SWICLE in relation to the native speaker corpora.

Table 3. The frequency of *I* involved in metatext units compared across corpora

Corpus	Approx. number of words in corpus	Raw frequency	Frequency per 100,000 words
SWICLE	204,630	275	134
LOCNESS BrE	95,508	20	21
LOCNESS AmE	149,767	55	37

These results raise several questions, for instance whether the proportions are similar across corpora also for impersonal types of metadiscourse, or if the great preponderance in the learner material is due to the general tendency among Swedish learners (as evidenced in this corpus) to use writer (and reader) visibility in their writing. In other words, do the data reflect the general picture regarding proportions in the use of metadiscourse, or are they dependent on the overuse of *I* in the SWICLE, noted above?

Table 4. The frequency of *we* involved in metatext units compared across corpora

Corpus	Approx. number of words in corpus	Raw frequency	Frequency per 100,000 words
SWICLE	204,630	57	28
LOCNESS BrE	95,508	26	27
LOCNESS AmE	149,767	25	17

In the British English material, there are more metatext units involving *we* than *I*, whereas the Swedish learner essays (in particular) and the American English material have the greatest figures in connection to *I*.

The Longman grammar (Biber et al 1999:334) comments on the 'unexpectedly high frequency' of *we/us* in their categories of news and academic prose as being "connected with the multiple uses of this pronoun in written prose: to make generalisations, and to refer to the author and reader". When comparing the overall figures for *we* and the frequencies of *we* involved in metatext elements, we can see that the latter use is far less frequent in the present argumentative essays.

There is a great difference in the learner material, where the metatext units involving *we* only amount to about a fifth of the units including *I* (57 versus 274). The overall occurrences of the pronouns *I* and *we* in the SWICLE, however, are roughly the same, even with a slight predominance of *we*. One conclusion to draw from this, is that the Swedish learners, in their use of personalised metatext, very much prefer the pronoun *I* to *we*. The native speakers, on the other hand, use the pronouns more or less to the same extent. The British English material has slightly more occurrences of *we* (26 instead of 20 in relation to *I*), whereas the American English material about twice as many metatext units including *I* (55, versus 25 for *we* occurrences).

If we take normalised figures into account, it is still clear that the Swedish learners overuse personalised metatext, particularly with first person singular *I* as subject, provided that we regard the native speaker data as the norm.

The results also show that there is a difference between the American and British English components of the LOCNESS corpus. The data suggest that writers in the American English variety use more metadiscourse (at least metatext) in their argumentative writing compared to writers who use the British English variety. It may be the case that writing conventions concerning metadiscourse differ in the two cultures, in which case Swedish learners would be more inclined towards (or influenced by?) the American English style, and a lot less so towards the British English.

5. What do the metatext units look like?
In this section, I will briefly bring up one characteristic of the learner essays that seems to me to be important. In many respects, the learner essays are more tentative, hedged, and polite in their use of metatext compared to the native speaker examples.[4] For example, *try* and *would like to* are very

---

[4] This may also partly depend on cultural differences in writing. Vassileva (1998:168), studying English, German, French, Russian and Bulgarian academic writing, notes that "[i]n German, the author's intentions are, more often than not, mitigated (downtoned, hedged) by means of modal verbs, predominantly *ich möchte* 'I would like' []". Writing in Swedish may also follow this 'mitigating' convention, which may be transferred by the learners to their

common in the learners' metatext units. *Try* is the most frequent verb after *say*, but it does not occur at all in the native speaker material. Examples including *I* are the following:

(10) In this essay *I will try to give* some examples of what imagination and dreaming have come to today with some references to the past, and *I will also try to answer* the question as to wether there is a place for imagination and dreaming in our modern world dominated by technology and science, or not. (Swicle)

(11) Should we accept all foreign cultures or discriminate some of them? In what way would our defence of other forms of Swedish culture and our attitude to foreigners and foreign things? *I will try to answer* these questions, as well as I can, but first *I will try to describe* the situation of the Swedish language today. (Swicle)

(12) To make a short summary of what *I have been trying to say* in this essay, technology will never make imagination and dreams unnecessary for two main reasons. (Swicle)

The first two examples announce what is going to follow in the discourse[5]: in (10) the writer will perform the discourse act of giving examples and answer an important topical question, and in (11) answer topical questions and describe a situation to the reader. These examples are placed at the beginnings of the essays they occur in.

Instead of giving anaphoric reference, example (12) points backwards in the discourse[6] and the writer announces that a summary of the main points that have been made in the previous discourse is to come. The *what I have been trying to say* gives a modest and even insecure impression, as if the writer doubts his or her ability to get his or her ideas across in writing.

Also note the occurrence of the phrase *in this essay* in both (10) and (12), which reflects the fact that essays in the learner material generally have a high degree of reflexivity, or awareness of the discourse as discourse. This is expressed particularly often in the first paragraphs, as in: "This paper is supposed to deal with immigrants coming to Sweden. To be more précis *I am going to write about* refugees, the biggest group of people who come here today." (Swicle). This helps explain why the verb *choose* is fairly frequent in the metatext units in the learner material; the writers explicitly comment on the choice of topic for the essay.

The polite formula *I would like to* also has high frequencies in the learner texts. Example (13) below introduces the topic and tells the reader what the writer is going to focus on in the essay. Note the extreme situation awareness in the adverbs *in my essay today*.[7]

(13) One of Anita Brookner's books is called 'A friend from England'. I read it last autumn, so I might have forgotten some names. If so, I've got to re-name them. In my essay today *I would like to concentrate on* the relationship between Heather and Rachel. I will also discuss in what way this was an important personal relationship. (Swicle)

The insecure and hedged quality of many learner texts is also noticeable here, in that the writer explicitly comments on the fact that he or she may have forgotten some names in the novel he or she is about to discuss. However, the solution to this problem, to make up new names, is also brought up. Another example involving *I would like to* is the following:

(14) To become a dog owner is one of the greatest gifts in life. I will support this statement by discussing three positive effects. *I would like to call* the first one

---

writing in English. In this contrastive study on authorial presence by means of first person pronouns, Vassileva found different cultural norms.

[5] This type has been called Announcement (Crismore et al 1993), preview (Crismore and Farnsworth 1990) and Advance Labelling (Tadros 1993).

[6] Types in which the writer tells the reader what he or she has already done in the discourse have been called Reminders (Crismore et al 1993), reviews (Crismore and Farnsworth 1990) and Recapitulation (Tadros 1993).

[7] Petch-Tyson (1998:112) examines the subcategory 'reference to situation of writing/reading' within the broader concept of writer/reader visibility in texts, including *here*, *now* and *this essay*. In a comparison of samples from the Dutch, Finnish, French and Swedish learner subcorpora within the ICLE project, both *here* and *this essay* were used about twice as often by the Swedish learners.

friendship. It is said that dogs are our best friends and I think that there lies a great deal of truth behind this statement. (Swicle)

In this example, the writer labels the first of three effects which are important supporting arguments for the writer's thesis that a dog is one of the greatest gifts in life. Naturally, the writer could have skipped the metatext element marked by italics above, and been content by just having 'The first one is friendship'. The writer could also have chosen to express more certainty by using modal *will* instead of *would like to*.

(15) is another example of *I would like to*, announcing to the reader that some examples supporting the writer's argument are to come:

(15) In this essay, which is of course wholly unscholarly, *I would like to point out* some examples, mostly by female writers where woman's place in society - or rather woman's possibility of controlling her won social position - is illustrated by the actions of the "archetypal English literary heroine". (Swicle)

The writer explicitly makes a point of the notion that his or her essay only provides the layperson's view on the topic (*in this essay, which is of course wholly unscholarly*), thus telling the reader something about how to interpret the text. The writer is careful not to present him- or herself as an expert in the field. This could be interpreted as if the learner writer does not feel confident in the role of an argumentative essay writer or with the possible objectivity and learnedness that traditionally accompany this task. A similar example is the following:

(16) Of course, you may wonder what the true reason for heavy taxation on these pleasures is. Is it really a concern about the average Swede's health and well being or is it, as many suspect, a concern about the government finances that is the reason? *I have not the competence required to answer this question*, but if you ask me if I think that taxes and restrictions are the right methods to keep people off the bottle and away from the cigarette, the answer is no. (Swicle)

A question is posed, but the writer chooses not to answer it on the grounds that he or she is not an authority on the area. In the following coordinated clause, however, the question is rephrased so that the writer feels that he or she rightly can answer it. Note the interactive expression *if you ask me*, which would also be classified as metatext.

6. How are the metatext units distributed in the texts?
The program WordSmith Tools (http://www.liv.ac.uk/~ms2928/index.htm) has been used to study the distribution of metatext elements. It gives the percentage of how far into the text file the search word occurs. The spread of the search word is examined in each text, starting at 1 and ending at 100 per cent. All percentages have been compiled and divided into tenths (1-10, 11-20, 21-30, etc). Measuring textual distribution in terms of percentages and not paragraphs can be criticised in several ways, but looking at the figures and tables in terms of percentages still gives a good picture of the spread of the metatext units through entire collections of texts.[8]

In the Swedish learner corpus (see diagram 6a below), there is a clear peak in metatextual units including first person *I* in the first parts of the essays. As many as 34 per cent of the overall number of metatext occur from 1 to 20 per cent into the essays. A certain rise is also noticeable at the very end, in the last percentage section.

There are few occurrences of metatext having *I* as subject in the British English material, and the great majority of units occurs in the very first section (see diagram 6c below). Although the data are scarce, the overall picture is similar to the distribution in the Swedish learner essays. In the American English material (diagram 6b), on the other hand, the highest peak is found in the last section, 91-100 per cent into the texts, but also in the first and third sections.

As expected, the results show that the beginnings and the ends of the essays are important sections for metadiscourse. Further analysis will reveal whether there are patterns in the use of individual verbs, for example whether some verbs have a preference for occurring at a certain point in argumentative essays. Thus, questions to be answered are: Which of the metatextual verbs occur
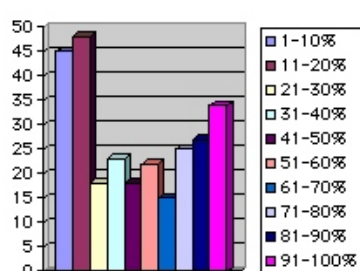
---

[8] However, it may be the case, for instance, that paragraph length differs across corpora, which could influence the figures.
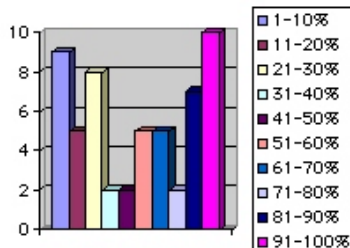
most frequently at the beginning or at the end of texts? Why is that? What do the forms look like? What functions do they have in the texts?
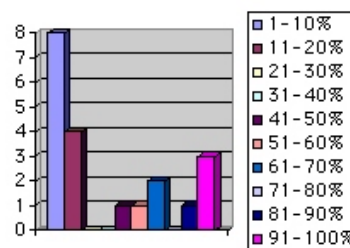
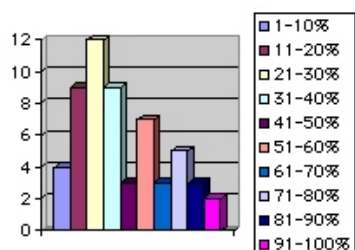*Diagrams 6 a-c and 7 a-c*. The distribution of metatext units involving *I* and *we* in the three corpora.

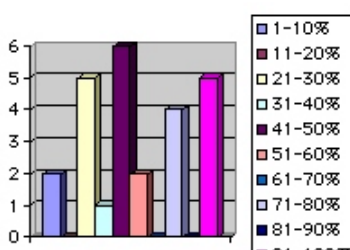*Diagram 6a*. The distribution of occurrences of *I* involving metatext in SWICLE

Diagram 6b. The distribution of occurrences of *I* involving metatext in US LOCNESS
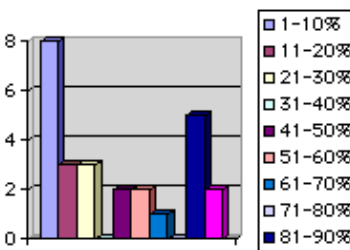
Diagram 6c. The distribution of occurrences of *I* involving metatext in UK LOCNESS

Diagram 7a. The distribution of occurrences of *we* involving metatext in SWICLE

Diagram 7b. The distribution of occurrences of *we* involving metatext in US LOCNESS

Diagram 7c. The distribution of occurrences of *we* involving metatext in UK LOCNESS

If we compare the patterns of distribution of metatext involving *I* and *we* in the individual corpora, the neatest match is in the British English data, in which the highest numbers are in the very first section (8 occurrences in both cases, see diagrams 6c and 7c above).

The figures in the American English data are only partly similar (diagrams 6b and 7b). There is a peak at the end for both pronouns, but it starts already at 81-90% for *I*, and the numbers are twice as high. The increase of metatext including *I* at the beginning do not correspond to a rise in the results including *we*. The top score for metatext involving *we*, instead, occurs at 41-50 per cent.

The distribution of metatext with *I* and *we* as subjects is basically reversed in the SWICLE material (diagrams 6a and 7a). In the case of *I*, there are two evident tops at the very beginning and at the very end of the essays. In the data involving *we*, on the other hand, the lowest numbers occur in those places, and the rise takes place at 21-30 and 31-40 per cent.

7. Conclusion

Some evidence that the three groups represented differ both quantitatively and qualitatively in their use of metatext as well as in their overall use of first person pronouns has been presented in this paper. Any definite answer to the reason why the differences between the corpora exist – in particular with regard to the learner essays versus the native speaker texts – will have to be left until the material has been analysed in detail.

However, one partial and tentative answer may be that the differences are related to different cultural conventions for writing. First of all, there is a possibility that the conventions for using metatext in argumentative writing are not the same in Swedish and English. In addition, there seem to be differences with regard to writer visibility and use of metatext among the two varieties of English in the native speaker corpus. For example, there is a clear dissimilarity in the use of *I* in the American and British English parts (43 versus 9 occurrences per 10,000 words, see table 1 above). This suggests that there may be cultural differences involved, and the division of the LOCNESS corpus into two parts made for the present investigation is endorsed.

To some extent, the presence of the discourse participants in texts and the use of pronouns are also a question of politics (see Wales 1996:84). The critical discourse perspective (see e g

Fairclough 1995) deals very much with ideology in texts, which may be done for example by asking questions such as: Is the agent behind the text visible? Is the position taken by the writer explicitly stated as belonging to the writer or is it rendered as a general truth or common knowledge? In this field, linguistic forms and structures are studied in terms of power and ideology, such as the agentless passive construction which, in contrast to personalised expressions, leaves causality and agency unclear – possibly with the aim of conscious hedging or even deception.

Acknowledgement
I would like to thank professor Karin Aijmer for reading and commenting on my paper.

References
Altenberg B 1997 Exploring the Swedish component of the International Corpus of Learner English. In Lewandowska-Tomaszcyk B, Melia P J (eds) *PALC'97: Practical applications in language corpora.* Lódz, Lódz University Press, pp 119-132.
Bäcklund I 1998 *Metatext in professional writing: a contrastive study of English, German and Swedish.* Texts in European Writing Communities 3. Uppsala, Uppsala Universitet.
Biber D, Johansson S, Leech G, Conrad S, Finegan E 1999 *The Longman grammar of spoken and written English.* London, Longman.
Bondi M 1999 *English across genres: language variation in the discourse of economics.* Modena, Edizioni Il Fiorino.
Chafe W L 1982 Integration and involvement in speaking, writing, and oral literature. In Tannen D (ed) *Spoken and written language: exploring orality and literacy.* Norwood, Ablex, pp 35-53.
Clark R, Ivanic R 1997 *The politics of writing.* London and New York, Routledge.
Connor U 1996 *Contrastive rhetoric: cross-cultural aspects of second-language writing.* Cambridge, Cambridge University Press.
Crismore A, Farnsworth R 1990 Metadiscourse in popular and professional science discourse. In Nash W (ed) *The writing scholar: studies in academic discourse.* Written Communication Annual: An international survey of research and theory: Volume 3. Newbury Park/London/New Delhi, Sage Publications, pp 118-136.
Crismore A, Markkanen R, Steffensen M S 1993 Metadiscourse in persuasive writing: a study of texts written by American and Finnish university students. *Written Communication* 10: 39-71.
Fairclough N 1995 *Critical discourse analysis: the critical study of language.* London and New York, Longman.
Granger S 1993 The International Corpus of Learner English. In Aarts J, de Haan P, Oostdijk N (eds) *English language corpora: design, analysis and exploitation.* Amsterdam, Rodopi, pp 51-71.
Hyland K 1998 Persuasion and context: the pragmatics of academic discourse. *Journal of Pragmatics* 30: 437-455.
Jakobson R 1980 *The framework of language.* Michigan, Michigan Studies in the Humanities.
Markkanen R, Steffensen M , Crismore A 1993 Quantitative Contrastive study of metadiscourse: problems in design and analysis of data. *Papers and Studies in Contrastive Linguistics* 28: 137-152.
Mauranen A 1993 *Cultural differences in academic rhetoric: a textlinguistic study.* Frankfurt am Main, Peter Lang.
Petch-Tyson S 1998 Writer/reader visibility in EFL written discourse. In Granger S (ed) *Learner English on computer.* London and New York, Longman, pp 107-118.
Quirk R, Greenbaum S, Leech G, Svartvik J 1985 *A comprehensive grammar of the English language.* London, Longman.
Taavitsainen I 2000 Metadiscursive practices and the evolution of early English Medical writing 1375-1550. In Kirk J M (ed) *Corpora galore: analyses and techniques in describing English.* Papers from the nineteenth international conference on English language research on computerised corpora (ICAME 1998). Amsterdam and Atlanta, Rodopi, pp 191-207.
Tadros A 1993 The pragmatics of text averral and attribution in academic texts. In Hoey M (ed) *Data, description, discourse: papers on the English language in honour of John McH Sinclair.* London, HarperCollins, pp 98-114.
Vande Kopple W J 1988 Metadiscourse and the recall of modality markers. *Visible Language* XXII: 233-27.
Vassileva I 1998 Who am I/who are we in academic writing?: A contrastive analysis of authorial presence in English, German, French, Russian and Bulgarian. *International Journal of Applied Linguistics,* Vol 8, No 2: 163-190.
Wales K 1996 *Personal pronouns in present-day English.* Cambridge, Cambridge University Press.