MÄLARDALEN UNIVERSITY
SWEDEN
School of Education, Culture
and Communication

A corpus linguistic investigation into patterns of engagement in academic writing in Swedish and English higher education settings

Degree Project in English HEN401

Håkan Almerfors

Supervisor: Olcay Sert

Examiner: Thorsten Schröter

Spring 2018

Title: A corpus linguistic investigation into patterns of engagement in academic writing in Swedish and English higher education settings

Author: Håkan Almerfors

Pages: 37

Abstract

Over the last few decades, the interpersonal dimensions of academic writing have received growing attention in the field of applied linguistics. As an important concept in academic writing, *engagement* has been a topic of interest to reveal how writers interact with readers to, for example, guide reasoning through arguments and to abide by conventions of politeness. Previous research has suggested that the higher students' academic level is, the more similar their use of engagement elements in writing will be. Previous research has also suggested that for non-native speakers, cultural factors as well as interlanguage, influence how engagement features are used in written English. The primary aim of this study was to investigate which engagement patterns could be found in L1 Swedish and L1 English students' academic writing in English, with the focus on linguistics as a subject. Using the methods of corpus linguistics, this project also strove to identify the ways patterns of engagement differed between L1 Swedish and L1 English students, and in what ways patterns of engagement varied between the students at B-levels and C-levels in written English of linguistics studies. The data for the study came from SUSEC, which is a corpus of written English that consists of texts collected at Stockholm University in Sweden and at King's College in England. In line with previous research, the results indicate that the L1 Swedish students use more elements of engagement than the L1 English students. Results also suggest that C-level students use fewer reader pronouns than B-level students, and that Swedish C-level students use more directives than Swedish B-level students. Overall, the comparison of students with two different first languages revealed several differences on how engagement is used, which can serve to inform future research on academic writing.

Contents

List of abbreviations

CIA – Contrastive Interlanguage Analysis

D – Directive

KC – King's College

KC2 – B-level King's College

KC3 – C-level King's College

L1 – First language

L2 – Second language

ptw – per thousand words

Q – Question

RP – Reader pronoun

SU – Stockholm University

SUBLING – B-level Stockholm University

SUCLING – C-level Stockholm University

SUSEC – Stockholm University Student English Corpus

1. Introduction and aim

Over the last decades, there has been a growing interest in applied linguistics for *metadiscourse*. Metadiscourse can be understood as the parts of a text that both refer to itself and interacts with presumed readers. Ädel (2006) has defined metadiscourse as "text about the evolving text, or the writer's explicit commentary on her own ongoing discourse" (p. 20). The parts of metadiscourse which deal with interaction with readers have inspired a focus of research that is often called *engagement*. Engagement can be described as "a reader-oriented aspect of interaction which concerns the degree of rapport which holds between communicative participants" (Hyland & Jiang, 2016, p. 29). Writers tend to use different *engagement elements* to directly communicate with their readers. Examples (1) and (2) are taken from the SUSEC corpus of university student English. They contain examples of two different types of engagement elements.

> (1) However, if **we** adhere to the assumption that expository texts should provide the receiver with factual information, then this word choice is exactly the case of the epistemic marking of a factually uncertain situation. (SUSEC_SUCLING_036)
>
> (2) **It is therefore necessary to** include the plural pronouns as a group of their own since they are not marked for gender in any way. (SUSEC_SUCLING_034)

In example (1), the writer invites the reader to follow a reasoning by using the inclusive *we* as a *reader pronoun*. In (2), the writer tells the reader what to do by using a predicative adjective as a *directive*.

Previous research has shown that the use of linguistic features such as engagement elements varies across academic disciplines (e.g. Biber, 2006; Hyland, 2001). Linguistics, as the scientific study of language it is, has been one of the oft-researched disciplines for engagement. Results from earlier studies (e.g. Hyland, 2005a; Hyland & Jiang, 2016; Myers, 1989) have shown that writers in this academic field tend to engage readers in certain ways. From the results of an investigation of academic articles, Hyland (2001) concluded that the fields of soft sciences, that is, humanities and social sciences, used a lot of reader-oriented elements, explicit references to shared assumptions and inclusive *we* pronouns. He also recognised that questions were more frequent in the soft sciences than in the hard sciences (e.g. physics and biology). In another study, Hyland (2009) concluded that directives were

less common in the soft sciences than in the hard sciences. There is evidence that suggests that cultural differences and cross-linguistic influence can affect engagement patterns. Mauranen (1993) analysed texts written by Finnish and Anglo-American academics. She concluded that, as opposed to the Anglo-American academics, the Finnish academics rarely used reflexive expressions to explain text but rather provided ample background information. Lafuente-Millán (2014) compared business research articles written in English by L1 Spanish and native speakers. He concluded that "Spanish researchers seem to transfer the syntactic rhetorical structures they use in their own language" (Lafuente-Millán, 2014, p. 219). He also pointed out that proficiency in the second language (henceforth L2) and transfer from the native language (henceforth L1) are factors that are likely to affect how academics write. In engagement studies it is common to investigate how writers engage readers to "head-off reader objection and bring them into the text" (Hyland & Jiang, 2016, p. 30). Engagement also fills other functions such as acknowledging counterarguments and achieving politeness. The present study is an investigation into how engagement is expressed in academic English student writing and how it compares between native speakers of English and L1 Swedish learners. Although several comparative studies have been carried out on the subject (e.g. Cao & Hong, 2014; Casal, Hitchcock & Lee, 2017; Mauranen, 1993; Ädel, 2006), these have not focused on engagement alone. The results from this study can therefore fill a gap in research with its exclusive focus on engagement.

The main aim of this study is to investigate which engagement patterns can be found in L1 Swedish and L1 English students' academic writing in English. The investigation is based on a corpus linguistic analysis of texts from the Stockholm University Student English Corpus (SUSEC). The authors of the essays were native English and L1 Swedish students. Additional aims are to find variations in use of engagement between the native English speakers and the Swedish speakers, as well as variations between different levels of studies across and within the different L1s. The research questions are:

- *What are the engagement patterns in the L1 Swedish and L1 English students' academic writing?*

- *In what ways do patterns of engagement vary between the L1 Swedish and L1 English students?*

- *In what ways do patterns of engagement vary between students at different levels in linguistic studies?*

2. Background

In this section, the theoretical framework that this study rests upon will be presented. First the field of metadiscourse is briefly described in section 2.1. This is followed by section 2.2 with a more detailed discussion of engagement, which is the focus of this study. It finishes with section 2.3 where some findings from previous research are presented.

2.1 Metadiscourse

Over the last few decades, *metadiscourse* has become a significant field of research in applied linguistics. Hyland (1998) has suggested that "[m]etadiscourse refers to aspects of a text which explicitly organise the discourse, engage the audience and signal the writer's attitude" (p. 437). It is this interpretation of the term that is used in this paper. An important feature of metadiscourse is that it is "integral to the contexts in which it occurs and is intimately linked to the norms and expectations of particular cultural and professional communities" (Hyland, 1998, p. 437). In the case of academia, this means the writer of a text needs to meet certain pragmatic expectations related to an academic discipline to be able to achieve a successful rhetoric. It is common to divide the field of metadiscourse into two approaches. The first is often called the "reflexive" model and focuses on the organisation of discourse in the text. The other is often called the "integrative" or "interactive" model and brings the analysis to an interpersonal level where the authors' efforts to connect with audiences are also considered (Hyland, 2015). Hyland (2015), who adheres to the latter approach, has written that it "sees metadiscourse as a set of features that together contribute to the interaction between text producers and their texts and between text producers and users" (p. 1). The concept of engagement, which this essay seeks to investigate, is most often discussed in research connected to the latter approach. In this integrative/interactive understanding, the concept of *audience* becomes very important, because it is towards such an entity writers are assumed to direct their rhetorical efforts. The concept of audience will be discussed in section 2.2.1.

Hyland (2015) has suggested a model for investigating metadiscourse in academic texts. In this model, he separates *interactive resources*, which help to guide the reader through the text, from *interactional resources,* which involve the reader in the argument. The former includes categories like frame markers (finally, to conclude), evidentials (according to X) and transitions (in addition, thus). The latter includes categories like hedges (might, possible),

boosters (in fact, it is clear that) and engagement markers (consider, you can see that). Ädel (2006) has described that it is useful in metadiscourse analysis to make a distinction between *metatext* and *writer-reader interaction*. The former involves how the text is organised with regard to discourse and text. The latter deals with direct interaction between writers and readers. Ädel separates *personal* metadiscourse, which refers directly to the reader/writer, from *impersonal* metadiscourse, which refers to writers/readers implicitly. In her interpretation, writer-reader interaction is always personal, while metatext can be both personal and impersonal. The parts of metadiscourse that involve the relation between writer and reader is the focus of research for what is often called *engagement*.

2.2 Engagement

The term *engagement* was coined by Hyland (2001) in an article about how writers address readers in academic articles. It is a focus of research which investigates how writers express their relationship to readers and the purposes this can serve (Hyland & Jiang, 2016). This section is divided into six sub-sections. In sections 2.2.1 and 2.2.2 respectively, the concepts of audience and politeness are discussed in relation to engagement. In section 2.2.3, the concepts of engagement and stance are explained. Next follows section 2.2.4, with some comments on engagement in academic research writing. In section 2.2.5, differences in use of engagement between speakers with different L1s are discussed. The review of engagement finishes with section 2.2.6 where Hyland's framework for investigating engagement is explained.

2.2.1 Audience

The concept of *audience* is of great importance in engagement studies. Bakhtin (1981) has suggested that all writing can be understood as dialogues between writers and their presumed readers. He has claims that there is always a discourse outside the discourse in the text, one where things like public opinion, previous research and common norms reside. Bakhtin also means that this discourse is stratified according to such things as genre and profession. White (2003) has suggested that a writer constructs a textual persona in the process of taking a stand and that this construction is shaped by socio-determined value positions in society. With regard to understanding writing as a dialogue with a reader, Hyland & Jiang (2016) conclude: "To understand writing as dialogic means examining discourse features in terms of the writer's projection of the requirements, perceptions and interests of a potential audience" (p.

30). In other words, Hyland and Jiang mean that writers not only direct their texts towards audiences, but that their rhetoric is defined by the discourse they assume these audiences expect. This type of audience is much like *the authorial audience* that Rabinowitz (1977) has suggested is a hypothetical construct that an author of a narrative directs the text towards.

2.2.2 Politeness

Politeness is an important aspect of engagement. In a discussion about *face*, Brown and Levinson (2006) have suggested that "while the content of face will differ in different cultures . . . we are assuming that the mutual knowledge of members' public self-image, or face, and the social necessity to orient oneself to it in interaction, are universal" (p. 312). In this understanding, a writer must include politeness to succeed in positively interacting with readers. In an academic text, engagement becomes a vehicle for this politeness. An example is how authors instead of asserting themselves strongly, try to bond with the reader. A sentence that starts with a phrase like "[p]erhaps we need to consider", politely invites the reader to participate and approve of what will come after. Myers (1989) describes how this need for politeness is expressed in academic writing:

> [C]laims and denials are indeed taken as possible impositions on the community, that the coining of new terms is also a possible imposition, that speculation must be done but must be apologized for, that hedging reflects social interactions rather than probability of statements, and that direct criticism is almost inadmissible. (p. 30)

In other words, academic texts are written to meet certain expectations when it comes to politeness. Politeness is also an important component of *stance*, which will be discussed in the next section.

2.2.3 Engagement and stance

There is a plethora of terms authors have used to describe how writers express their opinions and standpoints in text. It has been called such things as *evaluation*, *attitude*, *epistemic modality*, *appraisal*, *metadiscourse* and *stance* (Hyland, 2005b). In this paper, I will follow a terminology that is largely based on Hyland's understanding of metadiscourse and engagement, and in line with other studies in the literature (e.g. Hyland & Jiang 2016; Lafuente-Millán 2014; Ädel 2006).

Since the concept of *stance* is closely related to engagement, it is adequate to explain how it can be understood and how it relates to engagement. Stance can be described as expressions that are used by writers to articulate agreement or disagreement with ideas presented in their text (Ädel, 2006). Hyland (2005b) makes a separation between stance and engagement but consider them to be "two sides of the same coin" (p. 176), since they both describe the interpersonal dimension of discourse. He defines stance as how writers present themselves and "convey their judgements, opinions, and commitments" (Hyland, 2005b, p. 176). Hyland (2005b) suggests that engagement "is an alignment dimension where writers acknowledge and connect to others, recognizing the presence of their readers, pulling them along with their argument, focusing their attention, acknowledging their uncertainties, including them as participants, and guiding them to interpretations" (p. 176). He recognises four key resources for academic stance interaction (hedges, boosters, attitude markers and self-mentions) and five key resources for academic engagement interaction (reader pronouns, questions, directives, shared knowledge and personal asides) (Hyland, 2005b). The latter will be explained in more detail in section 2.2.4. Example (3) below illustrates what Hyland would consider a self-mention in an expression of stance.[1]

> (3) **I** will attempt to highlight the power and social influences which surround the articles and their communicative and social functions. (SUSEC_KC3_001)

The pronoun *I* in example (3) communicates the writer's desire to do something rather than to connect with the audience. Example (4) displays what Hyland would consider to be a *reader pronoun* in an expression of engagement because the *we* invites the reader to follow the argument.

> (4) To deal with questions of social literacies and literacy practices **we** need to consider modes. (SUSEC_KC3_003)

Example (5) below illustrates what Hyland would consider a booster in an expression of stance, as well as an appeal to shared knowledge in an expression of engagement.

---

[1] Throughout this study, emphasis in bold face has been added to the examples to point the readers' attention to the expression in focus. Corpus references to text numbers are in brackets (). In the example texts from the corpora, spaces have been removed where they were judged to have been added in the coding process.

(5) It is quite striking that the above modals, although **obviously** falling within the domain of epistemic knowledge, are clear signposts pointing to […] (SUSEC_SUCLING_036)

Obviously in example (5) can be considered a booster because it expresses certainty of the writer. It can also be considered an appeal to shared knowledge because it connects with the reader. This serves as a good example of how in Hyland's system, stance and engagement are two dimensions that often overlap. While stance is how writers position themselves in texts with interaction, engagement is how writers position their readers in interaction. In this essay, the latter dimension will be investigated.

2.2.4 Engagement in academic research writing

In an article about academic research writing, Hyland and Jiang (2016) suggest that "[t]he ability to craft a text which establishes solidarity, or at least a disciplinary affiliation, both supports a writer's community credentials and helps to head-off objections to their arguments" (p. 30). In other words, they propose that writers adapt their writing to certain disciplinary expectations. Regarding academic audiences, Myers (1989) argues that they can be divided into those with an *exoteric* profile, and those with an *esoteric* profile. The former includes a wider research community, which also would include university students, and the latter includes researchers that are involved in the same type of research. With regard to differences in use of stance and engagement between academic disciplines, Hyland (2005b) claims that:

> It is clear that writers in different disciplines represent themselves, their work and their readers in different ways, with those in the humanities and social sciences taking far more explicitly involved and personal positions than those in the science and engineering fields. (p. 187)

Furthermore, he suggests that a characteristic in the soft fields is that academics are "more interpretative and less abstract, producing discourses which often recast knowledge as sympathetic understanding, promoting tolerance in readers through an ethical rather than cognitive progression" (p. 187). In other words, Hyland suggests that engagement features are common in the soft fields and that this in part is dictated by the nature of the disciplines which they include. The next section will discuss in what ways native language and culture can affect use of engagement features.

2.2.5 Language transfer and interlanguage

The use of engagement does not only vary between disciplines. Research has shown that there are differences between how engagement is used in academic English texts produced by native and non-native speakers (e.g. Hyland, 2009; Lafuente-Millán, 2014; Mauranen, 1993). These differences are often connected to the notion that influence from the L1 carries into the L2. This process is often referred to as *cross-linguistic influence* or *language transfer*. Rod Ellis (1994) has defined language transfer as follows: "L1 transfer usually refers to the incorporation of features of the L1 into the knowledge systems of the L2 which the learner is trying to build" (p. 28). *Interlanguage* theory is another established approach in linguistics that can be useful when analysing differences in language use between native speakers and speakers with different L1s. It is based on the idea that in the process of acquiring an L2, a kind of intermediate language is constructed. This learner language is referred to as interlanguage (Ellis, 1997). In interlanguage, features like overuse and underuse are found to be frequent. *Contrastive Interlanguage Analysis* (henceforth CIA) is a method that has often been used to analyse learner corpora (Granger, 2015). Studies that use CIA regularly compare samples of learner language with samples of native language (Granger, 2015). The method has yielded interesting results, some of which are related to engagement. One is that certain features of interpersonal discourse, regardless of L1, tend to be overused used by English learners, which results in an informal writing-style (Gilquin & Paquot, 2008). Such discourse often mimic spoken language and can include features such as directives expressing personal opinions and questions directed towards readers. In (6) is an example of a question from the Swedish subcorpus that exemplifies such informality.

> (6) What would our world today look like if everyone looked the same?
> (SUSEC_SUBLING_019)

The influence of culture and rhetorical structures in the L1 on L2 writing is also a viable approach to understanding how engagement is used by non-native speakers. It will be discussed in relation to previous research in section 2.3.

2.2.6 Hyland's framework for investigating engagement

Hyland (2001) defines two main purposes for writers to use engagement. The first is to meet the reader's expectations of being included in the text. The second is to rhetorically position

the audience so that successful arguments can be made. Hyland's (2005b) five elements of engagement that were introduced in section 2.2.3, will now be discussed in detail.

Reader pronouns (henceforth RPs), like *we* and *our*, are direct and transparent ways for writers to connect with readers. The most frequent RP in academic writing is *we*. More than to achieve solidarity, the inclusive *we* allows space for the opinions of the readers and can guide them through arguments. Hyland (2001) claims that the RPs *you* and *your* are only common in philosophy. Example (7) below shows an inclusive RP.

> (7) However, if **we** assume that working memory does not cause processing differences […] (SUSEC_KC3_031)

Questions (henceforth Qs) are involving readers in the writer's discussion. They can position readers on the same level as participants in a venture to investigate an issue. Research questions are an example of such Qs. Many Qs also have rhetorical purposes, in which case they rather position readers to receive information since a reply is often given right after the question. This also includes questions that for some reason do not have an answer. In (8), the research question has no rhetorical purpose, which invites a reader to participate in the process of solving a problem.

> (8) Are there any social class-and/or-gender differences between the Swedish teenagers' choices of English swearwords? (SUSEC_SUCLING_005)

Example (9) below illustrates a Q whose function is rhetorical because the reader is introduced to a problem but is not allowed to participate in solving it.

> (9) […] how then can we verify the truth value of this information? . . . The only suitable answer to this question seems to be a profound genre analysis […] (SUSEC_SUCLING_036)

In example (10) is a Q with a rhetorical purpose since it is not expected to have an answer in the context it was presented.

> (10) How does a government account for these problems and reverse stigma attached to languages? (SUSEC_KC3_020)

Directives (henceforth Ds) can be understood as instructions to readers to do things or to follow a certain reasoning. They are often signalled by imperatives (such as *imagine* and *compare*), by modals of obligation (such as *must* and *should*) or by predicative adjectives

expressing importance (such as: It is *necessary* to view these results in the light of…). Hyland recognises three types of acts that are incentivised by Ds. *Textual acts* are used to guide the reader to another part of the text or to another text. Example (11) shows an incentive for a textual act.

(11) Because of the large number of words in the resulting database **(see Appendix 1)** (SUSEC_SUBLING_068)

Physical acts are Ds that tell readers how to perform actions in the real world. Example (12) illustrates an incentive for a physical act.

(12) To obtain the best results, **you need to stir the mix vividly**. (This example has been constructed for the purpose of this essay.)

Cognitive acts are Ds that guide the reader through an argument or to an understanding. Example (13) below illustrates how a reader is urged to agree with the writer's idea of how to think about a certain phenomenon.

(13) As gatekeepers of English as a second language, **it is important to provide and filter not only grammar and vocabulary** but also […] (SUSEC_SUBLING_022)

Since cognitive acts incentivised by directives informs the readers of how to think, they run the risk of being perceived as impolite.

Personal asides (henceforth PAs) can be understood as interruptions in the flow of text where readers are addressed directly. As shown in example (14), they often appear within brackets. PAs are used solely with the intent to connect. According to Hyland (2001), they are more frequently used in the soft fields.

(14) […] that the sound of the word **(which in this case is obvious since the words were read out aloud)** does not show if there should be one or two letters in the words. (SUSEC_SUCLING_004)

Appeals to shared knowledge (henceforth ASKs) are when writers ask readers "to recognize something as familiar or accepted" (Hyland, 2005b, p. 184). The readers are not only recognized, they are also attributed a role in the argument. According to Hyland (2001), ASKs are more common in the soft fields. In example (15) it is possible to see that the writer appeals to the readers' knowledge.

(15) **Of course Bernstein was not making his claims in a vacuum** - his assertions were made in the wake of […] (SUSEC_KC3_043)

In the present study, PAs and ASKs were not included. How Hyland's framework will be applied to the present study will be explained in section 3.

2.3 Previous research

Over the last decades, there has been a lot of research published that has investigated reader interaction in academic writing. In this section, some findings from such research will be presented. In section 2.3.1, findings related to engagement in academic writing are presented. This is followed by section 2.3.2, which contains some examples of what previous research on influence from native language and culture suggests.

2.3.1 Academic writing

In a corpus investigation, Hyland (2005a) researched 240 published academic articles as well as interviews with academics. The written corpus consisted of three texts in each of eight academic disciplines from ten leading journals. He found that in the research articles, there were approximately 5.9 cases of engagement elements per thousand words (henceforth ptw). Most common were RPs (a category called *reader reference* in his study) followed by Ds. PAs were most uncommon. The academic discipline that used the highest number of engagement elements was philosophy (16.3 ptw), followed by sociology (5.1 ptw) and applied linguistics (5.0 ptw). These results may confirm the notion, presented in section 2.2.4 above, that engagement is a more common feature in the soft sciences than in the hard sciences. The results from applied linguistics are particularly interesting with regard to the present study since it investigates students of linguistics. The most common engagement elements were Ds (2.0 ptw) and RPs (1.9 ptw), followed in falling order by ASKs (0.6 ptw), Qs (0.5 ptw) and PAs (0.1 ptw). In comparison, academics in marketing used fewer engagement elements in all categories. Even though Hyland's study did not include students in the research, which the present study does, the results are pertinent since it is likely that the more university students progress in the academic ranks, the more their writing will become similar to how professional academics in their field write.

In a cross-disciplinary study of stance markers, Aull & Lancaster (2014) compared the writing of first-year US university students to the writing of upper-level undergraduate US students

11

and published academics. Their corpus consisted of over 4,000 argumentative essays. They found "that texts produced by the most advanced writers were marked by a more dialogically expansive stance" (Aull & Lancaster, 2014, p. 173). They concluded that there was a tendency for student writers to develop more awareness for metadiscourse strategies the higher in the academic ranks they climb. From the results a research of graded papers, Lancaster (2016), suggested that the differences he identified between how undergraduate students of business and students of political theory used stance, were likely to be related to the specific challenges of their assignments. The business students were more focused on convincing readers of their analytical skill, while the political theory students were more focused on guiding the reader through reasoning. These results suggest that different fields produce different challenges, which in turn makes writers apply interpersonal discourse strategies such as engagement to meet these demands.

In a corpus investigation, Hyland (2002) compared research articles, textbooks and reports written in English by undergraduate Chinese students. He concluded that there were considerable variations in the use of Ds over different genres. The results also showed that the student reports contained fewer Ds than the textbooks and the research articles did, especially regarding imperatives. Hyland suggested this was due to cultural norms which made the students avoid claiming authority since it would risk provoking supervisors.

2.3.2 Influence from native language and culture

In a comparative research of 643 essays from the *International Corpus of Learner English* (ICLE), Ädel (2006) investigated metadiscourse. She found that essays written by Swedish students contained more writer-reader interaction than essays written by Anglo-American and British students. This was particularly the case regarding RPs such as *we* and *you*. These results suggest that we can anticipate a more frequent use of RPs in the Swedish corpus than in the British corpus. Another interesting finding in Ädel's study was that the Anglo-American students used more personal metadiscourse than the British students, but less than the Swedish students. On a general note, Ädel concluded that the L1 English students were more fact-oriented than the Swedish students, which were more audience-oriented. She suggested that possible explanations for the differences in use of metadiscourse could be lack of access to secondary sources, as well "as genre differences, lack of register awareness, learner strategies and differing cultural conventions for writing" (Ädel, 2006, p.95).

In a study of reader engagement by Lafuente-Millán (2014), three corpora were compared. One consisted of research articles written by native English speakers (henceforth ENG), a second of research articles written in English by Spanish speakers (henceforth SPENG), and a third of research articles written in Spanish by native Spanish speakers (henceforth SPA). The categories researched were RPs (in Lafuente-Millán's study called *inclusive markers*), Ds and Qs. ASKs and APs were not investigated. In summary, the results showed that RPs were in more frequent use among the Spanish speakers, especially when writing in Spanish. They also showed that the SPENG corpus contained considerably more Ds than the ENG and the SP corpora. Lafuente-Millán (2014) suggested that the distribution of results regarding RPs possibly had its explanation in cultural differences in politeness strategies and language transfer. He wrote:

> By giving priority to certain involvement strategies, such as the use of inclusive pronouns, Spanish BM (Business Management, explanation added) authors publishing internationally tend to rely on the rhetorical styles and interpersonal tactics preferred in their own cultures. (Lafuente-Millán, 2014, p. 219)

Regarding the high frequency of Ds in the SPENG corpus, Lafuente-Millán (2014) claimed that "for some reason, Spanish scholars in BM feel inclined to stress important ideas and concepts, useful specifications and recommendations, relevant results and contributions, as well as difficulties or limitations" (p. 216). In that light, the high frequency of Ds in the SPENG corpus may be connected to the tendency for learners to overuse interpersonal discourse that was suggested by Gilquin and Paquot (2008) (see section 2.2.5).

In a comparative corpus analysis of academic writing, Mauranen (1993) studied how Finnish and Anglo-American academics structured their texts and used reflexivity. The results showed that Finnish academics used fewer reflexive expressions than the Anglo-Americans and instead provided plenty of background information. She suggested that the main reason for this difference in rhetorical structure was that the Finnish academics avoided asserting themselves due to cultural norms. Mauranen (1993) claimed that "[a]uthoritativeness seems open to both a positive and a negative interpretation, and Finnish writers have often expressed a negative view, seeing authority as patronising and constraining" (p. 166). Similarly, from the results of a corpus investigation of Chinese undergraduate writing in English, Hyland (2009) suggested that the use of engagement in the corpus had been influenced by the Chinese culture, which puts emphasis on respecting authority and to not insult face.

Even if the research presented in this section has presented some interesting findings, there are several research gaps that can be filled. This is probably due to the fact that engagement is a relatively recent focus of linguistic studies. Hyland has included Chinese students in some of his research, but he has not compared student writers with different L1s. Just as in the present study, Ädel compared L1 Swedish to native English speakers, but her focus was not on engagement alone. This was also the case with Mauranen, Aull and Lancaster and Lancaster. Lafuente-Millán did focus on engagement and included writers with different L1s, but his research was on published academics. It is possible to conclude that none of these studies, at least to any great extent, have investigated how the use of engagement features in writing compares between university students with different L1s from different countries, which is one of the aims of the present study. It should also be added that the essays investigated in the present study were all written by students in the field of linguistics. This quality is likely to have been beneficial for the comparison of writers with different L1s, since differences in rhetorical structures caused by disciplinary expectations were minimised.

3. Material and methods

The research in the present study complies to the ethical principles established by the Swedish Research Council (Sw. 'Vetenskapsrådet') in 2002. They included demands of information, permission, confidentiality and use restriction. Regarding the material used from the corpora, ethical concerns were few since the texts are coded in such a way that it makes it impossible to trace them back to the students who wrote the papers.

3.1 Material: The SUSEC corpus

The corpus data researched in the present study is from the Stockholm University Student English Corpus (henceforth SUSEC). It contains essays written in English by students from King's College (henceforth KC) in London and students from Stockholm University (henceforth SU) in Stockholm. The texts from KC are written by native English speakers and the texts from SU by L1 Swedish speakers (Garretson, 2009, p. 1). The native speaker essays are written by second and third year linguistics students. The L1 Swedish essays are written by students between their first and fourth year of studies in English. The texts are classified into B- and C-level subcorpora. The former refers to second semester studies in linguistics,

and the latter third semester studies studies in linguistics. The texts are final drafts of student assignments or assignments on various topics in linguistics. Information about topics and writing instructions was not available.

For the present study, 225 essays from SUSEC were processed and analysed. The total number of words were approximately 800,000 words. 143 of the essays were from the L1 Swedish part of the corpus. The remaining 82 essays were from the native English speaker part. Table 1 shows a more detailed account of the distribution of the texts.

Table 1 Size of the various subcorpora components in SUSEC used in the research (Garretson, 2009, p. 2)

	Stockholm University		King's College	
Component	Texts	Words	Texts	Words
B-level	81	171,014	29	80,750
C-level	62	417,772	53	166,685
Totals	**143**	**588,786**	**82**	**247,435**

It should be noted that the limited size of B-level students in KC decreases the generalisability of the results from that subcorpus.

3.2 Methods

This study employed a corpus linguistic methodology to investigate engagement features in advanced level student writing, which means that the research was performed by using a body of texts that had been designed for the purpose of research. Hardie and McEnery (2012) have described a corpus as "some set of machine-readable texts which is deemed an appropriate basis on which to study a specific set of research questions" (p. 1). The term *corpus* refers to the actual database of texts. The method used in the present study can be called corpus-driven, that is, the data was interpreted to find patterns (McEnery & Hardie, 2012).

3.2.1 Methodological framework and analytical steps

The corpus data from SUSEC was uploaded into the concordance software *Antconc* (Anthony, 2018) and searched for elements of engagement in the categories *reader pronouns* (RPs), *questions* (Qs) and *directives* (Ds). These categories were taken from Hyland's (2001; 2005b) framework for investigating engagement as presented in section 2.2.6. The framework was slightly adapted to fit the scope of the present study. Most importantly, PAs and ASKs were

not investigated. The reason for that was that fewer categories allowed for a deeper analysis. After the collection-process had finished, each case was examined in its concordance line to decide whether it was an element of engagement or not. All cases that were judged to not indicate engagement were removed and the remaining examples were sorted into several sub-categories. The methods for this process and the exact sub-categories are described in more detail below. After the categorisation had finished, the results from the native speakers and the L1 Swedish students were correlated and analysed in accordance with the aims of the study. The principal aim was to find overall engagement patterns in the corpus. Secondary aims were to identify possible differences in patterns of engagement between the L1 Swedish students and the native English students, and differences in patterns of engagement between B-level and C-level students. The research in the present study has been inspired by CIA (briefly described in section 2.2.5), especially with regard to how it compares native speaker language with learner language. The discussion that follows next, is about the three types of engagement elements included in the research and descriptions of how they were investigated.

Reader pronouns are used by writers to connect with audiences as well as to guide them through arguments. The 6 RPs searched for were: *we*, *us*, *our*, *you*, *your* and *one*. Whether the impersonal pronoun *one* should be considered an RP is debatable. It was included in this research because even if it can be hard to classify it as an RP, it is often used as an RP. The search of *one* used as an engagement element was complicated since *one* is also a frequently used numeral. To facilitate the process only cases where *one* was followed by verbs were searched for. This choice was made after 100 concordance lines containing *one* had been analysed and it was found that in all cases where it was used as an engagement element, it was followed by a verb. Collocate searches for modal verbs and verbs were made in the two slots directly following cases of *one*. Example (16) shows how *one* is used as an engagement element followed by a modal verb. Example (17) illustrates how *one* is used as an engagement element followed by a verb in the singular third person present tense.

> (16) So one **must** ask how does a child acquire an adult like grammar? (SUSEC_KC2_001)
> (17) To understand the fundamental processes of global English, **one** needs to focus not only on the relationship between demographics […] (SUSEC_KC2_011)

Example (17) also contains the directive *needs to*. This is one of several occurrences in the text where two elements of engagement combine in the same function. In this study, both

elements were counted, which means that these overlaps must be regarded when reading the results. *We* is the most frequently used RP in academic writing. Since *we* is not always used as a RP, the analysis included the removal of all cases of exclusive *we* so that only cases of inclusive *we* were left. The latter are the RPs. In example (18) below is an exclusive *we* that was not counted as a RP, as opposed to the inclusive *we* in example (19).

> (18) […] and therefore in this study **we** aim to replicate these results by using […] (SUSEC_KC3_055)
> (19) […] but it is not surprising that **we** acquire speech easily since human beings have […] (SUSEC_KC3_036

The situation was similar for the other RPs and the analysis of the concordance lines included removal of all pronouns that were judged as exclusive. There were also several other reasons to why cases were removed. For example how the acronym *US* had to be removed from the search results for *us*.

Questions is a straightforward category of engagement elements. Searches were made for sentences containing a question-mark. Again, this method will not catch all Qs, but the results will be adequate for the scope of the present study. The findings were manually analysed and categorised into *non-rhetorical* and *rhetorical* Qs, as described in section 2.2.6.

Directives come in several guises. The search strings for directives are listed in Appendix A. When compared to some of Hyland's research (e.g. 2001; 2005a; 2016), they are considerably fewer. This means that the results are not comparable to Hyland's results. The fact that not all Ds are signalled by words or word strings further adds to the number of Ds that were not included in this study. In (20) is an example of a directive where a reader is incentivised to a textual act without the use of a modal of obligation, a predicative adjective or an imperative, which are the categories included in Hyland's framework.

> (20) Two out of three songs describe women as bad people (Appendix 5 and 6) and one song describe a woman who […] (SUSEC_SUBLING_032)

The reason why this choice was made was in part practical. To find all Ds in the 225 essays would simply have been too time-consuming. However, a considerable number of Ds were identified, and it was possible to compare native English speakers to L1 Swedish, which is the main aim of this study. In the process of choosing which directives to include, preference was given to those that were more frequent or appeared in several subcorpora. Some directives

(e.g. *assume*, *ought to*, *required to* and *safe to*) were included because they had appeared frequently in other studies.

Three categories of signalling words or word strings were searched for. The first was *imperatives*; the words searched for were: *assume/compare/consider/imagine/let us* (including *let's*)/*note/ see* and *think*. In example (21) below is a case where the search word *imagine* was not considered an element of engagement and was therefore excluded from the list. Example (22) shows a concordance line where *imagine* was considered an imperative element of engagement and therefore was included.

> (21) When people hear the term bilingual many **imagine** an individual who […] (SUSEC_KC2_009)
>
> (22) **Imagine** then, that a considerable number of […] (SUSEC_SUCLING_026)

The second category was *modals of obligation* and the words searched for were: *has to/have to/must/need to/needs to/ought to* and *should*. It can be hard to decide if a modal of obligation is engaging with the reader. In (23) is an example of a modal of obligation that was considered a directive, and in (24) is an example when it was not.

> (23) Morality **needs to** be incorporated into the metaphorical system. (SUSEC_SUBLING_069)
>
> (24) Here, it **needs to** be established that in creating the core vocabulary list […] (SUSEC_SUBLING_060)

The reason why *needs to* in (23) was considered a directive is that it clearly urges the reader to agree with something. In (24), *needs to* is used to communicate something the writer feels s/he has to do, rather than to direct the reader to doing something and was therefore not considered a directive. The third category was *predicative adjectives* that express writers' judgement of necessity/importance. Before the search for predicative adjectives was made, the files from SUSEC were POS-tagged (part-of-speech-tagged). That is, they were tagged into grammatical categories. The programs used for this end were *EncodeAnt* (Anthony, 2017) and TagAnt (Anthony, 2017). The search string *it_PP is_VBZ *_JJ to_TO* was used for searching the tagged data. This formula returned all word strings initiated by *it is* that was followed by an adjective. After manually checking the concordance lines, six predicative adjectives were identified. They were: *it is appropriate to/it is crucial to/it is essential to/it is important to/it is necessary to* and *it is relevant to*. The identified Ds were also categorised according to the three acts as presented in Hyland's framework. Ds that incentivise *textual acts* and guide

readers in texts, Ds that incentivise *physical acts* in real life, and Ds that incentivise *cognitive acts* which help readers understand reasoning. For examples of acts, see section 2.2.6.

3.2.2 Statistical significance

In the comparison of results, it was important to establish which differences could be considered as statistically significant. For this reason, a log-likelihood calculator was used. *Log-likelihood* is a recognised parameter for establishing statistical probability. The threshold value for a difference to be deemed as significant was $p < 0.05$ (critical value 3.84). All comparative results presented in this study have been tested for statistical significance. If significance was not established but results are referred to, it has been pointed out. Only some of the results were checked for dispersion. When this was the case, it was indicated in the text.

4. Results and discussion

Here, the results from the present study are presented and discussed in relation to previous research and literature. In section 4.1 is a presentation of the overall results. In sections 4.2, 4.3 and 4.4 respectively, the results regarding reader pronouns, questions and directives are presented. Last is section 4.5 with a summary of the analysis. The results in this section have been normalised to per thousand words. For the absolute frequencies, see Appendix B.

4.1 Overall results

Table 2 displays the overall normalised frequencies for engagement elements.

Table 2 Overall frequencies of engagement elements in SUSEC normalised to per thousand words

	SUBLING[2]	SUCLING	**Totals SU**	KC2	KC3	**Totals KC**	**TOTALS**
RPs	2.98	1.99	**2.28**	2.27	1.43	**1.71**	**2.11**
Qs	0.71	0.49	**0.56**	0.52	0.47	**0.49**	**0.54**
Ds	0.74	1.07	**0.97**	0.79	0.77	**0.78**	**0.91**
Totals	**4.42**	**3.55**	**3.81**	**3.58**	**2.68**	**2.98**	**3.56**

---

[2] Abbreviations for SUSEC subcorpora will be used throughout this study. They are: SUBLING = B-level students from Stockholm University SUCLING = C-level students from Stockholm University KC2 = B-level students from King's College KC3 = C-level students from King's College

.

The most common engagement elements were RPs, which had a normalised total of 2.11. Then followed Ds with 0.91 and Q with 0.54. The C-level students in SUCLING and KC3 had normalised totals of 3.55 and 2.68 respectively. This was less than the B-level students in SUBLING and KC2 that had the totals 4.42 and 3.58 respectively. This difference was directly connected to the fact that B-level students used significantly more RPs; SUBLING and KC2 had the scores 2,98 and 2.27 respectively, compared to SUCLING and KC3 that had the scores 1.99 and 1.43 respectively. In both SUCLING and KC3, the gap between use of RPs and Ds was significantly smaller than it was in SUBLING and KC2. With respect to first language, the SU-students used 3.81 engagement elements ptw, which was significantly more than the 2.98 of the KC-students. These results can give a general idea regarding the use of engagement, but there were also some interesting variations in different categories, which will be discussed in the sections that are to follow.

The afore mentioned study by Aull & Lancaster (2014) had indicated that more advanced writers would develop a stance more inclined towards a dialogue with readers. This was something that was not confirmed by the overall results. Instead, C-level students from both SU and KC used fewer engagement elements in their papers than B-level students. Another interesting results was that the students from KC used fewer elements of engagement than the Swedish students. When it comes to the results of the L1 Swedish students, explanations can perhaps be found in theories and research as presented by Ädel (2006), Hyland (2009), Lafuente-Millán (2014) and Mauranen (1993) (see section 2.3.2). What they have suggested is that there may be cultural reasons, be that academic or other, to why Swedish students use more elements of engagement than English students. The answer may also be found in interlanguage theory (see section 2.2.5) and as Gilquin & Paquot (2008) have suggested, that Swedish students tend to overuse interpersonal discourse in their learner language. If that is the case, the decline of engagement elements in the writing of L1 Swedish C-level students compared to L1 Swedish B-level students, may be because the latter have acquired better skills in English and that interlanguage and culture have become less influential. It does not explain why Ds were more used in SUCLING than in SUBLING; a possible explanation for that will be offered in section 4.4 where the results regarding Ds are discussed. The fact that the KC2 students used more RPs than the KC3 students will be discussed in section 4.2. In Hyland's study of research writing in applied linguistics from 2005, the results showed an

almost even balance between RPs and Ds. This was not the case in the present study where the students used almost twice as many RPs compared to Ds. However, the fact that there was a more even distribution of Ds and RPs among the C-level students compared to the B-level students, may suggest that the more experienced students used engagement in a fashion more similar to how researchers write.

4.2 Reader pronouns

6 RPs were investigated. They were: *we*, *us*, *our*, *you*, *your* and *one*. Table 3 shows the normalised frequencies for RPs.

Table 3 Frequencies for RPs normalised to per thousand words

	SUBLING	SUCLING	Totals SU	KC2	KC3	Totals KC
We	1.30	1.11	**1.17**	0.93	0.75	**0.81**
Us	0.30	0.20	**0.23**	0.43	0.11	**0.22**
Our	0.38	0.18	**0.24**	0.16	0.15	**0.15**
You	0.51	0.13	**0.24**	0.35	0.17	**0.23**
Your	0.14	0.02	**0.06**	0.06	0.02	**0.04**
One	0.34	0.35	**0.35**	0.35	0.19	**0.24**
Totals	**2.98**	**1.99**	**2.28**	**2.27**	**1.40**	**1.69**

In all subcorpora, *we* was the RP that had the highest normalised frequency. It is possible to see some patterns from the normalised frequencies shown in Table 3. One is that the students from SU used on average 2.28 RPs ptw, which was more than the students from KC that used 1.69 RPs ptw. Another is that B-level students used more RPs than C-level students. The B-level students from SU used on average 2.98 ptw and the B-level students from KC 2.27 RPs ptw, which was more than the averages 1.99 ptw and 1.40 RPs ptw of the C-level students from SU and KC respectively.

As has been suggested in the previous section, the reasons to why the texts from the L1 Swedish had more RPs compared to the texts written by native speakers may be cultural and/or to be found in interlanguage. Results from the previously mentioned study by Ädel (2006) showed that Swedish students used considerably more inclusive pronouns like *we* and *us* as well as direct references to readers (i.e. *you*, *your*). Example (25) displays how a Swedish B-level student used two RPs to include the reader in a discussion.

(25) Since the impressions **we** receive when **we** are children are clearly important this essay will focus on children's literature and how it communicates gender roles. (SUSEC_SUBLING_010)

4.3 Questions

Table 4 illustrates frequencies as well as normalised frequencies for all identified cases of Qs. The most prominent result was that SUBLING had the normalised frequency of 0.71, which was significantly more than SUCLING, KC2 and KC3 that had 0.49, 0.52 and 0.47 respectively.

Table 4 Frequencies for Qs as well as frequencies normalised to per thousand words for Qs

	SUBLING	SUCLING	Total SU	KC2	KC3	Total KC	TOTAL
Frequency	120	206	**326**	42	80	**122**	**448**
Normalised	0.71	0.49	**0.56**	0.52	0.47	**0.49**	**0.54**

All identified cases of Qs were categorised into rhetorical and non-rhetorical Qs. Table 5 displays the normalised frequencies of rhetorical (R) and non-rhetorical (NR) Qs.

Table 5 Rhetorical and non-rhetorical Qs normalised to per thousand words

SUBLING		SUCLING		Total SU		KC2		KC3		Total KC		TOTAL	
NR	R	NR	R	**NR**	**R**	NR	R	NR	R	**NR**	**R**	**NR**	**R**
0.56	0.15	0.32	0.17	**0.39**	**0.16**	0.35	0.17	0.22	0.26	**0.26**	**0.23**	**0.40**	**0.21**

SUBLING had the normalised frequency 0.56 for non-rhetorical questions. This was significantly higher than for all other subcorpora. With a normalised frequency of 0.22, KC3 used significantly fewer non-rhetorical questions than the other students. KC3 also used significantly more rhetorical Qs than the other students. In the process of identifying rhetorical and non-rhetorical questions, there seemed to be patterns to where they were located in the texts. Upon analysing which positions different Qs occupied in the essays, some patterns indeed emerged. Table 6 illustrates normalised frequencies of different Qs with regard to position in the text.

Table 6
Frequencies for Qs normalised to per thousand words with regard to position in the texts.

	SUBLING		SUCLING		Totals SU		KC2		KC3		Totals KC	
	NR	R	NR	R	**NR**	**R**	NR	R	NR	R	**NR**	**R**
Beginning	0.54	0.01	0.28	0.01	**0.35**	**0.01**	0.07	0.00	0.08	0.02	**0.08**	**0.02**
Mid or end	0.01	0.14	0.05	0.16	**0.04**	**0.15**	0.27	0.17	0.13	0.24	**0.18**	**0.22**
Totals	**0.55**	**0.15**	**0.33**	**0.17**	**0.39**	**0.16**	**0.34**	**0.17**	**0.21**	**0.26**	**0.26**	**0.24**

With normalised frequencies of 0.18 and 0.22 for non-rhetorical and rhetorical Qs respectively in the mid or end positions, the KC students showed a different pattern than the SU-students who had the frequencies 0.04 and 0.15 in the same positions. The SU-students had 0.35 for non-rhetorical questions in the beginning, which was significantly more than the KC-students who only had the frequency 0.08. The reasons for why these pattern emerged are not obvious, but part of an explanation might be found in writing instructions. When checking individual texts, many texts from SU and especially in SUBLING, contained research questions in the beginning. This was the case in 38 out 82 texts in SUBLING, and 39 out of 65 in SUCLING. Example (26) below shows a research question from SUBLING that was positioned in an introduction.

> (26) Is the information that is given to students today in schools through textbooks accurate and up to date? (SUSEC_SUBLING_022)

Why the texts in KC3 had so many rhetorical questions may be a sign of the fact that the KC3 students have become adapted to the discipline of linguistics. In a study, Hyland (2001) presented results that suggested that rhetorical questions were common features in academic writing of applied linguistics. Example (27) illustrates how a student in KC3 used a rhetorical question positioned in the middle of a text.

> (27) Why would it be though, that a newly independent state would choose English to be the chosen language? (SUSEC_KC3_020)

4.4 Directives

Three types of engagement elements for directives were searched for. They were *imperatives*, *modals of obligation* and *predicative adjectives*. The normalised frequencies for Ds are shown in Table 7.

Table 7 Frequencies for Ds normalised to per thousand words

	SUBLING	SUCLING	**Totals SU**	KC2	KC3	**Totals KC**
Imperatives	0.33	0.52	**0.47**	0.32	0.19	**0.23**
Modals of obl.	0.27	0.42	**0.38**	0.43	0.47	**0.46**
Pred. adj.	0.14	0.12	**0.13**	0.04	0.11	**0.09**
Totals	**0.74**	**1.07**	**0.97**	**0.79**	**0.77**	**0.78**

Comparing the SU-subcorpora, the C-level students used significantly more imperatives and modals of obligation than the B-level students. A closer look at the individual words that were searched for as markers for imperatives (see Appendix B – Table VI), the imperative *see* constituted 201 out of a total of 218 in the C-level subcorpus from SU. The high frequency of *see* influenced the results greatly and is key to understanding how SU-students used 0.47 imperatives ptw, compared to the KC-students considerably lower frequency of 0.23 ptw. Another interesting finding was that the KC-students used modals of obligations 0.46 times ptw, which was significantly more than the SU-students who used modals of obligations 0.38 times ptw.

The Ds were also classified into the acts they incentivise, that is, *textual acts*, *physical acts* and *cognitive acts*. Since no cases of Ds that incentivised physical acts were found, this category was ignored in the reporting of the results. Table 8 shows the normalised frequencies for acts.

Table 8 Frequencies for acts normalised to per thousand words

	SUBLING	SUCLING	**Totals SU**	KC2	KC3	**Totals KC**
Imperatives						
Textual acts	0.22	0.49	**0.41**	0.28	0.17	**0.21**
Cognitive acts	0.11	0.04	**0.06**	0.06	0.00	**0.02**
Totals	**0.33**	**0.52**	**0.47**	**0.35**	**0.17**	**0.23**
Modals of oblig.						
Textual acts	0.01	0.01	**0.01**	0.01	0.00	**0.00**
Cognitive acts	0.26	0.41	**0.37**	0.42	0.47	**0.46**
Totals	**0.27**	**0.42**	**0.38**	**0.43**	**0.47**	**0.46**
Pred. adj.						
Textual acts	0.00	0.00	**0.00**	0.00	0.00	**0.00**
Cognitive acts	0.14	0.12	**0.13**	0.04	0.11	**0.09**
Totals	**0.14**	**0.12**	**0.13**	**0.04**	**0.11**	**0.09**

In total, there were 299 cases where textual acts were incentivised and 466 cases where cognitive acts were incentivised (see Appendix B – Table VII). Noteworthy was that the students in SUCLING used a lot more textual acts than the other students, which was related to the frequent use of *see* as an imperative D. The textual acts in SUCLING were quite evenly dispersed and occurred in 45 texts out of 62. Example (28) shows how a student in SUCLING has used *see* as imperative Ds two times in the same sentence to incentivise textual acts.

> (28) I have looked at and listed all occurrences of swearwords in Sprängaren (**see** appendix 1), and also those in The Bomber, its English translation (**see** appendix 2). (SUSEC_SUCLING_026)

The significant number of textual acts inspired further enquiries. A possible explanation might have been that the native English speakers used different ways to incentivise textual acts. The abbreviation *cf.* short for the latin expression *confer/conferatur* is often used to direct people to other parts of texts. A search for *cf.* in SUSEC did not even the results out. Instead it confirmed that the students in SUCLING used a large number of directives that incentivised textual acts. Out of 43 hits, 6 belonged in KC2. The remaining 37 were found in SUCLING.

The students from KC2 and KC3 had the highest normalised frequencies for cognitive acts and in particular when incentivised by modals of obligation (KC2 0.43; KC3 0.47). Example (29) illustrates how a student in KC2 has used a modal of obligation as a D to incentivise a cognitive act.

> (29) With this point in mind one **must** ask then if a child cannot learn a language as powerful as a human language from positive exemplars of sentences then how do they learn language? (SUSEC_KC2_001)

4.5 Summary of the analyses

The results from the present study show that among the elements of engagement investigated in this study, most common were RPs, the second most common were Ds and the least common were Qs. The results also show that the L1 Swedish students used more elements of engagement than the native English students, which was directly related to the fact that the C-level native English students used fewer RPs than the other students. Another interesting finding was that Qs with non-rhetorical purposes were more common than Qs with rhetorical purposes, and that the former were more used by the L1 Swedish students than by the native

English students. It was also possible to see patterns in how Qs were positioned in the texts. KC-students had the highest number of their Qs in the middle or in the end of the texts. SU-students had most of their rhetorical Qs in the middle or in the end of texts but had most of their non-rhetorical Qs in the beginning of the texts. Regarding Ds, the students in SUCLING used significantly more imperatives than the other students, especially with regard to the word *see*, which was mostly used to incentivise textual acts. The students in KC2 and KC3 used significantly more modals of obligation that incentivised cognitive acts. In the next section, the results from the investigation will be related to the research questions presented in the introduction. The results will also be discussed in light of previous research.

5. Conclusion

This corpus study investigated engagement patterns in L1 Swedish and L1 English students' academic writing. SUSEC was used as primary source. It contains texts written by B- and C-level students in the field of linguistics from Stockholm University in Stockholm and King's College in London. The categories researched were reader pronouns, questions and directives. Several aspects of these categories were also investigated.

The first research question addressed finding engagement patterns in L1 Swedish and L1 English student's academic writing. Overall, the students used more reader pronouns than directives and questions. Another general pattern was that non-rhetorical questions were more common than rhetorical questions. The second research question was aimed at investigating the ways patterns of engagement varied between L1 Swedish and L1 English students. One such variation was that the Swedish students used more elements of engagement than the English students. Another was that Swedish students favoured non-rhetorical questions positioned in the beginning of texts, while the English students used rhetorical and non-rhetorical questions quite evenly and positioned them in the middle or the end of texts. Yet another variation was that the English students favoured directives with modals of obligation that incentivised cognitive acts to a significantly higher degree than the Swedish students. The third research question was aimed at investigating how patterns of engagement vary between students at B-levels and C-levels in linguistic studies. One such variation was that the C-level English students tended to use fewer reader pronouns than other students. Another variation

was that the C-level Swedish students used significantly more imperative directives that incentivised textual acts.

Previous research has suggested that Swedish students tended to make frequent use of engagement features, which was also the case in this study. The underlying reasons may be cultural, as suggested by Ädel (2006), Hyland (2009), Lafuente-Millán (2014) and Mauranen (1993). They can also be related to interlanguage, as suggested by Gilquin & Paquot (2008). Other possible explanations may perhaps be found in writing instructions and the nature of topics. Previous research (e.g. Aull & Lancaster; Hyland 2009) has suggested that the higher the students' academic level is the more similar their use of engagement elements in writing will be. In the case of linguistics, this would include an increased use of engagement elements, especially regarding directives (Hyland, 2005a). This was not confirmed with the English students since the use of directives decreased between B-level and C-level students. However, it was confirmed with the Swedish students since the use of directives increased significantly between B-level and C-level students. It was also established that the KC-students more often used directives to incentivise cognitive acts than the SU-students.

There are several shortcomings to this study. One is that the size of the corpus is relatively small. The limited size was perhaps compensated by the fact that it was bi-lingual and contained texts written by students in the same field of study. Another shortcoming is that information about the exact topics and writing instructions for the student assignments were not available. Access to that type of information could possibly have given important clues to some of the results, for example to why there were so many directives incentivising textual acts in the SUCLING subcorpus. The fact that some elements of engagement were not investigated, as is the case with appeals to shared knowledge and personal asides, has perhaps disallowed an integral understanding of the engagement patterns in the corpus. However, by limiting the scope of the research it was possible to investigate certain aspects of engagement more comprehensively. In the process of analysing the results, some flaws in methodology were detected. An example is that the search string *cf.* should have been included in the search for directives since it is a common way to incentivise readers to perform textual acts.

Further research is needed to confirm if the findings from the two corpora investigated in this study are generalisable to other databases. Ideally, a similar collection of texts would be analysed using the same methodology that includes information about writing instructions and topics for the assignments. Another suggestion for further research is to investigate how

factors like culture and interlanguage can affect the ways students use different features of engagement in academic writing. One approach could be to investigate in what manner Swedish students use engagement features in their writing in Swedish and to see if that translates into their writing in English. Another approach could be to research how Swedish and English use engagement features in other disciplines. It is also possible to see how investigating just one category of engagement features, for instance directives, would yield results with more depth, and in particular if you allow the inclusion of more categories than imperatives, modals of obligation and predicative adjectives. These are all interesting topics for research that may produce results that can help inform the teaching of academic writing.

Acknowledgements

For knowledgeable aid and good cooperation, I would like to thank the supervisor for my essay Olcay Sert. In addition, I want to acknowledge that the kind support and insightful guidance from Elisabeth Wulff-Sahlén has helped to improve the quality of this essay greatly. Last, I would like to thank Britt Erman at Stockholm University for kindly granting me access to use the SUSEC corpus.

Håkan Almerfors

Stockholm, June 13, 2018

References

Anthony, L. (2018). Antconc (Version 3.5.2). [Computer Software] Retrieved from:
http://www.laurenceanthony.net/software/antconc/

Anthony, L. (2017). EncodeAnt (Version 1.2.1). [Computer Software] Retrieved from:
http://www.laurenceanthony.net/software/encodeant/

Anthony, L. (2017). TagAnt (Version 1.2.0). [Computer Software] Retrieved from:
http://www.laurenceanthony.net/software/tagant/

Aull, L. L., & Lancaster, Z. (2014). Linguistic markers of stance in early and advanced
academic writing: A corpus-based comparison. *Written Communication, Vol. 31(2),*
151-183. [PDF] Retrieved from:
http://journals.sagepub.com/doi/abs/10.1177/0741088314527055

Bakhtin, M. M. (1981). *The dialogic imagination*. Austin-London: University of Texas Press.
[PDF] Retrieved from:
https://warwick.ac.uk/fac/arts/english/currentstudents/undergraduate/modules/fulllist/sp
ecial/english19thcentnovel/week_2_-_bakhtin_-_discourse_in_the_novel_1981.pdf

Biber, D. (2006). *University language*. Amsterdam-Philadelphia: John Benjamins Publishing
Company [E-book]

Brown, P., & Levinson, S. C. (2006). Politeness: Some universals in language use. In A.
Jaworski & N. Coupland (Eds.), *The Discourse Reader: Second Edition*. Chapter
22: 311-323. London-New York: Routledge [PDF] Retrieved from:
http://pubman.mpdl.mpg.de/pubman/item/escidoc:64421/component/escidoc:2225570/
Brown&SCL-Politeness1999.pdf

Cao, F. & Hong, H. (2014). Interactional metadiscourse in young EFL learner writing: A
corpus-based study. *International Journal of Corpus Linguistics*, *19:2*, 201-224. [PDF]
Retrieved from:
https://www.researchgate.net/publication/262580567_Interactional_metadiscourse_in_y
oung_EFL_learner_writing_A_corpus-based_study

Casal, J. E., Hitchcock, C., & Lee, J. J. (2017). Citation practices of L2 university students in
first-year writing: Form, function and stance. *Journal of English for Academic
Purposes, 33*, 1-11. [PDF] Retrieved from:

https://reader.elsevier.com/reader/sd/B1B9277F39366D1248F82120CAB4AE567573E
A5B39FE93AACE06AAE7E025A72952FF9FFD54CBA8C6C9785D0B3EA3EDBD

Chan, H. (2014). A corpus-based study of the expression of stance in dissertation acknowledgements. *Journal of English for Academic Purposes*, *20*, 176-191. [PDF]

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press. [PDF] Retrieved from: https://pt.scribd.com/doc/169111563/Rod-Ellis-The-Study-of-Second-Language-Acquisition

Ellis, R. (1997). *Second language acquisition*. Oxford: Oxford University Press.

Erman, B., & Garretson, G. *Stockholm University Student English Corpus* (SUSEC). King's College and Stockholm University

Garretson, G. (2009). *Manual to accompany the Stockholm University Student English Corpus (SUSEC)*. [PDF]

Gilquin, G., & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction, 1(1),* 41-61. [PDF] Retrieved from: https://www.researchgate.net/publication/233501092_Too_chatty_Learner_academic_writing_and_register_variation

Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research, Vol 1:1*, 7-24. [PDF] Retrieved from: https://benjamins.com/#catalog/journals/ijlcr.1.1.01gra/fulltext

Hyland, K. (1998). Persuasion and context: The pragmatics of academic metadiscourse. *Journal of Pragmatics*, *30*, 437-455.

Hyland, K. (2001). Bringing in the reader: Adressee features in academic articles. *Written Communication, 18(4),* 549-574. [PDF]

Hyland, K. (2002). Directives: Power and engagement in academic writing. *Applied Linguistics*, *23*, 215-239. [PDF] Retrieved from: file:///C:/Users/Hakan/Downloads/DirectivesAppliedling.pdf

Hyland, K, (2005a). Representing readers in writing: Student and expert practices. *Linguistics and Education*, *16*, 363-377. [PDF] Retrieved from: https://www.sciencedirect.com/science/article/pii/S0898589806000404

Hyland, K. (2005b). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, *7(2),* 173-192. [PDF]

Hyland, K. (2009). Corpus informed discourse analysis: The case of academic engagement. In M. Charles, S. Hunston & D. Pecorari (Eds.), *Academic writing: At the interface of corpus and discourse*, 110-128. London: Continuum

Hyland, K. (2015). Metadiscourse. *The International Encyclopedia of Language and Social Interaction*, *1*, 1-11. [PDF] Retrieved from: https://www.researchgate.net/publication/285591598_Metadiscourse

Hyland, K., & Jiang, F. (2016). "We must conclude that…": A diachronic study of academic engagement. *Journal of English for Academic Purposes, 24*, 29-42. [PDF]

Lafuente-Millan, E. (2014). Reader engagement across cultures, languages and contexts of publication in business research articles. *International Journal of Applied Linguistics*, *24(2)*, 201-223. [PDF]

Lancaster, Z. (2016). Expressing stance in undergraduate writing: Discipline-specific and general qualities. *Journal of English for Academic Purposes*, *23*, 16-30. [PDF] Retrieved from: https://www.sciencedirect.com/science/article/pii/S1475158516300340

Mauranen, A. (1993). Cultural differences in academic discourse: Problems of a linguistic and cultural minority. The Competent Intercultural Communicator. In L. Löfman, L. Kurki-Suono, S. Pellinen & J. Lehtonen (Eds.) *AFinLAn Yearbook, 23(51),* 157-174. [PDF]

McEnery, T., & Hardie, A. (2012). *Corpus linguistics*. New York: Cambridge University Press. [E-book]

Myers, G. (1989). The pragmatics of politeness in scientific articles. *Applied Linguistics*, *Vol. 10, No. 1*, 1-35 [PDF] Retrieved from: https://academic.oup.com/applij/article-abstract/10/1/1/215686

Rabinowitz, P. J. (1977). Truth in fiction: A reexamination of audiences. *Critical Inquiry, Vol 4, No 1*, 121-141. [PDF]

Vetenskapsrådet. (2002). *Forskningsetiska principer inom humanistisk samhällsvetenskaplig forskning*. Vetenskapsrådet. [PDF] Retrieved from: http://www.codex.vr.se/texts/HSFR.pdf

White, P. R. R. (2003). Beyong modality and hedging: A dialogic view of the language of intersubjective stance. *Text*, *23(2),* 259-284. [284] [PDF] Retrieved from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.133.7078&rep=rep1&type=pdf

Ädel, A. (2006). *Metadiscourse in L1 and L2 English*. Amsterdam-Philadelphia: John Benjamins Publishing Company [E-book]

Internet resources

Log-likelihood and effect size calculator. Available at: http://ucrel.lancs.ac.uk/llwizard.html

Appendix A

Search strings used in Antconc to search for directives:

<u>Imperatives</u>

assume

compare

consider

imagine

let us / let's

note

see

think

<u>Modals of obligation</u>

has/have to

must

need/needs to

ought to

should

<u>Predicative adjectives</u>

it_PP is_VBZ *_JJ to_TO

Appendix B

Table I Overall absolute frequencies of engagement elements in SUSEC

	SUBLING	SUCLING	Totals SU	KC2	KC3	Totals KC	TOTALS
RPs	510	830	**1340**	184	239	**423**	**1763**
Qs	120	206	**326**	42	80	**122**	**448**
Ds	126	446	**572**	64	129	**193**	**765**
Totals	**756**	**1482**	**2238**	**290**	**448**	**738**	**2976**

Table II Absolute frequencies for RPs

	SUBLING	SUCLING	Totals SU	KC2	KC3	Totals KC
We	223	464	687	75	125	200
Us	52	82	134	35	19	54
Our	65	75	140	13	25	38
You	88	53	141	28	29	57
Your	24	10	34	5	4	9
One	58	146	204	28	32	60
Totals	**510**	**830**	**1340**	**184**	**234**	**418**

Table III Absolute frequencies for rhetorical and non-rhetorical Qs

SUBLING		SUCLING		Total SU		KC2		KC3		Total KC		TOTAL	
NR	R	NR	R	**NR**	**R**	NR	R	NR	R	**NR**	**R**	**NR**	**R**
95	25	135	71	**230**	**96**	28	14	36	44	**64**	**58**	**296**	**154**

Table IV Absolute frequencies for Qs with regard to position in the texts

	SUBLING		SUCLING		Totals SU		KC2		KC3		Totals KC	
	NR	R	NR	R	**NR**	**R**	NR	R	NR	R	**NR**	**R**
Beginning	93	1	116	4	**209**	**5**	6	0	14	4	**20**	**4**
Mid or end	2	24	19	67	**21**	**91**	22	14	22	40	**44**	**54**
Totals	**95**	**25**	**135**	**71**	**230**	**96**	**28**	**14**	**36**	**44**	**64**	**58**

Table V Absolute frequencies for Ds

	SUBLING	SUCLING	Totals SU	KC2	KC3	Totals KC
Imperatives	56	218	**274**	26	31	**57**
Modals of obl.	46	176	**222**	35	79	**114**
Pred. adj.	24	52	**76**	3	19	**22**
Totals	**126**	**446**	**572**	**64**	**129**	**193**

Table VI Absolute frequencies of results for Ds related to element marker words

	SUBLING	SUCLING	Totals SU	KC2	KC3	Totals KC
Imperatives						
assume	1	0	**1**	0	0	**0**
compare	4	0	**4**	0	0	**0**
consider	9	2	**11**	0	0	**0**
imagine	0	1	**1**	1	0	**1**
let us / let's	2	5	**7**	2	0	**2**
note	5	9	**14**	2	1	**3**
see	35	201	**236**	21	30	**51**
think	0	0	**0**	0	0	**0**
Totals	**56**	**218**	**274**	**26**	**31**	**57**
Modals of obl.						
has/have to	12	32	**44**	3	9	**12**
must	13	51	**64**	11	36	**47**
need/needs to	7	27	**34**	9	20	**29**
ought to	0	0	**0**	0	0	**0**
should	14	66	**80**	12	14	**26**
Totals	**46**	**176**	**222**	**33**	**79**	**114**
Pred. adj.						
it is appropriate to	0	1	**1**	0	0	**0**
it is crucial to	2	1	**3**	1	0	**1**
it is essential to	1	3	**4**	0	2	**2**
it is important to	16	34	**50**	2	6	**8**
it is necessary to	5	13	**18**	0	9	**9**
it is relevant to	0	1	**1**	0	2	**2**
Totals	**24**	**52**	**76**	**3**	**19**	**22**
Directives Totals	**126**	**446**	**572**	**64**	**129**	**193**

Table VII Absolute frequencies for acts

	SUBLING	SUCLING	**Totals SU**	KC2	KC3	**Totals KC**
Imperatives						
Textual acts	38	203	**241**	2	31	**52**
Cognitive acts	18	15	**33**	5	0	**5**
Totals	**56**	**218**	**274**	**26**	**31**	**57**
Modals of oblig.						
Textual acts	2	3	**5**	1	0	**1**
Cognitive acts	44	173	**217**	34	79	**113**
Totals	**46**	**176**	**222**	**35**	**79**	**114**
Pred. adj.						
Textual acts	0	1	**1**	0	0	**0**
Cognitive acts	24	52	**76**	3	19	**22**
Totals	**24**	**52**	**76**	**3**	**19**	**22**