What’s in a Note?

Sentiment Analysis in Online Educational Forums

by

Najmeh Fakhraie

A thesis submitted in conformity with the requirements for the degree of Master of Arts

Department of Curriculum, Teaching and Learning

University of Toronto

© Copyright by Najmeh Fakhraie 2011
Abstract

This multi-disciplinary study examines the linguistic characteristics which influence communication and social interaction in computer-mediated communication (CMC). We begin by conducting a qualitative data analysis on a group of graduate students taking online courses. Through this, we look more closely at their perception of social interaction in their online learning environment (Knowledge eCommons). We then take individual student notes and analyze their linguistic characteristics. We look at the emotional cues in notes, the use of factual, objective language and other linguistic features. We study these notes through the use of sentiment analysis methodologies – which will be explained in detail in the first and second chapter. We have proposed a method for deducing note objectivity and have computed reliability testing of this method. Our analyses show that there is a high correlation between the use of objective language in a note and the value that students place on that note.
Acknowledgements

A thesis is a funny, odd kind of document. With one name on the cover, it is really an accumulation of multiple efforts, minds and ideas.

This particular effort began on a warm, sunny day on the 11th floor of OISE where I met Professor Jim Hewitt after having completed my first semester at OISE. I was shy, timid and very unsure and I remember we sat in the sunny, open lounge on the 11th floor and discussed knowledge building, online learning and scholarly pursuits of both. That initial conversation is what eventually led to this thesis. I hope that other students will have the opportunity to be initiated into the graduate life experience with the same kindness, warmth and welcome that he gave me.

Thanks to my friend and colleague, Siavash Kazemian, aka Haji. And to my siblings, Fati, Nazanin and Mohammad. No one can survive grad school without friendship, love and laughter and I was lucky to have been blessed with the best the world has to offer.

And ‘with special thanks to Amir Kermani for his technical support in preparing this document’ … but also for the late night laughs, burgers and heart-to-heart. For that, and much, much more, I will be eternally grateful.
# Table of Contents

Acknowledgements.............................................................................................................iii

Table ofContents................................................................................................................iv

List of Figures.....................................................................................................................vi

List of Tables.......................................................................................................................viii

List of Appendices...............................................................................................................ix

Chapter 1 Introduction.......................................................................................................1

1.1 Motivation for Study.....................................................................................................2

1.2 Research Questions.....................................................................................................3

1.3 Language and Communication..................................................................................4

1.4 Terminology................................................................................................................6

1.5 Knowledge eCommons..............................................................................................8

1.6 Why is this Interesting..............................................................................................12

2 Literature Review.........................................................................................................14

2.1 Communication and Interaction in Learning............................................................15

2.2 Communication in Online Learning Environments..................................................19

2.3 Sentiment Analysis....................................................................................................22

2.4 Academic Word List.................................................................................................26

2.5 Summary....................................................................................................................27

3 Methodology..................................................................................................................28

3.1 Overview.....................................................................................................................29

3.2 Survey Construction..................................................................................................30

3.2.1 Rating Scale.........................................................................................................32

3.2.2 The Survey..........................................................................................................33
3.3 Language Sentiment Analysis

3.3.1 Sentiment Detection Instrument

3.3.2 Sentiment Detection Guidelines

3.3.3 Sentiment Classification Instrument

3.3.4 Sentiment Scores

3.4 Academic Word List Computation

3.5 Putting it all Together

4 Results

4.1 Survey

4.1.1 General Questions

4.1.2 Research Question 1

4.1.3 Subquestion 1.1

4.1.4 Subquestion 1.2

4.1.5 Subquestion 1.3

4.1.6 Subquestion 1.4

4.2 Research Question 2

4.3 Research Question 3

5 Conclusion

5.1 Results

5.2 How is this Useful

References

Appendices
List of Figures

Figure 1: General outline of the project depicting a note and its linguistic and communicative characteristics.................................................................5

Figure 2: KeC homepage for a sample course, depicting the link to the courses’ different forums ....9

Figure 3: KeC forum for Week 3 of a sample course, with 3 subforums. .................................10

Figure 4: Recommend button in Knowledge eCommons which allows students to “recommend” their favored notes.................................................................10

Figure 5: sample webinar depicting the conversation between 4 individuals. ............................11

Figure 6: Two subtasks of sentiment analysis, sentiment detection and sentiment classification .....24

Figure 7: LIWC (linguistic inquiry and word count) software .....................................................44

Figure 8 – Sample LIWC output which shows the file name, positive emotion and negative emotion. ..............................................................................................................44

Figure 9: Sample note with objectivity/subjectivity and emotion scores ......................................45

Figure 10: Calculating sentiment for user Nom. 7 ......................................................................46

Figure 11: Academic Word List calculator which computes AWL from the student note ..........47

Figure 12: Looking for relationships between linguistic and communicative characteristics of notes ...............................................................................................................49

Figure 13: General Questions, number of online courses (n = 24) ..............................................51

Figure 14: Internet access for students in the sample (n = 24) .....................................................51

Figure 15: English fluency for students in the sample (n = 24) ....................................................52

Figure 16: Knowledge eCommons ease of use .............................................................................52

Figure 17: Projecting oneself in the environment (results, n = 24) ..............................................54

Figure 18: Co-presence, level of co-presence felt by students in the community .......................54

Figure 19: Instructor presence perceived by students .................................................................55
Figure 20: Perceived learning (level of learning students believed to have taken place) .......................56
Figure 21: students' views on the recommend button added to Knowledge eCommons .....................57
Figure 22: Student perception of the impact of the webinar on learning ........................................59
Figure 23: Student perception of private notes .....................................................................................59
Figure 24: Student perception of how avatars contributes to a sense of community in the classroom.
............................................................................................................................................................................60
List of Tables

Table 1 – Note and user attributes .................................................................................................................. 8
Table 2 – Rating Scale ..................................................................................................................................... 33
Table 3 - General questions ............................................................................................................................ 34
Table 4 – Projecting oneself in the environment ........................................................................................ 34
Table 5 – Co-presence ..................................................................................................................................... 35
Table 6 – Instructor presence ........................................................................................................................ 35
Table 7 – Perceived Learning ......................................................................................................................... 35
Table 8 – General Questions ........................................................................................................................... 36
Table 9 – Open ended questions .................................................................................................................... 37
Table 10 – List of variables for each user ..................................................................................................... 48
Table 11 - Statistically significant correlations with recommend ratio (p<0.05) ........................................ 70
List of Appendices

Appendix 1: Survey 87
Appendix 2: Rating Guidelines 100
1 Introduction

Social interaction has long been considered to play a significant role in the learning process (Vygotsky, 1978; Piaget, 1926; Slavin, 1995). Research in cognitive psychology has found that in order for the learner to retain information and to relate it to information already in memory, “the learner must engage in some sort of cognitive restructuring, or elaboration, of the material” (Slavin, 1995, p. 6).

Webb (1982) who has extensively studied the effects of cooperative learning argues: “students who gave or received explanations of how to complete the task scored higher than students who did not actively engage in group interaction” (p. 475). Repeatedly, studies have pointed to the educational benefits of giving and receiving explanations (King, 1992; Webb, 1989, 1991). Many of the aforementioned studies focused on face-to-face forms of collaboration. However, proponents of computer conferencing argue that many of these benefits can also be realized by engaging students in collaborative activity in online environments (Trentin, 2008; Anderson and Eloumi, 2004).

Despite the noted benefits cited extensively in the literature for online learning, some researchers have raised concerns that online learning is less socially robust. (Rovai, 2002; Krejins et al, 2007; Gunawardena et al., 1997). Computer-mediated communication is regarded as “less personal and possessing diminished social presence and social context cues when compared to face-to-face communication” (Rovai, 2002, p. 8). But in these mediums “students are not only looking for information; they are also looking for affiliation, support and affirmation” (Kreijns et al, 2007, p. 178). Wilson et al. (2006) contend that physical isolation inhibits the development of group trust and thus affects group effectiveness.

Through what medium does communication and interaction predominantly occur online? Through textual mediums: forums, emails, etc. It is written text that is the dominant form of
communication in online learning environments. Our aim is to study the nature of this text more closely in order to discern new dimensions to how communication occurs.

1.1 Motivation for Study

Text-based online environments are not solely used for educational purposes. On the web, there are a myriad of product rating forums, political forums, user guide forums, etc. To date, there have been many studies which look at the quality of communication in some of these environments: specifically political forums and product review forums.

But while text is the primary mode of communication in online educational environments, to date very little study has been done of the nature of the text, and how texts relates to communication and interaction in online learning. This study attempts to do just that.

The methodology employed in this paper is twofold:

A survey using a summated rating scale was employed for the assessment of Knowledge eCommons, and a rigorous quantitative and algorithmic data analysis was used with regards to the analysis of text. These divide into several methods that will be outlined in Chapter 3 explaining our methodology.

While quantitative data analysis has given us a view of the communicative landscape, further study is required to evaluate what these findings mean in terms of student relationships (that is, how these communicative features translate into real-life feelings that students develop for one another, their instructor and the environment) and this can be done using a comprehensive qualitative data study.
1.2 Research Questions

For the purpose of this study, we look at the questions presented below. Specific terminology used in these questions will be explained in the section that follows.

**Research Question 1:** What are student perceptions of social interaction in an online course environment?

Sub question 1.1 - Do students believe that online learning can be socially rewarding? (that is, can social interaction occur in that medium?)

Sub question 1.2 - What are their views on social interaction in KeC? Do they find it to be a socially robust environment?

Sub question 1.3 – Do students believe that they can successfully and comfortably project themselves in KeC?

Sub question 1.4 – In their view, what are the most socially robust aspects of the environment?

**Research Question 2:** Which linguistic characteristics of notes correlate positively with the recommend ratio? (the degree of favorable recognition students give to a note.) In other words, what linguistic characteristics of a note are well received by other students?

**Research Question 3:** Which linguistic attributes influence reply ratio? This question is meant to investigate what factors in language spark conversation and communication. What linguistic characteristic of a note entice readers to share their own views and keep the conversation going?
1.3 Language & Communication

Language is a dominant factor in connecting or isolating parties in a distributed environment. As noted by Tausczik and Pennebaker (2010) “words and language are the very stuff of psychology and communication” (p. 25). The traditional face to face social responses: smiles, laughs, frowns, disappointments, etc are all summarized within written messages and text becomes the dominant form of communication in these environments. Here we attempt to unravel text and find these relationships.

That is the motivation with which the project began: could we study and deconstruct text, to unravel information about student communication?

The student subjects studied for this project are graduate students studying at Ontario Institute for Studies in Education (OISE) at the University of Toronto. In this project, we begin by conducting a qualitative study on the feelings and perception of students on online learning in general, and Knowledge eCommons, the online learning environment used at OISE in particular.

This qualitative study presents an analysis of online discourse and the characteristics of messages that appear to promote extended conversations. From this vantage point, we conduct qualitative investigations of students’ attitudes, beliefs and experiences regarding online learning and online interaction.

In the second stage, we look at how certain language factors influence communication in online learning. We take the note as the unit of analysis, and conduct linguistic analyses on the note. The results of these we call the linguistic characteristics of a note – which will be thoroughly explained. Once the note is scored on its linguistic features, we look at the communicative characteristics of a note. These are the characteristics which show how others view that note. Our final aim is to look at this relationship, which is indicated by the “question mark” in Figure 1. That is, do linguistic features have a correlation with the communicative features? Are students prone to
respond to notes with particular linguistic features more than others? Are they more likely to prefer certain notes? Figure 1 is a general depiction of this study.

**Figure 1:** general outline of the project depicting a note and its linguistic and communicative characteristics.

Courses that were studied were graduate level courses offered on Knolwedge e-Commons. Knowledge eCommons is web-based educational software that creates a knowledge-building platform for computer mediated communication. This platform will be more thoroughly discussed in section 1.5.
1.4 Terminology

To more thoroughly analyze the research questions, some terms first need to be defined. They will be more thoroughly illustrated in the following chapter. The “linguistic characteristics” studied for each note follows:

**Sentiment analysis** - Sentiment analysis is a 2-dimensional characteristic of text concerned with two primary factors: the subjectivity/objectivity of text (in this case, the notes written by students) and the level of positive and negative emotion inherent in text. These two features together compose the two dimensions of the sentiment of a note.

- **Objectivity/subjectivity**: the level of factual knowledge reported is indicated by this variable. An objective note is one in which verifiable facts are referenced. A subjective note is based primarily on the feelings and thoughts of the individual writer.

- **Positive/negative emotion**: this variable computes the level of positive and negative emotion in a note.

Objectivity/subjectivity and positive/negative emotion are the two coordinates that compose the sentiment of a note. We thus compute the “sentiment score” of each note. This variable will be more thoroughly described in Chapter 2.

**Academic Word List**: the number of words used in a note from the academic word corpus. The academic word list is a corpus of words that appear regularly in English academic texts. Examples from this corpus include “Analyze”, “context” and “theory”. We look at the academic word list usage in each note and give each note an Academic Word List (AWL) score.

Having conducted this analysis on the linguistic feature of notes, we then look at how each note is viewed by other students. That is, which of these features seem to influence the students’
perception of that note? For example: do students prefer objective notes or subjective ones? To which kind of note are they most likely to respond?

Aside from the linguistic features of notes, we look at the “communicative features”, which signal communication. The communicative features of notes studied are: reply ratios, notes written and recommend ratio. An explanation of these words follows.

- Reply ratio (for each user): The Reply Ratio is the number of replies that a student receives per note written. This is used as one measure of each student’s impact on the class discourse.
- Notes written: Notes Written is simply number of notes written by an individual throughout the length of the course
- Recommend ratio: Knowledge eCommons provides students with a “ Recommend” button that students use to tag notes that they feel are valuable. The number of recommends each note has received is visible to all students. The recommend ratio is calculated by dividing the total number of recommends an individual receives by the total number of his or her notes. Like “Reply Ratio”, the Recommend Ratio is a measure of impact in communication.

Table 1 presents the linguistic and communicative characteristics we study.
Table 1

Notes and user attributes (linguistic and communicative characteristics)

<table>
<thead>
<tr>
<th>Note attributes (linguistic characteristics)</th>
<th>User attributes (communicative characteristics)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positivity – negativity</td>
<td>Reply ratio</td>
</tr>
<tr>
<td>Objectivity – opinion oriented</td>
<td>Notes written</td>
</tr>
<tr>
<td>Academic word Ratio</td>
<td>Recommend ratio</td>
</tr>
<tr>
<td>Number of Words per note</td>
<td></td>
</tr>
</tbody>
</table>

1.5 Knowledge eCommons

As previously noted, Knowledge eCommons (KeC) is a web-based educational environment that provides instructors a knowledge building platform – a CMC environment that supports knowledge building.

Although CMC (computer mediated communication) has a myriad of definitions, it refers to a mode of teaching and learning that utilizes a computer as a tool for communication, replacing the traditional face-to-face classroom (Groenke and Paulus, 2007; Walther, 2007). KeC was first launched in 2009 to support online graduate courses, but has since expanded to schools and intermediate level courses. Figure 2 presents an image of the KeC homepage.
Asynchronous communication describes delayed forms of communication (Anderson, 2004) – when there is an unlimited time lapse between one contribution by one user to the next, and time continuity between contributions can be hours or even days. Such features of the platform include:

- The class forum, which is a threaded discussion forum where students can write and respond to notes. There are academic and social forums created for the course. Class forums and threaded discussion has been shown to foster a level of learning in the classroom. The act of writing in online conferences may foster higher order thinking for reasons that have to do with the relationships between writing and cognition (Lapadat, 2002). Olson (1995) has argued that “writing enables us to say and think things that we could not, or at least have not, said and thought without writing” (in Lapadat, 2002).

- Asynchronous messaging service (where users can leave each other chat messages that can be responded to later).

- Class lounge: a forum for general discussions.
Figure 3 presents an image of the class forum.

Figure 3: KeC forum for Week 3 of a sample course, with 3 subforums.

Other asynchronous modes of communication that KeC supports are the recommend button.

Figure 4 illustrates the recommend feature of the forum on a sample note.

Figure 4: Recommend button in Knowledge eCommons which allows students to “recommend” their favored notes.
Knowledge eCommons provides students with a “Recommend” button that students use to tag notes that they feel are valuable. The number of recommends each note has received is visible to all students.

Synchronous communication on the other hand, is real-time communication, when one contribution is followed by the next instantaneously (within seconds or minutes). This is the way communication is predominantly carried out in face to face classrooms. Such features of the course were the webinar and chat room. Definitions for each feature follow:

- **Webinar**: weekly online conference meetings where students and instructor meet online for a predetermined time period and video and audio is transmitted real-time. Figure 5 presents an image of a sample webinar.

- **Chat room**: synchronous communication rooms where students can chat and talk to one another real-time. The use of the chat room is not compulsory but some student discussion leaders ask their group members to meet in the chat room.

![Sample Webinar](image)

**Figure 5**: Sample webinar depicting the conversation between 4 individuals.

Prior to this project, Knowledge eCommons classes in which student notes were to be linguistically analyzed were surveyed to assess what student perception was of the environment. This survey was meant to assess student perception of interaction in knowledge eCommons with respect
to their classmates and instructors. If feelings towards the learning environment were overtly negative, if students felt that the platform was not suitable for communication and was an impediment to it, or if students were struggling to communicate in the given language (English) – these are all factors that would influence the communication and interaction in the classroom, even prior to note analysis. If feelings regarding sociability were negative, examining the linguistic characteristic of notes in regards to social communication might prove problematic, since students did not think they could communicate to begin with. We wanted to assess student perception of interaction we therefore conducted a qualitative study using a survey instrument.

1.6 Why is this interesting?

In 1973, Mark Granovetter introduced the concept of “weak ties” in social network theory. Granovetter surveyed and interviewed hundreds of professionals for the purpose of his research, and he found that for 70% of those who had found their jobs through a contact, this contact was an individual who they did not consider a close friend, but rather a distant acquaintance (friends of friends or friends of friends of friends).

My first experience with online courses could not have been far from this: no weak ties, not even timid ones. I was a new graduate student at the University of Toronto, and although I lived on campus, the thought of not having to attend class was a very attractive one which I was too eager to try.

I was in for a surprise.

As the shy, quiet kid who liked to fade in the background, who was always more enthusiastic to write than to speak, the experience was a rewarding one. But as a new student who was hoping to expand my social circle through courses, the results were anything but. “No man is an island” the famous saying goes. In the case of online learning environments, he sure is.
I was further surprised to learn of the rapidly expanding terrain of online courses and online degrees, some of the statistics which I provide here. If online learning is increasing at such rapid speed, what will that mean for generations of students who will have less and less face-to-face classes to build their social, professional and personal ties?

The question almost becomes a post-modernist quandary. It was Jean Baudrillard (1995) who looked at how forms of communication any society or group employs in turn determine the social relations existent in that society. If we are increasing living in a state of hyperreality (a simulated version of reality) as he argues, then how do we create spaces in which the ties and bonds we need as professionals, as students, and more important as human beings are fulfilled and sustained?

All very big questions, which might exhaust dozens and dozens of theses. I thought I would start small and provide a landscape of what communication looks like in its current form in online learning environments, taking as unit of analysis the one thing that students write and read and ponder on the most in these mediums: text. If text is the main form of communication in these environments, doesn’t it deserve more study and observation?

I thought it did. That is what the rest of this manuscript will try to explore.
2 Literature Review

In the previous chapter, we outlined the need for an in-depth study of the relationships between communication and language, and posed the following research questions:

*Research Question 1:* What are student perceptions of social interaction in an online course environment?

*Research Question 2:* Which linguistic characteristics of notes correlate positively with the recommend ratio?

*Research Question 3:* Which linguistic attributes influence reply ratio?

To provide a thorough context for this investigation, the current chapter reviews the research literature relating to the three topics that bear most directly on this problem: communication and interaction in learning, communication in online learning environments and sentiment analysis. The chapter begins with an overview of research in the area of communication and interaction in learning. It is followed by a detailed analysis of the literature concerning communication in online learning environments and related online implementation issues. The final section presents different research perspectives on sentiment analysis, and the influences of sentiment analysis on student comprehension. These three sections are followed by a brief summary and conclusion.
2.1 Communication & Interaction in Learning

“Every function in the child's cultural development appears twice: first, on the social level, and later, on the individual level; first, between people (interpsychological) and then inside the child (intrapsychological). This applies equally to voluntary attention, to logical memory, and to the formation of concepts. All the higher functions originate as actual relationships between individuals.” (Vygotsky, 1978, p. 57)

There is a strong, vibrant literature in the modern learning sciences which emphasizes the role of social interaction in learning.

The common thread which connects these works is that they all study the relationship between interaction and achievement, the cognitive process and social-emotional mechanisms that bridge interaction and the learning process. Vygotsky and Piaget both argued that learning is socially constructed. Schmuck and Schmuck (2000), Johnson and Johnson (1998), Gardner’s work on interpersonal intelligence (1993), Robert Slavin (1995), B. F. Skinner (1968) and Paulo Freire (2000) are among other works analyzing the social aspects of learning.

Vygotsky looked at social interaction as a “primary source of cognition and behavior” (Huitt, W., & Hummel, J., 2003, pg. 1). He argued that the interaction of children around specific tasks to a large extent determines their command of those concepts. According to Vygotsky, “functions are first formed in the collective in the form of relations among children and then become mental functions for the individual.”(Slavin, 1995, pg. 4)

He defines the zone of proximal development (ZPD) as “the distance between the actual development level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers” (Vygotsky, 1987, pg. 86). As such, collaboration among learners in the same
range of development allows them to operate at levels more advanced than they would individually.

Interaction and collaboration increase individual developmental levels.

Writes Noreen Webb (1982), “to understand how group settings influence achievement, researchers need to focus on interaction as a mediating variable between the group setting and achievement” (p. 475). In her studies, she shows interaction to be a “key” predictor of achievement and a *mediating variable* between group dynamics and achievement (outcome). For example, looking at how group interaction facilitates problem solving for individual members of the group, she states: “students working in small groups are well-equipped to satisfy for giving and getting effective help” (Webb, 1989, p. 25).

King (1993) argues that “individuals understand and remember new material best when they elaborate on that material in some manner” (p. 111). He believes that elaboration makes learning material more meaningful and easier to understand. The role of the instructor is to facilitate this elaboration through robust collaborative group efforts.

Slavin (1995) calls the focus on cooperative learning methods “the greatest learning achievement of the 20th century” (p. 1) and emphasizes the tremendous benefits it has on student achievement – given the right conditions. Among the conditions he cites is that group goals and individual accountability need to be delicately balanced. Directly giving or receiving explanations reduces achievement, rather the group must be coordinated such that group members emphasize individual members’ understanding. In the case that one or members simply give away answers without this emphasis, “social loafing” (p. 8) – a process in which one or few students in the group end up with the entire work load – generally tends to occur. Such, the instructor must coordinate group goals so that all members are required to participate and understand the task given.
Structuring group interaction (p. 10) is another of the requirements Slavin (1995) cites. This entails “carefully structuring the interactions among students in groups.” For instance, it has been shown that summarization, explanations and reciprocal teaching are highly valuable learning tools. There is also the need to look at the type of learner and what best suits the individual student. Slavin’s findings show that cooperative learning helps both high achievers and those on the lower side of the spectrum.

Among other proponents of the social aspects of learning, Lave and Wenger (1991) depict a process they describe as “legitimate peripheral participation”: a highly interactive process through which learners partake a new task by interacting with a skilled mentor. This practice is not only favorable to the mentee but rather, both the mentor and the mentee are able to transform and refine their skills through this act of communication.

“Learning viewed as situated activity has as its central defining characteristic a process that we call legitimate peripheral participation. By this we mean to draw attention to the point that learners inevitably participate in communities of practitioners and that the mastery of knowledge and skills required newcomers to move toward full participation in the sociocultural practices of a community.” (Lave and Wenger, 1991, p. 28)

To them, learning is an “integral part of generative social practice in the lived-in world” (p. 35). In their opinion, the real problem lies in translating this into a theory and an analytical approach to learning.

According to Brown et al. (1989), “to learn to use tools [conceptual or physical] as practitioners use them, a student, like an apprentice, must enter that community and its culture. Thus, in a significant way, learning is a process of enculturation.” (p. 33)

They argue that learning is not only the mastery of abstract concepts, but “a function of the culture and the activities in which the concepts have been developed.” (Brown et al., 1989, p. 33)
Other research such as that of Bargh and Schul (1980), argues that from the teacher (tutor’s) point of view, there are two benefits to working in groups: first, content-specific gain, meaning that through the process of relaying the content to students, the tutor will be able to structure and organize that content better and that content will thus become “stronger and more accessible in memory.” Second is what they call “learning-to-learn” which facilitates the learning of the material at hand for the tutors. (p. 594)

This strong emphasis on the social aspects of learning and the vibrant social interaction that fits into the learning and teaching challenge also applies to online learning environments, which attempt to redefine the classroom online and transfer the learning process outside the physical barriers of a traditional class. Scholars argue that “theories of social constructivism should be applied in distance education and educational technology.” (Huang, 2002, p. 28) These environments need to deliver robust social interaction in order to progress the aims of learning.

And yet, the fact that numerous researchers and scholars in the field of online education and computer supported collaborative learning have warned of the lack of social interaction in online environments, signals that these mediums beg greater attention in this regard. Through the study of language and communication we attempt to shed new light on how social interaction occurs in these environments and thus, how it can be reinforced. What kind of language are students more inclined to respond to or read?
2.2 Communication in online learning environments

While the world of academia has been pondering the intricacies and complexity of the school classroom for the greater part of the 20th century, less is known of online classroom environments which have been developed and more widely used in recent years. And yet, they are being adopted at unprecedented rates. One of the greatest effects of the widespread use of the internet has been on education and our traditional perception of the classroom.

In the United States alone, between 2002 and 2008, the number of students taking at least one online course grew from 1.6 million to 4.6 million, a compound annual growth rate of 19%. This is while the overall higher education student body has grown at a rate of 1.5% during this same time period (Allen and Seaman, 2006, p. 5). ‘For the past several years, online enrollments have been growing substantially faster than the overall higher education student body.” (PEW/INTERNET, 2002, p. 67)

A critical look at these environments is necessary because they have significant differences with our perceptions of a traditional face-to-face classroom. The traditional social norms that dominate face-to-face interaction are replaced with new norms and patterns.

Computer-mediated communications are regarded as less personal and possessing diminished social presence and social context cues when compared to face-to-face communication (Rovai, 2002, p. 8). This is while in these mediums, “students are not only looking for information; they are also looking for affiliation, support and affirmation” (Kreijns et al, 2007, p. 178).

Drop-out rates tend to be higher in distance education programs (Rovai, 2002; Tyler-Smith, 2006). Kerka (1996) attributes this to the separation of students in an online environment which in turn reduces the sense of community and increases feelings of disconnectedness. Twigg (1997) on the other hand, attributes high dropout rates to feelings of isolation and distraction which are more
significant in distance education learning. Wilson (2006) contends that physical isolation inhibits the
development of group trust and thus affects group effectiveness.

These worries persist because historically, many have viewed teaching and learning as
inherently social endeavors and online constructive environments grounded in social constructivist
theories require a robust social environment (Gunawardena et al, 1997). But once the elements of
traditional “sociability” – face to face discussions, lectures, etc – are withdrawn from a classroom,
we must attempt to compensate with alternative methods.

Online learning environments promise more robust, sustainable learning communities
(Downes, 2007), democratized spaces (Díaz and Bontenbal, 2001), and support for more equitable
participation (Masters and Oberprieler, 2004). But “research on group learning shows that
asynchronous distributed learning groups (DLGs) utilizing computer supported collaborative
learning (CSCL) environments often lack the social interaction needed for these dialogue.” (Kreijns,
2003, p. 336) Gunawardena (1995) observes computer conferences and outlines her own negative
experiences: “the social interactions tend to be unusually complex because of the necessity to
mediate group activity in a text based environment. Failures tend to occur at the social level far more
than they do at the technical level” (p. 148).

This is while networking tools such as facebook and myspace have also become highly
popular among people (and learners) of all ages. But to date, the successful integration of social
networking and learning communities on a wide scale has not been achieved.

One common variable studied in the effectiveness of online learning communities is social
presence. Social presence is defined by “the degree to which participants are able to project
themselves affectively within the medium” (Garrison, 1997, p. 6).

Rovai (2002) believes many factors to be influential in the building of a community. These
include instructor immediacy, lurking, social equality, collaborative learning, group facilitation and
social presence. The latter – social presence - is emphasized as one of the most determinant factors in the building of an online community (Hewitt and Brett, 2007; Rovai, 2002; Aragon, 2003).

“Social presence involves fostering a commitment to shared goals among students, building trust and establishing productive norms of peer interaction.” (Hewitt and Brett, 2007, p. 1260)

In other parts of the literature, social presence in online environments is defined as “the degree to which a person feels ‘socially present’” (Leh, 2001, p. 110); “the sense of being present in a social encounter with another person” (McLellan, 1999, p. 40), and “the degree to which participants are able to project themselves affectively within the medium” (Garrison, 1997, p. 6). Social presence is defined as the “degree of salience of the other person in the (mediated) interaction and the consequent salience of the interpersonal relationships” This is interpreted as the degree to which a person is perceived as “real” in mediated communication.” (Richardson and Swan, 2003, pg. 3) and “the degree to which participants in computer-mediated communication feel affectively connected to one another” (Richardson and Swan, 2003, p. 1). Social presence indicates a level of comfort such that participants feel ease in the online environment and are able to participate in the discussions, forums and thus, learning process.

This connection to the environment has been called “co-presence”: “Mediated social presence is the moment-by-moment awareness of the co-presence of another sentient being accompanied by a sense of engagement with the other.” (Harms and Biocca, 2004, pg. 2) Aragon (2003) believes comfort and motivation to be key factors in social presence. How comfortable are students in participating? And to what extent to they feel motivated and drawn into this participation?

But social interaction alone is not the only factor which determines a high quality educational experience. “A community of inquiry must include various combinations of interaction among content, teachers, and students” (Garrison & Cleveland-Inns, 2005, p. 134). Garrison et al. (2005)
define three essential elements necessary to create a robust online learning experience: social presence, cognitive presence and teaching presence.

Berlanga et al. (2009) argue that in computer supported collaborative learning environments, social interaction and the gradual building of a community does not happen automatically. The main pitfalls they cite are taking social interaction for granted (assuming that it will take place because the technology is available) and the psychological dimension of the act of interacting (just because people interact this does not mean that will lead to the psychological state of interaction which requires trust, a sense of belonging, etc) (pg. 28).

The learning environment examined in this paper attempts to bridge these worlds and create a learning space embedded with robust features for social interaction and communication. Through this study we hope to find out more about the nature of social interaction in online environments.

Taking our cue from the literature, we used the survey to measure the following subconstructs, which we believe are the most important factors in social interaction in online environments. That is, we take them to be the building blocks (the subconstructs) of social presence and social interaction in online environments:

- Projecting oneself in the environment
- Co-presence (interacting with others and being aware of their presence)
- Instructor presence (responsiveness)
- Perceived learning in the environment

### 2.3 Sentiment analysis

This section will be a review of “sentiment analysis”, which is the major linguistic characteristic we study on students’ notes in Knowledge eCommons.
The use of modern text analysis goes back to the early twentieth century. In the 1950s, Gottschalk developed a content-analysis method to track Freudian themes in text samples. Pennebaker and his colleagues developed methods to analyze the text written by people experiencing emotional upheavals. (Pennebaker and Tausczik, 2010, p. 27)

“The history of the phrase sentiment analysis parallels that of opinion mining in certain aspects” (Pang and Lee, 2008, 10). Pang and Lee (2008) further state: “sentiment analysis” seeks to identify the viewpoint(s) underlying a text span.” Nasukawa and Yi (2003) state that “the essential issues in sentiment analysis are to identify how sentiments are expressed in text and whether the expressions indicate positive (favorable) or negative (unfavorable) opinions toward a subject.”(p. 70)

That is, sentiment analysis is concerned with extracting the sentiment, emotion, evaluations, judgments and feelings inherent in text (Wilson et al. 2005). The question is how these values are quantified and assessed algorithmically.

According to Pang and Lee(2008), the term “sentiment” referring to analysis of evaluative text and the rating of the judgment within the text appears in Das and Chena (2001), and Tong(2001). Today, such studies are done for various purposes. Cowell et al. (2006) study the sentiment of emails and chat messages. Feng et al. (2006) look at the sentiment in (non-academic) threaded discussion, and Malouf and Mullen (2008) examine the sentiment in online political debates. Pang and Lee (2004) study the sentiment in product and movie reviews.

Developing a sentiment detection algorithm has a myriad of benefits for various fields. For those involved in online marketing, it allows companies to automatically generate and group the positive or negative product reviews left for products. For movie reviews, for example as studied by Pang and Lee (2004) it allows the system to study all movie reviews for a given movie, and group them according to various sentiments expressed in the reviews.
Applications and appreciation for such studies have thus far extended to the realm of review related websites (e.g., product reviews, book reviews, etc), automatic advertising (e.g., detecting those advertisements which are not suitable for publishing), business and government intelligence.

For the purpose this study, we are looking at the sentiment inherent in educational notes as written in online threaded discussions on Knowledge eCommons. We think it is of crucial importance that this thread of computational linguistics be incorporate more strongly in online educational environments, where the source of communication is mostly text-based. The myriad of different communicative channels in face-to-face classrooms (physical gestures, body movement, facial expressions, etc) are absent in online classrooms, where text is the dominant form of communication. There may be information that we can extract from these texts, which give us a better (closer) perspective into the progress of students and the ways in which they communicate.

Using students’ notes in asynchronous discourse environments as our unit of analysis, we analyze the linguistic attributes of these notes based on well-established sentiment analysis techniques. Once the notes have been individually rated on sentiment, we then look at potential impacts of language on student communication.

Sentiment analysis is associated with two sub-tasks which are defined as sentiment detection and sentiment classification. This makes the sentiment of text a 2-dimesional variable. Figure 6 illustrates this 2-dimensional variable.

Figure 6: Two subtasks of sentiment analysis, sentiment detection and sentiment classification
Sentiment detection is categorizing units of text into documents that express opinions or personal perspectives vs. those that express objective (factual) accounts of matters (Pang and Lee, 2008). Sentiment classification is concerned with finding whether a unit of text has positive or negative opinion of a certain entity or occurrence (Pang and Lee, 2008). In this work, we process each written note in both of the mentioned dimensions.

For easier referencing, through the course of this work we will refer to sentiment detection as objectivity/subjectivity of text and to sentiment classification as the positive/negative emotion in text.

To distinguish between opinion oriented (subjective) notes from objective ones, we use the methodology provided by Wiebe et al. (2005). Objective notes are “material that is attributed to some source and is presented as objective fact” (p. 17). This is counter to opinion-oriented notes in which opinion or emotions are expressed (Wiebe et al, 2005). Wiebe et al. (2005) describe subjective notes as those that express “private states” – stemming from an individual’s thoughts, feelings, etc and thus not open to objective observation or verification. Pang and Lee (2004) describe subjective text as text which is based on the user’s opinion or feeling. Factual knowledge does not play a role in the text.

Finding algorithms to detect objective language in text is a thriving research field. However, the literature is focused on specific domains such as news articles (Wiebe et al., 2004; Stepinski & Mittal, 2007) and product reviews (Lui & Seneff, 2009; Ng et al., 2006). As with many other tasks in computational linguistics, sentiment detection is not necessarily domain independent (Pang and Lee, 2008) so we could not utilize methods from other domains to our task without extensive testing. We therefore decided to manually rate each note as objective or opinion oriented based on established definitions and using human judges. We explain the methodology of this further in chapter 3.
To measure positivity and negativity of language in notes, we employ text-analysis techniques (an algorithm) to determine the positive and negative emotion inherent in notes. This algorithm too is outlined in the next chapter. It functions by using a positive word list and a negative word list, respectively, to examine the frequency of positive (and negative) words in students’ notes. These lists were provided by a tool called LIWC (pronounced “Luke” - Linguistic Inquiry and Word count). “LIWC is a widely used, valid text analysis program that counts the frequency of words in psychologically meaningful categories, such as emotion.” (Carnelley & Rowe, 2010, p. 255) Studies have also shown that “LIWC accurately identifies emotion in language use” (Tausczik & Pennebaker, 2010, p. 32)

2.4 Academic Word List

As described by Coxhead (2000) “academic words (e.g., substitute, underlie, establish, inherent) are not highly salient in academic texts, as they are supportive of but not central to the topics of the texts in which they occur.” (p. 214)

By scanning 3.5 million words in academic texts, the author developed a word corpus of 2000 words which are most frequently used in academic text, but that are not among the first 2000 words in the General Service List (GSL) compiled by West (1953) which contains the 2000 most widely used and familiar words in English.

In this study, after computing the sentiment of each note, the other characteristic of the note that is examined is Academic Word List usage. The academic word count of each note was computed using Coxhead’s academic word list (2000). Similar to Hewitt (2008) we use this word list “to examine in detail the level of inquiry-based vocabulary usage in online courses.” (p. 2)
2.5 Summary

In this chapter, we review literature from the modern learning sciences which emphasizes the role of social interaction in the learning process. Learning scientists have long believed that social ties are necessary for learning to occur. Online learning environments however, while offering a promising new landscape, are in need of more investigation in this regard, given the isolating nature of these environments and the lack of social interaction which might ensue. That is precisely how this study hopes to contribute: in finding new keys to how students engage in online environments and how this interaction can be facilitated and maintained by the instructor and the tools the classroom provides.
3 Methodology

The previous chapters outlined the need for an in-depth study of the relationships between communication and language, and posed the following research questions:

Research Question 1 What are student perceptions of social interaction in an online course environment?

Research Question 2 Which linguistic characteristics of notes correlate positively with the recommend ratio?

Research Question 3 Which linguistic attributes influence reply ratio?

To provide a thorough context for this investigation, the previous chapter gave a literature review of the modern learning sciences and the background to sentiment analysis used in other fields. The current chapter reviews the methodology used in the context of this work. As described, the work was carried out in two stages: first, a survey instrument was used to question students on their perception of social interaction in online learning environments. Then, in the second phase, a qualitative study was carried out to more closely examine the effects of sentiment on communication in online learning environments.

What follows is a detailed outline of our methodology.
3.1 Overview

The methodology for this work will be described in two sections: the survey methodology and sentiment analysis methodology.

The source of data for this project was 2200 notes extracted from 11 distance education courses offered at Ontario Institute for Studies in Education (OISE) at the University of Toronto. The courses were hosted by the university and took place online in a learning environment named Knowledge eCommons (describe in Chapter 1). The notes were individually analyzed and rated. The methodology employed in this paper is twofold:

A survey using a summated rating scale was employed for the assessment of student perception of social interaction and social presence in Knowledge eCommons. This was done so that the researchers could gauge the student perception of the environment, their classmates and instructors. In the event that student perception towards these factors was negative, if they believed for instance, that Knowledge eCommons was an impediment to social interaction, or if they believed they could not communicate with their peers, this would give us serious considerations for future developments of Knowledge eCommons.

In the first stage, we found that not only did students not find the environment an obstacle to social interaction, but quite the opposite; they believed it was a constructive communicative environment that could foster social interaction. Once the results of that survey were studied, and researchers assessed adequate student perception of the learning environment, the second phase of the project began.

In the second phase, a quantitative and algorithmic data analysis was used with regards to the analysis of text (to rate linguistic characteristics of notes). These divide into several methods that will be outlined in this chapter explaining our methodology.
3.2 Survey Construction

This project began by first examining the social interaction and social presence of students within Knowledge eCommons, the online learning platform used for the purpose of these studies.

Fowler (2002) asserts that “sample surveys should be undertaken only after it is certain that the information cannot be obtained in other ways.” (p. 3) Accordingly, in this study, there is a strong rationale for using a survey for a number of reasons. First, we are analyzing a psychological construct that cannot be measured via physical means (rulers, calculators, etc). Also, this study involves questions of students’ private study habits and feelings and the participants must be given the opportunity to remain anonymous. Thus, direct involvement by the teaching instructor, faculty, etc cannot be utilized. There are psychological constructs being evaluated and we will quantitatively assess them using a self administered survey. Thus, for this research, an electronic survey was developed using Survey Monkey.

Fowler (2002) suggests that the reading and writing skills of the population, its motivation to cooperate, and computer use and skills are some of the most important criteria in the data collection method (p. 61).

Since the target population for this survey is graduate students taking an online course on Knowledge eCommons, they will have the necessary skills (linguistic, technological, cognitive) to participate in a self administered online survey. And thus, that is the mode of data collection that will be used. The skills required for participating in an online course include reading and writing notes, adding attachments, uploading photos, etc which are quite sufficient for completing an online survey.
Why self-administered?

“Ease of response is a priority to maximize returns” (Fowler, 2002, p. 62). A self-administered survey – questions that can be answered by simply checking a box - will be one of the easiest, least time consuming response methods for this particular group of respondents.

Also, it is believed that when dealing with sensitive topics, self administered procedures are thought to be best (Fowler, 2002, p. 63). Sensitive information is more accurately described in self-administered modes. Although this study does not deal with superbly sensitive personal issues, or ‘taboos’, it does ask questions pertaining to personal feelings and study habits, and thus a self-administered survey will be utilized.

Anonymity is also key – as many students would not wish to have their study habits and personal feelings about their peers or instructors be known to faculty, staff or the student body.

Why web based survey?

Today, a great majority of university students have daily access to the internet. “College students are frequently looking for email, with 72% checking email at least once a day” (Allen & Seaman, 2006). Many students at OISE solely take online courses for the duration of their semester, and thus have more access to the internet than the physical location of the school.

Amongst other reasons, Fowler(2002) contends “if a survey can be done over the internet, the costs per return potentially are the lowest of all” (p. 67). The potential high speed of returns and the ability of the computer to check for inconsistencies in the response are other advantages of online surveys.

Falenchuk (2008) points that web-based surveys provide improved opportunities to protect respondent confidentiality (as opposed to email surveys and personal interviews).
We did not send an initial invitation to students requesting to join the survey. Rather, we asked the instructors of the courses we were going to assess to send their students emails, with a link to the online survey, and asking them to participate (but not making it mandatory in any way).

Internet surveys also have the advantage of allowing participants time for thoughtful answers and checking records. Since this survey will involve questions regarding students’ behavior in online courses (how frequently they make posts or login to their course webpage, number of students in the class, etc) it is convenient to give them the chance to check more accurately for this information.

3.2.1 Rating Scale

Since this research will be attempting to measure psychological construct: social interaction, the use of a multiple-item scale is required. “Multiple item scales are preferred to yes-or-no questions because of reliability, precision and scope concerns” (Spector, 1992, p. 4).

Spector (1992) explains that summated scales are widely used across the social sciences to measure attitudes, opinions, and personalities (p. 1). They can be used to measure emotional states such as anxiety or happiness, or personality, job descriptors, etc.

In this survey, a five point Likert scale - a unipolar scale - is used because declarative items are used and “agreement items are declarative statements that one can agree with or not” (Spector, 1992, pg. 23). Table 2 depicts the rating scale utilized for this survey.

It should also be noted that a Neutral position is distinguished from a No Answer position. The former indicates that the respondents’ feelings are at midpoint (3) on the scale. Whereas the latter indicates that they have no answer. Also note that the numbers are evenly placed with one whole number difference between each choice to the next.
Table 2

*Rating Scale for the survey*

<table>
<thead>
<tr>
<th>Rating</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No Answer</td>
</tr>
<tr>
<td>1</td>
<td>Strongly Disagree</td>
</tr>
<tr>
<td>2</td>
<td>Disagree</td>
</tr>
<tr>
<td>3</td>
<td>Neutral</td>
</tr>
<tr>
<td>4</td>
<td>Agree</td>
</tr>
<tr>
<td>5</td>
<td>Strongly Agree</td>
</tr>
</tbody>
</table>

3.2.2 The Survey

The survey is composed of 3 sections: general questions, students’ thoughts on social presence and social interaction (where the survey instrument was used), and open ended questions. Explanations for each section and the relevant parts of the survey follow.

1. The first section consists of general questions about English fluency, access to the internet (minimum requirements for taking an online course) and Knowledge eCommons overall layout. These questions were asked because these basic factors can serve as a great impediment to interaction and could greatly jeopardize student ability to communicate in the course. These general questions are shown in Table 3.
2. The second section of the survey was questions regarding social presence and social interaction in the course. This is where the survey instrument was used with a 5 point rating scale. Taking our cue from the literature as outlined in section 2.3, we chose the following subconstructs (building blocks) as the building blocks of interaction in our survey.

- Projecting oneself in the environment. The following questions in the survey (presented in Table 4) were chosen to evaluate this subconstruct:

Table 4

Questions in the survey meant to assess “projecting oneself in the environment”

| | 
|---|---|
| I felt comfortable using the online technologies for the course. | 
| I felt hesitant sharing my personal opinion on the course forum. | 

- Interacting with other users (co-presence). The following questions in the survey were chosen to evaluate this subconstruct (demonstrated in Table 5).
Table 5

Survey questions for “co-presence”

<table>
<thead>
<tr>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Through our online communication, I had the opportunity to know some of my classmates.</td>
</tr>
<tr>
<td>I enjoyed reading my classmates’ writings on the forum.</td>
</tr>
</tbody>
</table>

- Instructor presence. The following questions in the survey were chosen to evaluate this subconstruct:

Table 6

Instructor presence in the survey

<table>
<thead>
<tr>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>The course instructors were active in the class.</td>
</tr>
<tr>
<td>My instructor responded to my emails and messages within reasonable time.</td>
</tr>
</tbody>
</table>

- Perceived learning. The entries presented in Table 7 are those survey questions which were meant to assess students on their perceived learning in Knowledge eCommons.

Table 7

Perceived learning questions

<table>
<thead>
<tr>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online web-based education is a good medium for social interaction.</td>
</tr>
<tr>
<td>I only take online courses when I am forced due to timing or distance considerations.</td>
</tr>
</tbody>
</table>
I was able to learn from the online discussions.

We also incorporated specific questions about features of Knowledge eCommons. We included questions regarding KeC features specifically, to know what students thought of them and how they could better serve student interaction in future versions. These general questions are presented below in Table 8.

Table 8

*Questions for assessing Knowledge eCommons features*

<table>
<thead>
<tr>
<th>Questions for assessing Knowledge eCommons features</th>
</tr>
</thead>
<tbody>
<tr>
<td>I didn’t find the recommend button useful.</td>
</tr>
<tr>
<td>The recommend button encouraged me to write better notes.</td>
</tr>
<tr>
<td>I was more likely to pay closer attention to those notes which were recommended by more of my classmates.</td>
</tr>
<tr>
<td>The course webinar helped me get to know the other students better.</td>
</tr>
<tr>
<td>Getting to see my classmates and instructors via the webinar made the class feel just like a real classroom.</td>
</tr>
<tr>
<td>I didn’t think the webinar feature of the course had any learning value.</td>
</tr>
<tr>
<td>Writing private notes to other students was a helpful feature of the learning environment.</td>
</tr>
<tr>
<td>I will keep in touch with some of the classmates and/or instructors that I met online.</td>
</tr>
</tbody>
</table>
Displaying a personal photo "avatar" (i.e., the image that appears on each note) contributes to a sense of community.

3. In the third section of the survey, we ask students to describe their experience with the course in 3 open-ended questions:

Table 9

Open ended questions in the survey

<table>
<thead>
<tr>
<th>Social presence, in an online environment, is defined as &quot;the degree of awareness of others in a communication interaction&quot;. Which features of Knowledge eCommons, if any, fostered a sense of social presence in the course?</th>
</tr>
</thead>
<tbody>
<tr>
<td>What kinds of activities and instructional methods in the course encouraged a sense of social presence?</td>
</tr>
<tr>
<td>Compared to a regular face-to-face course, how &quot;socially present&quot; were your classmates in this course? (To what extent did you feel you were getting to know your online classmates?)</td>
</tr>
</tbody>
</table>

This last section was added so that students could have the opportunity to give us their own feedback about Knowledge eCommons, not limited to questions we pose for them. In fact, as the results section will show, the responses gained in this section were instrumental in helping us understand the dynamics of the class.

The full survey is provided in Appendix 1. The results will be reported and analyzed in the following chapter.
3.3 Language Sentiment Analysis

As previously noted, sentiment analysis is associated with two sub-tasks which are defined as sentiment detection and sentiment classification.

Sentiment detection is categorizing units of text into documents that express opinions (personal perspectives) vs. documents which present objective accounts of matters (Pang and Lee, 2008). Sentiment classification on the other hand, is concerned with finding whether a unit of text is expressing positive or negative opinion (Pang and Lee, 2008). In this work, we process each written note in both of the mentioned dimensions. Each of these dimensions will be explained separately.

3.3.1 Sentiment detection Instrument

To distinguish opinion oriented notes from objective ones, we use the definitions provided by Wiebe et al. (2005). Objective notes are “material that is attributed to some source and is presented as objective fact” (p. 17). This is counter to opinion-oriented notes in which opinion or emotions are expressed (Wiebe et al, 2005). Wiebe et al. (2005) describe subjective notes as those that express “private states” – states of opinion and dialogue not open to objective observation or verification.

Peng and Lee (2008) demonstrated that differentiating between opinion oriented (subjective) texts and objective ones is indeed difficult. They cite numerous examples of this difficulty, such as this letter by Charlotte Bronte to George Lewis (pg. 20):

"Miss Austen is not a poetess … [she] has no eloquence, none of the ravishing enthusiasm of poetry."

While the first sentence “Miss Austen is not a poetess” is an objective sentence and verifiable, the second, “has no eloquence, none of the ravishing enthusiasm of poetry” is not, and
should be considered an opinion - even though the two sentences seem to convey the same information.

While to date, algorithms for scoring objectivity/subjectivity in certain fields (product reviews, movie reviews, etc) have been developed, no such algorithm exists for educational forums. This is while sentiment detection is a highly context-specific domain. “In general, sentiment and subjectivity are quite context-sensitive and, at a coarser granularity, quite domain dependent.” (Pang & Lee, 2008, p. 21)

We thus employ a method used by Cowell et al.(2006) and other scholars where objectivity was rated manually by three judges. Each note was rated by at least one judge and 20% of the notes were rated by all three judges to calculate inter-rater reliability. The degree of inter-rater reliability was calculated using Krippendorff’s alpha(2004). This was done to assess the rater reliability of the instrument we developed. Alpha was within the accepted limits as shown in the results section.

Krippendorff’s alpha is a reliability testing method which accounts for chance agreement among multiple coders (Stein et al. 2007, p. 106). The minimum acceptable level of 80% indicates a content analysis study to be considered reliable (Strijbos and Stahl, 2007). Krippendorff’s alpha was computed using SPSS.

To rate the notes, we had to devise guidelines to give out judges. These guidelines were printed and given to all three judges. No further explanations were provided individually by us to the raters in order to avoid rater bias. The guidelines will be thoroughly described in the following section.

3.3.2 Sentiment Detection Guidelines

The first criteria on the guidelines are as follows:
• *The note must be verifiable* (the text written must be connected to another source which is accessible, be it through search engine, documents, etc).

However, given that this is an educational forum (and this is where the context-specificity comes into play) a sentence such as this:

“E.T. goes home at the end of the movie” is not considered objective on its own. It has to fit with the greater academic conversation taking place.

Thus, our second criteria is:

• *This verifiable source must be related to the topic discussed in the forum.*

The way courses are structured in Knowledge eCommons, each week a group of moderators are responsible for posting questions related to the week’s readings and/or audio/video that the instructors assign. Thus, the note must be in response to the discussion, and a random verifiable note (such as “E.T. goes home at the end of the movie”) does not indicate an objective note (unless it fits into the academic discussion).

• The third criterion is that *the source to which the comment refers must be noted.*

Thus, “I remember some professor saying that knowledge building is a useful paradigm” does not indicate an objective note. A degree of referencing is required to show that the writer clearly and *intentionally* means for the note to be verifiable and connected to an authority beyond their own opinion.

Which brings us to our final criterion:

• Intentionality

The writer’s text must be intentionally verifiable. That is, they write by expressing factual knowledge and are aware of doing so.

To summarize then, our criteria were:
1. An objective note is defined as a note in which the writer uses objective sources or facts which are traceable and verifiable. Thus, citing any of the following sources makes a note objective:
   - published papers
   - organizations
   - quoting known experts
   - results of concrete experiments
   - any other credible source (leaving “credible” for judges to evaluate)

2. The fact(s) argued must be in some way related to the week’s discussions.

3. This verifiability must be conscious: referencing to vague individuals or sources does not constitute an objective note.

4. Quoting the instructor is only objective if is it done in a scholarly manner. So, for example, “Howard [instructor] would like to know what we will be doing” is not considered an objective note. However, “Howard argued in last week’s class that a teacher’s instructional practices must be evaluated by a peer” is considered objective.

   Quoting other students is not considered objective unless those students are considered an “expert” in the discussion.

5. Merely stating a source without explanation is not objective. (e.g., “I remember some professor saying that knowledge building is a useful paradigm”). The source needs to be verifiable.

   The guide we provided to raters is presented in Appendix 2.

   An objective note was rated “1”. Any note that is not objective is considered a subjective note and rated “0”, making this variable a binary variable. In future studies, we wish to take this
variable from a binary variable to a percentage. This will be feasible once a computer algorithm has been developed.

Now that the guidelines have been drawn, we will provide a few examples. The following sentences below constitute objective notes. The reasons are also given.

“I remember reading in the first chapter of “Introduction to Knowledge Building” that the teachers’ role is to circulate the students who work together and discuss ideas and monitor progress.”
[reason: clear referencing to an academic source]

“I did find an advantage that, according to the article, "Teachers can find a wide variety of WISE curriculum projects that have been carefully authored, tested and refined by our research group” (Slotta and Linn, in press)”
[reason: clear referencing to an academic paper]

“Last week Howard[course instructor] and I got into a discussion about pedagogical practices in various cultural divides. While I argued that these practices are universal, Howard pointed out that in field work he has done in rural parts of Afghanistan, he has witnessed up close that these practices can be highly context and cultural dependent.”
[reason: referencing course instructor who is a scholar in the field.]

This was an extremely labor intensive process that was made more intense due to our need to compute inter-rater reliability for our instrument which meant that not one but three judges were required. Studying sentiment classification methods done in other areas, and putting in our own criteria, we constantly fine-tuned our instrument until a coherent one was developed.
However, before we can develop algorithmic methods to compute objectivity in educational forums, these steps were highly crucial and necessary. That is why we decided to undertake such a cumbersome task.

For each note, the positive/negative emotion is also computed, which will be described next.

3.3.3 Sentiment Classification instrument

For conducting sentiment classification – which is defined as evaluating the positive or negative emotion inherent in a note - we employ text-analysis techniques. We do this by using two word lists: a positive word list and a negative word list. These lists allow us to measure the frequency with which positive and negative words are used in text. These lists were provided by LIWC (Linguistic Inquiry and Word count) which is a text analysis program with built-in positive and negative word corpuses.

Throughout ten years of research, LIWC developers have created a word corpus for positive and negative emotion (respectively) by the use of computational linguistics techniques and the comparison of these results with human raters. LIWC ratings of positive and negative emotion words correspond with human ratings of the writing experts (Alpers et al. 2005).

Tausczik and Pennebaker (2010) also note that the just like objectivity, emotion too is context dependent. In an examination of 2800 random texts, emotion words were negatively correlated with articles, prepositions, and relativity words. This is while emotion words were positively correlated with pronoun use, auxiliary verb use and negation use. All correlations were significant (p < 0.01). These relationships were also incorporated into the development of the word corpus.

Using the LIWC tool, each note was rated on positivity of emotion, on negativity of emotion and the difference was the emotive score of that note. Figure 7 is a snapshot of the LIWC tool.
Figure 7: LIWC (linguistic inquiry and word count) software

As input, this tool takes individual notes, and the output is produced in a text (.txt) file. A sample output file is presented in Figure 8. The output has three fields: file name, posemo and negemo.

Figure 8 – Sample LIWC output which shows the file name, positive emotion and negative emotion.

Posemo is the percentage of words in the note that belong to the positive emotion corpus, and negemo is the percentage of words in the note that belong to the negative emotion corpus. As sentiment classification is defined as the emotion inherent in notes, we subtract negative emotion (negemo) from positive emotion (posemo) to come up with the final emotion score for the note.
3.3.4 Sentiment Scores

Thus for each note, we have an objectivity score and an emotion score. Figure 9 shows a sample note, written by user “nom. 7” which has been scored in both dimensions. The scores are shown on the right hand side of the image.

Figure 9: sample note with objectivity/subjectivity and emotion scores.

Taking this user (user Nom. 7) as an example, we explain the procedure from here:

Once all the notes have been scored, we have a database which contains the objectivity score and emotion scores of all the notes written by each user.

We take all of nom. 7’s objectivity scores and compute an average. We also have his emotion score, and compute an average of that too.

Thus, nom. 7 has an overall objectivity/subjectivity score, a variable we’ll call avg_obj. Nom. 7 also has an overall positive/negative emotion score which we will call avg_emot. Figure 10 depicts Nom. 7’s sentiment score.
These are the two coordinates required for a sentiment score. Therefore, nom. 7 obtains a 2-dimensional sentiment score using these two variables, equivalent to: \([\text{avg}_\text{obj}, \text{avg}_\text{emot}]\).

This accounts for the sentiment computation of notes.

The other linguistic characteristic obtained from notes is the academic word list score which we describe next.

### 3.4 Academic World List Computation

The academic word corpus is offered by Coxhead (2005) and for the purpose of this project, students’ notes were individually rated for academic word list usage.

The list is available online and can be accessed through the *Using English for Academic Purposes* website. This list is a list of 2000 most frequent words used in English academic text. Assessing the academic word list usage of students in notes has been previously done by Hewitt (2008).

We rated each note on academic word list separately. A computer program was written that takes a note as input. It traces the program word by word to find words from the Academic Word List. It then computes a percentage (a number) as output. This number is the percentage of academic words used in the text.
This program is depicted in Figure 11 as the “Academic Word List” calculator. The output this program computes indicates what percentage of the words used in a note is from the Academic Word List.

![Academic Word List Calculator](image)

**Figure 11**: Academic Word List calculator which computes AWL from the student note

So for example in this sample note:

“I think the environment needs to be analyzed for positive reinforcements.”

The words “environment”, “analyzed”, “positive” and “reinforcement” are from the Academic Word List. Thus, the AWL score of this note would be: 36.3% or 0.363. In real life, notes are usually much longer and so the number of non-academic words (general words) used are much greater thus yielding a smaller AWL score.

### 3.5 Putting it all together

After our linguistic analysis is completed, for each user, we have a table that looks like Table 10.
Table 10

List of variables computed for each user

<table>
<thead>
<tr>
<th>Variable</th>
<th>--</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objectivity/subjectivity</td>
<td>--</td>
</tr>
<tr>
<td>Positive emotion/negative emotion</td>
<td>--</td>
</tr>
<tr>
<td>Academic word list</td>
<td>--</td>
</tr>
<tr>
<td>Average Number of words per note</td>
<td>--</td>
</tr>
<tr>
<td>Reply ratio</td>
<td>--</td>
</tr>
<tr>
<td>Notes written</td>
<td>--</td>
</tr>
<tr>
<td>Recommend ratio</td>
<td>--</td>
</tr>
</tbody>
</table>

Average number of notes per note, reply ratio, number of notes written and recommend ratio are computed automatically using Knowledge eCommons’ administrative features.

Figure 12 visually depicts the statistical analyses we did at this point in the project. We took the linguistic characteristics of notes to see whether there would be any correlations between these linguistic characteristics and communicative features of the notes. The communicative features we look at are the user’s reply ratio, recommend ratio and notes written as we have previously discussed.
In the following chapter we discuss and outline our results.
4 Results

This chapter, which will describe the results of our study, is presented in two sections. The first section will demonstrate the results drawn from the internet survey which was written with respect to Research Question I. The second section will illustrate the results of the sentiment analysis studies and will answer Research Questions II and III.

4.1 Survey

In the following subsections, we will outline the results obtained from the survey instrument. We will outline their perception of learning and social presence in the class, as well as their opinion on the class webinars, recommend button and other social feature of Knowledge eCommons.

4.1.1 General Questions

Figure 13 presents the accumulated results to the first question on the survey. This question asks the number of online courses students have taken to date, counting the current course. The general questions were designed to assess the minimum requirements for taking an online course: knowledge of English, access to the internet, etc.

The general questions in the survey assess:

- The number of online courses taken by the student
- Access to the internet
- Fluency in English
- Knowledge eCommons ease of use
Nearly 70% of the students have taken 3 or more courses, and approximately 90% have taken more than one online course.

Figure 14 demonstrates the results of the second question. 100% of students had access to the internet at home, and 68% at work.

As seen in Figure 15 81.3% of students were very fluent in English, with 100% choosing “very fluent” or “fluent”. The weighted average for fluency in English was 1.43, (1 being “very fluent” and 5 “not fluent”).
Next, students are asked about the ease of use of Knowledge eCommons, the online learning environment they used for the purpose of this study. Results can be found in Figure 16.

**Figure 15:** English fluency for students in the sample (n = 24)

<table>
<thead>
<tr>
<th>3. How fluent are you in English?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale</td>
</tr>
<tr>
<td>81.3% (13)</td>
</tr>
</tbody>
</table>

**Figure 16:** Knowledge eCommons ease of use

<table>
<thead>
<tr>
<th>4. Please rate Knowledge eCommons on ease of use:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale</td>
</tr>
<tr>
<td>75.0% (12)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5. Please rate Knowledge eCommons on visual appearance:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale</td>
</tr>
<tr>
<td>18.8% (3)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6. Please rate Knowledge eCommons on response time (speed):</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale</td>
</tr>
<tr>
<td>50.0% (8)</td>
</tr>
</tbody>
</table>

Ease of use of knowledge eCommons: the rating average for knowledge eCommons ease of use was 1.25. This indicates that students found KeC to be user friendly, an attribute which has been frequently cited as a crucial feature of online learning environments. The rating average for visual
appearance of knowledge eCommons stands at 2.0, and for response time (speed) at 1.5 (with 1.0 indicating “fast speed”).

The first section of the survey which was just discussed questioned students on general perceptions and requirements of online learning environments. The second part of the survey was a survey instrument used to assess the level of perceived social interaction in the environment which will be a study of Research Question 1. Specific questions were asking regarding students’ thoughts on particular features of the environment at the end.

4.1.2 Research Question 1. What are student perceptions of social interaction in an online course environment?

As discussed in chapter 3, the survey instrument was designed to assess the following constructs:

- Projecting oneself in the environment.

As also outlined in that chapter, the following questions in the survey were chosen to evaluate this subconstruct:

- I felt comfortable using the online technologies for the course.
- I felt hesitant sharing my personal opinion on the course forum.

Figure 17 demonstrates the results of these questions. The computed value for this variable was 1.17. (with “1” indicating full projection within the environment). Garrison(1997) and Woods(2002) stress the importance of student projection in creating a sense of sociality and social presence within the learning environment. Thus, survey results show that students in Knowledge eCommons believe a robust projection of themselves within the learning environment.
- Perceived interaction with other users (co-presence)

This subconstruct was evaluated using the following questions:

- Through our online communication, I had the opportunity to know some of my classmates.
- I enjoyed reading my classmates’ writings on the forum.

Co-presence within the environment was measured to stand at a 1.67 average according to survey results (with 1 being maximum co-presence). It is important to note that while this is a robust
number, none of these variables alone guarantee viable social interaction. Rather, a non-linear combination of these constructs is what provides this.

- Instructor presence

This construct was measured using the following questions:

- The course instructors were active in the class.
- My instructor responded to my emails and messages within reasonable time.

![Table 3: The course instructors were active in the class.]

<table>
<thead>
<tr>
<th>Scale</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Neutral</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40.0% (6)</td>
<td>40.0% (6)</td>
<td>20.0% (3)</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
<td></td>
</tr>
</tbody>
</table>

![Table 7: My instructor responded to my emails and messages within reasonable time.]

<table>
<thead>
<tr>
<th>Scale</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Neutral</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>7. My instructor responded to my emails and messages within reasonable time.</td>
<td>73.3% (11)</td>
<td>20.0% (3)</td>
<td>8.7% (1)</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
</tr>
</tbody>
</table>

*Figure 19:* Instructor presence perceived by students

Instructor responsiveness was considered a crucial component of social interaction in online learning environments. Research has shown that when students feel their instructors’ involvement within the virtual class, they are more likely to socialize themselves (Richardson and Swan, 2003). These researchers argue that students, who have high perceptions of social interaction in online learning environments, also score highly on their satisfaction with the instructor and their perceived view on instructor presence. The rating average for this construct was 1.57.

- Perceived learning in the environment

- Online web-based education is a good medium for social interaction.
• I only take online courses when I am forced due to timing or distance considerations.

• I was able to learn from the online discussions.

Figure 20: Perceived learning (level of learning students believed to have taken place)

We asked students if they felt an online environment was suitable for learning, and whether they perceived Knowledge eCommons as an environment in which learning takes place. The rating average for this construct was 1.84.

Richardson and Swan (2003) argue that perceived learning is an important factor in overall social presence in the online classroom, thus connecting the relationship between achievement (learning) and the social aspects of the classroom. Social interaction is not only required to give students a sense of community and membership, but rather, it affects learning outcomes as well. Without a strong sense of community and belonging, the learning goals for the course may not be achieved.
This questionnaire was not designed to quantitatively compute social interaction in KeC, but to provide a schematic of how students’ perceived social interaction in the environment, and whether they thought interaction existed in their virtual classroom.

We also incorporated specific questions about features of Knowledge eCommons. We included these questions regarding KeC features specifically, to know what students thought of them and how they could better be restructured or designed in future versions of KeC.

**Recommend Button:**

The students felt that the recommend button was “useful”, but they did not believe that it encouraged them to write better notes (over 50% did not agree that it encouraged them to write better notes). They were fairly neutral regarding whether notes that were recommended were those they paid more attention to, with 53.3% agreeing with this proposition and 40% disagreeing and or being statistically neutral. These results are shown in Figure 21.

**Figure 21:** students’ views on the recommend button added to Knowledge eCommons

This result is interesting in light of the findings of our qualitative analysis (covered in the next section) which show that those notes students recommended were not necessarily those to
which they responded to. This might be one explanation: they find it a useful tool, but they do not
find that it encourages them to write better notes. In other words, the recommend button may be a
good highlighting tool, but it does not function as a writing scaffold. To help students write better
notes, other forms of technology need to be utilized and tested.

Webinar:

The students believed the webinar feature of the course to have “learning value”, but it is
still unclear what value they gave it. They also believed that this feature made the class “just like a
real classroom” and helped them “get to know the other students better.”

As online classrooms become ubiquitous, mediums like the webinar, which create a hybrid
online/face-to-face environment need to be more avidly exploited and employed. Many researchers
in the field have stressed the importance of videoconferencing tools in online environments
(Cavanaugh, 2001; Anderson, 2008; Saw et al., 2008). From the student surveys and questionnaires,
we see that they place a great deal of emphasis on this feature, and we hope to expand its reach in
later versions of Knowledge eCommons.
**Figure 22:** Student perception of the impact of the webinar on learning

**Private notes:**

Private notes are being more and more utilized by the students, and this is a trend we are seeing grow course by course. They found this feature a useful tool, with 80% agreeing that it was a “helpful” feature of the learning environment.

**Figure 23:** Student perception of private notes

**Profile photos (avatar):**

Students believed that using the profile photos increased the sense of community within the class with 80% agreeing that it “contributes to a sense of community”.
4.1.3 Sub question 1.1 - Do students believe social interaction can occur in online learning environments?

One of the primary social features of a face-to-face classroom is “getting to know” other individuals which participate in the classroom environment. “getting to know” the teacher, other students, etc. We wanted to study whether students believed this was possible in an online environment and whether they believed they were interacting with and getting to know the other students. One of our greatest aims in creating a socially robust environment is to foster a space in which students are interacting, and thus getting to know one another, as is the culture of a face-to-face classroom. We wanted to know to what extent that was happening in KeC.

To make this inquiry, we asked the students to elaborate on: *To what extent did you feel you were getting to know your online classmates?*

The students’ responses follow:

- I felt that between the webinars and class lounge, my peers were socially present in the sense that I have gotten to know them well enough to speak to them on a more regular basis (i.e. outside of the course).

- I feel that the flex-mode offering encouraged a deeper level of engagement because we had various avenues to interact with our classmates and instructors.
• Some extremely well - some not at all. Those who regularly participate (perhaps half) demonstrate great netiquette and share snip bits of themselves throughout their posts. The other half, however, disappeared entirely.

• Pretty decent by the end of the course. Initially, it was not easy at all.

• I felt that the ones that were present on the webinar were very socially present. Those students who were not regular webinar attendees, I felt I had little connection to.

• Those who were active more online, I felt I had more of a connection to them

• Very well

There are two themes here: one the emphasis on the webinar, which we have seen before and will continue to see. The other is that “getting to know someone” is a reciprocal occurrence: students feel a connection to those students who are making an effort at being socially present within the class.

This may highlight considerations to be made by instructors and course coordinators as they encourage students to more enthusiastically participate in social dimensions of the course. For instance, having more frequent webinars, active participation in the webinars, student leadership of the webinars, etc may encourage students to interact.

4.1.4 Sub question 1.2- what class activities and/or instructor activities do students see as facilitating social interaction?

When asking students to name the class activities which they believe facilitate social interaction, they gave the following responses.

The webinar

• The webinar was a great activity.
• The webinars definitely encouraged social interaction. I felt like I didn't get to know those people who were unable to attend the webinars as well as those who I saw on a weekly basis.

• The webinar, if you were participating in real-time. When I participated (instead of watching the archived version), I was able to see and hear my classmates (very important), and directly discuss ideas. We were also able to IM while the webinar took place which helped to develop a social presence because it sometimes went off-topic.

• Webinar

• Integration of other technologies like "Etherpad" and webinars.

• The webinars in a way although some people were not quite forthcoming and took more of an observer's stance

The students view the webinar as a tool that can foster social interaction. One point that students stress repeatedly however, is that this tool is only socially helpful with respect to those students who are themselves active users of the tool. That is, those who only watched the recorded video, or did not actively participate in the webinar class are left out. Even those who took an “observers” stance were left out. As scholars have noted in the past, even when technology does afford us tools to socially communicate, without sound pedagogical and social planning, these tools will not have the desired outcomes. This points to the need to find ways of encouraging student involvement in the webinars, specifically students who are more hesitant to participate for various reasons: shyness, linguistic or technological barriers, etc.

**Forum**

• In the discussions about the readings each week and facilitation of the discussion together with a group member
• Having a discussion board for everyone to introduce themselves is a great way to start the course. I really liked the Krejins' article and thought it drew attention to the importance of social presence in courses. Additionally, some of the student moderators did some neat things (synchronous chats, posting avatars, etc.)

• The discussions and the journals that were someone's activity.

• encouraging discussion, encouraging responses, "like" button, comments on personal experiences

• responses to your contribution

• reading and software discussions

Students continuously cited the course forum as an activity which fostered social interaction. The course forum is the main academic portion of the class, where discussions on the weekly readings are held. It is encouraging to see that students themselves view these academic discussions as a way of socially interacting with their fellow classmates.

Rather than seeing the forum as a static entity where students can go to write their opinion, the students stress that it is the process of discussion that the forum entails which makes it a viable tool for social interaction. They thus use the words ‘encouraging’ discussion, and ‘responses to your contribution’.

As previously discussed, one of the greatest challenges to fostering social interaction in online learning environments is using existing tools towards more sociable ends. The forum is one such environment which has the potential to be a viable social tool or a stagnant forum.
Other

- Class chat that was initiated later in the course so that students were already a bit comfortable with each other. Relaxed hybrid meetings were also really nice.
- Online chats, wiki and cooperative learning.
- face-to-face meetings
- On-line chat discussions promoted by classmates in realtime
- Inclusion from the instructor, Bio activity, Journal activity, collaboration projects
- Collaborative assignments
- Collaborative assignments we did together.

There are a number of considerations to be made here: the students did not name “chat” as a feature of KeC which fostered interaction, but rather, as an activity “promoted” by group leaders. We have seen from experience that the chat rooms in online learning environments are typically not used by students, until that feature is incorporated into the weekly activities by group leaders, instructors, etc. Also, it is interesting to note that students would like to see instructors give them a sense of inclusion, or provide collaborative assignments.

4.1.5 Sub question 1.3 – Do students believe that their classmates can successfully and comfortably project themselves in KeC?

We wanted to question students on how well they were able to project themselves in KeC, especially when compared to face to face classrooms. So we asked them: *Compared to a regular face-to-face course, how "socially present" were your classmates in this course?*
• There were some that I felt I was getting to know very well, and others very little. In many ways it was like a normal class in that regard.

• especially those who were present during the discussions, I have learned to know as being 'social present', this was the majority of the group; in some f2f courses I had less contact with so many but maybe deeper contact with few.

• Classmates would include their personal experiences that enabled you to understand what else was going on in their lives besides the course

• More than in other courses. Instead of people posting to impress the prof, people eventually realized that it's important to post what you really feel and use 'normal person' language, not edu-babble.

• In this course, I felt a sense of social presence. However, only after at least 2 months of work did I develop a social presence.

• Having a face to a name/post is very helpful; this made me feel that my classmates were quite "socially present"; in some cases, I feel that I interacted more with my peers in this format that I would in person

• Very close to the level of face-to-face, but not fully there!

• this course was dismal for me initially as no one really wanted to showcase personal qualities. It was all about the work work work! Only in week 9 when an activity was specifically designed towards social presence did i feel that i got to truly know my classmates.

• the same

• some were too socially present
We must pay attention to a message students send over and over again: social interaction does not just occur out of nowhere. Rather, they argue that instructional methods and time were important factors which contributed to this sense of social presence.

They also continuously acknowledged the role that other students played in their feeling of connectedness to the environment. Classmates who participated actively in the webinars, who shared personal experiences, and who were more active in the class were those that other students connected with more easily.

Given their emphasis on the need for time to provide them the opportunity to get to know one another, we can ask: is the traditional time slots used for face-to-face courses enough for those classes offered online? If it take students longer to establish relationships that they would otherwise create in face-to-face classrooms, might there be a need to reassess the length of courses in online environments? Human relationships take lengthier time intervals to establish online and thus there may be a need to increase the length of online courses compared to face-to-face classes.

Cavanaugh (2005) and Lao and Gonzales (2005) stress the added time commitments that an online course requires of its instructors and teachers. Students frequently note that it also requires extra time commitments on their behalf. And this is solely for the academic activity in the course. Perhaps this signals that the regular course intervals are simply not enough for online courses.

4.1.6 Sub question 1.4 – In their view, what are the most socially robust aspects of the environment?

Recommend button:

The students offered two polarizing views on this, as demonstrated by the two following answers.
• The "thumbs up" weren't really used, although I liked the idea of them initially. It think netiquette (who gets more 'likes') (Clouder, 2011, Dorian, 2009) is in the back of people's heads and prevents them from being used more often.

• The recommend button was a great social tool

As previously discussed, the recommend button remains an enigmatic feature of the online environment. On one level, students seem to enjoy using this feature and they have demonstrated this in class and mentioned it frequently in the survey. At the same time, this feature was meant to encourage students to write better notes and to highlight those notes that were academically poignant and observant. In practice, students do not necessarily recommend those notes with academic worth, but rather, they may use it to highlight notes that are funny or feature an interesting story. They also do not perceive it as a tool which will help them write better notes (for instance, by vying to receive recognition from their peers).

We are further establishing this feature and its uses in future versions of Knowledge eCommons.

Avatars

The students mentioned the avatars many times as a feature of KeC which fostered social interaction. Some of their responses follow:

• The pictures of the people posting

• The images are nice

• Also, including a picture in your identity (shows up beside your name), helped to put name to faces more easily.

• The photo that appears with username.
• The avatars

• avatar creation helped me in getting to know the other students

Dalsgaard and Paulsen (2009) argue that the use of profile pictures and avatars has influence on the transparency provided in the learning medium and how in turn, transparency can help learning. They define transparency as “students’ and teachers’ insight into each other’s activities and resources.” (p. 2)

First and foremost, a profile picture gives a face to the members of the forum, which they would otherwise not have. Having a classroom of “real” faces will create a more realistic space. In future versions of knowledge eCommons, we wish to transform the avatar from a static entity to a feature which can convey students’ emotion, or level of understanding. Emotive Learning Technologies are the new frontier.

Forum & Journal

Open debate and dialogue is believed to be one of the greatest factors which contribute to social interaction in online learning environments. The creation of a more democratized space in which open debate and dialogue are nurtured can be one of the most promising aspects of these mediums. The journal was described by two students as influencing social interaction in KeC:

• Once we were prompted to "journal", the Class Lounge section really helped me learn about my peers and was a place that I could openly talk about my day-to-day issues with life (including school).

• sharing our biographies, journals and in the discussions during the week: so the discussion area
Info button

One student believed the “info” button to be a useful tool for social interaction. Clicking on this button will allow students to see the names of those who have read their notes. This student said:

- I like using the "info" button to see if/who has read my posts. (Although I'm fairly sure that very few others actually use this feature.)

Private notes:

Private notes are notes students write to each other, hidden to the instructor and other course participants. One student noted this attribute:

- The "private" notes turned out to be very helpful to my final collaborative project group.

We are hoping to further exploit the possibility of private notes, and have begun doing so in the newer current version of Knowledge eCommons. While the class lounge, forums and journal are public and visible to all, we believe that giving students the opportunity to more privately communicate within the space of the classroom (as opposed to sending email) will provide more room for social interaction.

4.2 Research Question 2

Which linguistic characteristics of notes correlate positively with the recommend ratio? (the degree of favorable recognition students give to a note.) In other words, what linguistic characteristics of a note are well received by other students?

As previously explored in our methodology, each note was rated on objectivity by at least one judge and 20% of the notes were rated by all three judges to calculate inter-rater reliability. The degree of inter-rater reliability was calculated using Krippendorff’s alpha(2004). The resulting alpha
was 0.8225 using the tool developed by Hayes and Krippendorff (2007) in SPSS. This value falls within the acceptable range of $\alpha > 0.8$ (Krippendorff, 2004).

The statistically significant correlations (for $p < 0.05$) with recommend ratio are presented in Table 11.

### Table 11

*Statistically significant correlations with recommend ratio ($p<0.05$)*

<table>
<thead>
<tr>
<th>Measures</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recommend ratio ~ objectivity</td>
<td>$r = 0.755$</td>
</tr>
<tr>
<td>Recommend ratio ~ words per note</td>
<td>$r = 0.719$</td>
</tr>
<tr>
<td>Recommend ratio ~ Academic Word Ratio</td>
<td>$r = 0.650$</td>
</tr>
</tbody>
</table>

The high positive correlation between recommend ratio and the use of objective language suggests that students value notes that are written in an objective fashion. Objective language also positively correlates with academic word ratio with a correlation coefficient $r = 0.694$ ($p < 0.05$). These correlations may be interrelated, since objective notes are often ones in which experts are quoted, and academic language is used.

Academic Word Ratio itself also correlated positively with recommend ratio. This signals that students favor notes which are academically oriented in content (which are objective) and language (which use words from the academic word list).

Words-per-note correlated positively with recommend ratio. Thus, notes that are highly recommended tend to be longer than other notes.

Number-of-words per unit of communication is an important, but often ignored feature of communication. For instance, number of words per utterance is deemed a significant factor in
selecting important utterances in speech data by human raters (Penn and Zhu, 2008). Since words-per-note also correlated positively with objectivity, we tested the correlation between objectivity and recommend ratio when controlling for words-per-note. We found a significant positive correlation at $r = 0.40$ for $p < 0.05$.

Thus, recommend ratio correlated positively ($p < 0.05$) with objectivity, Academic Word List Usage and number-of-words per note. These are the notes which students tend to favor, or find useful. We will have more to say on this in the next section.

The recommend ratio gives us a view of the types of notes students like to read. It doesn’t however show us what they are most inclined to respond to. This is covered in the following question.

### 4.3 Research Question 3

Which linguistic attributes influence reply ratio? This question is meant to investigate what factors in language spark conversation and communication. What linguistic characteristic of a note entice readers to share their own views and keep the conversation going?

Here, we did correlation studies on both the total emotion score (positive emotion – negative emotion), and each level of emotion (positive and negative) separately. We found no correlation with the total emotion score. However, the reply ratio had a significant positive correlation with number of words per note at $r = 0.39$ ($p < 0.05$) and a negative correlation with positive emotion at $r = 0.504$ for ($p < 0.05$). It is noteworthy to mention that we did not find any correlations between reply ratio and negative emotion. This suggests that understanding the effects of positive and negative language on students’ replies (or communication) is in further need of investigation.
It is also important to mention that no statistically significant correlations were found between reply ratio and recommend ratio. In other words, students did not necessarily respond to those notes which they favored. In the survey, students stated that notes they recommended were not necessarily those they responded to. This begs the question: what is the difference between the two (recommended notes and notes which are replied to)?

Recommend ratio correlated positively ($p < 0.05$) with objectivity, Academic Word List Usage and number-of-words per note. These are the notes which students tend to favor, or find useful.

By analyzing notes and responses, Hewitt (2003) argues that students tend to reply to those notes which have been more recently posted, rather than notes which hold crucial information, or are more important. He writes that: “students tended to respond to newer, rather than older notes.” (p. 34) where 80% of responses were aimed at those notes which were less than 48 hours old.

Given the results of this study, it may not be surprising to find that recommended notes are not necessarily those which are responded to the most. The recommend banner is a good tool for reminding students of notes with significance, but given the likelihood that they will respond to newer notes, it will not guarantee high reply ratio. Especially since students read notes in different times during the week (given the asynchronous nature of the environment) and by the time a note has accumulated significant recommends from students, it will have become an “old” note in the forum.

Finding other forms of highlighting recommended notes (for instance, moving them higher in the thread or highlighting them on a side banner) may give us ways of emphasizing more important notes for students to read and write to.


5 Conclusion

In this chapter, the concluding remarks are presented. The chapter is presented in two sections. The first section will discuss the results of our study, while the second section explains the greater aims of the project and its future goals.

5.1 Results

While text is the primary method of communication in asynchronous discourse environments, there have been few studies that employ text analysis techniques to study the effects of written communication on student interaction. Such parameters would be viable tools for studying class dynamics, assessment and troubleshooting. One of the primary reasons we are interested in studying text in relation to social interaction in online learning environments, is that the traditional (physical) means of social interaction (smiles, frowns, voices, gestures, etc) are replaced with modes of communication reduced to text.

In this multidisciplinary research we aim to look at linguistic characteristics of notes and how that related to various aspects of student communication. The online learning environment students use in this study is named Knowledge eCommons, a web-based educational environment that provides instructors a knowledge building platform – a CMC environment that supports knowledge building.

We begin by conducting a qualitative analysis of student perceptions of social interaction in online learning environments using a survey instrument. This is a qualitative investigation of students’ attitudes, beliefs and experiences regarding online learning and online interaction. We hope that through more discussion with students, we will be able to develop socially robust tools that will give students the opportunity to interact and socialize.
From these interviews we understand that students believe the webinar to be one of most interactive features of the course. The webinar is a newly introduced feature to KeC in which students meet online once a week with the instructors, creating a hybrid class. The benefits of such classes, which give students the chance to see, hear and interact with one another online has been previously outlined by other researchers (Cavanaugh, 2001; Anderson, 2008; Saw et al., 2008).

Private notes and journals were two of the other features which students often cite as giving the environment greater social cohesion.

One of the points we learned from the student discussions was the emphasis that students place on the interaction of other students. That is, they feel most socially related to those students that were making an effort themselves. They also stressed the importance of student team leaders who devised “novel” ways of using the class features, for example, the chat room.

This is further indication that placing social tools in online environments is not enough, and there must be personal, social and pedagogical usage for the tools, to urge students to incorporate them into their classroom experience. There is also evidence both here and in the literature that timewise, regular course intervals for face to face courses may not be suitable for online courses where social relations take longer to develop. Through studying the results of this survey, we hope to better modify our course features to suit student need and to better facilitate social spaces in the learning environment.

After the completion of the survey, we used linguistic analysis techniques to study note sentiment and its relationship to student communication.

Sentiment analysis is a 2-dimenational text analysis technique. It is concerned with two primary factors: the subjectivity/objectivity of text (in this case, the notes written by students) and the level of positive and negative emotion inherent in text. We take available communicative features
of student notes (recommend ratio, reply ratio and number of notes written) and see what correlations exist between these communicative features and linguistic attributes of the note.

Through these studies, we see that recommend ratio has a positive correlation with objectivity and words per note, indicating that students favor those notes which are longer and are written in the objective language. However, recommend ratio had no correlation with reply ratio. This signals that what students “like” in a note and what entices them to respond are two different phenomena and more investigation is needed to decipher this relationship. We offer reasons as to why this may be the case and provide solutions that can be utilized and/or developed.

5.2 How is this Useful?

As discussed in Chapter 1, nearly 40 years ago Mark Granovetter introduced the concept of “weak ties” in social network theory. By surveying and interviewing hundreds of individuals working in professional jobs, Granovetter found that of those who had found their jobs through a contact, nearly 70% stated that these contacts were individuals they did not consider close friends. Rather, these contacts were what they referred to as “distance acquaintances” - friends of friends or friends of friends of friends, individuals connected to them through 2 or 3 degrees of separation.

He thus stressed the importance of weak links in social network theory: “the weak tie between an individual and his acquaintance, therefore, becomes not merely a trivial acquaintance tie but rather a crucial bridge between the two densely knit clumps of close friends. These clumps would not be connected to one another at all were it not for the existence of weak ties. It follows then that individuals with few weak ties will be deprived of information from distant parts of the social system.” (p. 202). He also stresses that “social systems lacking in weak ties will be fragmented and incoherent.” (p. 202) We believe that in this, there are crucial lessons and warnings for online learning and teaching.
With the rapid expansion of online learning and online courses, as students rely more and more on these modes of delivery and abandon traditional face to face methods of learning, the lack of interaction and the lack of ties created in these mediums will become an even more important, pressing issue for their lifelong professional (and even personal) pursuits. As Granovetter stresses, these ties are crucial for integration into any community and without them, the community ceases to exist.

However, thus far online learning environments have failed to create such long term ties between individuals. As our own students indicated, few would be keeping in touch with their fellow classmates once the duration of the course is over. By studying textual communication within these environments, we are hoping to understand more about how students communicate. That will in turn help us devise better technological and pedagogical tools that facilitate more robust social interaction, and in turn, social, educational and personal ties. If for instance, students do favor notes that are objective in language and tone, but are not inclined to respond to such notes as such, how can we better organize the class to both get students to interact on a social and educational level? We have exciting new features in Knowledge eCommons which we think will take us a step further in this attempt.

Another motivation for this study is the creation of a tool that will be able to aggravate factual knowledge students produce.

The problems with threaded forums have been widely stressed by researchers. “linear discourse forms and isolated “conferences” entrap ideas” (Scardamalia, 2004). Hewitt (2001) argues that online discussions may be much more intertwined and interrelated than the threaded representation indicates, “but provide few facilitations for drawing discourse in meaningful ways.” (207) Suthers and Xu (2002) state that the representation of threaded discussions “is based on historical development of the discussion rather than its conceptual content.” (p. 2)
While wikis, blogs and multi-user editable documents all offer solutions to these problems, thus far no automated solution exist. We envision a tool that can collect the key ideas within a threaded conference to synthesize important academic notes and arguments students have made in the conference. How would an algorithm for instance, tell the difference between a joke a student makes, and their counterarguments to a course reading?

By devising an algorithm to determine objective notes, we are developing the first step towards this tool. The key academic ideas are located within those notes which use objective language. We hope to lead the way in creating a tool that will be able to automatically merge the most important academic notes within a threaded conference, thus removing many of the obstacles or hurdles to threaded conferences which continue to dominate online learning platforms.

In this study we used sentiment analysis techniques to reveal that students place higher value on the notes that are objective, lengthy, and make good use of academic vocabulary. We plan to develop and test advanced corpus-based sentiment detection and classification techniques in the educational forum domain. These techniques have produced better results than wordlists in news and product review domains. By adapting them to our domain, we will measure their additional merit in these studies.

For example, disagreement and controversy cannot be captured through the use of word lists alone. But these are important factors to study, since they play a significant role in the learning process. Berlyne(1966) argues that “disagreement with another person can be a source of conceptual conflict which in turn provokes the individual to examine the other person's ideas” (as cited in Johnson & Johnson, 1979, p. 53). As Johnson and Johnson (1979) note: “conceptual conflict has high arousal potential, motivating attempts to resolve it by seeking new information or by trying to reorganize the knowledge one already has.” (p. 53). We found that less than 5% of the notes had explicit disagreements with another student.
Finally, for the purpose of this study, we looked at the effects of linguistic characteristics on macro level student attributes, such as reply ratios and notes written. While these attributes are important, there are micro level indicators which are also in need of consideration. For example, what sort of responses does an objective note obtain? Do individuals who share sentiment tend to correspond more with one another? For all these questions and more, small scale studies will be conducted.
References


Understanding the dynamics of collaborative multi-party discourse. *Information Visualization, 5* 250-259.


Rovai, A. (2002). Building sense of community at a distance. International Review of Research in Open and Distance Learning, 3(1), 1-16.


Appendices

Appendix 1: Survey

**Part I. General Questions**

1. To date, how many online courses have you taken (including this one)?
   1 (this is my first)
   2
   3
   4 or more

2. Where do you have access to the course website? (mark all that apply)
   ○ Work
   ○ School
   ○ Home
   ○ Other (please specify):

3. How fluent are you in English?
   1 (very fluent) ........................................ 5 (not fluent)

5. Please rate Knowledge eCommons on visual appearance:
   1 (visually appealing) .................... 5 (visually not appealing)

6. Please rate Knowledge eCommons on response time (speed):
   1 (fast) ........................................ 5 (slow)
Part II. Your opinion

1. Social presence, in an online environment, is defined as "the degree of awareness of others in a communication interaction". Which features of Knowledge eCommons, if any, fostered a sense of social presence in the course?

2. What kinds of activities and instructional methods in the course encouraged a sense of social presence?

3. Compared to a regular face-to-face course, how "socially present" were your classmates in this course? (To what extent did you feel you were getting to know your online classmates?)
Part III. Course Questions

1. I felt comfortable using the online technologies for the course.
   - Strongly Agree
   - Agree
   - Neutral
   - Disagree
   - Strongly Disagree
   - No Answer

2. Through our online communication, I had the opportunity to know some of my classmates.
   - Strongly Agree
   - Agree
   - Neutral
   - Disagree
   - Strongly Disagree
   - No Answer

3. The course instructors were active in the class.
   - Strongly Agree
   - Agree
   - Neutral
   - Disagree
   - Strongly Disagree
   - No Answer

4. I felt hesitant sharing my personal opinion on the course forum.
   - Strongly Agree
   - Agree
   - Neutral
   - Disagree
   - Strongly Disagree
   - No Answer

5. I enjoyed reading my classmates’ writings on the forum.
   - Strongly Agree
   - Agree
   - Neutral
   - Disagree
   - Strongly Disagree
   - No Answer

6. Online web-based education is a good medium for social interaction.
   - Strongly Agree
   - Agree
   - Neutral
   - Disagree
   - Strongly Disagree
   - No Answer

7. My instructor responded to my emails and messages within reasonable time.
   - Strongly Agree
   - Agree
   - Neutral
   - Disagree
   - Strongly Disagree
   - No Answer

8. I was able to learn from the online discussions.
   - Strongly Agree
   - Agree
   - Neutral
   - Disagree
   - Strongly Disagree
   - No Answer

9. I only take online courses when I am forced due to timing or distance considerations.
   - Strongly Agree
   - Agree
   - Neutral
   - Disagree
   - Strongly Disagree
   - No Answer
Part III. Course Questions

10. I didn’t find the recommend button useful.
   ○ ○ ○ ○ ○ ○ ○ ○
   Strongly Agree  Agree  Neutral  Disagree  Strongly Disagree  No Answer

11. The recommend button encouraged me to write better notes.
   ○ ○ ○ ○ ○ ○ ○ ○
   Strongly Agree  Agree  Neutral  Disagree  Strongly Disagree  No Answer

12. I was more likely to pay closer attention to those notes which were recommended by more of my classmates.
   ○ ○ ○ ○ ○ ○ ○ ○
   Strongly Agree  Agree  Neutral  Disagree  Strongly Disagree  No Answer

13. The course webinar helped me get to know the other students better.
   ○ ○ ○ ○ ○ ○ ○ ○
   Strongly Agree  Agree  Neutral  Disagree  Strongly Disagree  No Answer

14. Getting to see my classmates and instructors via the webinar made the class feel just like a real classroom.
   ○ ○ ○ ○ ○ ○ ○ ○
   Strongly Agree  Agree  Neutral  Disagree  Strongly Disagree  No Answer

15. I didn’t think the webinar feature of the course had any learning value.
   ○ ○ ○ ○ ○ ○ ○ ○
   Strongly Agree  Agree  Neutral  Disagree  Strongly Disagree  No Answer

16. Writing private notes to other students was a helpful feature of the learning environment.
   ○ ○ ○ ○ ○ ○ ○ ○
   Strongly Agree  Agree  Neutral  Disagree  Strongly Disagree  No Answer

17. I will keep in touch with some of the classmates and/or instructors that I met online.
   ○ ○ ○ ○ ○ ○ ○ ○
   Strongly Agree  Agree  Neutral  Disagree  Strongly Disagree  No Answer

18. Displaying a personal photo “avatar” (i.e., the image that appears on each note) contributes to a sense of community.
   ○ ○ ○ ○ ○ ○ ○ ○
   Strongly Agree  Agree  Neutral  Disagree  Strongly Disagree  No Answer
### Appendix 2: Rating guidelines

#### Rating Guidelines and Explanations

**Rating method:**

- “1” for objective
- “0” for subjective

Please rate the following notes as “subjective” (0) or “objective (o).”

An objective note is defined as a note in which the writer uses objective sources or facts. Thus, citing any of the following sources makes a note objective:

1. published papers
2. organizations
3. quoting known experts
4. results of concrete experiments
5. any other credible source (leaving “credible” for judges to evaluate)

2. The fact(s) argued must be in some way related to the week’s discussions.

3. This verifiability must be conscious: referencing to vague individuals or sources does not constitute an objective note.

4. Quoting the instructor is only objective if it is done in a scholarly manner. So, for example, “Howard [instructor] would like to know what we will be doing” is not considered an objective note. However, “Howard argued in last week’s class that a teacher’s instructional practices must be evaluated by a peer” is considered objective.

Quoting other students is not considered objective unless those students are considered an “expert” in the discussion.

5. Merely stating a source without explanation is not objective. (e.g., “I remember some professor saying that knowledge building is a useful paradigm”). The source needs to be verifiable and/or traceable to a reasonable extent.