Chat Language

- In the continuum of speech and writing

Adam Lind
Abstract

The aim of this essay is to determine where on a continuum between speech and writing written computer-mediated communication (chat language) would be placed. The essay makes use of a methodology based on Biber (1988). This was done using a quantitative research methodology based on counting and comparing specific linguistic features in different texts. The data for chat language came from the NPS Chat Corpus. Other data used were transcripts of spoken discourse as well as a popular scientific text as material for comparison. This essay is mainly focused on four features: the use of pronouns, passives, ellipsis and the type/token ration of each individual text. Despite the limited size of the material sampled, the results showed that chat language had more in common overall with speech than with writing.

Keywords: Biber, Computer-Mediated Communication, CMC, Chat Language
## Contents

Abstract ............................................................................................................................................. 2

1. Introduction .................................................................................................................................. 1

1.1 General ..................................................................................................................................... 1

1.2. Aim and scope...................................................................................................................... 2

2. Theoretical background .............................................................................................................. 3

2.1 What is a chat? ..................................................................................................................... 3

2.2 Linguistic characteristics of online chat .............................................................................. 4

2.3 The methodology of Biber ................................................................................................... 5

2.4 Situational and functional differences between speech and writing according to Biber ........ 5

2.4.1. Spoken or written channel ............................................................................................ 5

2.4.2. The channel used .......................................................................................................... 6

2.4.3. Relationship of communicative participants to each other ........................................ 6

2.4.4. Spatial and temporal context ....................................................................................... 6

2.4.5 Integration ....................................................................................................................... 6

3. Grammatical features which are typical for speech and writing ............................................. 7

3.1. Passives ................................................................................................................................ 7

3.2. Ellipsis ................................................................................................................................... 7

3.3. Type/token ratio .................................................................................................................. 8

3.4. Use of pronouns .................................................................................................................. 8

4. Method and material .................................................................................................................. 9

4.1 Method ................................................................................................................................... 9

4.2 Material .................................................................................................................................. 9

4.3 Problems and limitations ...................................................................................................... 11

5. Results ....................................................................................................................................... 11

5.1 Ellipsis ..................................................................................................................................... 12

5.2 Pronouns and proper nouns ................................................................................................. 12

5.3 Type/token ratio .................................................................................................................. 13

5.4 Passives .................................................................................................................................. 15

5.5 Summary of results .............................................................................................................. 15

6. Discussion .................................................................................................................................. 16

6.1 Biber’s functional and situational features .......................................................................... 16

6.2 Discussion based on findings ............................................................................................... 18

   6.2.1 Ellipsis ............................................................................................................................ 18
6.2.2 Pronouns and proper nouns .......................................................... 19
6.2.3 Type token ratio ............................................................................. 20
6.2.4 Passives ......................................................................................... 20
7 Conclusion ............................................................................................ 21
References .............................................................................................. 23
Sources of empirical data ........................................................................ 24
Appendixes ............................................................................................... 25
1. Introduction

1.1 General

Chat, as described by Hård af Segerstad (2002) is a form of synchronous Computer-Mediated Language (henceforth CMC). The generation born in the early 1980s are in more ways than one unique. Prensky (2001) refers to this generation as 'digital natives'. A digital native according to Prensky is a person who was born into the digital world, children who essentially have been raised with a computer in their homes. According to Prensky the introduction of modern day computers is such a fundamental change to the individuals that it should be called a singularity (Prensky 2001 [www]). This "singularity" or what this author would define as a revolution, has created what is called computer-mediated communication, simply due to the fact that CMC cannot exist without computers.

The main purpose of this essay is to try and determine where on a continuum between written and spoken discourse chat language may be placed. Furthermore, it should be borne in mind that the present essay focuses on determining where chat language would be placed on a continuum, not CMC overall. There are other types of CMC, such as posting on internet forums or using software for phone-calls, such as Skype. This essay will however focus on synchronous chat. The methodology is based on the methodology used by Biber.

Biber (1988) has made an attempt to classify different types of texts, both spoken and written. Biber's work does however lack a classification of computer-mediated communication. This is not due to an oversight by Biber himself. It is rather due to the fact that CMC did not exist in such a way that it does in society today. Chat language was merely not present to the extent 1988 that it is in the year 2012.
1.2. Aim and scope

The purpose of this present essay is to try to determine, using similar theories and methods as Biber, where chat language would be placed on a continuum between spoken discourse and scientific writing. However, it should be noted that the present work will in no way be as extensive as Biber's study. The purpose lies within trying to classify where chat language comes in on the aforementioned continuum.

Furthermore the hypothesis is that speech and chat language have more features in common than chat language and scientific research articles as examples of writing.
2 Theoretical background

2.1 What is a chat?

According to Hård af Segerstad, teenagers use the very same strategies as they use for verbal communication when chatting. Teenagers also adapt the Swedish written language to meet the limitations and advantages of online communication' (2002: 125). Katzman et al. (1997) cited in Hård af Segerstad (2002) claim that in order to adapt to certain limitations with text-based CMC, participants in chat communication use abbreviations and emoticons. An emoticon is the arrangement of symbols to represent real-life facial expressions and moods (Katzman (1997) cited in Hård af Segerstad 2002). Synchronous chat communication in its simplest form is very simply two people sitting at a computer typing in real-time. The utterances are then shown on the respective screens in real time. A more advanced form of synchronous CMC is when several people participate in the same chat room (Herring ed. 1996).

A typical example of chat language taken from the NPS Chat Corpus follows below.

[1]
10-19-20sUser7: now im left with this gay name
10-19-20sUser7: :P
10-19-20sUser59: hey everyone
10-19-20sUser115: ah well
10-19-20sUser121: 10-19-20sUser7 is a gay name.
10-19-20sUser7: :)
10-19-20sUser115: hi 10-19-20sUser59
10-19-20sUser21: 26/ m/ ky women that are nice please pm me
10-19-20sUser115: there ya go 10-19-20sUser7
10-19-20sUser121: don't golf clap me.
10-19-20sUser7: fuck you 10-19-20sUser121: @
10-19-20sUser59: whats everyone up to?
10-19-20sUser121: i'll thunder clap your ass.
10-19-20sUser121: and i dont even know what that means.
10-19-20sUser84: that sounds painful
10-19-20sUser49: any ladis wanna chat? 29 m
10-19-20sUser21: 26/m
10-19-20sUser115: my cousin drew a messed up pic on my cast
10-19-20sUser23: 24/m
10-19-20sUser121: boo.
The coding of the usernames has been done by the creators of the corpus in order to privacy mask all usernames. The rooms in which the data has been recorded are age specific as well. The numbers in front of the user name stands for the following; 10 (month) - 19 (day) - 20s (presumed age of participants in chat room). All data gathering was done in 2006 (NPS Chat Corpus 2006)

2.2 Linguistic characteristics of online chat

Herring (2001), cited in D. Schiffrin et al. (2001) claims that CMC contains what Herring describes as 'non-standard features'. A small percentage of these features are errors or the lack of knowledge about formal language. The majority of these choices Herring says are more about mimicking spoken language and being economical about typing (Herring ibid: 617). The essence of what Herring says is that written chat language seen in online chat-rooms is for all intents and purposes supposed to mimic spoken discourse.

Biber (1988) discusses at length the different characteristics of speech and writing. Written text supposedly is more structurally elaborate, complex, formal and abstract; readers and writers are also separated by time and space. Furthermore the writer and the reader almost never meet each other and therefore there is rarely any interaction in text. Spoken discourse on the other hand is interactive and structurally simple. Speakers are however, according to Biber, not separated by time or space (Biber 1988). One might therefore suggest that it can be useful to look at the participants’ relationship with each other. Whereas such an analysis is not as relevant when studying the relationship between writer and reader

Hewings and Hewings (2005) bring up several features of CMC, describing CMC as having as many of the same features as any context in which a conversation might be held. The obvious factor is that it cannot be done face-to-face which is a questionable claim since that a written chat conversation can be done with a web camera permitting the transmittance of facial expressions. Hewings and Hewings also conclude that chat language is differentiated from ordinary standard language because there is a great deal of ellipsis in writing (see [6]). It is also mentioned that features of face-to-face conversation that cannot be seen are represented in text. Hewings and Hewings exemplify this by using the word "eeeeeerrr" as an example of lengthening (2005:258).
2.3 The methodology of Biber

The present study is however based on Biber's (1988) research. The general consensus as described by Biber seemed to be that speech was primary and writing is secondary to speech. Partly, this opinion seemed to be based on the argument that writing is only the visual representation of speech (Biber 1988).

Biber's study was carried out by using what he called dimensions. According to Biber, dimensions are not polar opposites each located at the end of a line. They should instead be regarded as phenomena that vary in terms of their position. Speech and writing are not binary oppositions of each other according to Biber. They are rather dynamic and within a specific dimension a specific kind of speech might be closer to a kind of writing than to other kinds of speech. A dimension is a predefined set of co-occurring features. In dimension 2 for example “Narrative versus Non-narrative Concerns” Biber lists third person pronouns, public verbs and other factors like verbs in the past tense (Biber 1988:92).

Biber (1988) also claims that several dimensions must always be taken into consideration when performing this type of study. According to Biber, linguistic variation in text is multidimensional and cannot be determined by using only one dimension (ibid).

2.4 Situational and functional differences between speech and writing according to Biber

Biber states that there is no characterization for speech and writing that is entirely true for all genres. Biber does however in spite of this choose to list a couple of features that are stereotypical for each genre. Speech according to Biber is simple, dependent on context and fragmented. Writing on the other hand is the opposite of speech. Writing is supposedly more structurally complex. It is less dependent on a shared base of knowledge, situation or time. It is more planned and also contains a higher amount of new information than speech (Biber 1988:37).

2.4.1. Spoken or written channel
This is a quite obvious channel but Biber feels the need to express its existence. Speech is a vocal medium whereas writing is a physical medium (Biber 1988:38).

2.4.2. The channel used

Biber differentiates between different channels. The first channel is referred to by Biber as the lexical/syntactic channel and what is meant by this is that writing is limited to the channel of morphemes. What is meant by this is that meaning in writing is given from how the words are arranged. In speech meaning can be given by intonation. However there are factors that can affect how one chooses to interpret the limited channel of writing. Biber writes that notes on a table might point towards an object underlining, bold or text in italics can further expand the written channel i.e. writing in bold or italics can be used to further emphasize a point. On the contrary, speech, according to Biber has several channels; an example of this is gestures. Speech can however be limited as well. Whilst speaking in a dark room it is not possible to communicate via body language (Biber 1988:38).

2.4.3. Relationship of communicate participants to each other

The extent of interaction dictates that in a spoken situation the participants can react instantly. This reaction might then lead to inquiries, requests for clarifications or a simple nod to indicate further interest in the subject of the conversation. With shared knowledge about the participants, Biber means that the knowledge the participants have about each other can influence the direction for the interaction. The participants having knowledge about each other is more common in speech than it is in writing (Biber 1988:40-41).

2.4.4. Spatial and temporal context

Speakers and listeners according to Biber normally share a physical space whereas readers and writers do not. Speakers and listeners also share time during the communicative interaction whereas readers and writers do not (Biber 1988:42).

2.4.5 Integration

According to Biber the writer can compress a great deal of information using relatively few words since the writer is under no time constraint and the reader has more time to interpret the
text he or she is presented with. Speech is however created and interpreted under time constraints. The opposite of integration is called fragmentation (Biber 1988:43).

3. Grammatical features which are typical for speech and writing

Like Biber, the present work will focus on establishing a specific type of text on a continuum. The features which will be used for this are the ones that follow.

3.1. Passives

A passive clause is created by the usage of the auxiliary verb "be" used in conjunction with a verb in the past participle (Estling Vannestål 2007:155). An example of this would be:

[2] *A gas chromatograph can be utilized to examine the samples' organic substance content.* (Science Illustrated 2012 [www]).

Passives are also more commonly used in formal and scientific writing. One of the most common reasons for using the passive voice is in order to avoid referring to a person performing an action. It can also be used in situations where the active voice might seem tactless (Greenbaum & Nelson 2009).

The study which was made by Biber shows that passives were more common in scientific texts than in conversation (Biber 1988). It can also be used to refocus the attention to the action of a sentence, not the one performing it (with the semantic role of agent).

3.2. Ellipsis

Ellipsis is a result of economy in spoken discourse since it is the omission of words. Ellipsis is a phenomenon where words are left out which otherwise are considered being grammatical units that are important for a complete sentence and grammatical structure. Omission of these words is more common in speech since the meaning of a clause can be understood from the context in which an utterance is made. There are also different types of ellipsis, such as omission of verb phrases or subject (Greenbaum & Nelson 2009:166).
An example of this could be the utterance taken from the NPS Chat Corpus [3] "10-19-20sUser155: am a student u". This sentence written in standard English would look like: "I am a student, what about you?" Three words have been omitted: I, what and about.

### 3.3. Type/token ratio

Type / token is a concept used in several different disciplines, such as philosophy as well as linguistics. According to the *Stanford Encyclopedia of Philosophy*, C.S Pierce made the distinction between types and tokens by saying that the type is the amount of different words whereas token is the total amount of words (Stanford Encyclopedia of Philosophy 2006[www]). The type token ratio can also be used as a measurement tool for variation within texts (Speech Therapy and Information 2009 [www]).

The following sentence taken from the NPS Chat Corpus exemplifies this. [4] "10-19-20sUser121: you're not supposed to be here you're over your 20s".

The abbreviated form of "you are (you're)" has been used twice giving the sentence a type/token ratio of 9/10. Allwood (1998) claims that there is twice the amount of types in written language than there is in spoken discourse.

It is possible to put this statement into perspective with Biber's theory about integration in spoken and written communication. Since written language operates under no time constraints the writer has the possibility to select words, which increases the variation for the text and therefore avoids repetition.

### 3.4. Use of pronouns

A pronoun is a word that is used instead of a noun. The name comes from the Latin word *pro nomen* which means “instead of noun” (Estling Vannestål 2007). Pronouns have been selected because their usage should be higher in spoken discourse than in scientific writing and should thus aid in the establishment of the continuum. Furthermore, the usage of proper nouns will also be analyzed. Quirk (*et al.* 1985:335) claims that the meaning of a pronoun is undetermined and can only be understood in the context it appears in.

Using the pronoun "it" in a written text might therefore make it hard for the reader to interpret what "it" refers to; the writer does has more time to create proper references in the written discourse. A higher grade of interaction between speaker and listener also mean that any
unclear references such as "it" can be clarified. Clarification in written texts is however not possible between reader and writer.

4 Method and material

4.1 Method

The method used for the present study is a quantitative one. It consists of counting the features which have been mentioned above. Each text used will be given a specific value based on how often a specific feature appears per 100 words. The type/token ratio as well as the amount of pronouns will be calculated using computer software called Wordsmith (Scott 1996, 1998). The type/token ratio will be counted in an attempt to find out how varied the language is in the different samples. Since the utterances in the chat language sample is very small the relevance of type/token ratio can be discussed. However, calculating type/token ratio will still aid in establishing a part of a continuum. The purpose of this essay is to calculate and compare features. These features are all relevant since they aid in creating a comparison between genres. Ellipsis and passives will however be manually counted by hand.

Wordsmith as described by its creator is a piece of software used for analyzing words in text. In the words of the software's creators; "The tools have been used by Oxford University Press [...], by language teachers and students, and by researchers investigating language patterns in lots of different languages in many countries world-wide" (Lexically Net [www]). What Wordsmith can be used for is the creation of corpora; Appendix 1 has as an example been created using wordsmith.

4.2 Material

In order to analyze and establish where chat language would end up on a continuum other material was required to serve as a reference. The collection process of this material simply consisted of locating texts that would serve as reference material. Furthermore, since the material for the empirical data was readily available in the form of a partially-annotated corpus, all efforts could be focused on analysis.
The example of written discourse was however written down in a standard text document so that *Wordsmith* would be able to analyze it. In retrospect it is hard to determine whether or not the process of digitalizing the text saved time or not.

Furthermore, passive clauses had to be counted by hand and which became a process that consumed more time than originally intended.

*The dream of Mars, a new Era* (Science Illustrated [www]) was used as a scientific reference text. The text itself will be used for comparison as well as establishing one "end" of the continuum using the four features.

Popular science differs some from what can be called research-orientated science. Popular science bridges the gap between the general audiences as well as simplifying the language. Standard scientific literature tries to convince their peers about the validity of the research made. Popular science informs people outside the specific scientific community (Wikipedia [www]). Popular science also focuses more on the actual results themselves rather than the research methodology or the goals of said research (Brooks 2006).

For the purpose of this essay, spoken and scientific discourse was chosen as reference materials. However as Biber mentions, there are no absolute polar opposites in the comparison of texts. Texts are dynamic and one cannot determine a text as being truer to a specific medium (Biber 1988). Keeping this in mind, the present study might just as well have used spoken discourse and newspaper articles as references. The scientific discourse was however chosen because it does to some extent belong to the genre of scientific texts. It was expected for speech and popular scientific writing to be sufficiently different in order to create a viable comparison. This genre was assumed to have different characteristics from spoken discourse, such as the increased use of passive clauses.

The text which has been used as reference material for spoken discourse has been taken from *Exploring Spoken English* (Carter et al. 1997). The text itself has not been chosen on the basis of any other criteria than appearing to be normal speech. The text is based on a recording of a family that is cooking rice. The sample is an authentic and spontaneous example of speech in
England. The sample originates from the University of Nottingham and was used to create a corpus of English discourse (Carter et al. 1997).

The chat language data has been taken from the NPS Chat Corpus. The NPS Chat Corpus is distributed for educational and research purposes. The material used was selected because it appeared as fairly standard chat language. According to its creators the gathering of material has been done in age specific chat rooms for a specific amount of time on a specific date (NPS Chat Corpus 2006).

4.3 Problems and limitations

The present essay is limited in that it has focused on a relatively small number of features that will be counted. For the purpose of this essay four different features will be used, whereas Biber (1988) used ten times more features. Needless to say, had time not been a factor to the extent that it currently is, a more extensive survey using an extended set of features as well as larger and more varied text samples would have been preferable.

The material used from the NPS Chat Corpus also required some ‘cleaning’ for it to be able to be accurately analyzed using Wordsmith 4.0. First of all, the timestamps i.e. the nicknames that had previously been masked by the creators of the corpus were removed. It should also be mentioned that removing timestamps from several hundreds of lines of texts is a more arduous task than one might imagine. Secondly all the lines which were used to create the timestamps needed to be removed. Wordsmith counts 10-19-20 as three different words when they were used in the NPS Chat Corpus as one lexical item.

5. Results

In the following section the results gathered using manual counting as well as automatic counting using Wordsmith will be presented. The findings regarding the use of ellipsis, pronouns, proper nouns, passives and type/token ration will all be accounted for. Furthermore, the results are placed on a continuum in order to create a graphical representation of the findings.
5.1 Ellipsis

In the reference sample for spoken discourse, 19 words out of 375 words were omitted; this is 5%. Out of 62 utterances 11 utterances had ellipsis with a total of 18%.

Example: In the utterance [5] "what next" (Carter et al. 1997:65) the speaker has omitted the verb "is".

In the reference material for scientific texts no ellipsis was found.

A common utterance in the NPS Chat Corpus is the "24/m" or a similar number. The utterance 24/m in the context of a chat conversation stands for age and gender. Normally, this would be expressed using a full sentence; "I am a 24 year old man". However due to limited time it is abbreviated like 24/m. These utterances are for the purpose of this essay considered as highly ellipted. Example: [6] "10-19-20sUser23: 24/m" (NPS Chat Corpus 2006)

It was mentioned by Herring that chat language has non-standard features in it. Thus attempting to find anything within chat language that breaks away from what can be defined as standard grammar can be difficult.

In the NPS Chat Corpus out of 706 items a total of 92 contain omitted words (ellipsis) which is 13% of all utterances. Out of 1976 words that were analyzed, 740 words were omitted. An omission of 740 words per 1976 is approximately 37% of all omitted words.

If ellipsis within all the samples were placed on a continuum it would look approximately as follows:

SciT 0%  Speech 18%  Chat language 37%

5.2 Pronouns and proper nouns

The reference text for spoken discourse contains a total of 53 different pronouns which is a total of 15% of the words. Proper noun usage made up approximately 1% of the words used.
The reference material for scientific texts contains a total of 18 pronouns which is a total of 1%. Proper noun usage was at a total of 3%.

Example: [7] "The first person to present the idea was Carl Sagan" (Science Illustrated 2012[www])

The sample text taken from the NPS Chat Corpus contains a total of 284 different pronouns which is a total of 14% of all the words. What is remarkable however is the usage of the actual nicknames in the chat. All nicknames have for the purpose of this essay been treated as proper nouns.

Example [8] "I like it when you do it, 10-19-20sUser83". In this example it is clearly visible that the nickname, made anonymous, is used as a proper noun.

The list (Appendix 1) was compiled using the software Wordsmith 4.0 showing that a total of 8% of all uttered words are the actual nicknames of the participants in the chat. The creators of the NPS Chat Corpus have replaced all nicknames with a timestamp; these timestamps are registered as a "#" in the Wordsmith list (Appendix 1).

If pronoun usage within all the samples were placed on a continuum it would look approximately as follows:

<table>
<thead>
<tr>
<th></th>
<th>SciT 1%</th>
<th>Chat language 14%</th>
<th>Speech 15%</th>
</tr>
</thead>
</table>

If proper noun usage within all the samples were placed on a continuum it would look approximately as follows:

<table>
<thead>
<tr>
<th></th>
<th>Speech 1%</th>
<th>SciT 3%</th>
<th>Chat language 8%</th>
</tr>
</thead>
</table>

5.3 Type/token ratio

Out of the 62 utterances in the spoken discourse sample, 9 had a type token ratio of less than 1.0. Out of those 9 utterances only 1 had an utterance length that was less than 10 words. The
mean utterance length was 6. The total number of tokens used was according to *Wordsmith* 357 with a total of 152 types which in turn means that 43% of all words were unique. The total amount of utterances with a ratio of less than 1.0 is 15%.

The reference text for scientific material has a mean utterance length of 20. According to *Wordsmith* the total amount of words in the reference text was 1986 with a total of 711 different types. 711 types out of a total of 1986 tokens are 36% unique words. A total of 38 different utterances contained a type/token ratio which was less than 1.0. The total percentage of utterances with a ratio less than 1.0 is 40%. All of the utterances which contained a ratio less than 1.0 contained more than 10 words.

The analysis of the data from the NPS Chat Corpus concluded that out of 706 items 19 contained a type token ratio less than 1.0. 19 out of 796 are 3%. The mean utterance length of the NPS Chat Corpus is 2.7 words and this is half of the mean utterance length of ordinary speech. The NPS Chat Corpus has a total of 34% unique words.

Example: [9] "10-19-20sUser23: too hot in the summer and too cold in the winter". In this example User23 has repeated the words *too, in and the*. The repetition of these words gives the sentence a total ratio of 8/11.

If type/token ratio within all the samples were placed on a continuum it would look approximately as follows. This continuum represents the total amount of utterances with a type/token ration of less than 1.0.

<table>
<thead>
<tr>
<th>Chat language</th>
<th>Speech</th>
<th>SciT</th>
</tr>
</thead>
<tbody>
<tr>
<td>3%</td>
<td>15%</td>
<td>40%</td>
</tr>
</tbody>
</table>

If mean utterance length within all the samples were placed on a continuum it would look approximately as follows.

<table>
<thead>
<tr>
<th>Chat language</th>
<th>Speech</th>
<th>SciT</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.7</td>
<td>6</td>
<td>20</td>
</tr>
</tbody>
</table>
If the usage of unique words within all the samples were placed on a continuum it would look approximately as follows.

Chat language 34%  SciT 36%  Speech 43%

5.4 Passives

The reference text for spoken discourse contains no passives.

In contrast to the reference for spoken discourse the reference material for written text was found having passive clauses. Out of the total 184 clauses, in the text a total of 24 were determined to be passive clauses, this is an estimate of 13%.

Out of the 706 utterances or items that were analyzed in a specific part of the NPS corpus only one post contained a passive clause [10] "anyone can be bought if theres enuff money 10-19-20sUser83:)" this is a percentage of 0.1 (NPS Chat Corpus 2006).

If passives within all the samples were placed on a continuum it would look approximately as follows:

Speech 0% Chat language 0.1%  SciT 13%

5.5 Summary of results

The above results show that chat language is highly ellipted and with its omission level of 37% it omits twice as many words as normal speech. The pronoun usage of chat language is at approximately the same level as speech and fairly different from scientific text. What is remarkable however is the usage of pronouns. The type/token ratio of chat language is closer to speech when the amount of utterances with a ratio of 1.0 or lower has been counted. Chat language and speech have a similar mean length of utterance. Chat language and scientific
text were closer to each other than speech and chat language in the amount of unique words compared to the total amount of words. The number of passives in chat language and speech is very similar.

6 Discussion

6.1 Biber's functional and situational features

Firstly, it would appear that the Biber analytical framework has been useful in the analysis of chat language. In section 1.4 some of Biber's theories regarding situational and functional differences between speech and writing are accounted for. Stereotypical speech is within the spoken channel and it also has the benefit of being multi-channel (see section 3.4.2). The extent of interaction dictates that the speakers can instantaneously react to utterances; their knowledge about each other can also steer a conversation in a distinct direction. Furthermore, speakers share time and sometimes space. Speech is also highly fragmented i.e. the message of any utterance is most likely not compressed in to a small amount of words.

Writing has all the characteristics that speech does not. Writing uses the textual medium to convey its message and writing is also fairly limited to the use of channels. Furthermore there is rarely any interaction between author and reader; the author and reader are also separated by time and space. Finally, a message in writing is often highly integrated since the writer operates under no time constraint.

Keeping Biber's situation and functional differences in mind, as well as Herring's claim that chat language mimics spoken language it is possible to describe chat language as a combination of several factors. Chat language is within the written medium in terms of conveying any messages and with that comes limitations. Although chat language is supposed to mimic the characteristics of speech it cannot make use of body language / paralinguistic features to the same extent as speech can. Chat language does on the other hand have some benefits in written language that it is rare to see anywhere else.

Biber was right in that body language might be somewhat restricted to the extent that one cannot flail one's arms around in a chat room. However, facial expressions are quite simple to show in text. The question as to why this has arisen within the area of chat language can also
be asked. Looking at example 1, there are emoticons used in order to represent facial expressions.

Normal written language according to Biber does not operate under any time constraints. However the nature of chat language actually makes it operate under a certain form of time constraint. Unfortunately, the NPS Chat Corpus does not contain any timestamps that show timings of utterances relative to one another in a chat exchange so one cannot from that corpus draw any conclusion about the time between utterances. However, anyone who has used a chat should be accustomed to the way that a chat room works. One cannot expect immediate answers simply because the recipient might be talking with several people simultaneously. This might in turn cause a very high restriction on the amount of time the communicative partners have to create an utterance. One can draw a parallel in imagining a person sitting in the middle of a circle participating in five different conversations at the same time.

This time restriction affects chat language and despite it being a written medium it is highly fragmented. It is simply not a realistic expectation that people in chat rooms plan their writing as well as someone sending an email or a letter. It is possible then that the usage of "emoticons" has arisen from the need for expressing feelings in a very short amount of time. In addition, users of chat language also have the benefit of being able to use italics, bold text and further features. It should however be noticed that this can vary depending on what client one uses for chat language.

In addition, when it comes to the temporal and spatial context chat language appears to be a sort of hybrid between text and speech. Biber (ibid.:42) claims that speakers share time and more often than not space whereas writing shares neither. Unfortunately Biber does not describe what time lapse there must be between two utterances for them to be considered separated between time and space. Chat language could however be considered to be sharing at least the temporal context. However, space is not shared to the same degree since the participants might be sitting literally anywhere in the world that has access to computers and internet connections.
Additionally chat language has a high extent of interaction which is unusual for written discourse. Since communication through chat language is mostly done in real time it is open for participants to make further inquiries about any specific subject. In comparison shared knowledge from the participants about each other highly differ. It is quite possible for any individual to acquire a nickname and log in to a chat room anywhere in the world. If one does not know what the channel is about and enters blindly there is a distinct possibility that the room is filled with people with whom one has nothing in common. However the participants may also have planned to enter a specific virtual room at a specific time and may therefore have plenty in common to talk about.

6.2 Discussion based on findings

6.2.1 Ellipsis

Ellipsis found in the chat language was high. This might depend on several factors. Omission of words is a useful strategy when attempting to produce extensive numbers of utterances in a fairly short amount of time. Chat language does operate as mentioned earlier under some sort of time constraints. Under these circumstances a planned discourse seems to be rare. Omissions were also done where the utterance itself was understandable for a reader with some grasp of specialized chat language terminology (see example 6).

The fact that no omission was found in the reference sample for scientific text was not a surprise. More surprising was however the major difference between the spoken discourse and chat language. Spoken discourse had an omission of 18% whereas chat language had a total of 37%. What this might depend on is debatable but it is possible that users of chat language operate under even heavier time constraints than the type of speech which was used for comparison. It takes more time to produce 100 words of written text than it does to say 100 words. However debatable that might be, writing does take more time than talking. However, this varies depending on how fast one can type. It is possible that the extreme time constraints have created an environment in which heavy abbreviation and omission is required.
6.2.2 Pronouns and proper nouns

The usage of pronouns in the scientific text was held at a minimum, as was expected. On the pronoun continuum, chat language could be placed close to speech. What was remarkable about chat language however, is the large amount of actual proper nouns used in the discourse.

It is understandable why proper nouns are used in the scientific text; pronouns are to be held at a minimum in order to make a text look more professional and make referential links more clear. It was however unexpected to find proper nouns used so frequently in chat language. The usage of pronouns in chat language and speech is quite similar, they are a mere 1% apart. However, proper noun usage is within a difference of 7%.

There might however be a reasonable explanation as to why there is such a discrepancy. Within the context of an ordinary oral discourse between a limited amount of people there are several tactics that can be used to direct a message to a specific individual. If the utterance is in direct relation or an answer to another person's utterance it is often quite obvious to whom the utterance is intended. If that does not work, looking at the intended person or a mere physical 'poking' of the other individual might be sufficient. Chat language on the other hand is very restricted in its use of communicative channels, as Biber suggests. One cannot obviously use body language in written discourse.

There might also be another explanation as to why there is such an extensive use of proper nouns in the samples of analyzed chat language than the one mentioned above. Participants in a chat room are often more than one. In fact the labels of the chat room participants in the NPS Chat Corpus range up to and above 100. What this means is that there might be well above 100 participants operating at the same time in the same room. In such an environment to utter a personal pronoun might probably not be sufficient to obtain the attention of the intended recipient.
6.2.3 Type token ratio

The type token ration of chat language in comparison with ordinary speech and scientific text might at first glance seem remarkable. The answer is however to be found in the mean length of utterance. The mean utterance length in the analyzed chat language is a mere 2.7 words per utterance. Such a low mean utterance length will mean that the vocabulary of the chat language users is in no way higher in the text samples. The percentage of unique words in the scientific text and the chat language text is quite close. The amount of unique words will however not be interpreted as users of chat language having a particularly high overall written proficiency. Nor should it be interpreted as users of chat language having a varied vocabulary. This in turn does not mean that users of chat language are less proficient in English. The point is that several people, maybe even hundreds have their utterances recorded in the NPS Chat Corpus. Thus there is no way one can know individual proficiencies in the NPS Corpus. Nor is it possible to determine whether or not the article from Science Illustrated was written by a single person.

A type token ratio closer to 1.0 would however be more expected in writing than it is in speech. Once again, this has to do with the time constraint. Chat language as well as speech will most likely more often than not operate under heavier time constraints than other written media. Under those circumstances it cannot be expected that the users of the discourse select proper words. The result is that chat language seems to have a low lexical variation in its sentences. Creating a sentence with 3 words and a ratio of 1.0 will be considered easier than creating one with more than 20 words. The result is therefore inconclusive, the short length of the utterances in the chat language makes it hard to determine how varied the language is.

6.2.4 Passives

Passive clauses are much more common in scientific texts than in the other two forms of texts analyzed here. Passive clauses are more common in scientific writing. A passive clause can sometimes make a sentence more difficult to understand. Since passive voice is more
commonly used in scientific texts than in everyday speech (as concluded by Biber 1988) active voice being used in chat language is not something that should be perceived as odd.

Herring claimed that CMC is attempting to mimic speech and it would appear as if CMC is mimicking the active voice of speech as well. Furthermore, the passive is concerned with avoiding to refer directly to the person performing an action. Since the chat language analyzed for this essay is highly informal the use of the passive has simply not been relevant.

7 Conclusion

In summary, it can be said that the aim of this essay was to establish chat language on a continuum using speech and scientific writing as references as well as discuss the "nature" of chat language using Biber's situational and functional differences. Within the time constraints of this paper, the analytical goals have been achieved.

Where chat language ends up is however not a question that is easily answered. Just as Biber mentioned, speech and writing are not different poles of a continuum; they are more varied than so (Biber 1988).

As an example of the dynamic nature of all communications is how chat language is much closer to scientific text when it comes to the usage of unique words than chat language is to speech. Chat language and scientific text are also closer to each other when it comes to usage of proper nouns. On the other hand chat language is closer to speech in every other factor.

However for one to make a final conclusion as to which category of discourse that chat language is closest to one must have more than two types of reference samples. In conclusion the claim being put forward is that chat language in its structure and use is closer to speech than writing. This is based on both the findings and the discussion regarding Biber's functional and situational features.

The method and materials used for this essay had some strengths and weaknesses. One of the most obvious weakness is the small data sample used for research. The materials are not as strong as they could be and this author suggests that any future researchers attempt to acquire
and analyze larger samples of text. Furthermore, the type/token ratio did not prove as useful as originally thought. The amount of participants in the chat room as well as the short utterance length made the actual results inconclusive.

The methodology used has its benefits. Making use of computer-based and updated software eliminates the factor called human error. On the other hand, having the researcher doing a lot of manual counting reintroduces that factor. The optimal way would be to have the software do all the work. The results would then only have to be checked by the researcher to look for calculation errors.

Due to unforeseen events the present study has been done under pressure. The factor of time has been ever so apparent through the construction. The study did however obtain the expected result. The writer of this essay, being a frequent user of chat language had expected chat language to be quite similar to speech. The only remarkable aspect was the high amount of proper nouns used in the chat language discourses.

A possible direction for future research would be to adopt Herring's (1996:3) framework. Herring claimed that CMC has unique features. A future research project would be to determine exactly what these unusual features are.
References


**Sources of empirical data**


Appendixes
Appendix 1 is a text file containing the wordlist of NPS Chat Corpus created by using *Wordsmith 4.0*.

Appendix 2 is a text file containing the raw NPS Chat Corpus data.

Appendix 3 is a text file containing the popular science text.

Appendix 4 is a text file containing the wordlist of popular science text created using *Wordsmith 4.0*.

Appendix 5 is a text file containing the sample of spoken discourse.

Appendix 6 is a text file containing the wordlist of spoken discourse created by using *Wordsmith 4.0*. 