

Spoken Language Features in Internet Discussion Groups

Stefan Nyström
Department of English
Lund University
Linguistics, A60
Autumn 2003
Supervisor: Satu Manninen

Table of Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 1 |
| 2 | Background | 2 |
| 2.1 | <i>Internet discussion groups</i> | 2 |
| 2.2 | <i>Spoken and written language</i> | 3 |
| 2.3 | <i>Cohesion</i> | 4 |
| 2.4 | <i>Discourse markers</i> | 5 |
| 2.5 | <i>Ellipsis</i> | 9 |
| 3 | Design of the study | 12 |
| 4 | Results | 13 |
| 4.1 | <i>Instances and types of nominal ellipsis</i> | 13 |
| 4.2 | <i>Instances and types of verbal ellipsis</i> | 14 |
| 4.3 | <i>Instances and types of clausal ellipsis</i> | 14 |
| 4.4 | <i>Ellipsis as a cohesive device</i> | 16 |
| 4.5 | <i>Frequency and use of discourse markers</i> | 18 |
| 5 | Discussion | 20 |
| 6 | Concluding remarks | 21 |

1 Introduction

Discourse on the Internet takes place in a medium that lies on the borderline between the written and the spoken word. Even though texts are undoubtedly written down, terms such as *chat*, *talk* and *speak* are commonly used to describe the exchange of messages. This duality indicates that there is room for research in order to establish a clearer picture of the nature of the language used in Internet discussion groups.

Several authors have discussed the differences between spoken and written English, but in many cases they have focused on contrasting the different situations that speech and writing are produced in (Hughes, 1996), or features such as lexical density (Halliday, 1985). Crystal (2001) takes the same approach when he discusses Internet texts, and in his summary of differences between speech and writing, he focuses on vocabulary, context and stylistic choices. Based on this, he comes to the conclusion that what he calls Netspeak is closer to writing than to speech (Crystal, 2001:47). Nevertheless, when participating in a chat, or viewing logs from a chat session, the resulting text does seem quite far removed from writing – more so than text from, for instance, a news group¹.

This paper looks at Netspeak from a somewhat different viewpoint. By studying certain features that can be considered characteristic of spoken English, a chat group and a news group are examined to see whether there is a difference regarding the occurrence of features typical of spoken English, and if so, what these differences are. The presumption is that the chat group will contain more features typical of spoken English.

To answer the questions, the paper first discusses the characteristics of Internet texts in section 2.1 and what features may be considered typical of spoken English in section 2.2. In sections 2.3 to 2.5 these features – ellipsis and discourse markers – are described, and so is their contribution to cohesion in the texts.

The theory and classification of discourse markers are primarily based on the definitions and categories from Fraser (1999), with certain additions to address variants encountered in chat groups as suggested by Bergman (2003). The description and classification of ellipsis follows that described by Halliday & Hasan (1976), as does most of the theory of cohesion. The result of the research is presented in section 4, discussed in section 5 and summarized in section 6.

¹ The concepts of chat and news groups are explained in section 2.1.

2 Background

2.1 Internet discussion groups

While public discussion forums on the Internet exist in a wide variety, the basic principle behind almost all varieties is the same. Messages are sent to a central computer, or *server*, which is then responsible for distributing the message to those who have signified an interest in receiving it. The user who participates in a synchronous chat, or real-time chat, sees the messages on the computer screen almost immediately after they are sent and can respond or see responses from other participants continuously. In an asynchronous, or postponed-time chat, the messages are stored for later retrieval on a central computer and can be viewed on demand. Both real-time and postponed-time chats are usually divided into rooms or *chat groups* based on the intended topic for discussion. In a real-time chat you select the room or chat group you are interested in and ‘enter’ it in order to start receiving messages. In a postponed-type chat the same type selection is made, but the terminology is different. An important characteristic to note is that these discussions take place between several people at once, and so might be likened to a discussion at a table or in a crowded room rather than a face-to-face conversation (Crystal 2001).

One particular form of postponed-time chat is the Usenet, a loosely connected set of servers that carry discussion groups on a multitude of topics. These topics are organized into a hierarchy of *news groups*, each catering to a specific interest or subject. The time between when a message is sent and when it is received can range from minutes to hours, depending on where in the world the sender and recipient are located. This has an effect on the pace of discussions, and also on the length of the messages that are sent. Crystal (2001:144ff) cites statistics that average messages lengths range from 3.5 to 8 lines in different postponed-time chats.

Real-time chats, on the other hand, typically have much shorter messages. Crystal (2001:156) reports an average of 4.23 words per message. The time from when a message is sent and it is received ranges from fractions of a second to several seconds. Crystal (2001:31) notes that there seems to be a practical limit to what participants can tolerate at about five seconds. Whatever their cause, such as busy network connections, slow servers or problems on the client equipment, these delays are referred to as *lag*. Crystal states:

The widespread experience of lag, and the knowledge of its causes, must be one of the factors which influence the overall length of chat group messages. People are under pressure to keep their messages short, over and above the natural tendency to save time and effort while sending (Crystal, 2001:156).

This lag is different for different participants, and varies unpredictably. Coupled with the short sentences and the large number of participants in a chat group, this is the cause of perhaps the most typical feature of real-time chats: *participant overlap*. If the messages are at all possible to separate into different parallel conversations or *threads*, these threads will be mixed to a degree that is very confusing to the casual observer (Crystal 2001:157).

2.2 *Spoken and written language*

The most obvious difference between speech and writing is their different physical forms: speech uses waves of sound pressure travelling through the air, and writing prototypically uses markings on some surface. From a linguistic point of view, many of the differences arise out of the varying conditions in which they are created. Unlike writing, speech is essentially produced in a linear fashion through time. This means that much of the careful choosing of words and phrasing that goes on when we write texts is impossible, since you cannot go back and change a word once you have uttered it. Rather, when speaking we continually refine what we have just said by adding qualifiers or simply saying what we meant instead (Hughes, 1996). The need to think and talk at the same time means that speech contains more comments, looser constructions and vagueness than written language (Crystal 2003). In addition to this, the speaker is usually involved in direct interaction and thus needs to consider the continuous negotiation of subject and turn taking.

When listing differences between speech and writing many previous researchers have focused on properties outside the text. Crystal (2003:291ff) mentions facial expression and gestures, the social functions of speech, prosody and the context in which it is produced. Hughes (1996:33) is somewhat more concrete and notes that speech has a lower lexical density than writing, that it contains repetition and echoing and that it is characterised by “reformulation and refinement, sometimes by cooperation between speakers”.

The problem with these approaches is that they are vague, and do not easily lend themselves to measurements. Therefore it is difficult to base a quantitative study on such criteria.

This paper needs a quantitative way to measure how close a text is to speech, and for this countable features need to be found. Carter and McCarthy (1997) mention *discourse* markers and situational *ellipsis* as examples of features that are more frequent in spoken English. Using the occurrence of ellipsis as a characteristic of spoken English is supported by McCarthy (1998:76) who points to his studies of the million-word CANCODE corpus, and Hughes (1996:33) who agrees that “speech has a tendency toward ellipsis”. Schourup (1999:234) confirms the association between discourse markers and spoken English noting that one of the characteristic features of discourse markers is orality.

2.3 Cohesion

One common property of both discourse markers and ellipsis is that both features usually contribute to *cohesion*, the semantic relation whereby different clauses and sentences are joined together to form a text. The simplest way to define cohesion is that it “occurs when the interpretation of some element in the discourse is dependent on that of another” (Halliday & Hasan, 1976:4). The basic element of cohesion is a *tie*. This is a simple cohesive relation between two elements, such as in this example:

- (1) Wash and cook six cooking apples. Put the apples into a fireproof dish.

(example from Halliday & Hasan, 1976:3).

In this case, the term *the apples* in the second sentence forms a cohesive relation with *apples* in the first by repetition. The tie refers to an earlier item, and is thus called an *anaphoric* tie. A tie that refers to an element that comes later in the text is called *cataphoric*. For completeness, there is a form of relationship that refers to elements outside the text that is called *exophoric*. This relationship is not cohesive, but is useful when discussing ties.

There are several ways to create a cohesive relation. According to Collins & Hollo (2000:163ff), methods include *lexical cohesion*, *grammatical cohesion* and *generic signposts*. Lexical cohesion uses forms of repetition to create ties. This can be poetic forms of repetition such as alliteration, synonyms or collocation (fixed phrases such as *it's raining cats and dogs*), or it can be repetition of simple nouns or proper names. The latter is illustrated in this example:

(2) Henry presented her with his own portrait. She had always wanted a portrait of Henry.

(example from Halliday & Hasan, 1976:284)

Here, both the common noun *portrait* and the proper name *Henry* are repeated, creating two ties.

Grammatical cohesion may be achieved through the use of proforms – grammatical items which may be either co-referential or substitutions – or by ellipsis. If the reference is to the same item as in the co-text, the link is said to be co-reference, in substitution the proform refers to a similar but different entity. If the linkage is by ellipsis, all mention of the original item is omitted the second time (Collins & Hollo, 2000). To illustrate:

(3) Q: Would you mind getting me a copy too?

A1: Sure, I'd love to. ellipsis: get you a copy

A2: Of course not. substitution: not = I wouldn't mind

(adapted from Collins & Hollo, 2000:168).

When encountering ellipsis, the information that is missing must somehow be supplied from the surrounding co-text. It is quite possible to have ellipsis that presupposes items in the same sentence, but only when the presupposed item lies outside of the sentence does this create a cohesive tie (Halliday & Hasan, 1976:146). Ellipsis is discussed in more detail in section 2.5.

Generic signposts are other features of the text or dialogue that indicate that elements belong together. In the case of writing this can be chapters, paragraphs and other visual cues. In the case of spoken dialogue, this can be adjacency pairs or discourse particles (Collins & Hollo, 2000:165). Discourse particles is another name for discourse markers, which are discussed in section 2.4.

2.4 *Discourse markers*

Discourse markers are words or expressions that relate two segments of discourse; one “lying in the segment they introduce, the other lying in the prior discourse” (Fraser 1999:938). They can be single words such as *well* and *however*, or they can be short expressions like *back to my original point* or *in any case*. Discourse markers serve several functions, including turn-taking, introducing new topics and referring backwards and forwards in the discourse (Aijmer 1996:203). The following examples illustrate the concept of discourse markers, with the discourse markers set in bold type:

- (4) I want to go to the movies tonight. **After all**, it's my birthday.
- (5) A: Harry is quite tall. B: **In contrast**, George is quite short. (example from Fraser 1999:944)

Fraser (1999) uses the notation S1 and S2 to indicate the two segments, and states that the expected form to see is that of the examples, < S1. DM + S2. > (Fraser 1999:938). S1 indicates the segment in the prior discourse, and S2 indicates the segment introduced by the discourse marker.

The fact that discourse markers relate two different segments of discourse is the basis for the first criteria for deciding what is a discourse marker: *connectivity*. In example 5, the discourse marker *in contrast* establishes a relation between the segments S1 and S2. Example 6, however has no segment S1, and the corresponding form would be * <DM + S2>. This means that 'frankly' only signals an attitude toward the following segment and it is consequently not a discourse marker but a commentary marker (Fraser 1996:180).

- (6) Frankly, I don't really care. (example from Fraser 1990:390)

Another criterion is that a discourse marker does not change the propositional meaning of the segments it relates (Fraser 1990:390), something that Schourup (1999:232) refers to as *non-truth-conditionality*. In both example 4 and 5, the segments have precisely the same meaning without the discourse marker. It can, however, guide the interpretation of the segments (Fraser 1999). In example 5, the discourse marker shows that the two segments are intended to illustrate a contrast. The final condition that a discourse marker should satisfy is *optionality*. If the discourse marker is removed, the sentences must still be grammatically correct. This can also be seen in examples 4 and 5.

Except for these three conditions that are considered necessary there are others that usually apply but are not universally considered necessary (Schourup 1999:232). One of these is *initiality*. A discourse marker normally occurs in the initial position in the segment it introduces. Another is *orality*. Discourse markers occur both in written and spoken English, and many of the discourse markers listed below (*moreover, consequently, in contrast*) seem to be linked to writing because they indicate a high level of utterance planning. Nevertheless the association to a particular channel is rarely strict, and while not all discourse markers are exclusively associated with speech, Schourup (1999:234) notes that "most forms claimed to be discourse markers occur primarily in speech".

One problematic area is that there is disagreement concerning which words and expressions actually are discourse markers. Schiffrin (1987) includes *and, but, or, so, because, now, then, y'know* and *I mean*. Fraser (1990) considers *oh* to be an interjection, and discards it together with *y'know* and *I mean* because they lack connectivity; they do not relate the following segment to an earlier segment. On the other hand, Bergman (2003) analyses chat room data and based on Fraser's definitions adds the contrastive discourse marker *even though*, the implicative marker *since* and two new categories: markers of information that contain *y'know, oh* and *then*, and markers of agreement that contain *OK*. Redeker (1990) includes time adverbials such as *after that*, and *all this time*. This paper chooses to follow the definitions and lists of Fraser (1999) since that is the most specific and extensive list of discourse markers provided in the surveyed literature. Bergman's (2003) extensions have been used as well, since I agree that these words function as discourse markers.

Fraser (1999) and Bergman (2003) divide discourse markers into two classes according to function. The first class relates messages and the second class relates topics. The first class can itself be divided into subclasses, and the main subclasses are *contrastive markers, elaborative markers* and *inferential markers*. There are also other small subclasses that are not explained further.

Contrastive markers signal that S2 is in some way a contrast to S1. This can be because it contradicts the message in S1, or that it is some sort of comparison. This is illustrated by example 7:

(7) A: Chris is a happy bachelor. B: **But** Chris is female. (Example from Fraser 1999:947)

The subclass includes the following discourse markers: *(al)though, but, contrary to ..., conversely, despite (doing) this/that, however, in comparison, in contrast, in spite of (doing) this/that, instead of (doing) this/that, nevertheless, nonetheless, on the contrary, on the other hand, rather (than (doing) this/that), still, though, whereas, yet* (Fraser 1999) and *even though* (Bergman 2003).

Elaborative markers signal that S2 elaborates on S1 by adding more information along the same line as S1, or by adding further arguments to support S1. This is illustrated by example 8:

- (8) You should always be polite. **Above all**, you shouldn't belch at the table.

(Example from Fraser 1999:947)

This subclass includes the following discourse markers: *above all, also, analogously, and, besides, better yet, by the same token, correspondingly, equally, for another thing, further(more), in addition, in any event, in particular, I mean, likewise, more to the point, moreover, namely, on top of it all, or, otherwise, similarly, to cap it all off, too, well, what is more* (Fraser 1999).

Inferential markers signal that S2 somehow follows from S1, for instance that S1 is an argument for that S2 is true. This is illustrated by example 9:

- (9) The bank has been closed all day. **Thus** we couldn't make a withdrawal.

(Example from Fraser 1999:947)

This subclass includes the following discourse markers: *accordingly, all things considered, as a consequence/conclusion, as a result, because of this/that, consequently, hence, in any case, in this/that case, it can be concluded that, of course, on that condition, so, then, therefore, thus* (Fraser 1999) and *since* (Bergman 2003). Fraser mentions a few discourse markers that he does not label. These include *after all, because, for this/that reason* and *since*. Bergman (2003) groups them together with the inferential markers even though the relationship is reversed, so that S1 follows from S2, and calls the subclass *implicative* markers.

Markers of information and markers of agreement are minor classes introduced by Bergman (2003). Markers of information signal that the speaker was "familiar with the information" in S2 (Bergman 2003:19). This subclass includes *y'know, oh* and *then*. Markers of agreement signal that S2 is a statement that is said in agreement with S1 (ibid). This subclass includes the discourse marker *OK*.

The class of discourse markers that relates topics signals discourse management rather than relationships between segments (Fraser 1999:949). This is illustrated by example 10:

- (10) I'm glad that is finished. **To return to my point**, I'd like to discuss your paper.

(example from Fraser 1999:949).

This class includes the following discourse markers: *back to my original point, before I forget, by the way, incidentally, just to update you, on a different note, speaking of X, that*

reminds me, to change to topic, to return to my point, while I think of it, with regards to (Fraser 1999).

2.5 Ellipsis

According to Collins & Hollo (2000:155) ellipsis is “the omission of various obligatory clause or phrase elements, which must be recoverable in their precise form from either the immediate context or the surrounding co-text or on the basis of our knowledge of the grammar of English.” Halliday & Hasan (1976:144) define ellipsis as that which occurs when “something that is structurally necessary is left unsaid”. This is illustrated in the following example:

- (11) (i) Are you ready?
 (ii) In a minute.
 (iii) Well, where are you?
 (iv) Coming. (example from Collins & Hollo 2000:155)

In informal dialogue, terse responses like these are common. Structurally speaking, B’s responses do not constitute complete sentences and the missing information needs to be filled out by (in this case) the hearer. Clauses that require information to be supplied from a different item in the text in order to be completed are said to *presuppose* that item. Ellipsis is usually anaphoric, in that it presupposes items that have occurred earlier.

Halliday & Hasan (1976) divide ellipsis into three types: nominal ellipsis, verbal ellipsis and clausal ellipsis. Nominal ellipsis is ellipsis within the nominal group, or noun phrase. Table 1 from Halliday & Hasan (1976:147) illustrates the terms used in this paper for the different parts of a nominal group:

| | | | | | |
|---------|------------|---------|------------|--------|------------------|
| those | two | fast | electric | trains | with pantographs |
| deictic | numerative | epithet | classifier | head | qualifier |

Table 1: the nominal group

In nominal ellipsis, the head and one or more of the modifiers are removed, and one of the remaining modifiers takes on the role of head. This is sometimes the epithet, rarely the classifier and usually the deictic or numerative. Cohesion is created since a previous nominal group is needed to supply the information that the elliptical nominal group presupposes. This example illustrates different forms of nominal ellipsis.

- (12) Here are my two white silk scarves.
- a) where are yours?
 - b) I used to have three.
 - c) Can you see any black?
 - d) Or would you prefer the cotton? (Halliday & Hasan, 1976:150)

In 12a) the deictic *yours* functions as head, and all other elements are presupposed. In 12b) the numerative *three* functions as head, and scarf, silk and white are presupposed. In 12c) the epithet *black* functions as head with *silk* and *scarf* presupposed (but the number and owner are not) and in 12d) the classifier *cotton* functions as head, presupposing only *scarf*. The last form is rare, since a classifier usually is a noun that would easily be interpreted as a head in its own right (Halliday & Hasan, 1976:150).

Verbal ellipsis occurs when elements from the verbal group, or verb phrase, are omitted. The general structure of a verb phrase is shown in table 2.

| | |
|-----------------|--------------|
| have been | swimming |
| verbal operator | lexical verb |

Table 2: the verbal group

Generally, a verbal group requires 1) a lexical item (the lexical verb) and 2) systemic selections that must be made when a verb group is used. These systemic selections include finiteness, polarity (positive or negative), voice (active or passive) and tense (past present or future) (Halliday & Hasan, 1976:167). If some of these are not represented, the verbal group is elliptical. The systemic selections do not necessarily correspond to separate words; in a sentence such as *John saw Mary through the window* the systemic selections are made, even though they are all embedded in the single word *saw*. This example illustrates verbal ellipsis:

- (13) a) Have you been swimming? – Yes I have.
 b) What have you been doing? – Swimming. (example from Halliday & Hasan, 1976:167)

There are two different types of verbal ellipsis: *lexical ellipsis* and *operator ellipsis*. Any case where the lexical verb has been omitted is called lexical ellipsis, such as 13a). It is possible that other parts of the verbal group be omitted as well, but as long as the lexical verb is omitted it is called lexical ellipsis. This is often called ellipsis from the right, since the

lexical verb is always the rightmost word in the verbal group. The opposite, ellipsis from the left, is formally called operator ellipsis. Here the operator in question is the *verbal operator have been*, as in *have been swimming*. In 13b) the entire operator has been subjected to ellipsis. In most cases of operator ellipsis, everything is presupposed except for the lexical verb. If there is operator ellipsis, the subject of the sentence is always presupposed as well (Halliday & Hasan, 1976:167).

Clausal ellipsis covers other kinds of ellipsis that affect the structure of the clause. Depending on which part of the clause is presupposed, it can be either *modal ellipsis* or *prepositional ellipsis*. Table 3 illustrates the division of the clause as used in this paper.

| | | | |
|---------------|----------------|-----------------------|-------------------------------|
| The duke | was | going to plant | a row of poplars in the park. |
| Modal element | | Propositional element | |
| Subject | Finite element | Rest of verb group | Complements & adjuncts |

Table 3: *the clause*

The modal element contains the subject plus the finite element of the verb group. The propositional element consists of the residue (*ibid*).

The typical use of propositional ellipsis is answers “where the mood and the polarity are the principal components of the message: typically, responses to statements and yes/no questions” (Halliday & Hasan, 1976:198). This is illustrated by the following example, where the answer only contains the polar element.

(14) Q: Are you coming? A: Yes [I am coming]. (example from Halliday & Hasan, 1976:209)

The typical use of modal ellipsis is answers to questions of the form “what (did, does, etc) ... do?” (Halliday & Hasan, 1976:198) as illustrated by this example that shows an answer where only the WH-element is present:

(15) Q: How much does it cost? A: Five pounds. (example from Halliday & Hasan, 1976:209)

3 Design of the study

The purpose of this study was to compare two different sets of source data – one from a real-time chat, one from a postponed-time chat – and to see if there was a difference regarding the occurrence of features that are typical of spoken English. These sets were collected from two different sources.

The real-time source, referred to as the chat group, was collected from the Microsoft Network chat group *politics* on November 10th, 2003. Microsoft Network is a portal service containing among other things e-mail, chat groups, news groups and a search facility. 360 consecutive messages were collected, and after removing messages that were either administrative notices or contained less than one intelligible word of content, 269 messages were studied in detail. In the study, the messages will be referred to as P1 – P360². The chat group *politics* was chosen for two reasons: it was the chat group with the highest number of active users at the time and it would allow comparison with Bergman (2003) who has studied the same group.

The postponed-time source, referred to as the news group, was collected from Google Groups and covered messages from November 11th 2003 until November 26th 2003. Google Groups is an archive of Usenet messages from 1981 to the present, and is probably the most complete archive of this kind in existence (Google Groups FAQ, 2003). The messages were selected using two criteria. First, they should be posted to the news group *talk.politics.misc* since this was considered to be the closest equivalent to the chat group *politics*. The reason for choosing similar groups was the assumption that this would minimize the influence of external factors, such as background and age of the participants. Second, they should belong to a thread of at least 50 messages, so that a continuous discussion had had time to develop. In the end 50 messages each were selected from two threads, “Hitler was no clear and present danger to the United States” and “We can’t afford Democrat liars any longer”. In the study, these messages will be referred to as H1-H50 and D1-D50 respectively. After removing duplicates and quoted material 405 sentences were studied in detail.

The messages were analysed with respect to discourse markers using Fraser’s (1999) categories and definitions. A parallel analysis was also done using a wider definition of discourse markers as suggested by Bergman (2003) and described in section 2.4. The use of

² The numbering scheme includes the messages that have been excluded from the main study, since some of these will be referred to in the discussion in section 5.

ellipsis was analysed using Halliday & Hasan's (1976) categories and definitions. This was used both to categorise the different types of ellipsis found according to the coding scheme suggested by Halliday and to see how they contributed to cohesion.

A second analysis was later made to look for other cohesive elements than ellipsis and discourse markers. These results are discussed in section 5.

4 Results

4.1 Instances and types of nominal ellipsis

Nominal ellipsis was found in 3 instances in the news group (0,7% of the sentences) and not at all in the chat group. Table 4 summarises the results of the analysis for nominal ellipsis.

| Types of nominal ellipsis | Chat | News |
|--------------------------------|------|------|
| <i>Numerative as head</i> | 0 | 2 |
| <i>Epithet as head</i> | 0 | 1 |
| <i>percentage of sentences</i> | 0% | 0,7% |

Table 4: nominal ellipsis

These examples illustrate where nominal ellipsis occurred:

- (16) Bush managed to turn increases of less than or around 1,000,000,000 [dollars]³ under the Clinton administration into increases of ... what ... over 500,000,000,000 [dollars]? (D3)
- (17) Could it be that he didn't collect enough [money] to cover what he's spending? (D20)

In the first sentence, the numeratives *1,000,000,000* and *500,000,000,000* function as heads. In the second sentence the epithet *enough* functions as head. In both these cases the common noun that originally functioned as head has been subjected to ellipsis. Since the entire discussion concerned money and budget, the various types of references to money were presupposed. In these particular cases the presupposed item did not occur in the co-text, but only in the context of the discussion.

³ For the purpose of clarity, the elements subjected to ellipsis have generally been filled out with a suggested replacement. This is indicated by putting the replacement in square brackets.

4.2 Instances and types of verbal ellipsis

Verbal ellipsis was found in 6 instances in the chat group and 9 instances in the news group, or in both groups 2.2% of the sentences. Table 5 summarizes the results of the analysis for verbal ellipsis.

| <i>Type of verbal ellipsis</i> | <i>Chat</i> | <i>News</i> |
|--------------------------------|-------------|-------------|
| <i>Lexical ellipsis</i> | 3 | 8 |
| <i>Operator ellipsis</i> | 3 | 1 |
| <i>percentage of sentences</i> | 2,2% | 2,2% |

Table 5: verbal ellipsis

The majority of the cases of verbal ellipsis in the news group are lexical ellipsis with 8 instances compared to 1 instance of operator ellipsis. In the chat group lexical and operator ellipsis occur equally often, or 3 times each. In both groups, lexical ellipsis occurs either in short phrases indicating emphasis, in tag questions or in simple answers:

- (18) Because the French don't control the media, Jews do [control the media] (H7)
(19) All you demmies have left is your lies, isn't it? (D40)
(20) Many do jp (P215)

Operator ellipsis occurs in yes/no-questions where the initial do is omitted, as illustrated by these examples:

- (21) You know for sure the economy wouldn't have recovered just as fast under another president? (P214)
(22) Think KCK will share them with us? (P221)

This does not agree with Halliday & Hasan (1976:193), who state that typically, "we find operator ellipsis in answers to questions".

4.3 Instances and types of clausal ellipsis

By far the most common form of ellipsis was clausal ellipsis, found in 13.4% of the sentences in the chat group and 11.6% of the sentences in the news group. The results of the analysis for clausal ellipsis are summarized in table 6 with a listing of how many instances of each type of clausal ellipsis, and what percentage of the sentences contained these kinds of ellipsis.

| Type of clausal ellipsis | Chat | | News | |
|--|-------------|------|-------------|------|
| <i>Propositional ellipsis</i> | 4 | 1,5% | 7 | 1,7% |
| <i>Modal ellipsis</i> | 7 | 2,6% | 6 | 1,5% |
| <i>General ellipsis (only WH-element remains)</i> | 4 | 1,5% | 9 | 2,2% |
| <i>General ellipsis (only polarity element remains)</i> | 5 | 1,9% | 5 | 1,2% |
| <i>General ellipsis, other single clause element, yes/no question/answer</i> | 3 | 1,1% | 1 | 0,2% |
| <i>General ellipsis, other single clause element, WH- question/answer</i> | 2 | 0,7% | 2 | 0,5% |
| <i>General ellipsis, other single clause element, otherwise</i> | 12 | 4,5% | 18 | 4,4% |
| <i>percentage of sentences containing clausal ellipsis</i> | 13,4% | | 11,6% | |

Table 6: clausal ellipsis

Modal ellipsis occurred in 7 sentences in the chat group, and 6 sentences in the news group. The typical use of modal ellipsis, according to Halliday & Hasan (1976:198), is answers to questions of the form “what (did, does, etc) ... do?” The results of this study do not agree; in no case does the modal ellipsis occur in an answer. The most common form is instead a contracted question or a fragmentary sentence, such as these examples:

- (23) forget what? (P117)
(24) hate to go out (P247)
(25) Want to know where that Clinton revenue surplus was[?] (D48)

The typical use of propositional ellipsis is answers “where the mood and the polarity are the principal components of the message: typically, responses to statements and yes/no questions” (Halliday & Hasan, 1976:198). Out of the eleven instances, only three use propositional ellipsis differently – to reinforce a statement or in tag questions as these examples illustrate:

- (26) Ah tried to get them torpedoman, Ah really did. (D18)
(27) You do realize that I’m posting satirical statements that are the opposite of what I actually believe in order to show that the invasion of Iraq was justified by the very beliefs of those who are arguing against me, don’t you? (H47)

General ellipsis of the clause resulting in single clause elements is expected in question-answer sequences and other forms of responses. Such single clause elements are found as questions or answers in 14 instances in the chat group (5,2%), and 17 instances (4,2%) in the news group, as shown by these examples. First a single clause element used as a yes/no-question, followed by an answer where only the polarity element remains, then a WH-

question in a single clause element and finally a full WH-question with an answer where only the WH-element remains:

- | | | |
|------|---|--------|
| (28) | natural too | (P119) |
| | yes | (P122) |
| (29) | DAVE: How far west? | (P64) |
| (30) | What do you Democrats have against marriage?? | (D42) |
| | None. | (D43) |

Single clause elements outside of question-answer groups are found in 12 instances (4.5%) in the chat group and 18 instances (4.4%) in the news group. This category contains the many short comments and fragmentary sentences found throughout both the chat group and the news group. The majority of these are sentences where a form of *pronoun+ be* (sometimes + ‘a’) has been left out such as in these examples:

- | | | |
|------|--|-------|
| (31) | KC, [that is] unfounded | (P81) |
| | Nope, [that has been] founded and proven | (P86) |
| (32) | [That is a] Good point. | (H10) |
| (33) | [That is] Amazing, huh? | (D3) |

The remaining four instances were

- | | | |
|------|---|--------|
| (34) | NOT. | (D12) |
| (35) | OK, [I say] as you say, “nevermind”. | (D33) |
| (36) | What do you democrats have against marriage? [Give me] specifics, please..... | (D42) |
| (37) | dee you never have shown much knowledge. [You have] just [shown] partisanship | (P256) |

4.4 *Ellipsis as a cohesive device*

All instances of ellipsis were analysed to see if there was an associated presupposition in the co-text, and if so, where that presupposed element was located. Table 7 summarizes the results of this analysis.

| Target | News | | Chat | |
|----------------------------------|------|-------|------|-------|
| <i>Same sentence</i> | 8 | 14,3% | 3 | 8,6% |
| <i>Other sentence in message</i> | 10 | 17,9% | 2 | 5,7% |
| <i>Other message</i> | 22 | 39,3% | 14 | 40% |
| <i>No presupposition</i> | 16 | 28,6% | 16 | 45,7% |

Table 7: location of presupposed elements in ellipsis

In the chat group 16 instances of ellipsis, or 45,7%, were not resolved by a presupposed item in the co-text at all. The corresponding numbers for the news group were 16 instances and 28,6%. These were sentences such as the following examples:

- (38) whaTAS THAT CITY NEXT TO CHICAGO [CALLED?] (P36)
(39) [Give me] specifics, please... (D42)

In the chat group 3 instances of ellipsis, or 8,6%, were resolved by a presupposed item in the same sentence. The corresponding numbers for the news group were 8 instances and 14,3%. This case was found in sentences such as the following example:

- (40) Because the French don't control the media, Jews do [control the media]. (H7)

None of the preceding cases of ellipsis contribute to cohesion, since they do not presuppose elements in other sentences. The remaining cases, however, do. Ellipsis resolved by elements in a different sentence in the same message was naturally more common in the news group since almost no messages in the chat group contained multiple sentences. This occurred in 3 sentences or 8,6% in the chat group, and 8 sentences or 14,3% in the news group:

- (41) Germany declared war on the United States. Iraq didn't [declare war on the United States]. (H12)

Finally, 14 instances (40%) of ellipsis in the news and 22 instances (39,3%) in the chat group were resolved by a presupposed item in another message:

- (42) this isn't the danny I know lol (P124)
blue maybe I am [the danny you know] (P130)

In summary, the fraction of elliptical items that are cohesive (that are resolved by a presupposed item outside the sentence) is slightly larger in the news group than in the chat group: 57,2% in the news group versus 45,7% in the chat group.

4.5 Frequency and use of discourse markers

The results of the analysis of discourse markers are presented in table 8.

| Discourse marker | Type | Chat | News |
|------------------|-----------------|------|-------------|
| After all | Implicative | | 1 |
| Also | Elaborative | | 1 |
| And | Elaborative | 4 | 13 |
| Because | Implicative | 1 | 3 |
| But | Contrastive | 5 | 5 |
| Hey | Information 1) | | 1 |
| In either cases | Implicative 1) | | 1 |
| In the Xth place | Elaborative 1) | | 2 |
| Oh | Information | | 5 |
| OK | Disagreement 1) | | 4 |
| So | Implicative | 5 | 9 |
| Then | Implicative | | 1 |
| Well | Elaborative | 3 | 3 |
| Summary | | 18 | 36 (49) |
| Sentences | | 269 | 405 |
| %ages | | 6,7% | 8,9 (12,1)% |

Table 8: discourse markers

1) This classification is made in this essay

The chat group uses only five different discourse markers: *and* (four times), *because* (once), *but* (five times), *so* (five times) and *well* (three times). The news group uses the same discourse markers (thirteen, three, five, nine and three times respectively) and also several more: *after all*, *also*, and *then*, once each.

In addition to those categorised by Fraser (1999) and Bergman (2003), four new discourse markers were found in the news group. The first, *hey*, is classified by this paper as a marker of information, since it indicates that S2 is a piece of information triggered by S1. This is illustrated in the example below:

- (43) How would you boost the economy? (D13)
 By helping to get rid of Bush and his voodoo economic plans, duh. (D16)
 > This is the usual non-response from the kookist Libs. (D18)
Hey, it worked when we got rid of Bush senior. (D18)

The second discourse marker *in either cases* occurred once. It is merely a variant of *in any case*, which Fraser (1999) classified as an inferential marker. Accordingly, *in either cases* is classified in this paper as an inferential marker since it indicates that S2 is a conclusion that follows from S1 as this example illustrates:

- (44) [...] it was obvious that either Germany would take over Western Europe (UK included) or that Russia will. **In either cases**, it was very very bad news for the US [...]. (H32)

The third discourse marker, *in the first/second place*, is used twice to indicate that S2 parallels and augments S1. Hence, it is classified in this paper as an elaborative discourse marker. This is illustrated by the following example:

- (45) In the first place, Hitler was a clear and present danger [...]. **In the second place**, his submarines were already sinking American merchant vessels. **In the third place**, Germany declared war on The United States. (H2)

The fourth and final discourse marker that deviates from the classifications of Fraser (1999) and Bergman (2003) is *OK* with its alternative spelling *okay*. It occurred four times in the news group, as this example illustrates:

- (46) **Okay** let me see your logic here. Hitler was a threat, therefore anyone you say is a threat is a threat. WWII was justified, therefore the Iraq invasion was justified. (H48)

Bergman (2003:21) classifies this as a marker of agreement. In all cases observed in this study, however, while it has superficially marked agreement, the sarcastic tone in the surrounding co-text clearly indicated that it functioned as a marker of disagreement.

In summary, the chat group has a lower density of discourse markers than the news group, regardless of whether you follow Fraser's (1999) list strictly or you use the more extensive list suggested in section 2.4: 6,7% of the sentences in the chat group compared to 8,9% in the news group for the Frasier (1999) list and 12,1% using the more extensive list. Worth noting is that there were no topic-change discourse markers found in any of the groups. See section 5 for a discussion on this.

5 Discussion

The initial presumption, that the chat group should contain more features typical of spoken English, does not hold. There are fewer discourse markers in the chat group *politics* than in the news group *talk.politics.misc*, and there are not significantly more cases of ellipsis in the chat group *politics*. How can this be explained?

In the chat group, there are many fragmentary sentences that do not constitute discourse markers or ellipsis but contribute to the conversational nature of the chat. Some examples are:

- | | | |
|------|---------|--------|
| (47) | lol | (P48) |
| (48) | omg lol | (P51) |
| (49) | haha | (P152) |

They do not fit into the theory of ellipsis presented by Halliday & Hasan (1976) and seem more related to the prosodic features of conversation. Therefore ellipsis according to their definition may not be a good way to describe the language used in chat groups. As noted in the results, several of the different categories of ellipsis were not used in the expected way, which also supports the conclusion that chat groups use ellipsis differently.

Several of the functions for discourse markers are absent or rare in chat groups. The function *topic shifting* (McCarthy 1998) is not made explicitly in chat groups, but instead topics shift gradually (Crystal, 2001:163). This may explain why no topic change discourse markers were found⁴. The same argument applies to *returns to interrupted topics* (McCarthy 1998): these do not usually occur (Crystal 2001:163). Another function, *turn-taking*, is mentioned by Schiffrin (1987) and Aijmer (1996). Crystal notes that “the notion of turn-taking ... is undermined” (2001:162) and that chat groups allow extraordinary disruptions in turn-taking (2001:178). Since messages belonging to different conversations are interleaved, and many messages cannot be attributed to a specific conversation, there is little use of DM for explicit turn taking.

Generally speaking, there were few discourse markers found in the chat group. On average, 6,7% of the sentences in the chat group contained a discourse marker and only five

⁴ Note that gradual topic shifting is a characteristic that is also present in spoken conversation and thus might be seen as supporting the notion that chat is closely related to speech.

different discourse markers were used. Compared to the results of Bergman (2003) who found 10 different discourse markers, in 5% of her sentences, this shows a slightly higher frequency but a lower variety. This study agrees with her result that there are few discourse markers in chat room conversation.

The relatively high fraction of ellipsis that did not contribute to cohesion, 45.7% in the chat group and 28,6% in the news group, stood out. Looking closer at the sentences containing general ellipsis of the clause, the majority of these had an ellipsis of a form of *pronoun + be* (sometimes + 'a'). This kind of ellipsis is not really accounted for by Halliday & Hasan (1976), but it is my opinion that this is a fragmentary sentence created by the need to be brief in order to combat lag, keep up with conversation in the real-time environment and comply with the general requirement for brevity in chat room conversations (see end of section 2.1).

The fraction of elliptical items that are resolved within the same message in the news group is more than twice than in the chat group; 32,2% versus 14,3. This is expected, since a majority of the news group messages consist of more than one sentence, but only nine of the chat group messages. The unexpected result in this examination is that the news group creates cohesion by using ellipsis to a greater extent than does the chat group.

During a second examination of the source material, a preliminary review was made to see if there were other cohesive features present that separated the chat group from the news group. It was found that the chat group makes extensive use of repetition to create cohesion. Out of the 269 messages examined, 81 (30%) started with a nickname or an abbreviation of a nickname, indicating that the message was part of a certain conversation or directed at a certain person. This occurs in only 5 of the 100 messages in the news group, or 5 of the 405 sentences (1,2%).

6 Concluding remarks

The initial assumption this paper set out to prove was that chat groups contained more features characteristic of speech than news groups, and therefore were closer to speech. From the data analysed, it was not possible to confirm that assumption when looking at discourse markers or ellipsis. In fact, the news group contained more discourse markers, and a larger variety. Both groups used approximately the same amount and types of ellipsis.

Since both ellipsis and discourse markers are cohesive devices, the analysis also looked at the groups to see if there were differences in how cohesion was created. A closer

examination of how ellipsis was used did not help confirm the initial assumption; the chat group made less use of ellipsis in a cohesive fashion than the news group. However, the chat group differed markedly from the news group in that it used a large amount of repetition of the names of the participants to signal that messages belonged to a certain discussion. This occurred only several times in the news group.

The work done for this paper has indicated areas that would warrant further study. The problem with using ellipsis to describe the chat group certainly raises the question how to describe the language in chat groups. Even though the differences found were not those expected, there were some. What other features distinguish chat groups from news groups? Finally, cohesion in chat groups and news groups is hardly covered in literature. What are the mechanisms by which news groups and chat groups generate cohesion?

References

Primary sources:

- MSN chat group *politics*. See <http://groups.msn.com/>. Extracted on 10 November, 2003.
- News group *talk.politics.misc*. Archived at <http://groups.google.com/>. Threads starting 11 November, 2003 and ending 26 November, 2003.

Secondary sources:

- Aijmer, Karin. 1996. *Conversational Routines in English: Convention and Creativity*. London: Longman.
- Bergman, Jill. 2003. *The use of discourse markers in chat room conversation*. 60-point essay. Lund, Department of English at Lund University.
- Carter, Ronald. & McCarthy Michael. 1997. *Exploring spoken English*. Cambridge: Cambridge university press.
- Collins, Peter. & Hollo, Carmella. 2000. *English Grammar - an Introduction*. London: Macmillan.
- Crystal, David. 2001. *Language and the Internet*. Cambridge: Cambridge University Press.
- Crystal, David. 2003. *The Cambridge Encyclopedia of the English Language 2nd ed.* Cambridge: Cambridge University Press.
- Fraser, Bruce. 1990. An Approach to Discourse Markers. *Journal of Pragmatics*, 14, 383-395.
- Fraser, Bruce. 1996. Pragmatic Markers. *Pragmatics*, 6, 167-190.
- Fraser, Bruce. 1999. What Are Discourse Markers? *Journal of Pragmatics*, 31, 931-952.
- Halliday, M. A. K. & Hasan, Ruqaiya 1976. *Cohesion in English*. English Language Series 9. London: Longman.
- Halliday, M. A. K., 1985. *Spoken and Written Language*. Victoria: Deakin University Press
- Hughes, Rebecca. 1996. *English in Speech and Writing. Investigating Language and Literature*. London: Routledge.
- McCarthy, Michael. 1998. *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press.
- Redeker, Gisela. 1991. Ideational and Pragmatic Markers of Discourse Structure. *Journal of Pragmatics*, 14, 367-381.
- Schourup, Lawrence. 1999. Discourse Markers. *Lingua*, 107, 227-265
- Schiffrin, Deborah. 1987. *Discourse Markers*. Cambridge: Cambridge University Press.

Web resources:

- Google Groups FAQ 2003-12-28, See <http://www.google.com/googlegroups/help.html>.