

The Genre Structure of Bulletin Board Messages

Maite Taboada

Department of Linguistics, Simon Fraser University

mtaboada@sfu.ca

Abstract

Messages posted on bulletin boards on the Internet represent a new form of communication. They draw both from casual conversation and from other written genres in their structure and characteristics. In this paper I explore the genre structure of bulletin board posts. Genre is defined a purposeful, goal-oriented activity, which evolves in clearly defined stages. Messages posted on bulletin boards correspond to a variety of genres: argumentation for an opinion, personal attacks on other posters, pros and cons of a certain product, demands for advice, etc. Because the goals of each genre are different, the genres will also be different in structure. This paper provides definitions of four of these genres and an exploration of their structural characteristics. The descriptions are useful in computational applications such as topic classification and information retrieval.

KEYWORDS: Bulletin boards, linguistics, genre, systemic functional linguistics, rhetorical structure theory, text classification

1. Introduction

Messages posted on bulletin boards can be classified into different types, or genres. This classification can be done on an intuitive basis, by examining the overall purpose of the message, and certain expectations of what it is one achieves by posting a message (providing information, asking for information, arguing for a certain point of view, etc.). Ideally, this intuitive classification should be automated for search and retrieval purposes.

The problem is how to formalize the classification of individual messages based on their characteristics. Research in computational linguistics and information retrieval has provided a number of methods to perform automatic text classification. These are often based on the presence of certain keywords (Lewis, 1992), extraction patterns or signatures (Riloff and Lorenzen, 1999); linguistic features (Biber, 1988) and style markers (Stamatos et al., 2000); the similarity of a given document to a number of “model” or “neighbouring” documents (Friedman, 1999; Yang

and Liu, 1999); and, in general, statistical techniques (Yang, 2000).

I propose to use structural characteristics of the messages as one of the bases for text classification, together with some lexical and grammatical characteristics. The first step towards automatic classification is to examine the structural characteristics of bulletin board messages in general, and of different genres within those messages.

Such classification could be used when searching for messages. One could add the genre of the document to search criteria, and search only for messages that are “informative”. Bulletin board administrators might want to search for “personal attacks”, to be removed from the board. In a political discussion thread, it would be very useful to be able to search for messages that offer “pro” and “con” arguments for a particular position or point of view.

The characteristics of different genres of bulletin board messages described here will not suffice to perform automatic text classification. Structural characteristics are not as reliable as the presence of keywords, for instance. However, structure can become one more tool in the complex task of classifying text automatically.

The paper first makes some remarks about genre, and its relation to structure, in Section 2. Section 3 describes the general characteristics of bulletin board messages. Section 4 is the main analysis section, describing the characteristics of four typical genres within bulletin board messages. Section 5 provides some discussion of applications, and conclusions.

2 Genre and Structure

The definition of genre used in this paper borrows from the writings of Bakhtin, formalized in the work of systemic functional linguists. For Bakhtin, language is realized through individual concrete utterances by participants in the various areas of human activity. These utterances reflect specific conditions and goals of each such area through different aspects namely content, linguistic style (selection of lexico-grammatical resources) and compositional structure. All three aspects are linked to the whole of the utterance and determined by the nature of the particular sphere of communication where they are produced. “Each separate utterance is individual, of course, but each sphere in which language is used develops its own relatively stable types of these utterances. These we may call speech genres.” (Bakhtin, 1986:60). Speech genres are numerous and diverse, “because the various possibilities of human activity are

inexhaustible, and because each sphere of activity contains an entire repertoire of speech genres that differentiate and grow as the particular sphere develops and becomes more complex” (Bakhtin, 1986:60).

This notion of genre has been refined and made more specific by researchers working in Systemic Functional Linguistics. Martin defined genre as “a staged, goal-oriented, purposeful activity in which speakers engage as members of our culture” (Martin, 1984:25). Under such definition, there are very many types of genres, a few of which are: literary genres (short stories, romance, tragedies, sitcoms); popular written genres (instructional manuals, newspaper articles, recipes); and educational genres (lectures, tutorials, essays). Eggins (1994:26) points out that there are also a number of everyday spoken genres: buying and selling things, telling stories, gossiping, making appointments, chatting with friends.

Genres are mainly defined through their social purpose and the stages necessary to fulfill that purpose. Stages are the steps interlocutors take in order to reach their goal(s), and they constitute the genre’s schematic structure. For instance, the high-level generic structure of a telephone conversation consists of an opening, a problem-solving stage (or main body) and a leave-taking stage. Mitchell (1957) proposed the following schematic structure for a market auction in a Moroccan marketplace. The symbol ^ indicates sequence.

(1) Auctioneer’s Opening ^ Investigation of Object of Sale ^ Bidding ^ Conclusion

Each one of those steps is a stage, and they appear in the order indicated. Stages may be recursive: it is quite likely that bidding will include several instances of offer and counter-offer.

Genre is used to describe language use in the context of culture. Another construct, register, represents the context of situation. Register is concerned with the aspects of the situation that have an impact on the linguistic expressions used. Halliday (1978; 1989) proposed three situation variables (also in Eggins, 1994:52):

- Field: what the language is used to talk about.
- Tenor: the role relationships between the interactants.
- Mode: the role language is playing in the interaction.

The following configuration could describe the present paper: field: aca-

demic discussion of text analysis; tenor: specialist-to-specialist communication; mode: written.

We can then use genre to define overall purpose and staging, and register to define the situational variables. Whereas situational variables may be similar in some of the messages in this study, the structural characteristics and the overall purpose are different, as we will see in Sections 3 and 4.

Egins and Slade (1997:231-235) propose six steps for a generic structure analysis. Their study was one of casual conversation, but the same steps can be used to study and describe any genre, written or spoken. The six steps are:

1. Recognizing a chunk. A “chunk” is a part, a segment or series of segments of a text that has a global or macro-structure.
2. Defining the social purpose of the chunk and labelling the genre. This step not only includes classifying the overall function of the genre and its name, but it also involves identifying the way the text type constructs social reality. In this step we would define the social practices the text refers to, and the attitudes and values formed by and reflected in it.
3. Identifying and differentiating stages within a genre. Stages are constitutive elements of genres. They can be identified through the use of functional labels.
4. Specifying obligatory and optional stages. The obligatory elements are defining of the genre and they are key elements in recognizing a genre. Optional elements are not defining features and can occur across genres.
5. Devising a structural formula. This includes writing down the stages in a linear sequence, together with some notation that would show their ordering and whether they are obligatory or optional. In Egins and Slade’s (1997) notation -- widely used after Hasan’s (1984) analysis of nursery tales and service encounters -- the symbol ^ denotes order of the stages with respect to each other, and stages enclosed in parentheses are optional. For example, a structural formula for narratives would read: (Abstract) ^ Orientation ^ Complication ^ Evaluation ^ Resolution

^ (Coda).

6. Analyzing the semantic and lexico-grammatical features for each stage of a genre. Lexico-grammatical choices (of words, linguistic patterns and discourse characteristics) will differ across genres, but also across the different functional stages within a given genre.

In summary, Eggins and Slade's system will have three basic steps: (1) identifying the genre, (2) finding a structural formula that will represent most instances of the genre and (3) analyzing the linguistic characteristics of each stage. In this paper, I concentrate on the first two steps, identifying first four different genres within bulletin board messages (each message being a 'chunk'). In the second step, structural differences outline the differences among the four genres. I will only briefly discuss linguistic features for each genre (step 3); a more complete analysis of those features is left for future research.

3 Bulletin Board Messages: General Characteristics

Electronic bulletin boards emerged first as an electronic version of the "community bulletin board" (Bulletin Board Corp., 2003) that one can find at the grocery store or the post office in a small community. From this point of view, bulletin boards are places to exchange information and offer or request services. The first dial-up bulletin board, the Chicago Bulletin Board System, was created in 1978 by Ward Christensen and Randy Suess (Christensen and Suess, 1989)¹. These first bulletin board systems were places to exchange files and information. In the 1990s, with the advent of the HTTP protocol and the use of browsers², some bulletin boards migrated to Internet web sites, and were renamed message boards or web boards. There, they became places to not only exchange information and services, but to carry on discussions on different topics: products, politics, the stock market, hobbies, etc.

Messages posted on bulletin boards range in content and especially in format, from the letter-like style to the conversation-like stream, making it difficult to classify them as either writing or speech (Herring, 2001). Crystal (2001:42) describes web pages, e-mail, chat groups and virtual worlds according to what spoken or written language criteria they fulfill. The picture that emerges from Crystal's description is that the communication that takes place on a bulletin board shares characteristics of

both written and spoken language (loosely structured, but not face-to-face; space-bound, but not graphically rich).

S. Yates (2000) argues that most computer-mediated genres have an origin in the office memo, and J. Yates and Orlikowski (1992) trace a chain of evolution, from business letters to memos and from those to e-mail communication. The messages are indeed similar to letters in that they often have the familiar structure of letters: opening, body, closing. For instance, Example (2)³ could be, from a structural point of view, a written letter, or an e-mail message. It has a salutation (*Rad & others*), a main body, and a closing (*Good luck: Joe*). Its language would not qualify as that of a formal written letter. The structure, however, is very similar⁴.

(2) By: JOEGELT

18 Nov 2000

Rad & others

100% correct (imho) on this Basher call !!

Have many other Basher sightings worth investigating.

Good luck :

Joe

Bulletin board messages include many other characteristics that one would not find in a letter, such as a lack of openings and closings (including no signature). In that sense, they are closer to the structure of informal e-mail communication. Herring (1996) studied messages sent to a listserv, and found they were quite similar to letters, except that few (13% of the ones studied) had a salutation, possibly because headers including “to” fields are inserted automatically by the system. Example (3) has only a brief opening (*doc*), on the same line as the message, and no closing. It also makes use of capitalization for emphasis.

(3) By: angiela22

04 Apr 2001

doc - you do great research, thanks. We REALLY need to know about revenues!

Additionally, messages exhibit some of the characteristics of conversation, in that they are turns in an ongoing exchange, and are often informal in style. Example (4) is a turn in response to two other posts, whose authors are mentioned (*Steve and Putter*). The language is quite informal.

(4) Date: 4-22-2001 22:48:00

Subject: OT: take the high road

Author: trustme2069

Steve I agree with you on this one for sure. And I do understand Putter asking for e-mail, because its too easy to get caught up in a P***isg contest with these jerks. They come they go they change their name or steal ID's from others, but within 2 or 3 posts, we all know where they are headed...to the ignore list.trustme

The label *persistent conversation* has been applied to this type of communication (Erickson, 1999), because, although it is conversational in many aspects, it is also preserved for future review. Bregman and Haythornthwaite (2003) consider that capturing the communication at the time of delivery is an essential attribute of persistent conversation (as opposed to reproducing or summarizing an instance of communication after it has taken place). It is also crucial that the logging of the proceedings is performed as part of the activity, not to serve a secondary purpose. Court recordings are recorded for archival and reviewing processes, but the judicial process may be performed without them. Message archival, on the other hand, defines bulletin boards to a large extent. Without archival, they would become mailing lists or simply private e-mail messages.

On the other hand, bulletin board messages are unlike conversation in that conversational openings often receive no response. Herring and Nix (1997)⁵ found that 18% of messages in a chat room were not responded to. In a study of three asynchronous listserv discussions, Herring (in press) reports that 34% of participants who posted messages received no response.

These messages are at the same time clearly anchored in the written domain, with links to previous written documents (quoting other posters' messages). Herring (1999; 2001) argues that quoting creates the illusion of conversational coherence, recreating an adjacency pair, as if the previous turn (the quoted material) and the current turn had happened at the same time. Tanskanen (2001) observed that quoting was frequently used to show disagreement with an earlier message, or to initiate a correction of the quoted material. Example (5) uses quotes to refer to specific parts of another poster's message, and to disagree with those. An interesting aspect of this message is that the author used quotation marks (single and double), rather than the usual on-line > symbol (or a line to the left of

the paragraph) to indicate quoted material.

(5) By: soup

03 May 2002

cherokeeruby: It's not 'just a ticker change.' The ticker was changed as a symbolic representation of precisely what you state in the rest of your sentence: 'new beginning, a new company' in which the contributions of the old CPQ and old HWP are both recognized.

"Besides just a ticker change it could be an attempt to signal a new beginning, a new company, so to speak."

Taking a look at the new stock symbols, HPQ, we CPQ shareholders are not making as big a change as HWP shareholders--we only have to change the 'C' to 'H' whereas they only retain the 'H' and have to change 'WP' to 'PQ.'

I have been in communication with someone at HWP's IR off and on during the past 2 weeks [I think their IR is MUCH more responsive than our current CPQ IR]. Yesterday I mentioned to the fellow how much I liked the new stock symbols and he replied that it is going to take people a while to get used to them in his division.

On this point I totally disagree with you: "Would not surprise me a bit if in the not too distant future the name of the new company was changed to HPQ Corp"

Changing a corporate name is expensive and not a move any corporation takes lightly. Changing stocks symbols is a big move, too. Had there been any serious interest in changing the corporation's name, it would be done right now. "Hewlett-Packard" is a world recognized brand name. Changing it to "HPQ" would be a disaster in my mind. Now, changing it to ComPac might not be so bad ;> Lynn

Other studies of computer-mediated communication have found that it lies in the area between spoken and written language. Ko (1996) examined 28 linguistic features of computer-mediated communication, as compared to both written and spoken English. The results showed that computer-mediated language is closer to speech, although still sharing many of the char-

acteristics of written language. However, the language is in some aspects “more speech-like than actual speech” (Ko, 1996:20), e.g., in that it uses more wh-questions and second person pronouns than a sample of speech.

The messages examined for this study are all related to publicly traded companies. Different sites, such as Yahoo! (<http://messages.yahoo.com/>), Raging Bull (<http://ragingbull.lycos.com/>), Motley Fool (<http://boards.fool.com/>), Silicon Investor (<http://www.siliconinvestor.com/>), or Stockhouse (<http://www.stockhouse.com/>) offer boards, typically classified by company or sector. Messages found on these boards vary from requests for advice, “insider” tips, and press releases, to attacks on the company’s management or against other posters’ opinions. Many posters draw attention to a recently announced press release, or any other event that could affect the price of the stock. Messages are typically short, reflecting the celerity with which the stock is being traded (see Table 1 for average message length).

Part of the definition of a genre includes its purpose, reflected in structural characteristics. Some of the messages examined seem to have a dual purpose: they are often directed at individuals, in response to a previous post, but they will be widely read by all the contributors and “lurkers” on the board. Typically, messages contain both very personal addresses (dismissing or criticizing other posters’ opinions), but also a soap-box-like argumentative style (defending an opinion, bashing or pumping the stock). One example is presented in (6) below, where the poster makes it clear that he is addressing all the readers on the board (*I thought it good...to remind myself as well as others on this board...*).

(6) Date: 23-4-2001

Subject: TXN a player in the optical space

Author: gurdjieff2000

TI’s Breakthrough Technology Revolutionizes Optical Wireless Networking per yahoo and PRitem. I thought it good to diverge from thisatmosphere of doom and gloom to remind myself as well as others on this board, that TXN isa MAJOR player in the semiconductor industry,and these prices several years from now willbe but a dream. These drastic sell offs occurin technology every 5 years or so, and onething for sure, the dynamics that caused this stock price correction will reverse one day,and the industry will be back in favor. Youshould ask yourself this

question, where do you want to be at that time!!!

The dual purpose is a general characteristic of all bulletin board messages. In the following section, I argue that there are further possible classifications within the bulletin board descriptor, each type having a more specific purpose.

In terms of register, all messages studied have a similar configuration:

- Field: specialized discussion about stocks and company performance.
- Tenor: specialist-to-specialist (sometimes novice-to-specialist), informal, low affective involvement among participants.
- Mode: computer-mediated communication, written to be read, -visual, -aural, +rapid feedback.

The field description is clear and straightforward. In all the messages, the topics of discussion are stocks, stock performance, and company performance. In terms of tenor, the messages are among specialists, and sometimes from novice to specialist (requests for advice). Although the tenor is informal, there is low affective involvement among the participants, given that most likely they do not know each other. Finally, the mode is communication mediated by a computer, and all the communication is in the form of text written to be read. There is no visual or aural feedback, but the written feedback is usually prompt⁶. The only variation in the four genres studied lies in the tenor, which sometimes is not completely friendly, as is the case with attacks.

Register is not enough of a descriptor to help us characterize the messages. They all have similar register configurations. The different types of messages are distinguished because they have different purposes, and, according to genre theory, those different purposes should be reflected in different structural characteristics. The next section investigates those structural characteristics.

4 Structural Characterization of Four Bulletin Board Genres

Bulletin board messages have general structural characteristics in common. In this section, we will first of all examine those, and then we will study

four different types of messages, which can be differentiated based on further structural characteristics.

One first approach to structural description of all bulletin board posts is the familiar letter style: opening, body, closing. As mentioned in the previous section, electronic posts also usually contain a link to previous discourse, either via a reference or through quoting other posters' messages, in part or in full. They also often lack introductions and closings, in which they are more similar to turns in spoken language. One initial characterization is presented in Example (7). The three-part structure is still present, with an opening (*Hi Guus*), a closing (*Teddybull*), and a main body. The labels are inserted in bold type.

(7) By: teddybull06

Mar 2003, 10:48 AM EST

Opening

Hi Guus,

Body

I didn't know about Boss/WH-deal, but it does not surprise me completely as the press release of Cryp stated that WH would exclusively use CERTAIN casino games, so not ALL of them. This shows increasing margin pressure in the egaming business, which has been going on for the last 2 years. Only few of the parties will survive, probably Boss and Cryp (approx twice the size of Boss(?)) are both survivors. Cryp is still the market leader by far and jumped just in time into Poker it looks like. Earnings growth looks pale, but cash in hand is more than o.k.. Let's wait for some new licencees, big names can improve the growth potential. Analyst coverage diminished, last CC was short and quite... contarry indicator? If net margins would drop to 10%, Cryp delivers USD 3 to 4 mln earnings a year, P/E just above 10 for a debt free - cash overloaded company. Does not worry me at all.

Closing

Teddybull

In the body of the message, we can have a further breakdown into backward- and forward-looking material, with a link to what Guus, the interlocutor, has said before, and a contribution by the current author.

(8) Link to previous material

I didn't know about Boss/WH-deal,

New contribution

but it does not surprise me completely as the press release of Cryp stated that WH would exclusively use CERTAIN casino games, so not ALL of them. This shows increasing margin pressure in the egaming business, which has been going on for the last 2 years. Only few of the parties will survive, probably Boss and Cryp (approx twice the size of Boss(?)) are both survivors. Cryp is still the market leader by far and jumped just in time into Poker it looks like. Earnings growth looks pale, but cash in hand is more than o.k.. Let's wait for some new licencees, big names can improve the growth potential. Analyst coverage diminished, last CC was short and quite... contrary indicator? If net margins would drop to 10%, Cryp delivers USD 3 to 4 mln earnings a year, P/E just above 10 for a debt free - cash overloaded company. Does not worry me at all.

The new contribution part can, in turn, be divided according to rhetorical purposes. The author is making a point, and he or she⁷ uses evidence to prove it. For now, however, we will concentrate on the following division of the main body of a message:

- Link to previous material
- New contribution
- Link to following discourse

The first and last components can always be recursive, weaved into the new contribution. Herring (1996) proposes a similar characterization for messages to a listserv, which were divided into: link to an earlier message; expression of views; appeal to other participants. One problem with such a characterization is that it is overly general. It only identifies the messages as being dialogic. According to Bakhtin, all language is dialogic in some sense or another:

The dialogic orientation of discourse is a phenomenon that is, of course, a property of any discourse. On all its various routes toward

the object, in all its directions, the word encounters an alien word and cannot help encountering it in a living, tension-filled interaction. Only the mythical Adam, who approached a virginal and as yet verbally unqualified world with the first word, could really have escaped from start to finish this dialogic inter-orientation with the alien word that occurs in the object. (Bakhtin, 1981:279).

Given that link to previous and following discourse are not distinguishing structural characteristics, we need to find other ways of identifying different types of messages. I have selected four types of messages within the financial domain, based on their purpose (part of the definition of a genre), and then examined their structural characteristics. The four genres are:

- Request for advice or information
- Argumentation for and against an opinion
- Bashing and pumping the stock/company
- Attacks against other posters in the message board

These are all posted on bulletin boards related to the financial domain. The messages are collected from boards organized by company (e.g., Microsoft, Nortel, General Electric, etc.), or by type of investment (real estate, futures, utilities, etc.). This is an interesting domain, because emotions run high. It is very important to have the most up-to-date information, since many investors use that information to make decisions on buying or selling stock in the particular company.

The messages were collected randomly, and later divided into the four different genres (messages that seemed to belong to other genres were discarded). Two sets of collections took place: one during the first few months of 2001 (February-April), and another one in 2003 (March-June). Thirty messages were collected for each genre, to a total of 120 for the study. Table 1 provides the total number of messages and size of the corpus in words. Table 2 breaks down the average word length per genre. Word counts were performed automatically, and may not be completely accurate, since they assume a space between words.

Total number of messages	120
Total number of words	13,879
Average words per message (overall)	115.66

Table 1. Size of the corpus

Genre	Average word count
Requests for advice	63.59
Argumentation	212.52
Bashing and pumping	96.10
Attacks	98.76

Table 2. Average length of messages (in words)

4.1 Requests for advice or information

Example (9) is a typical request for advice. These are usually from posters who present themselves as novices, looking for advice from the experts⁸.

(9) Date: 4-22-2001 20:35:00

Subject: can't find industry rankings anywhere!!

Author: justcallmeangst_inthemorning

somehow I CANNOT find the rankings of individual stocks in each sector.....i know it is somewhere in yahoo finance.....can you help me?thanks,ANGST!!!!!!!

From a structural point of view, the messages contain two characteristic stages:

- Background information
- Question

The order is flexible (background + question or question + background). However, a question never occurs in between two instances of background information. Requests for advice are the easiest to distinguish of the four types of messages: they are usually short, and they always contain a question: a general one, such as *can you help me?* in Example (9), or a request for specific information, such as *btw, has anyone tried selling those fractional shares you received?*

Requests for advice are the types of messages that are closest to the original idea of bulletin boards (electronic or not): a participant has a need, for services or information, and he or she draws on the common available expertise to satisfy that need. Answers to the requests were not included in the study, but they are often prompt and helpful.

The genre of requests for advice is characterized by its purpose (demand information), and by characteristic stages (background information plus question). Its most relevant grammatical device is the use of questions. The structural formula is represented in (10).

(10) Background information ^ Request / Request ^ Background information

4.2 Argumentation

Messages that argue for a particular position or point of view, or that defend an individual or a company, are quite frequent in bulletin boards. They are usually meant to provoke and often to sway other readers' opinions on the particular company or product being discussed.

Example (11) illustrates some of the typical characteristics of an argumentation post. It concedes that another post had good points, but it provides further evidence for the author's point of view (that PLLL is a solid company, and that their investments are sound).

(11) Date: 4-22-2001 22:43:00

Subject: Wilcox Risks

Author: z24blackjet

Seismo,I thought your post was very thoughtful and balanced. However, I haven't sensed that anyone on this board has misrepresented the Wilcox prospects.We might get excited and talk about Scooter's party, but that's just normal enthusiasm for our investment in PLLL. I think we all recognize that the Wilcox results must be positive for PLLL to hit \$10 per share in the near future (6-12 months), unless NG prices go through the roof. I also think everyone realizes the Wilcox drilling carries higher risks than the wells currently in production.Sans Wilcox, we've just got another \$7 stock on our hands by October. Which, by the way, is good enough for me. PLLL is too smart (and experienced!) to bankrupt themselves on a solo Wilcox play. I like the way this company thinks and operates. - Z

From a structural point of view, argumentation messages usually have the

following stages. Elements in parentheses are optional.

- Link to previous discussion
- (Statement of author's point of view)
- Objection to previous argument
- Statement or restatement of author's point of view
- (Examples)
- (Disclaimer)

Grammatical characteristics include the use of concessive relations between major parts of the text:

- X might be true, but I think Y
- Positive comments about X. However, X is not the case; Y is. Further evidence for Y.
- Argumentation messages also show a high number of counterfactuals (If X were true, then Y would have happened).

Example (12) is an excellent illustration of an argumentation post, which is presented with numbers attached to each clause. The message starts with a concession to a previous poster (thanks... I agree with you that...). The next element, clause 3, is a statement of disagreement, clearly signalled by the discourse marker however. The statement is followed by evidence the author produces in order to further his case, this time signalled by for example. The message ends with a summary of the opinion (I think), in clause 17, followed by a disclaimer (anything can happen).

(12) Date: 4-22-2001 23:09:00

Subject: weekend update

Author: tzmw

1 cmn07, Thanks for sharing your insights.

2 I agree with you that the daily chart shows an overbought condition.

3 However, it can stay overbought for an extended periods of time.

4 For example, if you take a closer look at the weekly chart,

5 you will see it's started to show an overbought condition around 3000 and kept staying overbought until 5000.

6 Anyway, based on the weekly chart,

7 we now have had two white soldiers,
8 if I read it correctly.
9 I am hoping for another one,
10 then it would be very bullish.
11 Farmer_MO perhaps can help us to explain a little
more in detail.
12 We might get it
13 if the market sentiment is just right.
14 Who knows.
15 The STOCH is turning up,
16 so is the RSI(weekly chart).
17 I think its trend is more likely to remain intact,
18 at least for now.
19 Again, anything can happen in this kind of market.
20 I've learned to expect the unexpected.

The message can be analyzed in more detail if we use rhetorical relations (Mann and Thompson 1988), which are represented in Figure 1. Rhetorical relations are relations between parts of a text that capture the effect the writer wants to achieve on the reader. They are most often hypotactic (a main and a subordinate part), although some of them are paratactic (two main parts combined, similar to coordination). I will not provide here a detailed discussion of rhetorical relations; the names and the representation in the figure should be intuitive enough to capture the relations represented. Numbers in the figure are those of each unit in Example (12). The figure shows that the main organization of the text is one of Concession (*although I agree with part of what you said, I have a different opinion*). Then the post is organized around an opinion, expressed in 3, and supported with different pieces of evidence. The opinion is again summarized in 17.

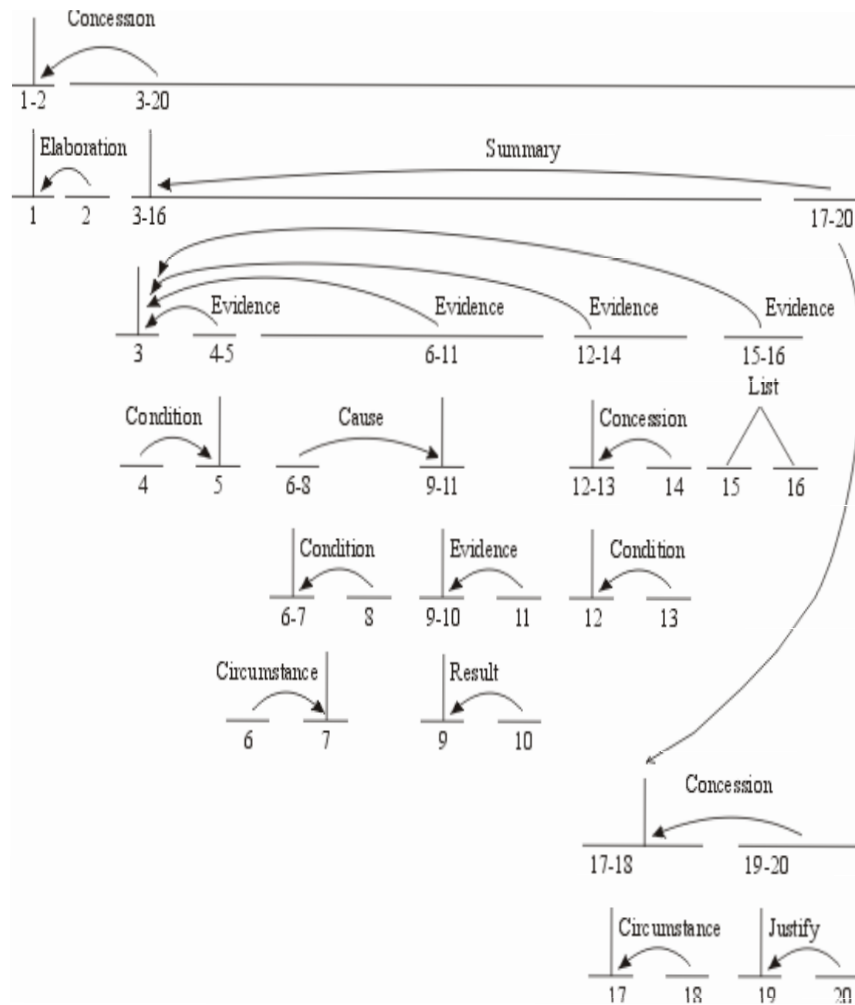


Figure 1. Rhetorical analysis of an argumentation post

A rhetorical analysis shows the highly structured nature of argumentation messages. Although not all are as complicated as the example shown in (12), each displays some of those elements (concession, conditions, statements of opinion). Thus, the main structural characteristic of argumentation posts is their organization around an idea or opinion. A general structural formula is provided in (13).

(13) Link ^ (Statement) ^ Objection ^ (Re)statement ^ (Examples)
^ (Disclaimer)

In terms of grammatical and lexical devices, the most important characteristic of argumentation messages is the presence of discourse markers and conjunctions that indicate the rhetorical relations present (*however, if, although, etc.*).⁹

4.3 Bashing and pumping

While most investors want the stock they invest in to increase in value (“buy low, sell high”), some investors borrow stock at a high price, sell it, and make a profit by paying off the borrowed stock at a low price (“shorting”). Thus, it is easy to find instances of both pumping and bashing on a company’s board. The terms ‘bash’ and ‘pump’ are commonly used, by other posters, to characterize this type of messages.

These messages are different from argumentation posts in that very little support for the opinions expressed is offered. Sometimes press releases or other sources of information are used, but often with a spin (or “de-spin”, exposing public relations jargon). They are also different in that opinions are exclusively positive or negative towards a company. Argumentation posts contain a wider range of opinions (on the market, on a particular business strategy, etc.). Bashing and pumping posts are often exclusively all negative or all positive towards a company.

In terms of structure, the messages contain either all-positive or all-negative opinions, sometimes accompanied by evaluation or justification on the opinion. The stages are:

- Negative opinion / positive opinion
- (Justification)

Example (14) is a typical pumping message. The author uses recent news (supposedly about Phillips entering into a partnership with other companies; it is unclear from the message) to declare Phillips a good company.

(14) Date: 4-22-2001 20:44:00
Subject: netster...re phillips news.....ot
Author: waytoomanyzs
this is great news. phillips motorola and att could there

be better partners ? phillips has been on the leading edge technologically ever since i can remember. if they see promise in adco i believe it adds more creditability to it.

Bashing is illustrated in Example (15). The author uses justification (poor customer service) to discredit a company. He or she also explicitly links the customer dissatisfaction to the owning of shares in that company, stating that he/she will sell the shares as soon as possible.

(15) Date: 4-23-2001 1:43:00

Subject: Bad Long distance Bad wireless it ne

Author: kensas

I agree... I've had nothing but problems with my cell service. They signed me up for one plan only to bill me at a much higher plan. Now they tell me the plan I signed up for doesn't exist. They promise to call me back and never did. Waited over 30 min for customer service to pick up! Thanks for the e-complaints site... Just sent them my story...I plan to sell all my WCOM on a decent uptick. I guess being a dissatisfied customer prevents me from being a satisfied investor...

Most messages are entirely positive or entirely negative. A few use negative statements about other companies in a similar sector to then conclude positively on a different company, or vice versa. In Example (16), the author uses a negative opinion by somebody else (that Apple will not launch new products in a long time) to prove the other poster wrong. The author also compares Apple's performance to that of other companies (Dell, Compaq, Gateway and Microsoft) to highlight Apple's achievements.

(16) Date: 4-23-2001 1:38:00

Subject: A look at the next 3 quarters

Author: dsheehy

1) No new designs. There has been no significant changes in the design form factor of the iMac since its introduction other than color/translucency. Time and effort on re-design seem to have been focused on individual projects like the iBook and then the Cube. Given the known (and speculated) product development cycle for those two

products, I doubt we'll see anything in the next 6 months. If new designs are coming, it won't be until March, 2001. Darn, that's hilarious. A perfect example of overly negative spin. Well, we got our new designs, or new OS, and we're very happy with them. Now what exactly have Dell, Compaq, Gateway and Microsoft done in this time? Oh, that's right - lay off employees. While we get a nice new boost from these wonderful products. The Ti Powerbook is the hottest thing in computing today. Too bad Plato didn't see that one coming - despite everyone here knowing about "Mercury" for around a year or so!

From the point of view of the lexico-grammar, bashing and pumping messages can be identified because they contain a large number of evaluation words and patterns (Hunston and Thompson, 2000). Although adjectives by themselves help identify a text as being positive or negative (Turney, 2002), the most reliable method is to find patterns of evaluation and Appraisal (Martin, 2000; White, 1998).

The structural formula for these messages is as follows:

(17) Negative/positive opinion ^ (Justification)

4.4 Attacks

The final type of message is the direct attack towards other members of the board. These are relatively common, and often seem to have as ultimate purpose to discredit other posters, and, by extension, their opinions.

Personal attacks are an unfortunate side effect of the medium. A certain amount of anonymity¹⁰ allows posters to express uncensored opinions that they might not express in face-to-face communication. This behaviour has been characterized as flaming. O'Sullivan and Flanagin (2003:84) define flames as "intentional (whether successful or unsuccessful) negative violations of (negotiated, evolving, and situated) interactional norms." According to O'Sullivan and Flanagin, in the past, flaming has been recognized only based on the content of the message: there are a number of content analysis studies of the frequency of flaming in on-line communication¹¹. However, content by itself, and the intentions of the author, are not enough, in their opinion. They argue that uptake is crucial in flaming behaviour, i.e., a third party observer might miss instances of

flaming, or see flaming in messages that are not considered as flames by the participants.

The messages classified as “attacks” here are all taken as such from the point of view of an outside observer. In most cases, it is very clear that they are attacks mainly meant to discredit somebody’s point of view. It is not clear, however, whether they qualify as flames, because it is possible that they do not violate the interactional norms of the boards. Expertise and credibility are at a premium on these boards, where much of the information exchanged is personal opinions or assessments. Thus, if an expert discredits a member of the board, the expert could be benefiting the whole community, by exposing somebody whose opinions should not be taken into account.

The crucial problem, then, is to distinguish attacks posted by experts, and thus worthy of attention, and attacks posted by authors with hidden agendas (pumping or bashing the stock through the bashing of other posters). In (18) we can see an example of an attack message.

(18) Date: 4-23-2001 13:07:00

Subject: Positive comments on JCP.....

Author: TailSpin77

Well wallstreet--you come across as a total dimwitted dipstick. JCP does not need morons like you in or around their stores. I would suggest security be called whenever morons like you get near the store. Go play a game of pocket pool with yourself..

The distinguishing structural characteristics of attacks are:

- Attack
- Elaboration
- (Attack/dismissal)

Those are not very different from bashing messages, except for their target: persons for attack messages versus companies or products for bashing messages. Attacks are short (98.76 words on average in the sample studied), but so are “bashing and pumping” messages (96.10 words on average). Attacks are best distinguished using lexical and grammatical properties:

- They use swear words

- They use “you” and frequent vocatives, especially at the beginning of the message
- They contain rhetorical questions (e.g., *who the hell do you think you are?*)

The Appraisal framework would also be useful in distinguishing these messages. Most frequent are values of Judgement to do with social esteem, and the capacity of the person under attack (Martin, 2000). The following structural formula describes the attacks examined in this study:

(19) Attack ^ Elaboration ^ (Further attack) ^ (Dismissal)

5 Discussion

Bulletin board messages are not homogeneous, modified versions of either conventional letters or e-mail messages. The register, in terms of field, tenor and mode, defines and shapes the characteristics of these messages. However, if we consider register configurations alone, we find that most bulletin board messages are similar in content, interpersonal style and medium. Register is not enough of a descriptor for the variability observed in the messages.

Genre, on the other hand, provides us with the tools to make distinctions between different types of messages. The definition of a genre includes its purpose, and since bulletin board messages have different purposes, they should be classified into different genres. The crucial question is one of how genre is reflected in the structure or content of the messages, and how we can use those differences to perform a genre-based classification.

The purposes of the messages are varied. First of all, in terms of audience, they are mixed: some of them are personal (especially attacks), but they will be read by an audience, the other members or readers in the same board. There is a continuum, from the very personal (attacks on other posters), to the general-audience target of argumentation messages. Of course, “general audience” refers to the audience within the board, possibly also other readers who encounter the board by chance.

Purpose defined the four major genres chosen for this study. From that starting point, we set out to describe the different structural characteristics. As can be seen from the descriptions in Section 4, the messages develop in different stages according to their purposes. In each genre, there

is a characteristic stage that helps define the genre: questions in requests for advice; statement of point of view in argumentation messages; statement of positive or negative opinion in pumping/bashing messages; and attacks in attack messages. Around that main stage we find other obligatory or optional stages that contribute to the overall purpose.

Structure is more evident in genres that have a long-standing written tradition. Argumentation messages are characterized by complex rhetorical relations, often signalled by discourse markers. We have also seen that arguments, bashing/pumping and attacks all share some similarities, in that they include the writer's personal opinion, and expressions that reflect the writer's appraisal (of an idea, a company, a person).

A practical application of genre-based analyses is in the area of topic detection. Topic detection and classification is a computational method of automatically identifying the main topic of a text, so that it can be retrieved or classified with other texts in the same topic area (Yang and Liu, 1999; Yang, 2000). A method for topic detection can benefit from some knowledge of the structure of the text. If we were to classify the messages included in this study, we could use the structural formulas outlined above.

Crucial to topic detection is the identification of the different stages. That identification needs to rely on lexical and grammatical patterns. For instance, the concessive relations we saw in argumentation messages are indicated by discourse markers (*however, although, nonetheless*). However, relying on structure exclusive will not provide sufficient identification. Not all the messages in each of the genres fit the structural descriptions above. Generic characterizations describe, but they do not determine the possible realizations within a genre:

[T]his relationship between context and text is theorized as probabilistic, not deterministic: an interactant setting out to achieve a particular cultural goal is most likely to initiate a text of a particular genre, and that text is most likely to unfold in a particular way—but the potential for alternatives is inherent in the dialogic relationship between language and context. (Eggin and Martin, 1997:236).

Therefore, a method to classify texts should use a combination of structural and lexical methods.

To conclude, this study has presented a new way of exam-

ining the characteristics of bulletin board messages, by classifying them in terms of genre. The study analyzed four different genres within financial boards: requests for advice, argumentation, pumping and bashing a company, and attacks on other posters. There are a number of other genres in the area of financial boards (for example, commercials), and possibly more in other bulletin boards in general.

Endnotes

¹ Jason Scott is currently producing a documentary on the history of the Bulletin Board System. Two web pages are associated with this project: a main page <<http://www.bbsdocumentary.com/>> and a timeline page <<http://timeline.textfiles.com/>>.

² The first browser, NCSA Mosaic, was released in 1993, according to Slashdot <<http://slashdot.org/article.pl?sid=03/03/14/1534218&tid=95>>.

³ Messages are reproduced as posted, including typos, spelling and grammatical mistakes. I have included dates, subject line and author information in the order and format displayed on the board.

⁴ The main structural difference with a letter is the date, which is automatically generated, i.e., not provided by the author.

⁵ Cited in Herring (1999).

⁶ The description of mode is similar to the one Eggins (1994:54) proposes for e-mail communication.

⁷ It is often difficult to establish the gender of contributors to message boards, although Thomson and Murachver (2001) suggest that there are some defining characteristics.

⁸ The issues around voice and self-presentation in on-line communication are too complex to discuss here. For a treatment of voice and power in Internet discourse, see Mitra and Watts (2002).

⁹ For a discussion of rhetorical relations and discourse markers, see “Taboada (2004, forthcoming).” A general description of discourse markers can be found in Schiffrin (1987).

¹⁰ Most posters use nicknames and/or false information in their personal profiles.

¹¹ Content analysis (Krippendorf, 1980; Neuendorf, 2001) is a method used in the social and behavioural sciences to analyze text based on the presence or absence of certain features, defined by the researcher. One example is the analysis of speeches during political campaigns (Hart, 2000).

References

- Bakhtin, Mikhail. "Discourse in the Novel." Trans. Caryl Emerson and Michael Holquist. *The Dialogic Imagination: Four Essays by M. M. Bakhtin*. Ed. Michael Holquist. Austin: University of Texas Press, 1981. 259-422.
- . *Speech Genres and Other Late Essays*. Trans. V. W. McGee. Austin: University of Texas Press, 1986.
- Biber, Douglas. *Variation across Speech and Writing*. Cambridge: Cambridge University Press, 1988.
- Bregman, Alvan, and Caroline Haythornthwaite. "Radicals of Presentation: Visibility, Relation and Co-Presence in Persistent Conversation." *New Media and Society* 5.1 (2003): 117-40.
- Christensen, Ward, and Randy Suess. *The Birth of the Bbs*. 1989. Available: <<http://www.chinet.com/html/cbbs.html>>. April 30 2003.
- Corp., Bulletin Board. *Why Create a Bulletin Board? 2003*. Available: <<http://www.bulletinboards.com/about.htm>>. April 30 2003.
- Crystal, David. *Language and the Internet*. Cambridge: Cambridge University Press, 2001.
- Eggins, Suzanne. *An Introduction to Systemic Functional Linguistics*. London: Pinter, 1994.
- Eggins, Suzanne, and James R. Martin. "Genres and Registers of Discourse." *Discourse as Structure and Process, Discourse Studies: A Multidisciplinary Introduction*. Ed. Teun A. van Dijk. Vol. 1. Thousand Oaks, CA: Sage, 1997. 230-56.
- Eggins, Suzanne, and Diana Slade. *Analysing Casual Conversation*. London: Cassell, 1997.
- Erickson, Thomas. "Persistent Conversation: An Introduction." *Journal of Computer-Mediated Communication* 4.4 (1999): <<http://www.ascusc.org/jcmc/>>.
- Friedman, Jerome. *Flexible Metric Nearest Neighbor Classification*: Department of Statistics. Stanford University, 1994.
- Halliday, Michael. "Context of Situation." *Language, Context and Text*. Eds. Michael Halliday and Ruqaiya Hasan. Oxford: Oxford University Press, 1989. 3-14.
- . *Language as Social Semiotic*. London: Arnold, 1978.
- Hart, Roderick P. *Campaign Talk: Why Elections Are Good for Us*. Princeton: Princeton University Press, 2000.
- Hasan, Ruqaiya. "The Nursery Tale as Genre." *Nottingham Linguistics Circular* 13 (1984): 71-102.
- Herring, Susan. "Computer-Mediated Discourse." *The Handbook of Discourse Analysis*. Eds. Deborah Schiffrin, Deborah Tannen and Heidi E. Hamilton. Malden, Mass.: Blackwell, 2001. 612-34.

- . "Interactional Coherence in Cmc." *Journal of Computer-Mediated Communication* 4.4 (1999): <<http://www.ascusc.org/jcmc>>.
- . "Two Variants of an Electronic Message Schema." *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives*. Ed. Susan Herring. Amsterdam: Benjamins, 1996. 81-106.
- . "Who's Got the Floor in Computer-Mediated Communication? Edelsky's Gender Patterns Revisited." *Computer-Mediated Conversation*. Ed. Susan Herring. Cresskill, NJ: Hampton Press, in press.
- Herring, Susan, and Carole Nix. "Is 'Serious Chat' an Oxymoron? Academic Vs. Social Uses of Internet Relay Chat." Presented at the American Association for Applied Linguistics. Orlando, FL, 1997.
- Hunston, Susan, and Geoff Thompson, eds. *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press, 2000.
- Ko, Kwang-Kye. "Structural Characteristics of Computer-Mediated Language: A Comparative Analysis of Interchange Discourse." *Electronic Journal of Communication* 6.3 (1996): <<http://www.cios.org/www/ejcmmain.htm>>.
- Krippendorff, Klaus. *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks, CA: Sage, 1980.
- Lewis, David. "Feature Selection and Feature Extraction for Text Categorization." *Proceedings of the DARPA Workshop on Speech and Natural Language, 1992*. 212-17.
- Mann, William C., and Sandra A. Thompson. "Rhetorical Structure Theory: Toward a Functional Theory of Text Organization." *Text* 8 (1988): 243-81.
- Martin, James R. "Beyond Exchange: Appraisal Systems in English." *Evaluation in Text*. Eds. Susan Hunston and Geoff Thompson. Oxford: Oxford University Press, 2000. 142-75.
- . "Language, Register and Genre." *Children Writing: Reader*. Ed. Frances Christie. Geelong, Victoria: Deakin University Press, 1984. 21-29.
- Mitchell, T. F. "The Language of Buying and Selling in Cyrenaica: A Situational Statement." *Hesperis* 26 (1957): 31-71. Reprinted in *Principles of Neofirthian Linguistics*, Longman, London. 167-200.
- Mitra, Ananda, and Eric Watts. "Theorizing Cyberspace: The Idea of Voice Applied to the Internet Discourse." *New Media and Society* 4.4 (2002): 479-98.
- Neuendorf, Kimberly. *The Content Analysis Guidebook*. Thousand Oaks, CA: Sage, 2001.
- O'Sullivan, Patrick B., and Andrew J. Flanagin. "Reconceptualizing 'Flaming' and Other Problematic Messages." *New Media and Society* 5.1 (2003): 69-94.
- Riloff, Ellen, and Jeffrey Lorenzen. "Extraction-Based Text Categorization:

- Generating Domain-Specific Role Relationships Automatically.” *Natural Language Information Retrieval*. Ed. Tomek Strzalkowski. Dordrecht: Kluwer, 1999. 167-96.
- Schiffrin, Deborah. *Discourse Markers*. Cambridge: Cambridge University Press, 1987.
- Stamatos, Efstathios, Nikos Fakotakis, and George Kokkinakis. “Automatic Text Categorization in Terms of Genre and Author.” *Computational Linguistics* 26.4 (2000): 471-95.
- Taboada, Maite. “Discourse Markers as Signals (or Not) of Rhetorical Relations.” *Journal of Pragmatics*. forthcoming.
- . *Building Coherence and Cohesion: Task-Oriented Dialogue in English and Spanish*. Amsterdam: Benjamins, 2004.
- Tanskanen, Sanna-Kaisa. “Avoiding Conflict in Computer-Mediated Discussions, or, Fear of Flaming.” *English in Zigs and Zags*, (Anglicana Turkuensia 23). Eds. R. Hiltunen, et al., 2001. 227-42.
- Thomson, Rob, and Tamar Murachver. “Predicting Gender from Electronic Discourse.” *British Journal of Social Psychology* 40 (2001): 193-208.
- Turney, Peter. “Thumbs up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews.” *Proceedings, 40th Meeting of the Association for Computational Linguistics, 2002*. 417-24.
- White, Peter. *Telling Media Tales: The News Story as Rhetoric*. Unpublished Ph.D. Dissertation. Sydney: University of Sydney, 1998.
- Yang, Yiming. “An Evaluation of Statistical Approaches to Text Classification.” *Journal of Information Retrieval* 1.1/2 (2000): 67-88.
- Yang, Yiming, and Xin Liu. “A Re-Examination of Text Categorization Methods.” *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '99)*, 1999. 42-49.
- Yates, JoAnne, and Wanda Orlikowski. “Genres of Organisational Communication: An Approach to Studying Communication and Media.” *The Academy of Management Review* 17.2 (1992): 299-326.
- Yates, Simeon. “Computer Mediated Communication: The Future of the Letter?” *Letter Writing as a Social Practice*. Eds. David Barton and Nigel Hall. Amsterdam: Benjamins, 2000. 233-51.