

THE DEGREE OF SPOKENNESS IN TWO MODES OF CMC

Veronika Kloučková

INTRODUCTION

COMPUTER-MEDIATED communication is popularly seen as written speech with people writing as if they were speaking. The language of the Internet was by some scholars identified as a distinct, third modality. But can we generalize? Does this apply to all forms of CMC? David Crystal and Noemi Baron came to the conclusion that with the arrival of the language of the Internet which “differs in fundamental respects from traditional speech and from writing” (Crystal 2005, 1), “the general public has become increasingly attentive to relationship between written and spoken language” (Baron 2009, 1).

The classification of CMC is a topical issue that many scholars are dealing with. Until the present time, a unified classification system which would embrace all modes of CMC has not yet been devised. However, the language of the Internet is usually categorized across two dimensions; first, the extent to which the communication proceeds synchronously and second, the nature of the interaction which includes three possibilities: one-to-one, one-to-many and group communication, also called multi-party communication or “polylogue” (Herring 2008, 15). With the advancement of communication technologies new types are emerging, the visual channel is becoming relevant as a third dimension.

There is little agreement among researchers on the question of real differences between spoken and written language, some advocate a strict division of written and spoken language as two distinct and not interchangeable norms (Urbanová 2003, 13), and others suggest that “there is no absolute difference between speech and writing” (Crystal 2001, 25). The approach adopted for the sake of this paper suggests a written/spoken continuum where the extremities contain features typical of speech or writing. The difference between speech and writing is

then an issue of conventional distinctions rather than an absolute difference, although it is unquestionable that the activities of speaking and writing differ substantially. The situation is further blurred by the fact that we are currently facing the phenomenon of “conversionalisation” (Urbanová 2003, 13) of written language. “The fact that users experience CMC in fundamentally similar ways to spoken conversation, despite CMC being produced and received by written means” (Herring 2011, 2) is not only typical of the language of the Internet but partly also of written communication in general.

METHODOLOGY

The aim is to examine specific features typical of spoken language in two modes of CMC and to situate these two modes on the written-spoken continuum according to these features. The asynchronous (not real-time) discussion forum and the synchronous (real-time) chatroom interaction that this study examines share two basic characteristics: they are both multi-party and text-based. They differ in the degree of synchronicity, or virtual co-presence.

Each of the modes is represented by a corpus. The first one, *synchronous chat corpus* was created on the basis of NPS Chat Corpus (Naval Postgraduate School, Monterey California), posted messages were gathered from various online chats in accordance with their terms of service. The second corpus, *asynchronous discussion corpus* was created on the basis of several online discussions, with various topics, a part of them belonging to asynchronous quick and the other part to asynchronous slow conversations. Both corpora count approximately 5,000 words and they are referred to as *corpus S* and *corpus A*. The data for spoken and written language was obtained from a publication by Geoffrey Leech: *Word Frequencies in Written and Spoken English: based on the British National Corpus* (2001).

Basically, there are two main points of view from which we can assess the difference between spoken and written language and to situate the CMC in relation to this dichotomy. As has been mentioned above, the process of producing either writing or speech differs substantially, the first set of criteria will then be called *process criteria*. Further, the features of the final prod-

uct, speech or writing, can be evaluated and these features will be called *product criteria*. The set of the process criteria was compiled according to several publications, namely Wallace Chafe's *Discourse, consciousness, and time: the flow and displacement of conscious experience in speaking and writing* (1994), David Crystal's *Language and the Internet* (2001) and Josef Vachek's *Two Chapters on Written English* (1959). The product criteria were identified on the basis of Simeon J. Yates's "Researching Internet Interaction: Sociolinguistics and Corpus Analysis" (2001), Douglas Biber's *University Language: A corpus based study of spoken and written registers* (2006), and Naomi S. Baron's papers "The myth of Impoverished Signal" (2009) and "Language of the Internet" (2003).

THE EXAMINED PHENOMENA

Let us now present the individual features and discuss the position of CMC, in our case represented by corpus S and corpus A. The process features will be discussed at first, and then they will be related to the product features of spoken language, written language, synchronous chat and asynchronous discussion. The process research criteria are summarized in Table 1.

The product criteria are easier to evaluate from the empirical point of view and each of them can be assigned to one of the process criteria. In other words, what we explain concerning the process (of writing, speech or online communication) results in a product (writing, speech, CMC) that can be analyzed empirically. We could assign one or more product criteria to each process criterion and analyze them; however, such an extensive analysis is beyond the scope of this paper for purely practical reasons. Therefore, three product criteria, that are relatively easy to evaluate, were selected (in Table 2).

It is important to understand that we cannot assess the properties of CMC in general. The varieties of CMC are multiple and they can differ in numerous aspects as we explained in the introduction of this paper. Although we frequently use the terms computer-mediated communication or language of the Internet, there is no singular coherent entity behind these terms. CMC or language of the Internet includes all communication that happens in the digital form. If we are to discuss the criteria

mentioned above, it is necessary to specify a particular variety or varieties under examination. The conclusions drawn from our research are then applicable exclusively to those varieties; they cannot be interpreted as general features of CMC. As indicated above, this paper focuses on the cases of synchronous chat and asynchronous discussion represented by corpus S and corpus A. Let us now proceed to the analysis of process and product criteria.

1. Time and Space

Speech is transient and time-bound whereas writing is permanent and space-bound. The asynchronous discussion forums are rather permanent. Unless the web site hosting the given discussion is deleted, it is accessible at any time and open to anyone. Chat is transient but it does not fade as fast as speech does because the usual setting of chat services allows the users to scroll up the screen and see either a part or the whole of the given conversation. But as soon as the user logs off, the previous conversation cannot be accessed anymore; with a new login an entirely new conversation begins.

2. Tempo

The tempo of speech can be considered as a baseline from which the tempo of writing and reading are evaluated. In speech, there is no time lag between the production and the reception, in writing, the time lag is always present. Reading is considerably faster than listening and it provides the reader with the option of scanning and skimming. The asynchronous discussion resembles writing in this respect. As far as the reception is concerned, the participants can read or just very quickly skim the preceding interaction. The time lags in production are present in both CMC modes, although it is appropriate to note that in synchronous chat, the interaction might get close to the speech tempo, especially in systems which allow the keystroke-by-keystroke transmission. The time lag between sending of the message and its reception is also dependent on the limitations of the available technology.

3. *Planning*

In a conversation, there is little time for elaborate thinking over, ideas and suggestions follow quickly one another, interruptions and overlaps are common. In writing, on the other hand, the lack of spontaneity is replaced by rethinking and working over, errors are eliminated because they are perceived as inadequacies. In this respect chat very much resembles spoken conversation (due to the tempo of the interaction) because errors and typos occur frequently and participants rarely re-think their messages. Serious flaws or errors are usually spotted once they appear on the screen and if needed, they are corrected by posting another message. Asynchronous discussion forum, on the other hand, resembles writing in that the contributions are usually carefully structured, errors and overlaps do not occur so often. The forums are relatively durable and they might be read by a wide audience; that is the reason why the contributors usually strive to make a good impression on their readers.

The first product criterion is the frequency of the definite article. The usage of the definite article is connected to pre-planning and structuring of a discourse. The definite article is used more frequently in written than in spoken language. *The* is the most frequent word in corpus A, in corpus S, it is the fifth. Chart 1 shows the results of the word counts effectuated in all four corpora. We can conclude that as far as the usage of the definite article, asynchronous discussion is situated between written and spoken language; and the definite article is by far the most scarcely used in synchronous chat.

4. *Prosody and Naturalness*

Speech is greatly enriched by a number of prosodic features; some scholars consider speech more natural than writing mainly because it is unconditionally learned before writing. Certainly, there are some words and constructions that more natural of spoken conversation, for instance colloquial expressions, nonsense words and contractions. Writing, in contrast, disposes of a unique set of graphical features, among them organization into pages and paragraphs, or punctuation. Multiple instances of subordination and elaborated syntactic patterns

are natural to writing. In chat interaction we face a paradoxical situation where attempts are made to compensate the spoken naturalness by abbreviations and acronyms and other graphical features that originate in written language and certainly are not natural or typical of speech. Hobbs suggests that “when we move from face-to-face conversations to dialogs over computer terminals, the communication is purely verbal. The work done nonverbally now has to be realized verbally” (1980, 65). These verbal and graphical aids are generally called the e-paralinguistic features. They are used to substitute for the extralinguistic cues of spoken conversation and sometimes also to effectuate time economy. The following examples show the most frequent graphic means to express emphasis, intensification or emotion (emoticons, abbreviations and acronyms, multiplied characters, capitalization, conventionalized interjections and other symbols):

<27> oh no :(
 <59> and ill bring da **weeeeeeeeeeeeeeeeeeeeeeeeee**
 <39> i desire.....
 <44> i dont like work **THAT** much
 <22> **grrr** i can't take it anymore
 <27> **ahhahahaha!!!!**
 <22> well i work with a bunch of **f*cktards**
 <73> *shakes head*¹

The asynchronous discussion has at its disposal all the graphic means accessible to normal off-line writing. The individual transmission units are often carefully structured into complex sentences and organized into paragraphs. On the other hand, we can find a number of e-paralinguistic features also in asynchronous discussions but their variety is lower and usage less frequent than in chat. The following examples were taken from corpus A:

Jaceylacey
 Ruben, I can't recall any other religions that call for death to non believers ☺
charlotte2113: Still smoking???

1. Corpus S, privacy masked (the usernames were replaced by numbers)

[...] Is there anyone else that is doing the same???

really_an_oriole_fan, New Jersey

We have elected a gifted **POLITICIAN**, full stop, Sorry.

emsd

oops sorry -double post!

leonkyle: Me too

[...] then let me know **hun** take care **xx**

wildfilly:

good luck all and TC

5. *Situatedness*

In writing, the language producer and receiver usually do not share the same time and place. Co-presence and interaction are the key conditions for a successful spoken conversation, together they account for situatedness. Written language is desituated in the sense that the external conditions during the production and the reception have little influence on the language itself. Spoken language often becomes vaguer because participants rely on the context. The situation for chat is comparable to spoken language, the participants of the interaction share the same time and even though they do not share the same place, they share the virtual environment, which works as a powerful unifying element.

The relevant product criterion is deixis. Deictic expressions are used in speech where they are defined by the situational context and due to the lack of the situational context, they are avoided in writing. Chart 2 shows the occurrence of deictic expressions. In chat as well as in asynchronous discussions deictic expressions are used very frequently. This is an interesting fact because the implication is that even though the language of the asynchronous corpus is mostly written-like and desituated, a strong feeling of shared online context allows for the implementation of temporal and spatial deixis.

6. *Function*

Spoken language is better suited for social/emotional function while written language is better suited for recording facts and communicating ideas (informative function). Spoken language can also communicate intellectual content but it renders the emotional content in an easier way. The function of chat is obviously social or emotive rather than informational. But we can encounter great differences in the case of professional or academic online synchronous debate. As far as asynchronous discussion is concerned, the emotive function is also present but the information exchange is usually vital. Again, the topic and the purpose of the discussion are decisive.

The usage of first and second person personal pronouns is typical of speech because it is tied to the emotive function. In written language, as it has closer to the informative function, we usually find less first and second person personal pronouns, except for letters. Chart 3 displaying the frequency of personal pronouns of the first and second person shows that the score for synchronous CMC is very high which confirms that chat interaction is a highly personalized and interactive mode of communication where the emotive function is essential. Asynchronous discussion is also situated significantly closer to speech, which can be due to its resemblance to letters, as was discussed above.

SPECIFIC CMC FEATURES

Until now we have examined synchronous chat and asynchronous discussion solely from the point of view of their similarities or differences with speech and writing. Such a view, however, would be rather simplistic. If we want the analysis to be complete we also need to discuss the properties of online communication that are not to be found in the off-line context. We have already started doing so by covering the e-paralinguistic features. We will now turn our attention to conventionalized acronyms and emoticons, the lack of simultaneous feedback, multiple conversations, disrupted turn adjacency and last but not least, we will briefly discuss the influence of gender on the language of the Internet.

CONVENTIONALIZED ACRONYMS AND EMOTICONS

Most of the e-paralinguistic features are not entirely new: capitalization, multiplication of characters and other means named above have been used in writing before, though in a modified and certainly more modest way. Conventionalized acronyms and emoticons, however, are an invention of the computer-mediated communication users which has never been used before. Their main function is the disambiguation of certain utterances for emoticons and time economies as well as a sign of group identity for conventionalized acronyms.

THE LACK OF SIMULTANEOUS FEEDBACK

The lack of simultaneous feedback is one of the most important features of online communication and one of the key differences between chat and face-to-face communication. The absence of a simultaneous feedback channel prevents the participants of the synchronous online communication from adapting their posts to the immediate response of the other parties. This inevitably leads to misunderstanding and tends to shift the communication in the direction of seemingly “abrupt, cold, distant, or antagonistic interaction” (Crystal 2011, 22).

DISRUPTED TURN ADJACENCY

In chat, mostly due to time lags, overlapping exchanges and disrupted turn adjacency are very frequent. Topical fragmentation over time is also typical of spoken multi-party conversation. However, in chat it is possible for the contributors to engage in multiple chains of conversation or to skip a part of conversation for external reasons such as leaving the computer. Also, the contributors are typically involved into multitasking: while chatting they are engaged in other online or offline activities. Contrary to our expectations, disruption in turn-taking does not seem to provoke misunderstanding. The participants are able to follow a number of different conversations on different topics and also to distinguish adjacent pairs which are not directly following one another “partly by remembering the linguistic context but also by using their knowledge of the situation” (Crystal 2011, 27).

Chart 4 displays the structure of an extract from a chat conversation. In the asynchronous discussion, the individual posts usually logically follow one another. If a discussion participant wishes to refer back to a more distant post, they usually resort to certain artificial links created either by means of addressing directly the author of the given post or simply by inserting the part of the post that they are responding to.

ONLINE LANGUAGE AND GENDER

In a recent study of instant messaging, Naomi Baron (2008) advocates the importance of the role that gender plays in the online linguistic behavior. According to her research, women tend to adopt the involved, or emotive, style while men incline to the usage of the informational discourse. Generally, female language tends to be more formal, so that in the online setting it is closer to the norms of standard written language. Men, on the other hand, are inclined to use speech style in the online environment. The only exception is the use of emoticons which largely prevails by women. Since emoticons are seen as accounting for the missing prosodic features, in this respect women are closer to the speech style than men. The conclusion of her study suggests that synchronous online communication, instant messaging in particular, is less speech-like than was assumed, and more importantly, gender plays a significant role. The aspect of gender was mentioned to complete the discussion of specific CMC features; however, it lies beyond the scope of this paper.

CONCLUSION

Table 3 summarizes the analysis underlying the present paper. The objective of this paper has been to situate two modes of computer-mediated discourse on the written-spoken continuum. David Crystal describes the chat situation as causing “the most radical linguistic innovation affecting several basic conventions of traditional spoken and written communication” (2001, 130). In most respects, the synchronous chat was closer to spoken conversation and asynchronous discussion was closer to written language (time and space aspect, tempo, planning, emo-

tional versus informational function). However, both CMC modes were closer to speech in that the participants are sharing a common, virtual context. Asynchronous chat is mostly written-like as far as the production is concerned but the analysis of the product revealed that in some respects it is situated between written and spoken language. The degree of spokenness in synchronous chat is substantially higher than in asynchronous discussion. Asynchronous discussions are more like writing than speech but in some respects they also resemble spoken language and most importantly, they share the use of the e-paralinguistic features. We can safely claim that chat exhibits a number of spoken language properties, however, in some respects it also shares written language properties and above all, it features a number of unique properties specific to the online setting, namely the use of conventionalized acronyms and emoticons, the lack of simultaneous feedback and the overlapping adjacency turns. We also noted the influence of gender on online linguistic behavior as one of the possible directions in future research.

Table 1: Process Criteria

Process criteria	Speech	CMC	Writing
1 time, space	transient, time-bound, dynamic	?	permanent, space-bound, transportable, static
2 tempo	quicker, no time lag	?	slower, time lags
3 planning	spontaneous, overlaps, interruptions	?	revision, working over, rethinking
4 prosody	pitch, melody, timbre, sentence stress	?	no but graphical features
5 naturalness	unmarked, natural, specific structures	?	has to be learned, specific structures
6 situatedness	co-presence, interaction, vaguer (shared context)	?	desituated, more precise (no shared context)
7 function	better suited for emotive function	?	better suited for informational function

Table 2: Product Criteria

Product criteria	Speech	CMC	Writing
1 the definite article (planning)	less frequent	?	more frequent
2 deictic expressions (situatedness)	present	?	avoided
3 personal pronouns, first and second person (function)	more frequent	?	less frequent (except for letters)

Table 3: Process and Product Criteria, Summary

Process and product criteria	Speech	Synchronous chat	Asynchronous discussion	Writing
1 time, space	transient, time-bound, dynamic	S	W	permanent, space-bound transportable, static
2 tempo	quicker, no time lag	W	W	slower, time lags
3 planning +definite article	spontaneous, overlaps, interruptions, definite article used less	S	W	revision, working over, rethinking, more definite articles
4 prosody	pitch, melody, timbre, sentence stress	e-paralinguistic features	to some extent e-paralinguistic features +graphical features	no, but graphical features
5 naturalness	unmarked, natural, specific structures	S	W	has to be learned, specific structures
6 situatedness +deictic expressions	co-presence, interaction, vaguer (shared context), deictic expressions	S	S	desituated, more precise, deictic expressions avoided
7 function +personal pronouns	better suited for emotive function, personal pronouns: more first and second person	S	W (+S)	better suited for informational function, less personal pronouns of first and second person

Chart 1: Frequency Counts of the Definite Article

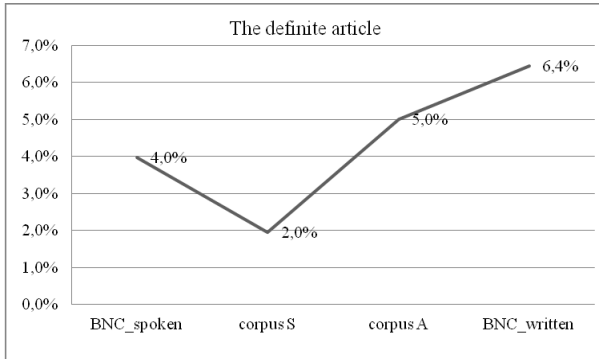


Chart 2: Frequency Counts of Deictic Expressions

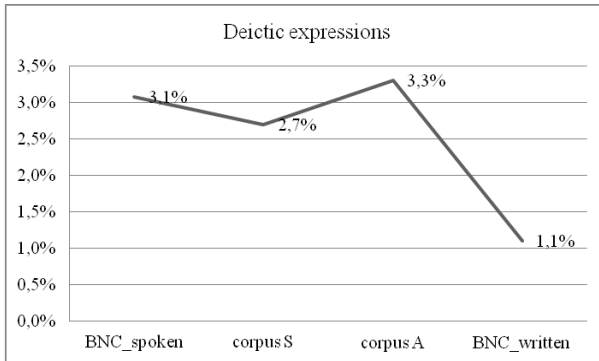


Chart 3: Frequency counts of personal pronouns, 1st and 2nd person

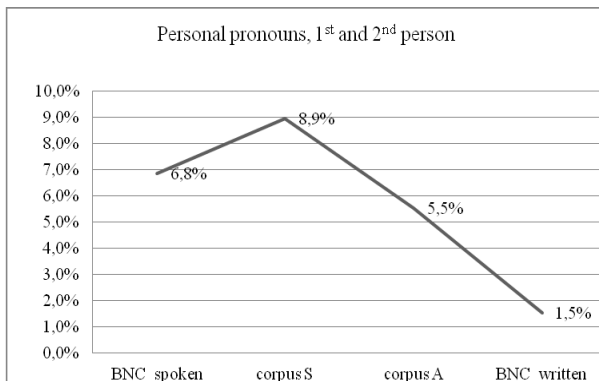
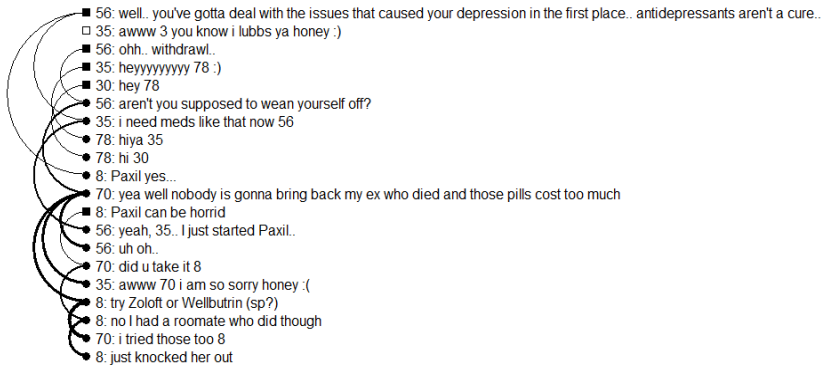


Chart 4: Overlapping adjacency sequences in chat conversation, taken from corpus S, visualization by ChatLine Software



CORPORA

Geoffrey Leech, Paul Rayson, and Andrew Wilson. 2001. *Word Frequencies in Written and Spoken English: based on the British National Corpus*. Longman, London. List 2.4:

Distinctiveness list: contrasting speech and writing (ordered by log likelihood). Accessed January 30, 2013.
<http://ucrel.lancs.ac.uk/bncfreq/flists.html>.

Corpus S

Eric N. Forsyth and Craig H. Martell. 2007. "Lexical and Discourse Analysis of Online Chat Dialog." *Proceedings of the First IEEE International Conference on Semantic Computing (ICSC 2007)*. 19–26. Accessed January 28, 2013.
<http://faculty.nps.edu/cmartell/NPSChat.htm>.

Corpus A

My Sun. "The only way to stop extreme Islamists..." Accessed May 28, 2009. http://www.thesun.co.uk/discussions/posts/list/9/The_only_way_to_stop_extreme_Islamists_8230~185194.page.

The New York Times. "Washington Memo: Despite Major Plans, Obama Taking Softer Stands." Accessed May 28, 2009.
http://community.nytimes.com/article/comments/2009/04/19/us_politics/19lobby.html.

- So feminine*. "Parenting: Still smoking???" Accessed January 20, 2010.
http://forum.sofeminine.co.uk/forum/maternite1/_f806_maternite1-Still-smoking.html.
- The Well*. "Deadsongs: The Wheel." Accessed January 20, 2010.
<http://www.well.com/conf/deadsongs.vue/topics/223/The-Wheel-page01.html>.
- The Well*. "Deadsongs: Me And My Uncle." Accessed January 20, 2010.
<http://www.well.com/conf/deadsongs.vue/topics/132/Me-And-My-Uncle-page01.html>.

BIBLIOGRAPHY

- Baron, Naomi S. 2003. "Language of the Internet." In *The Stanford Handbook for Language Engineers*. Edited by Ali Farghali, 59–127. Stanford: CSLI Publications. Accessed January 30, 2013.
<http://www.american.edu/cas/lfs/faculty-docs/upload/N-Baron-Language-Internet.pdf>.
- Baron, Naomi S. 2008. *Always On. Language in an Online and Mobile World*. OUP.
- Baron, Naomi S. 2009. "The Myth of Impoverished Signal: Dispelling the Spoken Language Fallacy for Emoticons in Online Communication." In *Emotion and ICTs*. Edited by Jane Vincent and Leopoldina Fortunati. London: PeterLang. Accessed January 30, 2013. http://www.american.edu/cas/lfs/faculty-docs/upload/Baron_Emoticons-1-the-Myth-of.pdf.
- Biber, Douglas. 2006. *University language: a corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
- Chafe, Wallace L. 1994. *Discourse, consciousness, and time: the flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.
- Crystal, David. 2001. *Language and the Internet*. Cambridge: CUP.
- Crystal, David. 2005. "The scope of Internet linguistics." *American Association for the Advancement of Science meeting*. Accessed March 6, 2010.
http://www.davidcrystal.com/David_Crystal/internet.htm.
- Crystal, David. 2011. *Internet Linguistic: A Student Guide*. Routledge.
- Herring, Susan C. 2007. "A faceted classification scheme for computer-mediated discourse." *Language@Internet*. Accessed February 10, 2010. <http://www.languageatinternet.de/articles/2007/761>.
- Herring, Susan C. 2008. "Language and the Internet." In *International Encyclopedia of Communication*. Edited by W. Donsbach. Blackwell Publishers.

- Herring, S. C. 2011. "Computer-mediated conversation Part II: Introduction and overview." *Language@Internet*, 8, article 2. Accessed February 10, 2010.
<http://www.languageatinternet.org/articles/2011/Herring>.
- Hobbs, J. R. 1980. "Interactive discourse: Influence of social context." In *The 18th annual meeting of the Association for Computational Linguistics and Parasession on Topics in Interactive Discourse: Proceedings of the conference*. Edited by N. K. Sondheimer. Philadelphia: University of Pennsylvania.
- Holmer, Torsten. 2008. "Discourse Structure Analysis of Chat Communication." *Language@Internet* 5(9). Accessed March 15, 2010.
www.languageatinternet.de/articles/2008/1633.
- Kloučková, Veronika. 2010. *Language of the Internet: One Language Variety?* Masaryk University, Brno.
- Scott, Mike. *WordSmith Tools: software for finding word patterns*. Oxford: OUP. Accessed January 22, 2010.
<http://www.lexically.net/downloads/version5/HTML/index.html>.
- Urbanová, Ludmila. 2003. *On Expressing Meaning in English Conversation: Semantic Indeterminacy*. Brno: Masarykova univerzita v Brně.
- Vachek, Josef. 1959. *Selected Writings: Two Chapters on Written English*. Praha: Academia.
- Yates, Simeon Y. 2001. "Researching Internet Interaction: Sociolinguistic and Corpus Analysis." In *Discourse as data: a guide for analysis*. Edited by Margaret Wetherell, Simeon Yates and Stephanie Taylor. London: SAGE.

ABSTRACT

The paper examines the place of computer-mediated communication on the written-spoken continuum. It is based on an analysis of two CMC modes: an asynchronous (not real-time) discussion forum and a synchronous (real-time) chatroom interaction. The criteria for the analysis were chosen according to the characteristic features of written and spoken language and both CMC modes are represented by two comparable corpora. The study reveals a substantial degree of conversational strategies in both CMC modes, although significantly prevailing in the synchronous one. The results of the present paper contribute to the justification of the application of conversation analysis tools to the study of CMC which, in fact, is a written discourse.

AUTHOR

Veronika Kloučková is a Ph.D. student in English Linguistics at the Department of English and American Studies, Masaryk University. Outside the department, she translates and interprets from French. She is interested in pragmatics, sociolinguistics and different aspects of the language of the Internet; her main research area is synchronous multiparty online communication from the pragmatic point of view.