

UNIVERZITA KARLOVA V PRAZE

Filozofická fakulta

Ústav anglického jazyka a didaktiky



DIPLOMOVÁ PRÁCE

Anna Mišutková

**Angličtina na Facebooku: ke specifickým rysům angličtiny v
internetové komunikaci**

Facebook English: on the Specific Features of English Netspeak

Praha 2014

Vedoucí práce: PhDr. Markéta Malá, PhD.

Ráda bych poděkovala PhDr. Markétě Malé, PhD. za odborné vedení, ochotné poskytnutí cenných rad a připomínek, vytrvalou podporu a za všechny čas, který nad mou prací strávila.

„Prohlašuji, že jsem diplomovou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.“

V Praze dne 15. ledna 2014

Souhlasím se zapůjčením diplomové práce ke studijním účelům. / I have no objections to this MA thesis being borrowed and used for study purposes.

ABSTRAKT

Cílem této diplomové práce je představit jazyk elektronické komunikace ('netspeak') jako jeden ze současných proudů vývoje jazyka. Práce vychází z hypotézy, že 'netspeak' představuje samostatnou multimodální jazykovou varietu, která sdílí některé rysy s neformální mluvenou angličtinou. Na vybraném vzorku textů anglicky mluvících studentů britských univerzit ze sociální sítě Facebook byla provedena detailní analýza tohoto jazyka. Kvantitativně-kvalitativní rozbor shromážděných dat a jejich srovnání s ústní formou standardní angličtiny pomocí mluvené složky Britského národního korpusu, nalezení společných i odlišných znaků a nových prvků (odchyly od standardního pravopisu, novotvary, atd.) potvrdily tuto hypotézu.

Klíčová slova: netspeak, CMC, Facebook, emotikony, osobní komunikace, neformální konverzace

ABSTRACT

The aim of this thesis is to describe the language of electronic communication ('netspeak') as one of the present trends of the development of English. The thesis is based on the hypothesis 'netspeak' represents an independent multimodal linguistic variety sharing some features with informal face-to-face conversation. For this purpose, the language of selected samples of the texts of English-speaking students of British universities obtained from the social network Facebook was studied. The qualitative and quantitative analysis of these data and their comparison with spoken form of standard English, namely with the spoken demographically sampled part of the British National Corpus, confirmed the hypothesis.

Key words: netspeak, CMC, Facebook, emoticons, face-to-face communication, informal conversation

LIST OF TABLES

Table 1: Summary of the features of speech and writing	18
Table 2: 25 most frequent words in the Facebook and the BNC corpora	56
Table 3: Top 25 key words of the Facebook corpora	59
Table 4: The most frequent emoticons in the Facebook corpora	61
Table 5: The most frequent adjectives in the Facebook and Spoken-BNC corpora	62
Table 6: 25 most frequent verb lemmata in the Facebook corpora and in the Spoken-BNC	63
Table 7: Selected action verbs (lemmas) in the Facebook corpora and Spoken-BNC	65
Table 8: Selected modal verbs in the Facebook corpora and Spoken-BNC	66

LIST OF FIGURES

Figure 1: IRC chat room conversation	21
Figure 2: Usenet message	23
Figure 3: Default box for adding a status update on Facebook	40
Figure 4: Sample 1 (Facebook personal wall comment)	41
Figure 5: Sample 2 (Facebook group wall comment)	48

LIST OF ABBREVIATIONS

AIM	America Online Instant Messenger
ASCII	American Standard Code for Information Interchange
AOL	America Online
BBS	Bulletin Board System
BCC	Blind Carbon Copy
BNC	British National Corpus
CMC	Computer-mediated Communication
COCA	Corpus of Contemporary American English
EMC	Electronically-mediated Communication
ICT	Information Communication Technologies
IM	Instant Messaging
IRC	Internet Relay Chat
MSN	Microsoft Messenger
SMS	Short Message Service
SNS	Social Networking Sites
TCU	Turn Construction Unit

TABLE OF CONTENTS

1 Introduction.....	9
2 Theoretical background	10
2.1 Computer-mediated communication	10
2.2 Types of CMC	13
2.2.1 E-Mail	14
2.2.2 Chat rooms.....	15
2.2.3 Electronic mailing lists	15
2.2.4 Internet forums.....	16
2.2.5 Instant Messaging (IM).....	17
2.2.6 Social networks.....	18
2.3 CMC vs. writing and speaking	18
2.3.1 CMC and speaking	20
2.3.2 CMC and writing.....	24
2.3.3 Linguistic features of CMC	25
2.3.3.1 Graphic and graphological level of CMC	26
2.3.3.1.1 Emoticons.....	26
2.3.3.2 Grammatical level of CMC.....	28
2.3.3.2.1 Economy of effort	29
2.3.3.4 Lexical level of CMC.....	30
2.3.3.4.1 Acronyms	30
2.3.3.5 Discourse level of CMC.....	31
2.3.4 CMC as a continuum	31
2.4 Social networks	33
2.4.1 Facebook.....	35
3 Practical part.....	38
3.1 Methodology.....	38
3.1.1 Communicative options of the platform	39
3.2 Linguistic analysis	41
3.2.1 Qualitative analysis of Sample 1: Personal wall comment	41

3.2.1.1 Graphetics and graphology	41
3.2.1.2 Morphology.....	43
3.2.1.3 Syntax.....	44
3.2.1.4 Lexicology	45
3.2.1.5 Discourse.....	45
3.2.2 Qualitative analysis of Sample 2: Group wall comment	47
3.2.2.1 Graphetics and graphology	48
3.2.2.2 Morphology.....	50
3.2.2.3 Syntax.....	51
3.2.2.4 Lexicology	53
3.2.2.5 Discourse.....	53
3.2.3 Quantitative analysis.....	55
3.2.3.1 Wordlists analysis.....	56
3.2.3.2 Graphetics and graphology	60
3.2.3.3 Morphology and syntax	62
3.2.3.4 Lexicology	67
3.2.3.5 Discourse.....	68
4 Conclusion	70
5 Résumé.....	73
6 References.....	77

1 INTRODUCTION

The aim of this thesis is to describe the language of electronic communication (netspeak) as one of the present trends of the development of English.

The theoretical part explains the terminology connected to Internet linguistics, such as ‘language variety’, ‘netspeak’ and ‘computer-mediated communication’ (CMC). The characteristic features of common CMC genres are compared to the written and spoken form on the linguistic levels of phonology, graphetics and graphology, morphology, syntax, lexicology and discourse. The comparison indicates that synchronous CMC genres incline to spoken communication while asynchronous ones resemble written form. The last chapter of the theoretical part focuses on the phenomenon of social networking represented by Facebook.

The practical part describes the methodology of the research and the hypothesis of the thesis. The hypothesis claims that netspeak represents an independent multimodal linguistic variety sharing some features with informal face-to-face conversation. For this purpose, the language of selected samples of the texts of English-speaking students of British universities obtained from the social network Facebook studied. The qualitative and quantitative analysis of these data and their comparison with spoken form of standard English, namely with the spoken demographically sampled part of the British National Corpus, confirmed the hypothesis.

2 THEORETICAL BACKGROUND

2.1 Computer-mediated communication

English constantly changes, develops and evolves. The change comes from the inside as well as from the outside of the linguistic world and its results are traceable after several years or decades. One of the extralinguistic factors which triggered the beginning of a new language variety are the technical inventions of the 20th century. New electronic media for long distance communication gradually appeared and besides the mass media such as radio and television, it was the invention of the modern computer and the Internet which opened a new path for human interaction. However, the journey from the first computer network ARPANET to social networking sites such as Facebook was long and linguistically variegated and the status quo of English on the Internet is the product of this complex process.

In the following chapter the terminology and the theoretical background linked to this emerging language variety will be treated in greater detail. In addition, the findings and conclusions of the hypotheses of several studies will enlighten the questions and issues arising in this field. The literature available is based on recent articles as well as time-proven works of the linguists operating in this young domain of science.

First and foremost, the term *English*, or more generally *Language on the Internet* should be attributed a definition. There are numerous interchangeable expressions denoting the same phenomenon, most of them are self-contained and do not need further explanation. David Crystal who is considered one of the fathers of Internet linguistics introduces a new coinage *netspeak* and explains that “the term ‘netspeak’ is an alternative to ‘Netlish’, ‘Weblish’, ‘Internet language’, ‘cyberspeak’, ‘electronic discourse’, ‘electronic language’, ‘interactive written discourse’, ‘computer-mediated communication’ (CMC), and other more cumbersome locutions” (Crystal 2001: 19). Other authors add the terms such as ‘Internet slang’, ‘Net lingo’ and ‘electronically-mediated communication’. It can be observed that the scope of Internet linguistics is rather broad and the units such as electronic discourse, interactive written discourse, variety and communication often overlap. Due to the vagueness and inconsistencies in their definitions their mutual relationship can be of synonymy as well as of hyponymy. More innovative

expressions such as ‘Netlish’ or ‘Cyberspeak’ give evidence that the discipline is in its beginnings and its terminology has not been conventionalised yet.

Computer-mediated communication (hereafter CMC) is a well-established term connected with the communicative activities among computer users. However, the arrival of new electronic gadgets, such as smartphones and tablets made this term inaccurate. These devices possess the ability to connect to the Internet and to fully profit of the same functions as the computers do. Naomi Baron in her book *Always on: language in an online and mobile world* (2008) comments on this issue:

“With the development of mobile devices such as the BlackBerry and mobile phones, which aren’t really computers, the term CMC became something of a stretch. Many researchers began speaking of information communication technologies (ICTs), alluding to the machines themselves (computers, personal digital assistants, mobile phones) rather than to the information they conveyed. What we now needed was an umbrella term for various types of language transmitted via the gamut of ICTs.” (Baron, 2008: 11-12).

She suggests using a broader term ‘electronically-mediated communication’ (EMC) covering all up-to-day electric power-driven devices, including the ones to come. However, due to the popularity of the term computer-mediated communication, CMC will be preferred in this paper.

What are then the characteristics of CMC, is it a valid homogenous language variety or a mere blend of heterogeneous features? The answer can be found provided the boundaries of other language forms are delimited. To be able to do so, a definition of a language variety will be needed. *The Oxford Companion to the English Language* (McArthur and McArthur, 1992: 1081) identifies two broad types of variety:

“(1) *user-related varieties*, associated with particular people and often places, . . . [and]
(2) *use-related varieties*, associated with function, such as *legal English* (the language of courts, contracts, etc.) and *literary English* (the typical usage of literary texts, conversations, etc.).”

The definition (1) is focused on a certain location or a group of people, in which case we speak about a ‘dialect’ while the definition (2) is function-centred and influenced by the subject matter. It is not easy to characterise CMC on the basis of these two definitions as it

is locationless, or rather spread all over the world and it serves various functions. Then a different approach is needed and in this case a pertinent variable is the platform of communication. The question how to convey one's ideas is answered by two traditional dichotomous varieties of language which are speaking and writing. Before comparing and contrasting them with netspeak, the parameters according to which they are recognised either as oral or as written medium will be listed.

Language varieties of written language can be characterised on the basis of the following five types of features (cf. e.g. Crystal, 2001: 8-9):

- A. “graphic features: the general presentation and organization of the written language, defined in terms of such factors as distinctive typography, page design, spacing, use of illustrations, and colour; for example, the variety of newspaper English would be chiefly identified at this level through the use of such notions as headlines, columns, and captions.
- B. orthographic (or graphological) features: the writing system of an individual language, defined in terms of such factors as distinctive use of the alphabet, capital letters, spelling, punctuation, and ways of expressing emphasis (italics, bold- face, etc.); for example, American and British English are distinguished by many spelling differences (e.g. *colour* vs. *color*), and advertising English allows spelling modifications that would be excluded from most other varieties (e.g. Beanz Meanz Heinz).
- C. grammatical features: the many possibilities of syntax and morphology, defined in terms of such factors as the distinctive use of sentence structure, word order, and word inflections; for example, religious English makes use of an unusual vocative construction (*O God, who knows . . .*) and allows a second-person singular set of pronouns (*thou, thee, thine*).
- D. lexical features: the vocabulary of a language, defined in terms of the set of words and idioms given distinctive use within a variety; for example, legal English employs such expressions as *heretofore*, *easement*, and *alleged*, as well as such phrases as *signed sealed and delivered* and Latin expressions such as *ex post facto*.
- E. discourse features: the structural organization of a text, defined in terms of such factors as coherence, relevance, paragraph structure, and the logical progression of ideas; for example, a journal paper within scientific English typically consists of a fixed sequence of sections including the abstract, introduction, methodology, results, discussion, and conclusion.”

As for the language varieties of spoken language, there are two more features which could be added, the phonetic and phonological ones. Although speech is becoming part of the Internet in the form of voice and video calls, it still has a minor function and is

restricted to special software such as VoIP programmes and Skype. That is the reason of excluding them from this thesis.

2.2 Types of CMC

If netspeak is an independent language variety functioning as a unified system, it is possible to describe it with the help of the features listed above. However, each distinct situation on the Internet carries some unique traits. Their emergence is conditioned by the technological innovations and whereas some continue to exist since the beginning of the CMC, others did not last for a long time, their popularity decreased and they were replaced by other forms. In the chronological order of their appearance on the Internet they can be arranged as:

- e-mail
- chat rooms
- electronic mailing lists
- Internet forums
- IM (Instant Messaging)
- social networks

The list is not exhaustive, however, all basic types of Internet communication are enumerated. The nature of some Internet forms (for example blogs and wikis) is more monological and less conversational and as this thesis deals with more interactive types of communication, they are not included. A separate chapter could be written about virtual worlds but because their connection to a specific game makes them monothematic and virtual-reality-focused, they are not mentioned either.

These forms can be further subdivided in two classes of synchronous vs. asynchronous media. The synchronous media, such as chat or IM require that all participants are online at the same time and the communication is thus a swift exchange of messages. Contrarily, the asynchronous media such as e-mail or a forum expect slower pace, from seconds to hours to days and most often the participants are not present at the

same time. In some cases which happen mainly on the Internet forums, a reply can take months or even years or does not have to be formulated at all.

The second usual classification of the types of CMC is into one-to-one and one-to-many communicants. The best examples to illustrate is IM vs. chat, nonetheless, there are many hybrid channels which can serve both types, such e-mail. More on this issue will be available in the following chapters.

2.2.1 E-Mail

'E-mail' is a remarkable term from the linguistic point of view thanks to the abundance of its spelling variations. Users can encounter the options, such as 'e-mail', 'email', 'mail', 'eMail', 'EMail' and 'E-mail' and the usage is governed by factors such as region (British vs. American English) and media type (online vs. printed). David Friedland clarifies that the two most common forms of spelling 'e-mail' and 'email' are "in standard use at this point, although e-mail retains a vast majority of usage in edited, published writing according to my research using the Corpus of Contemporary American English (COCA)".¹ The morphology of this term is not unified either, in some regions 'e-mail' is considered a non-count with undifferentiated quantity.

The word-formation process in the e-mail domain created the following words:

- spam defined as 'an unsolicited bulk message'²
- hoax meaning 'a deliberately fabricated falsehood made to masquerade as truth'³
- phishing which is explained as 'the act of attempting to acquire information such as usernames, passwords, and credit card details (and sometimes, indirectly, money) by masquerading as a trustworthy entity in an electronic communication'⁴

¹ Friedland, D. English Language & Usage Stack Exchange. <<http://english.stackexchange.com/questions/1925/email-or-e-mail>> Accessed on September, 30 2013.

² 'The Spamhaus Project - The Definition Of Spam'. Spamhaus. <spamhaus.org> Accessed on September, 30 2013.

³ MacDougall, Curtis D. (1941). *Hoaxes*. New York: Macmillan.

⁴ Ramzan, Z. (2010). 'Phishing attacks and countermeasures'. IN Stamp, M. & Stavroulakis, P. *Handbook of Information and Communication Security*. Springer.

- (B)CC which is an acronym meaning (blind) carbon copy; a tool for sending an overt or covert copy of an e-mail

2.2.2 Chat rooms

Chat rooms (or chatrooms) can be dated back as early as the 80s of the 20th century. It is possible to classify them into the group of synchronous media. There are various chat platforms including one of the oldest ones called IRC (Internet Relay Chat). It was created in 1988⁵ and it is based on one-to-many communication.

In order to begin an IRC conversation, the user has to have a client software or to access web-based environment. Subsequently he or she chooses a channel(s) of their choice which is usually topic-centred. After entering the channel, the user can see the list of online members and can start the communication with one or more participants. Chatting is public, i.e. visible to online users on the same channel, although there is also an option of private messaging. The written record is available on the screen for a limited period of time and as the conversation develops, older messages are deleted.

A distinct vocabulary can be found also in chat communication. The terms such as ‘flaming’ and ‘trolling’ acquired a new meaning and they refer to a negative type of behaviour among the communicants who interact in an intentionally hostile and insulting manner. Gathering of several flammers in one chat room can result in ‘flame wars’ and is usually difficult to resolve.

2.2.3 Electronic mailing lists

Another instance of one-to-many type of Internet communication are mailing lists. In contrast to chat, they belong to the asynchronous type of communication. The earliest known automated mailing list is ‘Listserv’ from 1986⁶ but nowadays other web tools are more popular. They share many features with classic e-mail as their interface is typically identical (the form of mailboxes) and no special additional software is needed. They are

⁵ ‘History of IRC’. IRC. <<http://www.irc.org/history.html>> Accessed on September 30, 2013.

⁶ ‘History of LISTSERV.’ LSOFT. <<http://www.lsoft.com/corporate/history-listserv.asp>> Accessed on September 30, 2013.

known also as distribution lists and its participants form a closed group focused on a selected topic.

Baron (2003: 15) summarises their function: “Today, listservs are commonly used by professional organizations, academic classrooms, or groups sharing common interests, enabling individual members to voice opinions or raise questions.”

The system provides some advantages in comparison with the manually-controlled group e-mail. First, it is a time-saving tool for the sender as a single address must be typed into the field ‘To’. Second, the privacy of its users is protected, non-members of the mailing list (and sometimes even the members) do not have access to the full list of recipients. It is a privilege of the administrator who manages the list. This can be seen as a disadvantage as well, as the email composers might not be sure how to adjust the form and the contents of their email if some of the recipients are unknown to them.

2.2.4 Internet forums

An Internet forum, or a message board, is an online discussion site where people can hold conversations in the form of message posting. It is traditionally considered to be an asynchronous medium. Historically, it is based on BBS (Bulletin Board system), a software which was necessary to take advantage of the communicative functions of the system.⁷ Today's forums⁸ do not require BBS, however its name is used as a synonym for an online forum. An alternative precursor of modern Internet forums was Usenet, with identical functions.

Forum users are usually acquainted with the terms such as:

- ‘post’, which is a posted message
- ‘(topic) thread’, which is a collection of posts about a single topic, visually separated from the rest of the threads
- ‘moderator’ who is a person who moderates a discussion.

⁷ ‘vBulletin Community Forum - FAQ: What is a bulletin board?’. VBULLETIN. <vBulletin.com> Accessed on September 30, 2013.

⁸ Morphologically seen, the plural of the Latin word *forum* is dominated by the form *forums* resulting in 1 600 million hits on Google while the original plural form *fora* has only 166 million hits. (author's research, Google September 2013)

The main difference between message boards and chats is the permanence of the posts which favours the first mode. Even the posts dated several years back are retrievable and if the thread is not locked by an admin or a moderator, it is possible to resume the conversation.

Compared to mailing lists, although forums do not ‘flood’ the mailboxes by new posts, the users of mailing lists have the advantage of being immediately informed. Some more sophisticated forums offer a special function of notification about new messages, so that the users do not have to log in on the forum to check it.

2.2.5 Instant Messaging (IM)

IM is a type of online channel offering real-time, i.e. synchronous communication. Unlike chatting, the conversation is one-to-one. Instant messaging takes place between two users who typically need a special client installed on their computer. ‘Instant messengers’ are the programmes whose interface is used to run the communication and their stratification is mainly regional.

Before the coming of Facebook in 2004 which also includes a type of a messenger and whose worldwide popularity reduced the number of traditional IM users, there was no dominant IM on the market. The best-known messengers are AIM (America Online Instant Messenger) and Yahoo! in the USA; and MSN (Microsoft Network) in Western Europe. The region of Central Europe was influenced by a programme called ICQ which later merged with AOL.

An inherent feature of any instant messenger is the contact list with an alternative name ‘buddy list’, which is the list of people added by the user. The contacts usually know each other outside the electronic world and the number of ‘buddies’ is relatively stable. These two facts differentiate IM from chat whose users do not know each other in real world and their number in chat rooms constantly changes. The number of chatting people is also much higher than the contact list of any instant messenger.

2.2.6 Social networks

Social networks can be described as an aggregate of all the services mentioned. Although the first social network was launched in 1997⁹, the real boom came almost 10 years later. Two best known social networking sites of the present era are Facebook and Twitter. More on their history, functions and linguistic features can be found in a separate chapter of this paper.

2.3 CMC vs. writing and speaking

Having briefly characterising these communicative channels, in order to understand the similarities and differences with other language varieties, the salient features of two language varieties, writing and speaking will be analysed. A lot of research has been done in this area and as Baron (2010) observes, “there is a considerable literature analyzing the relationship between spoken and written language (e.g., Baron, 2000; Biber, 1988; Chafe & Danielewicz, 1987; Chafe & Tannen, 1987; Crystal, 1995; Tannen, 1982a, 1982b)”. She adds that the difference between the two modalities is relatively predictable. For instance, the structure of written language is usually more complex than in speech whereas more contractions and more first and second person pronouns are typical of speech (Ibid.: 3). Crystal's Chapter 2 of his *Language at the Internet* (2001: 28-30) provides a brief and yet complete overview of the features of the two varieties and Table 1 is based on his findings.

Speech	Writing
<ul style="list-style-type: none">• time-bound• dynamic• transient• both participants are present• a particular addressee	<ul style="list-style-type: none">• space-bound• static• permanent• the writer is distant from the reader• the reader is unknown

⁹ 'The Brief History of Social Media'. University of North Carolina at Pembroke. <<http://www.uncp.edu/home/acurtis/NewMedia/SocialMedia/SocialMediaHistory.html>> Accessed on September 30, 2013.

<ul style="list-style-type: none"> • no time-lag between production and reception • spontaneity and speech prevent complex advance planning • pressure to think while talking promotes looser constructions, repetition, rephrasing and comment clauses. • Intonation and pause divide long utterances into manageable chunks but the sentence boundaries are often unclear 	<ul style="list-style-type: none"> • a time-lag between production and reception • there is anticipation of effects and problems with interpretations because of many recipients in diverse settings • repeated reading, close analysis, careful organization and compact expression, intricate sentence structure • the units of discourse are easy to identify through punctuation and layout
<ul style="list-style-type: none"> • presence of extralinguistic cues, i.e. facial expressions, gestures and immediate feedback • vague lexicon containing deictic expressions 	<ul style="list-style-type: none"> • no direct feedback and insufficient context • no deictic expressions as they could cause ambiguity
<ul style="list-style-type: none"> • features of informal speech, e.g. contracted forms, lengthy coordinate sentences, nonsense vocabulary, obscenity, slang 	<ul style="list-style-type: none"> • multiple subordination in one sentence, elaborate syntactic patterns, long sentences • certain items of vocabulary are never spoken (e.g. longer names of chemical compounds)
<ul style="list-style-type: none"> • it maintains social and phatic function • nuances thanks to prosody and non-verbal features 	<ul style="list-style-type: none"> • written records are suitable for memory and learning as they are easier to keep and can be read at a chosen speed
<ul style="list-style-type: none"> • errors cannot be withdrawn • interruptions, overlapping speech 	<ul style="list-style-type: none"> • errors can be eliminated • invisible interruptions
<ul style="list-style-type: none"> • uniqueness of prosodic features of intonation, loudness, tempo, rhythm, pause and tone 	<ul style="list-style-type: none"> • uniqueness of pages, lines, capitalization, spatial organization, punctuation • prosodic features are expressed through question marks, italics etc. • several genres cannot be read aloud (e.g. timetables, graphs, complex formulae)

Table 1: Summary of the features of speech and writing

The features of the two modalities presented in Table 1 complement each other and their boundaries are clear. It is important to note that this is not always so and in reality the difference between speaking and writing is much more subtle. However, this division provides a good model for the comparison with the language on the Internet which shares certain features with both forms.

2.3.1 CMC and speaking

Tannen (1982: 36) defines speaking as language produced in its spontaneous form, as opposed to written language. Its alternative name 'face-to-face' communication is often disputed due to its inaccuracy. For instance, Fortunati (2005: 53) proposes the term 'body-to-body' communication as "the communicative act involves not just the face but the entire body, its gestures and postures". Although body communication is a constitutive part of any conversation, 'face-to-face' remains the preferred term.

The word 'netspeak' could be reformulated as speaking on the Internet. Yet, there are substantial differences between netspeak and face-to-face communication. For example, the recipient's feedback is rather restricted and cannot be expressed simultaneously. The typical result is that the messages are sent in one piece and they are unidirectional. Crystal (2001: 33) labels it as an 'on-off' system. It is especially visible in a longer electronic communication that the sender does not have any feedback while composing the message and the recipient has to wait patiently. Marvin (1996: 6) notes that "[in] face-to-face conversations, a listener waits for an ending to a speaker's long statement, and stays alert for opportunities to speak, perhaps inwardly thinking, 'When will this person stop?' In typed conversations of the MOOs [a multiplayer real-time virtual world], a long statement requires a long wait on the part of the reader, during which the reader wonders, 'When will this person start?' " The lack of simultaneous feedback is connected with time delays or lags. While in face-to-face conversation, the silence between taking turns is usually in milliseconds and even a few-second pause is a signal of some kind (e.g. of an attitude), the pace of an Internet communication is much more unbalanced and unpredictable. An online response can take seconds to months, moreover its delay or even lack is ambiguous. It can be caused by a technical problem or a signal of expressing an attitude. The second reason usually carries a negative connotation.

Turntaking can be not only delayed but also disrupted. Susan Herring (1999) claims that CMC is "interactionally incoherent, and specifically, that processes of turn-taking and topic maintenance are subject to disruption and breakdown in computer-mediated environments". She develops the idea that turn adjacency is disrupted because of the system which posts messages according to the received order without considering the logic

of turn-taking. This claim was examined in her study where a number of messages interrupted a question and its reply. She illustrates it in IRC chat channel #PUNJAB (Figure 1, taken from Paolillo, 2011) where Jatt's question in line 9 is answered by Ashna as late as line 14 and there are four lines of irrelevant contents in between. To complicate the matter Jatt initiates the second conversation, another feature not possible to find in face-to-face interaction.

1 <ashna> hi jatt
2 *** Signoff: puja (EOF From client)
3 <Dave-G> kally i was only joking around
4 <Jatt> ashna: hello?
5 <kally> dave-g it was funny
6 <ashna> how are u jatt
7 <LUCKMAN> ssa all
8 <Dave-G> kally you da woman!
9 <Jatt> ashna: do we know each other?. I'm ok how are you
10 *** LUCKMAN has left channel #PUNJAB
11 *** LUCKMAN has joined channel #punjab
12 <kally> dave-g good stuff:)
13 <Jatt> kally: so hows school life, life in geneal, love life, family life?
14 <ashna> jatt no we don't know each other, i fine
15 <Jatt> ashna: where r ya from?

Figure 1: IRC chat room conversation

Herring (1999) concludes that “[v]iolations of sequential coherence are the rule rather than the exception in CMC. It follows from the high incidence of disrupted exchanges in CMC that turns that end up physically adjacent are often pragmatically irrelevant to one another”. Thus, from the pragmatic point of view the Gricean maxim of relevance is violated and it is not an uncommon phenomenon.

The incoherence of electronic communication is balanced by the means which are typical neither for speaking nor for writing. Herring (1999) labels them as ‘medium accommodating properties’ and they provide innovative methods of signalling attention and feedback. For instance, these are *nods*, *giggles*, *hehs*, *grins*, *smiles* and ? (questioning look). They are applicable for synchronous communication such as chat where it is expected that both participants are online.

Another distinct feature of CMC which can be seen as an advantage compared to an ordinary conversation is persistence. Herring (1999) claims that “even the least persistent synchronous interface is more persistent than spoken language, which disappears immediately once it is uttered.” In Table 1 Crystal characterises speaking as transient and even though in some synchronous media the conversation is very dynamic, the visual presence of the text makes CMC more comparable to writing which is permanent. However, electronic discourse is not absolutely permanent, which is especially valid for new hybrid media such as Facebook. More on this issue will be said in the chapter devoted to social networking media.

The next significant point in the analysis of spoken and written communication are their prosodic properties. Table 1 mentions that one of inherent properties of speaking is intonation and pause which serve as dividers of longer sequences, so that the hearer can process the emitted information more easily. A drawback of shorter chunks in speech is that its boundaries are frequently blurred. On the contrary, identification of sentential and higher units in writing is clearly marked through visual cues of punctuation and formal organisation of the text. How CMC deals with these properties, is related to the choice of the type of CMC.

For instance, chat and instant messaging which aim to supply a real conversation are technically adjusted to prompt the users to break their utterances into smaller pieces. It is done by means of the window size where the conversation takes place, i.e. the window is usually divided into a large multi-line space for message reading and a much smaller single-line window for message composing. The sender can also send the message very quickly, either by clicking on a nearby button ‘SEND’ or by pressing ‘ENTER’.

Apart from technical reasons, there are other variables. Baron (2010) examines the hypothesis claiming that IM turns are chunked. After analysing an IM Corpus of 22

students of American University in Washington she concludes that one of the pertinent factors is gender, revealing that “while IM conversations between male dyads tended to resemble spoken discourse according to this dimension, IM conversations between females bore more similarities to traditional written language” (Baron, 2010: 1).

Although intonation, loudness, stress, speed, rhythm, pause, tone and tempo as elements of prosody and paralanguage of speech are not to be found as such in CMC, they are retrievable in an alternative textual form. Herring (2001:5) speaks of unconventional orthography which supplies prosody, laughter and other non-language sounds, and uses for illustration the Usenet message (Figure 2, adapted from Herring, 2001: 5).

Al,
hahahahahahhahahahahahahahahahahahaa
sniff waaaaaaaaaaaaaaaaaaaaaaaaahhhhhh
I laughed, i cried. . . . that post was GREAT! :-)
Amusedly,
-Mirth-

Figure 2: Usenet message

On the one hand, the interjections *haha...haa* and *waa...hh* containing reduplicated letters together with asterisked sound **sniff** aim to show an expression of profound amusement. The lowercase spelling of the personal pronoun *i* and the capitalisation of the evaluative adjective *GREAT* emphasise informality and overall joyfulness of the message. The non-standard usage of punctuation marks, e.g. four dots after which comes the summary of author's evaluation, followed by the exclamation mark and an emoticon also resemble an oral utterance.

On the other hand, the structure of the post contains features of written communication following the established patterns of letter writing. Yet, the opening and closing expressions *Al* and *Amusedly* and the writer's identification *-Mirth-* are rather informal. *Al* is a familiar name for *Albert*, *Allison* or a similar name, *Amusedly* parodies the

formal leaving formula *Yours Sincerely* and the signature *-Mirth-* is enlivened by two dashes. Herring summarises these observations: “Significantly, this results in a linguistic variety that despite being produced by written-like means, frequently contains features of orality.” (Ibid.: 6)

Besides the properties of prosody and paralanguage, netspeak seeks to imitate facial expressions, gestures, body movement, posture and spacing. Haptics, kinesics and proxemics which typically accompany conversation are expressed via emoticons in the Internet language. More on emoticons will be said in a chapter devoted to CMC features.

All described phenomena show that CMC is partially able to substitute face-to-face communication. Herring (2010: 1) notes that although in present the claim that CMC can be considered a conversation is perfectly valid, in past years “the view of many scholars was that, at best, ‘conversation’ could be a metaphor for CMC, but not a literal description of it, since it was not produced orally or received auditorily like speech, and conversation was, by definition, spoken and heard”. Later in her article she adds an observation that: “In casual parlance, Internet users often refer to textual exchanges as conversations, using verbs such as ‘talked,’ ‘said,’ and ‘heard’ rather than ‘typed,’ ‘wrote,’ or ‘read’ to describe their CMC activities” (Ibid.: 1-2). Another argument favouring this claim are the views of some scholars who label CMC as ‘speech writ down’ (Horowitz and Samuels, 1987), ‘written speech’ (Maynor, 1994), and ‘visible conversation’ (Colomb and Simutis, 1996).

2.3.2 CMC and writing

CMC is a medium whose form bears a resemblance with writing. The communicants neither speak nor hear; they write and read texts composed of the characters available on a computer keyboard. Similarities can be found also in the speed of the input. No matter how fast the sender types the message, it will always be much slower than a spoken word and thus closer to writing. A speed-related property is preplanning which happens even with synchronous electronic communication. Users can always pause typing in order to think what to write next or in order to immediately correct their messages before sending them to the other party. This lack of spontaneity can be balanced for example by informality and warmth of the message achieved by various linguistic means. More on the lexicogrammatical practices will be said in the chapter about CMC.

A significant difference between netspeak and writing consists in spatial durability. Printed texts are permanent and no more editable whereas online material is dynamic with an easily changeable form or content. Per contra, as already discussed in the previous paragraphs, electronic messages carry more persistence, which is typical for written communication.

In writing, counterpart features of the prosody are for instance capitalisation, punctuation marks and other font-related means, such as font size and type, italics and bold. These tools are accessible and used even more extensively in electronic communication. The set of characters known as emoticons which is an exclusive feature of CMC can serve as an example. Crystal (2001: 42) explains a possible reason of its absence in traditional writing by stating that “[a] rapidly constructed Net message, lacking the usual courtesies, can easily appear abrupt or rude. A smiley defuses the situation.” He believes that emoticons are just a temporary matter until the users become accustomed to the medium and express themselves in a more carefully and explicitly organised manner.

Emoticons contain a property which unites written and electronic medium. Smileys and other graphical clusters of characters occurring in CMC, similarly to selected genres of writing such as graphs, timetables and complex formulae are alike in that they cannot be read aloud.

2.3.3 Linguistic features of CMC

CMC is a platform with a unique grammatical structure. Susan Herring (2012) points out that CMC as a purely text-based medium lacks phonology and the functions of sound are expressed via innovative features of typography and orthography. She refers to the grammar of electronic language as to ‘e-grammar’, emphasising that this term does not simply imply the existence of a single grammar for all CMC varieties (Herring, 2012: 1).

CMC as a heterogeneous system of multiple varieties cannot be analysed via traditional description of linguistic levels (e.g. cf. Crystal and Davy, 1992). However, some of its most prominent features present in most CMC subtypes will be mentioned.

2.3.3.1 Graphic and graphological level of CMC

Graphetics and graphology of CMC is a strong domain of e-grammar as it contains numerous innovative features. The layout of each subgenre of CMC is related on the medium and software used and they are so distinct and variegated that only some general characteristics can be provided. The comparison with the written form reveals that traditional structuring of the text to chapters and numbered pages is substituted by hyperlinked expressions and icons which immediately take the user into required part of the communication. Majority of CMC genres marked by the presence of nontextual content, such as pictures, video and audio clips, flash games etc.

Focusing on orthography, phenomena such as reduplicated letters, nonstandard capitalisation or lower case and asterisked expressions were already described in relation to speaking since they frequently imitate prosodic elements of face-to-face conversation. Herring (2012: 2) mentions other keyboard symbols used in electronic communication, for instance repeated punctuation (!!!) or substitution of numbers and letters for words or parts of words (e.g. 4 'for,' 2day 'today,' ur gr8 'you're great'). The shortening is probably influenced by the language of text messaging which is shaped by the limited number of characters. Although these innovations result from the playfulness and creativity, Kapidzic's recent research (2010) shows that only a small number of nonstandard spellings, such as *wanna* for 'want to' and *msg* for a 'message', have been conventionalised. The described spelling changes influence also the area of word formation.

2.3.3.1.1 Emoticons

Emoticons are another borderline area. Their scope reaches orthography, syntax, semantics, pragmatics and other linguistic disciplines. Because of the complexity of their functions and their potential ambiguity, a separate subchapter is devoted to the phenomenon.

Dresner and Herring (2010) define emoticons as non-verbal indicators of emotion and provide an overview of different definitions used throughout the CMC literature mentioning Rezabek and Cochenour (1998: 201) and later Walther and D'Addario (2001) who see emoticons as "visual cues formed from ordinary typographical symbols that when

read sideways represent feeling or emotions.” Subsequently, Wolf (2000: 828) cites the Hackers’ Dictionary definition of an emoticon as “an ASCII glyph used to indicate an emotional state”. Internet linguists, such as Crystal (2001: 36), describe emoticons as ‘combinations of keyboard characters designed to show an emotional facial expression’, and Baron (2000:242) refers to them as ‘emotion markers’.

The word ‘emoticon’ was created by the process of blending two words ‘emotion’ and ‘icon’ and its first appearance is dated to the year 1982 and attributed to Scott E. Fahlman. Interestingly enough, the original meaning of the emoticon :-) concerned only the act of joking and the emoticon :-(marked non-jokes. Since then the smiling emoticon has gained a more general meaning, marking not only jokes, but also a friendly and positive tone of the message. Moreover, the frowning smiley completely lost its original interpretation and is used as a means of expressing sad and negative feelings.

The form of emoticons is culturally dependent. Baron (2009b: 7) makes a comparison between American emoticons which “are read sideways and emphasize the mouth” and Japanese symbols called *kaomoji* which “are read horizontally and focus on the eyes. For example, the closest equivalent to the American smiley, that is :-), is the kaomoji ^-^-^”.

In view of many authors, the primary function of emoticons is to convey non-linguistic information which in face-to-face communication exists in the form of facial and body expressions. However, Dresner and Herring (2010: 60) argue that the primary function of many emoticons is to convey “pragmatic meaning, and thus this function needs to be understood in linguistic, rather than extra-linguistic, terms”.

Their framework is provided by the theory of speech acts (Austin, 1962; Searle, 1969, 1979) and their hypothesis claims that “in many cases emoticons are used not as signs of emotion, but rather as indications of the illocutionary force of the textual utterances that they accompany. As such, they help convey the speech act performed through the production of the utterance. These uses of emoticons do not contribute to the propositional content (the locution) of the language used, but neither are they just an extra-linguistic communication channel indicating emotion.” (Dresner and Herring, 2010: 62).

The hypothesis is illustrated by the use of the winking smiley ;-)) which frequently indicates that the writer is joking about the content of his or her message. Joking itself is

not an emotion as it can be accompanied by various feelings. “Rather, joking is a type of illocutionary force, something that we do by what we say” (Ibid.: 63).

Besides this one, Dresner and Herring (Ibid.: 62) introduce two other basic functions of emoticons and all three are listed below:

- 1) emotion indicators which are mapped directly onto facial expression
- 2) indicators of non-emotional meanings, mapped conventionally onto facial expressions
- 3) illocutionary force indicators that do not map conventionally onto a facial expression.

Baron (2009b: 5) addresses another issue of emoticons. She observes that using smileys in CMC has a contradictory effect. On the one hand, they help to clarify the meaning which is impoverished by the absence of visual and auditory cues. On the other hand, since emoticons came to acquire multiple meanings, they are more prone to be misinterpreted. This incongruity causes that no definite conclusion about their impact has been formulated yet.

Emoticons are not the only representation of physical actions. Herring (2001: 11) states that they can be visualised also textually, by typing literal expressions such as <grin> and *yawn* which “may serve as contextualization cues for a playful or relaxed discourse frame”. These expressions are typical for the genres of chat and instant messaging where informality plays an important role. Later Herring (2012: 5) categorises them as performative utterances as they carry out an action by being typed in a specially bracketed form * * or <>.

2.3.3.2 Grammatical level of CMC

The morphosyntax of netspeak deviates from standard syntax in various degrees. Asynchronous media, such as e-mails and discussion forums, which tend to contain more lengthy utterances are in their structure very similar to offline writing genres. Contrarily, synchronous media, such as instant messaging and chat can often be described as ‘telegraphic and fragmented’ (Herring, 2012) due to their brevity and informality. One of

convincing examples is the issue with utterance breaking (described in the section about speaking).

There are several reasons why reduction and elision happen. First, increased typing speed and thus shorter utterances in synchronous CMC imitate face-to-face conversations. The second reason is technical as users of selected CMC channels are allowed to type only a limited number of characters (resembling text messaging where one SMS can contain 160 characters at most). Third, the creativity and playfulness of this new platform enables users to be innovative and expressive. Last but not least, the economy of effort which is valid across several linguistic levels is a vital factor.

2.3.3.2.1 Economy of effort

This phenomenon is analysed from different angles by numerous authors. For instance, Herring (2001: 5) explains that the features of the non-standard syntax are rarely “errors caused by inattention or lack of knowledge of the standard language forms”. Rather, the majority of them are perceived as deliberate choices of users, functioning as economisers of typing effort.

Cho (2010: 3) summarises the literature available for the concept of linguistic economy and mentions Ferrara, Brunner and Whittemore (1991) and Murray (1990) who deal with synchronous media. They identify phenomena such as ellipsis by omission of pronouns and determiners and clipping. Additionally, the common use of abbreviations, ellipsis and orthographic reduction (e.g., *bb ppls* for *bye bye peoples*) in Internet Relay Chat (IRC) is described by Werry (1996).

Cho also points to the lack of similar research (with an of Frehner's work, 2008) for asynchronous media where responses are temporally more flexible and thus with less pressure on the composer. In his paper (2010: 14-16) he provides several new insights in this electronic genre. He notices a common habit in writing e-mails between colleagues and business partners, which is the omission of both greeting and leave-taking formulas. In the cases where the formulas were present, the degree of formality is on a low level, containing nicknames and expressions such as *hi* and *see you*. In contrast, the majority of

memoranda, also sent electronically, contain initial and final formulas and Cho concludes that their frequent presence is conditioned by their phatic function.

Another variable contributing to the popularity of linguistic economy is, according to Cho, “an increasing amount of information accessible to readers” (Ibid.: 20). This fact triggers an expectation that users will read more and consequently produce information for others more rapidly. Crystal shows a second, socially-based motivation and claims that: “The chief use of slang is to show that you’re part of the gang!” (Crystal 1997: 53). Hence, by using modern features of netspeak, the user is recognised as a member of an Internet group.

2.3.3.4 Lexical level of CMC

Netspeak shows its vivacity also in the field of word formation. Herring (2012) remarks the emergence of several productive prefixes, such as *e-*, *cyber-* and *hyper-*, semantic shift, e.g. *flame* ‘unleash invective on a computer network,’ from *flame* ‘to act conspicuously homosexual’; *spam* ‘Internet junk mail,’ from *spam* ‘a type of canned meat’ and conversion, for example *text* and *spam* as verbs.

2.3.3.4.1 Acronyms

Another prominent feature of CMC are acronyms. Neil Randall (2002) realised a survey concerning acronyms as an innovative feature of the language on the Internet. He created three focus age groups (<20, 20-34 and 35+) and analysed their knowledge of several common Internet-based acronyms. The results were following: 86% of the users under 20 years of age are familiar with the meaning of the abbreviation LOL (laughing out loud), 71% G2G (got to go), and 62% TTYL (talk to you later). In the second group of the 20-34 years old, only 60% know LOL and less than 20% G2G or TTYL. Among the 35+ group 28% know the meaning of LOL. He concluded that teenagers have almost intuitive knowledge of the meanings and functions of Internet language (Randall 2002: 5).

Moreover, Randall opens a question of classification of these expressions as acronyms. In his view TTYL is an example of an initialism, rather than of an acronym. He

explains that “A true acronym [...] not only joins the first letter of a phrase, but in doing so creates a pronounceable word that eventually becomes a recognized word in the language” (Ibid.: 27). In this very rigid definition, only words such as *laser*, *scuba* or *sonar* are a good fit, however, the term ‘acronym’ is widely used to refer to any abbreviation formed from initial letters.

Meaning change is noticeable also with acronyms, a phenomenon analysed by Baron (2009b). In her focus group composed of American University undergraduate students one of the most popular acronyms in instant messaging is LOL (meaning ‘laugh(ing) out loud’). She points out that although laughter is usually linked to positive emotions and humour, in CMC conversation it acts as “a phatic marker, meaning the equivalent of OK, Cool, or Yeah” (Baron 2009b: 5).

2.3.3.5 Discourse level of CMC

Electronic discourse is not an easy genre to analyse because of its great variability in length and hybridity of styles. As this emerging type of communication contains many oral-like forms of shorter size, their features cannot be generalised. Yet, some discourse patterns, such as coherence, relevance and logical progression of ideas can be observed. The specific findings will be analysed and compared against features of conversation in the practical part of this thesis.

2.3.4 CMC as a continuum

The comparison of CMC with speaking and writing and the subsequent characterisation of its inherent features in the previous chapter indicates the current position of netspeak in linguistics. However, it is impossible to define it as a single language variety due to the subgenres it contains. Kalman and Rafaeli (2007) focus on the traditional classification of CMC media into synchronous and asynchronous and their implications on the perception of CMC genres. They show that the dual division between communicative platforms, such as chat and instant messaging vs. email and discussion board is no longer valid and instead of a dichotomy, there is a continuum in synchronicity. Their second note concerns the factors exerting influence over the level of synchronicity.

They emphasise that both the parameters of the medium itself and the decisions taken by users shape the final degree of synchronicity.

The same authors speak about four current trends whose confluence is gradually effacing the difference between the two types of media, viz. 'digitization', 'media convergence', 'always-on' and 'portability'. The first one, digitalization is technologically an easy process when more and more written communication is digitised. The convergence of media refers to the fact that the differences between the media are blurred, e.g. emails can be read on mobile phones or a text message sent via computer. The third and fourth trend are interconnected and observe the tendency to be permanently online using portable devices. According to Kalman and Rafaeli, all of these indicate that temporality of the online communication is modulated by the decision and preference of the user.

A similar research was led by Baron (2009a). The author analysed the linguistic characteristics of university student use of instant messaging and some of her findings were at odds with general views of the medium. One of them concerned the question of synchronicity and Baron's conclusion was that instant messaging often carries features of asynchronous media and email of synchronous media. In social practice, the dichotomy of synchronicity is cancelled and its dividing line is blurred. This is done by means of multitasking, involving either other computer-based activities, such as multiple IM conversations, web browsing or games or miscellaneous activities, such as "chatting with a friend on the phone, listening to music, or munching on potato chips" (Baron 2009a: 6).

The factors influencing synchronicity of the used media were indicated by students participating in Baron's research and could be encapsulated into three items: the quality of 'gossip' in the conversation, the seriousness of the conversation and individual communications habits (minority of students considered it rude to hold multiple simultaneous conversations) (Ibid.: 12). The summary of the research confirmed users' control over keeping their conversations alive or suspended, which concurs with Kalman and Rafaeli's conclusions.

The question of the dichotomy of synchronicity of CMC is closely linked to the relationship of CMC towards speaking and writing. Baron (2010) lists authors researching this triple relationship since 1991, e.g. Ferrara, Brunner & Whittemore (1991), Maynor

(1994), Yates (1996), Collot & Belmore, (1996), Baron (1998, 2000 and 2003), Crystal (2001), Danet (2001), Hård af Segerstad (2002) and Herring (2002, 2003).

Baron's (1998) own conclusions towards the end of the 1990s was that CMC was a mixed modality, in some features similar to speech, in others to writing. Her current views (2010: 5) are that CMC as a collection of several modes has a resemblance with face-to-face communication, due to its conversation-like features as well as social functions it fulfils. She specifies the role of CMC: "Indeed, textual CMC has become a prime site for conversation, supplementing F2F and telephone for personal, professional, and commercial interactions." Herring (2010: 6) agrees that a shift from old to new conversational media is apparent, however, she does not think that face-to-face conversation is in danger of becoming extinct.

In contrast is Crystal's perspective, claiming that "netspeak has far more properties linking it to writing than to speech ... netspeak is better seen as a written language which has been pulled some way in the direction of speech than as spoken language which has been written down" (2001: 47). He concludes that netspeak is not just an aggregate of the two modes, it is a genuine third medium.

Baron (2003) adds to this an alternative definition of CMC as "a kind of linguistic centaur, incorporating features from both traditional writing and face-to-face discourse but ending up being more than a simple amalgam of the two". She also notices that "a number of distinctive linguistic conventions characterizing many people's use of language on the Internet are beginning to seep into traditional spoken and written language" (Baron 2003: 23).

Finally, Perez-Sabater (2008: 74) summarises that CMC is a heterogeneous modality containing multiple genres and subgenres "that carry over distinctive linguistic features of traditional off-line genres". Thus, it seems that a perception of CMC as a distinct medium is acceptable for majority of linguists.

2.4 Social networks

The phenomenon of social networking sites (hereafter SNSs) is a matter of the 21st century. However, its history dates back as early as to 1997. According to Boyd and Ellison

(2007: 217), the first social network *SixDegrees.com* was launched in 1997. In the period between 1997 and 2001, the sites such as *AsianAvenue*, *BlackPlanet* and *MiGente* started to offer new social tools, for instance profiles containing personal and professional information. Business-focused online networks, such as *Ryze.com* began their activity in 2001. In 2002 Ryze launched *Friendster* as its social counterpart and its main rival was an online dating site *Match.com*. Paradoxically, the increase in popularity of Friendster caused its fall due to technical issues. A new SNS called *MySpace* profited from the failure of Friendster and soon began to expand. Unlike the previous SNS which were profile or profession-centric, MySpace was culturally-centric, with the focus on music. Three discernible social groups formed the base of this site, musicians and artists, teenagers and ‘post-college urban social crowd’. Its gradual decline was inversely related to the success of Facebook which came in 2004.

According to the Wikipedia statistics (September 2013)¹⁰ Facebook has over 1 billion active users, followed by its main rivals Google + with 235 million active members (established in 2011), Twitter (founded in 2006) with 200 million and LinkedIn (professional networking site launched in 2003) with 160 million user base. What is behind the current boom of SNS? To find out, the definition and characteristics of SNSs will be elaborated.

Boyd and Ellison (2007: 211) define social network sites as “web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system”. These three functions are equally important as they focus on the presentation of self, relationships with a circle of users and the information provided by other communicants.

Baron (2008) gives an insight into the first one, the construction of a profile visible for the community. She points back to the notion of ‘presentation of self’ introduced by the sociologist Erving Goffman fifty years ago. It is a social construct built on an argument that “people consciously or unconsciously present themselves to others as if they were actors on a stage: Do I want to appear assertive? Vulnerable? Sophisticated?

¹⁰ ‘List of virtual communities with more than 100 million active users.’ Wikipedia.

<http://en.wikipedia.org/wiki/List_of_virtual_communities_with_more_than_100_million_active_users>
Accessed on September 30, 2013.

Available?” (Baron 2008: 71). The role of actors can be even abused by entering untrue information, which is possible due to the relative anonymity of the Internet. Although social networks require its users to complete the information displayed on their profiles accurately, the cases of fake profiles are numerous.

The uniqueness of social networks, as Boyd and Ellison notice, is anchored not in the opportunity for socialisation with unknown people but in the option of articulation of individual users' micro-networks. There is a shift from interest-centred to people-centred environment. The keystone of SNSs lies in a visible profile where one can “type oneself into being” (Boyd&Ellison 2007: 211). The profile page contains information about the user, his or her contacts and sometimes the record of their mutual interaction. The communicants are often called ‘friends’ or ‘connections’ although their relationships are of various kinds and do not always mean friendship.

Besides the profile page, most SNSs offer multiple options of contacting one's connections. The choice ranges according to the level of synchronicity, from private messages copying ordinary webmail through public messages with or without a specific addressee to a synchronous instant messaging. Not all SNSs have these three options, the second one currently being most common.

Some SNSs can be determined geographically. Kopytoff (2004) illustrates it on Orkut, a social network originally launched only in English which soon became popular with Brazilian users (speaking Portuguese) who now form the majority of user base.¹¹

2.4.1 Facebook

Baron (2008) gives a quick overview of the history of Facebook. The first milestone is November 2003 when Mark Zuckerberg, a Harvard student (at the time) designed a web-based programme for Harvard students. His sources of inspiration were Friendster and traditional university facebook. The original site was called *TheFaceBook.com* and was launched in February 2004. The access to this site was first granted only to Harvard students, later it was expanded to other schools and universities in the USA and other

¹¹ Kopytoff, V. ‘Google's Orkut puzzles experts / Internet watchers ponder reason for social network site.’ SFGate. <<http://www.sfgate.com/business/article/Google-s-Orkut-puzzles-experts-Internet-2633049.php>> Published on November 29, 2004. Accessed on September 30, 2013.

countries and finally, in September 2006 “the lid was taken off of Facebook” and anyone with an email address could be registered (Baron, 2008: 81).

Now follows a description of chosen features of Facebook which are linguistically pertinent. As this social network is full of dynamism and constant changes due to various reasons, the status quo of Facebook (September 2013) will be analysed and any previous or future version will not be taken into consideration.

One of the advantages of Facebook is its support of channels of various synchronicity. Interesting enough, ‘Messages’ which provide the same functions as regular webmail (including own e-mail address `nickname@facebook.com`) can turn into a fully synchronous instant messenger with one click. Users then chat via a chat window and are able to see which contacts are online. Chat includes the features such as the restriction over the visibility of the user by his or her Friends (i.e. the user simply chooses either a list of Friends who can see him or her online or bans certain contacts) and an alert that a Friend is contacting the user (a notification ‘... is typing’).

An important communicative channel of Facebook are various fan pages, open and closed groups on any topic one can think of and events. They work on the basis of Internet forums, i.e. one user is administrator and the groups have a definite number of members. Any member or invited person can create these social communities and events although the access to them works on various degrees.

The most usual communicative option of this SNS is ‘status’ and ‘comments’. Under each status or comment there is a special space devoted to several kinds of information. For instance, temporal (and optionally geographical) data, number of comments and ‘likes’. ‘Like’ is a highly popular function thanks to the easiness of its use. Any textual and nontextual comment can be attributed ‘Like’ to express a positive evaluation towards its content.

Comments in events, groups and fan pages also contain information about the number and the identity of the users who have seen them. Similarly, Messages display a notification when the message is read by the recipient. These features also contribute to the conversation-like character of CMC on Facebook.

Conversely, an option (which was added in June 2012)¹² enables editing of comments. It distances Facebook from speaking by the fact that the sender can correct himself or herself without the reader seeing the original utterance. At least, the reader of the comment is notified about the change by a small note 'Edited' below the respective comment.

Another remarkable feature of Facebook lies in the fact it serves as a support of existing relationships. The research of Ellison, Steinfield, and Lampe "The benefits of Facebook 'friends': Social capital and college students' use of online social network sites." (2007) indicates that this SNS is used to maintain existing offline relationships or solidify offline connections, as opposed to meeting new people. Hence the anonymity between the users is not so high as with other online communicative interfaces.

Perez-Sabater is one of the few authors who examine this social network from the linguistic point of view. In her study she explains the reasons of the scarcity of research. It is not only the novelty of this type of CMC but also its complexity built on multiple parallel genres present in one platform (Perez-Sabater, 2013: 83-90). Thus, the new language variety which is neither speech nor writing is still evolving in accordance with the development of the medium itself. More research is expected once Facebook and other SNSs are conventionalised with established rules.

¹² Golijan, R. 'You can now edit Facebook comments as much as you want.' NBC News. <<http://www.nbcnews.com/technology/you-can-now-edit-facebook-comments-much-you-want-840938>> Published on June 22, 2012. Accessed on September 30, 2013.

3 PRACTICAL PART

3.1 Methodology

The main hypothesis of this paper is based on the assumption that netspeak English can be considered an independent language variety displaying its own distinct features. It is expected that English on Facebook exhibits features similar to a face-to-face communication and is more distanced from the genres of the written form. The idea will be tested on a selected corpus of Facebook texts written by British university students, native speakers of English. The Facebook data will be compared with spoken texts from the British National Corpus.

Although some linguistic features are expected to be typical of Facebook communication in general, there are likely to be differences between sub-types of Facebook texts. These differences are assumed to be connected with the tenor, i.e. the participants and their mutual relations, and the communicative goals or social functions of the communication. Two types of Facebook communication will therefore be examined: the first set of data originates from a communication located on the personal wall of a university student; the second one is acquired from a group communication placed on the wall of the respective group. While the interactions on the personal wall occur between the users who know each other outside electronic world, this is not ensured in the groups. Therefore, it is envisaged that the former type of communication will be more oral and informal than the latter type.

For the analysis proper, a basic unit has to be set; however, there are several problems connected with this task. The novelty of CMC and particularly of Facebook and subsequently the lack of systematic linguistic description makes it difficult to define the terminology. Although the input is textual, the sentences are not its basic constitutive elements. Therefore, the unit of the analysis will be defined in terms of the oral communication. Similarly to a conversation where a turn construction unit (TCU) (Sacks, Schegloff & Jefferson, 1974) is the fundamental segment speech, in the environment of Facebook the basic unit is a comment. It can be defined as an utterance of one user introduced by his or her name, which is visually independent. A comment can be of various

types, e.g. a status update, a wall comment (either on the wall of a user or of a group) or a reaction on a comment.

Following the hypothesis, the comments will be analysed and compared to an oral genre, namely to informal conversation. Biber et al.'s description of the grammar of conversation (1999) will provide the background necessary for the comparison and contrast. This approach classifies the TCU into two types. The first one is a clausal type, i.e. “a structure consisting of an independent clause together with any dependent clauses embedded within it” (Biber et al., 1999: 1069). The second one is non-clausal, containing grammatically fragmentary components of speech including inserts which form situation dependent isolates. These disjunctive elements can be classified into several types, such as prefaces, noun phrase tags or unembedded dependent clauses and they are considered to be complete grammatical units (Biber et al., 1999: 1038-1108). This approach will be applied also to the Facebook comments.

The analysis itself will include two sections, i.e. the qualitative analysis of two Facebook samples and the quantitative one studying the whole corpus. Each section will comprise five traditional linguistic levels (cf. Crystal&Davy, 1992) of graphetics and graphology, morphology, syntax, lexicology and discourse. The analysis of the samples will try to be as detailed as possible while the analysis of the whole corpus will be focused on specific phenomena. These will be chosen according to the results of the qualitative analysis. Additionally, the quantitative analysis will compare Facebook data with the spoken demographically sampled part of the British National Corpus.

For the purposes of the corpus study a concordance programme AntConc, developed by Laurence Anthony is chosen.

3.1.1 Communicative options of the platform

This section will deal with the environment in which a comment is created on Facebook. A comment can be inserted into a white box whose small size indicates that messages are usually short. However, contrarily to other channels of CMC, such as text messaging or Twitter, the message is not limited by the number of characters. This means

that the length of the comment can range from a single character to many pages.¹³ Comments do not necessarily have to contain only linguistic material, as the Figure 3 printscreen of an update status ready for editing shows.

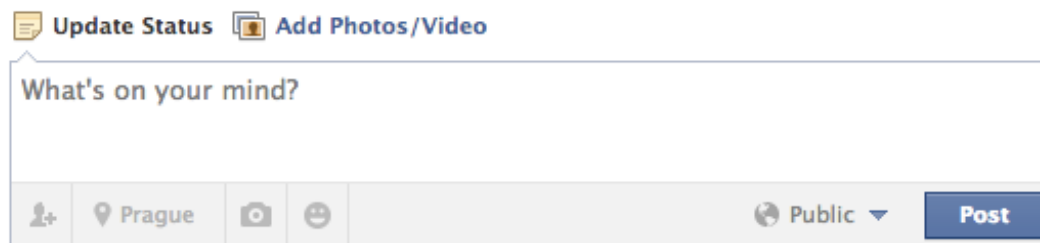


Figure 3: Default box for adding a status update on Facebook

The topmost part of the window shows two options. Linguistically relevant is the first one ‘Update Status’. The white box below contains a hint in the form of a question ‘What's on your mind?’. In the bottom part of the comment there is a number of add-ons which can enhance the message. Proceeding from left to right, the first icon enables tagging other people in the comment, i.e. mentioning their names which will be automatically linked to their profiles and about which they will receive a notification. The adjacent icon marks the location of the user which is editable. Two other symbols enable adding a file, or choosing an emoticon from a predefined list. The last icon (marked ‘Public’ in the picture) adjusts the visibility of the comment for other users.

The described communicative environment indicates that although composing a textual message is one of the main functions of this platform, the wide range of extra features prepare the grounds for creating a dynamic multimodal communication. It is especially tagging of the users which adds a social dimension. As the users are notified about being tagged, they are more eager to provide reactions and to create an interaction similar to a face-to-face conversation.

¹³ Facebook shortens longer messages and displays only six first lines. Clicking on ‘See more’ opens the whole comment in a separate pop-up window.

3.2 Linguistic analysis

3.2.1 Qualitative analysis of Sample 1: Personal wall comment

Sample 1 which will be analysed is a status update posted on Facebook by a university student and the reactions of her friends, i. e. it is produced on the initiator's personal wall. In this short exchange there are three participants, two female and one male and it is expected that they know the composer of the status update (hereafter speaker¹⁴ Jane) outside the electronic world. For reference the original document is copied here as Figure 4.



Figure 4: Sample 1 (Facebook personal wall comment)

3.2.1.1 Graphetics and graphology

On the level of graphetics and graphology, the status and its reactions contain only ASCII characters¹⁵, nontextual items such as graphical emoticons and embedded images

¹⁴ Although the participants are not involved in face-to-face conversation, they will be referred to as *speakers* because of the similarities with informal face-to-face conversation

¹⁵ ASCII or American Standard Code for Information Interchange is a character scheme based on English alphabet. The characters are available on a standard keyboard in Euro-American zone, cf. <<http://www.asciitable.com>> Accessed on October 31, 2013.

are absent. Nevertheless, emoticons are present in the textual form, e.g. :d, =[and =p. Their forms can be described as less typical due to several reasons.

All the emoticons are written using only lower case characters and those symbols which can be written without the Shift key, i.e. :d, =[, =p instead of the more usual forms :D (smile), :((sad face) and :P (sticking out tongue). While this is probably due to economy reasons (capital letters are generally avoided in the comments and the emoticons are so widespread among the Facebook users that the use of a lower case letter is not likely to cause confusion), the spelling may also contribute to personal style and originality of expression.

The function of the emoticons present in this conversation is not identical (cf. Dresdner and Herring, 2010): :d and =[indicate positive and negative emotions, respectively; =p, a face with the tongue sticking out, can be seen as a marker of irony or mocking. The question of their contribution to the communication remains open. For instance it is not clear whether :d expresses only amusement over the propositional content or making fun of the second speaker.

Another graphological point is the distinct use of punctuation marks: punctuation may be non-standard (functioning also as emotive markers) or missing. For example, the introductory sentence *is sick of men right now* is missing a closing punctuation mark, probably because it is not needed for understanding. Sentence-final punctuation marks tend to be used only if they signal a particular discourse function, such as the question mark for a question (*awww how come?*), double exclamation mark for an emotionally-coloured exclamation (*[...] you moved away to germany!!*)¹⁶ and ellipsis dots for an unfinished sentence (*We are trouble-causers...*). A full stop is used only in case of dividing two sentences in one comment (*Ah true. We are trouble-causers...*) and to conclude the conversation (*trouble causers not worth the trouble.*).

The majority of the sentences start with a lower case letter, which, similarly to the emoticons used here, may have effort- or time-saving reasons. There is also one instance of low-case personal pronoun *i* and of a place name *germany*. The presence of typing errors, such as missing space between the emoticon and a word in =*pyou*, absent plural marker in

¹⁶ The double exclamation marks in Jane's last comment have an atypical syntactic position, splitting the sentence into two parts connected by *and*.

several year and spaceless two words with a spelling error *frogotwho* all attest to the increased speed of writing. This is connected to the spontaneity of the Facebook communication, resembling that typical of spoken conversation.¹⁷

Emoticons, lower case letters, missing punctuation and spelling errors are the phenomena characteristic for synchronic type of CMC. However, here thanks to time records below each utterance it can be observed that the conversation is asynchronous. The time delay of one, four and fifteen minutes and even three days confirms this. Thus, it seems that the expansion of features of synchronous e-communication to asynchronous one has an upward tendency.

3.2.1.2 Morphology

On the morphological level, the presence of the interjections *awww* and *Ah* (at the beginning of the second and the fourth comment) brings a feature of orality and informality. Their meaning is highly context-dependent.¹⁸ They provide a feedback for Jane's comment: *awww* expresses sympathy of speaker Christian with Jane's feelings, while *Ah* followed by the word *true* indicates agreement. Additionally, it seems that similarly to emoticons both expressions hide amusement and even irony of speaker Christian and therefore may have a different interpretation.

As to the ratio of the parts of speech, the sample is in agreement with a real conversation (cf. Biber et al., 1999). It is in particular high frequency of personal pronouns with direct addressing and self-reference *you* (four times) and *i/me* (twice) which have deictic reference. Additionally, there are single instances of the pronouns *they* and *We*. The adjectives used, *worst*, *worth*, and *sick (of men)*, are evaluative. The time expression *right now* suggests again the immediacy of the communication similar to spoken conversation. As far as the verb categories are concerned, all the verbs are active, and only simple present and past tenses are used.

¹⁷ When characterising the external determinants of conversation, Biber et al. (1999: 1048) note that conversation „takes place in real time ..., so that speakers are continually faced with the need both to plan and to execute their utterances in real time, ‘online’ or ‘on the fly’.”

¹⁸ *Aw* is “used to express disapproval, protest or sympathy” whereas *ah* is “used to express surprise, pleasure, admiration or sympathy, or when you disagree with somebody” (Oxford Advanced Learner's Dictionary online <<http://oald8.oxfordlearnersdictionaries.com/dictionary> > Accessed on December 27, 2013.

3.2.1.3 Syntax

The comments contain clausal, elliptical and non-clausal units. There is only one complex sentence comprising two dependent clauses (a time and a content clause). The structure of Jane's last comment resembles the complexes found in spoken conversation due to the 'add-on' strategy of online spoken production (Biber et al., 1999: 1068): the coordinated clauses are linked *asyndetically* or by the conjunction *and*, the adverbial dependent clause (with a clausal object) is attached at the end of the coordinated clauses. A formally complete clause is marked as incomplete in Christian's comment *We are trouble-causers ...*

The elliptical clauses can be completed unambiguously based on the grammatical structure (*true*) or on the shared context (*you are*; Jane's first post *is sick of men right now*).¹⁹

The non-clausal units are found typically at the beginning of the comments. They comprise the interjections (*awww*, *ah*), or a conjunction introducing an isolated dependent clause (*cause ...*). They may, however, constitute the body of the comment, viz. the non-sentence *How come?* Similarly, the sentences *trouble causers not worth the trouble.* and *Ah true.* are fragmentary. The emoticons in clause-final position can be considered a specific type of tag, providing (additional) clues to the interpretation of the preceding clause, similarly to paralinguistic features, intonation, stress or voice quality in spoken conversation. Alternatively, the emoticon can preface a clause (following a conjunction in Jane's comment).

Subsequently, *awww*, *ah true* and *you are =p* as the reactions on the previous comments are by their position prefaces identical with non-clausal elements of a real conversation. The comments consist also of final disjunctive elements in the form of tags which close the respective turns. These are usually emoticons, e.g. :d.

The comments are rather simple. Three out of six turns contain a single clause and other three are not significantly longer. This simplicity is a feature shared with informal conversation where, the more shared context, the less overt grammatical structures are

¹⁹ Apart from ellipsis, the fact that Facebook communication, like spoken conversation, relies on shared context, may be illustrated by pronoun reference as a type of grammatical reduction (Biber et al., 1999: 1042).

needed. This is reflected by the occurrence of prefaces and noun phrase tags (Biber et al., 1999 : 1043).

The lack of complexity is apparent also at the phrase level. The noun phrases are not accompanied by modifiers and they typically comprise only the nominal (or pronominal) head, *men, trouble, trouble-causers, germany*, pronouns *they, nothing, me ...*²⁰ Similarly, the sentence *trouble causers not worth the trouble.* is the repetition and specification of a noun phrase from a previous turn *We are trouble-causers....*

3.2.1.4 Lexicology

The signs of informality and word play can be observed also on the lexical level. Speaker Jane makes use of homonymy in the word *cause* which functions as a subordinator (informally shortened from *because*) and also as a noun/converted verb. Besides the low lexical density, the vocabulary is repetitive and restrictive (e.g. the words *trouble* and *cause/causers* occur several times).

Looking closer at the vocabulary employed in this discussion, an important group of verbs is missing. These are modals verbs and other markers of expressing politeness. The conversation, especially by speaker Jane is rather direct and even offensive. On one hand there are no softeners and hedges, on the other hand lexical vulgarisms, such as swear words, imprecations or curses are not used either.

3.2.1.5 Discourse

The last level to describe, the discourse one, crosses the boundaries of the sentential and non-sentential elements and aims to analyse the functions and mutual relationships of higher units formed by individual comments. The structure of the status update and its reactions is closely connected to their function. The initial comment is evaluative and expresses the need to communicate an opinion. It needs an additional explanation and that is why the content of the first reaction is a *wh*-question. This question is not neutral either,

²⁰ This is a feature observed by Biber et al. (1999: 1044) to be typical of conversation: “participants in conversation make the most use of pronouns, characteristically reducing the noun phrase to a simple monosyllable; ... they find little call for the more elaborated forms of noun phrase structures, containing noun heads with various forms of modification.”

as its author is a man, i.e. the criticised group member. Speaker Christian shows pity towards speaker Jane and wants to hear the reason of such an emotive statement. Jane's explanation is even more emotional but Christian remains calm. He explicitly identifies himself with the group of the accused (pronoun *We*) and reiterates Jane's answer by nominalising the action. Jane continues in her aggression and expands on his statement, loosely ended by ellipsis dots. She mocks him with the help of emoticon =p and a concrete argument. There is no further reaction from Christian and it may have several reasons, e.g. he did not read her comment, he opted for no comment or he contacted her by other channels. Three days later speaker Nicole enters the communication by postmodifying Christian's group label *trouble-causers* with a reduced relative clause *not worth the trouble* but this opinion remains without a reaction.

The described progression of the electronic conversation is in discord with the non-electronic one in a few points. Seen from an external perspective of an observer, the communication is unfinished as the reaction of speaker Christian to Jane's comment, as well as Jane's reaction to Nicole's remark are unknown to the reader. While in face-to-face communication, the non-verbal expressions of the speakers often provide a clue for a correct interpretation and no further words are needed, in CMC there is no such possibility. In this respect, no or delayed response connected with the lack of immediate feedback typical for asynchronous types of CMC approximates the whole conversation thread to a traditional writing genre.

The comments convey two interlinked functions.²¹ The referential one describes the mental state of the first speaker (*is sick*) and includes the deictic space reference (*right now*). The expressive function is illustrated by the excessive usage of exclamation marks and emoticons in the reactions under the comment.

An issue which emerged after the analysis of the discourse level is gender. Research focused on the questions of gender influence was done in several studies by Baron²², however, no such data have been examined on Facebook. In this conversation, there are representatives of both sexes, moreover, the topic is gender related. Similarly to the study

²¹ Cf. Jakobson's model of communicative functions, Jakobson, R. (1960) *Linguistics and Poetics*. IN T. Sebeok, ed. *Style in Language*. Cambridge, MA: M.I.T. Press, 350-377.

²² The question of the importance of gender was analysed in a different CMC environment (instant messaging) by Baron and she concludes that "gender is a relevant variable in the construction of IM conversations" (Baron 2009b: 1).

by Baron in which the results show that emoticons are used predominantly by women (as polite markers and softeners), it is valid for this particular conversation. Jane's comments contain three emoticons and two exclamation marks whereas Christian uses only one smiley and no exclamatory sentences.

These findings can be supported by the tone of the turns chosen by the participants. Speaker Jane seems to use the words with a negative connotation (e.g. *sick*, *trouble* and *worst*) and adds an element of mocking (e.g. the emoticon =p) while her male counterpart Christian is amused and ironic (e.g. the emoticon :d and the sentence *Ah true.*). Still, no general conclusions can be drawn, that is why this phenomenon will be continuously observed in other comments.

3.2.2 Qualitative analysis of Sample 2: Group wall comment

The second sample of CMC communication which was selected for the qualitative analysis originates in a Facebook group called *Erasmus society* which is one of the societies created for students at the University of Kent in Canterbury, England. The group is open to anybody and there are no privacy restrictions. Although the society is focused on international students (such as the ones enrolled in the Erasmus exchange programme but also on regular foreign students taking their degree at this university), the example contains the utterances of native speakers of English.

For reference, the full copy of the sample is attached below as Figure 5.

There are five speakers in total, four female and one male. The main speaker (Emily) is one of the organisers of various social events and she often initiates the communication in this Facebook group. Other speakers (Sophie, Abhi, Astrid and Zoë) are members of the society and it is probable that they do not know each other personally (at the time of writing the comment). This conclusion is drawn from the fact that they are not linked as friends on Facebook and did not contact each other via this medium although they use it.



Figure 5: Sample 2 (Facebook group wall comment)

3.2.2.1 Graphetics and graphology

Compared to the previous sample, the graphological level of this communication is richer as it contains graphic emoticons. However, it is also more uniform and impersonal with only one kind of emoticon, 😊, used always with the same function, viz. indicating the positive charge of the message.

Another nonlinguistic expression is xxxxxxxxx and xx at the end of the comments of the first two speakers is a reduced emoticon (x resembles a kiss). The function of this smiley is general affection, the more xs, the more friendly tone of the message.

The orthography of this conversation shows some similarities with Sample 1 in the extensive use of the exclamation mark !. For example, the introductory comment contains

four instances and in total all comments except for one include this mark. There are several distinct functions provided by exclamation marks used here. The invitation contains two imperative sentences which openly but politely give its readers the instructions what to do, i.e. *bring your friends* and *comment or like this post*. They both attempt to convey the emotion of elation which culminates by the doubling of the exclamation marks in the last sentence.

A different function can be attributed to the exclamation mark in the reactions and in the invitation. Apart from the imperative sentences it is used in the invitation to add emphasis in ... *we were thinking of having another unofficial social at the Ballroom!!* The exclamation mark is combined with a question mark to emphasise the polite request *Can I pop along for two maybe and say hellooo?!* Abhi's exclamatory reply is assuring. The function of *!* in Zoë's comment, on the other hand, is to communicate an apology. Despite the variety of functions this punctuation mark performs, it seems that all of them make the communication look less formal, lighter and more vivid. As their role imitates the prosodic features, such as loudness, higher pitch and intonation, an analogy with a real conversation can be seen.

Minor spelling errors, such as the lack of a space behind a comma in *again,there* or a typo *unoficcial* (both in the first comment) indicate that thorough proofreading is not common in this type of CMC.

Next, there are two capitalised words *ALSO* and *AWESOME* which catch the reader's attention and alleviate the monotonicity of a lengthy comment. Similarly, reduplicated letters *h* in *yeahhhhh* and *o* in *hellooo* make the text look more informal. None of these expressions with non-standard spelling are employed in Sample 1, most probably because the functions of both samples differ. Whereas Sample 1, being in harmony with its status of a personal message, expresses an emotion and an opinion and request friends' support, this sample, aimed at a larger group united by the same interest, communicates an event invitation. Hence, in order to obtain as many readers as possible it makes use of specific marketing strategies (similarly to printed media) and one of them is the innovative language, easy to understand for its reader.

3.2.2.2 Morphology

Informality of the conversation is evident also on the morphological level. For example, it is the contracted form of the negative particle in *isn't* (Emily's comment) or the reduced form expressing a planned future action *gonna* (Zoë's comment). At the same time, there are instances of full form of verbs, e.g. *We were thinking* or *We have booked* (Emily's comment). Additionally, in Zoë's sentence *probably be late though!* the subject-auxiliary verb part *I will* is again omitted.

The representation of parts of speech is similar to Sample 1. The distribution of personal and possessive pronouns corresponds with the functions of the comments. For instance, the invitation contains five personal pronouns *we* and *us* which refer either to the organisers *we were thinking* or also includes the recipients *we all had*. The focus is also on the recipients as there are three instances of *you* and *your* addressing the whole group, which is specified by the noun *guys*. The recipients' replies are predictably ego-centred with three examples of the pronoun *I*. Unlike Sample 1, third-person pronouns do not occur here.

Elements such as substitution proforms or deictic expressions are infrequent. It goes hand in hand with the fact that the participants share only limited amount of background information and mostly do not know each other. Moreover, there is an unspecified number of the readers of the invitation and the message needs to be clear to be successfully communicated. The most important piece of information is the physical and temporal location of the event and the sentences abound in respective adverbials, e.g. *the Ballroom* and three instances of anaphoric *there*, *this week*, *last week*, *until 10.30* and *8pm onwards*.

There is only one interjection *yeahhhhh* but there are several evaluative adjectives, such as *AWESOME* and *Sweet*. Another similarity with the real speech (as defined by Biber et al., 1999) consists in that the frequency of modified noun phrases, attributive adjectives and pronouns of other types is considerably low. Conversely, there is a significant number of the verbs of action, e.g. *have booked*, *move*, *bring*, *meeting*, *pop along* and *come* and modal verbs, e.g. *can* and *might*. Similarly to face-to-face conversation, the verbs *say-think-know* occur here, e.g. *were thinking of* and *say hello*. The verb tenses used are more varied than in Sample 1, including present simple, present perfect, the simple and progressive past, and *will*-future.

3.2.2.3 Syntax

Similarly to the previous sample, there are instances of clausal, elliptical and non-clausal units.

The structure of the invitation post (by Emily) differs syntactically from the subsequent comments. The comments are shorter (between two and ten words), comprising one or two clauses or non-clausal units (*Of course!*), accompanied by emoticons. The invitation (99 words) contains more clausal units comprising four multiple sentences. There are types typical of speech and of writing as well.

The clausal units consist of several compound sentences with additive function whose clauses are connected with 'and' coordinator. Two of them starting with *We have booked...* and *So yeahhhhh...* are multiple as they contain three clauses linked asyndetically and by *and* coordinator. Similarly to Sample 1, these sentences reflect the principle of add-on strategy, and resemble oral communication. At the same time, this type of multiple sentence enables keeping a different intentional modality for each of the clauses and therefore it is possible to connect by coordination a declarative and an imperative sentence starting with *If you are interested...* .

However, some more sophisticated patterns typical for written style are incorporated in the invitation. These are multiple complex sentences, e.g. a subordinate clause of conditional-concessive type with the subordinator *Although*. Moreover, there are infinitival structures present, such as *we can decide whether to move somewhere else or stay* where the object of the clause is expressed by a non-finite dependent clause containing two infinitives. Other non-finite forms include two gerunds *of having* and *in meeting*, which increases the complexity of these multiple sentences.

The possibility of expansion of elliptical sentences relies on the immediate (grammatical) context: Zoë's comment *probably be late though!* lacks the subject and the auxiliary verb. These items are easily retrievable from the previous context and pose no problems to the comprehensibility of the comment. Ellipsis is also a frequent phenomenon of oral communication.

The non-clausal elements include the preface to the invitation *ALSO. also also also* emphasised by capitalisation and reduplication. It explicitly links Emily's post to another invitation the same user posted ten minutes earlier, highlighting – at the same time – the present invitation.

The invitation contains another non-clausal preface which combines the connective and emphatic functions, *So yeahhhh*, highlighted further by the non-standard spelling of the interjection. Another non-clausal evaluative unit, *Sweet*, is placed at the end of the invitation. In connection with the neighbouring emoticon it enables a smooth structural transition from the main part of the invitation to its end. It is followed by the emoticon *xxxxxxxxxx*. As has already been said, the form of the comment resembles an e-mail and this emoticon concurs with this idea. Placed finally, it functions as a closing greeting, in the same way as the expressions *cheers* and *sincerely* in emails and letters. Looking at the invitation structurally, despite the absence of the conventional opening greeting, all other obligatory parts of an email are included. Namely, there is a covert but contextually obvious recipient (the society members), a sender (whose name is at the beginning of the message, thus a final signature is not needed) and a closing formula (*x* emoticon).

Another grammatical isolate with an important syntactic and pragmatic function are the emoticons, all of which are placed at the end of clauses. In total, there are four smileys, two textual and two graphic ones. Although their position is the same, they differ in the range of their functions.

Sophie's reply contains two smileys placed side by side. The graphical one marked textually as :-) is linked back to the previous question and signals friendliness and a positive attitude of the speaker. The textual one *xx* copies the structure of the invitation and serves as a closing formula. These two extralinguistic elements are probably preferred over the textual representation because of their brevity.

The emoticon placed in Zoë's reply stands in the final position of the first sentence. As the following sentence does not contain a capitalised first letter, this smiley is the only visual marker of the split between two sentences. The emoticon highlights the positive attitude of the sentence. Its more usual comment-final position is not possible here as its scope is limited to the preceding sentence due to the difference in the sentence type (declarative vs. exclamatory).

The structure of the reply comments reveals some salient characteristics of this type of CMC. First, it is informality in the ending area of the clauses, e.g. Astrid's reaction *I might come - but probably not before 11...* is visually divided by a dash emphasising the information which follows. Moreover, it is ended with the suspension points signalling an unfinished thought, probably an explanation which the speaker does not consider appropriate to articulate.

3.2.2.4 Lexicology

Although the invitation comment contains rather complex syntax, the lexis balances it by its less formal character. The lexical choice of this CMC is influenced by the informal environment and less personal atmosphere of the communication. There are a few casual phrasal verbs, such as *bring on* (speaker Emily) or *pop along* (speaker Sophie) and a colloquial form of addressing *guys* (speaker Emily).

Interesting word-formation changes are apparent also with the verbs *comment* and *like* in *please comment or like this post* (speaker Emily). In case of the verb *to comment* there is a semantic specification to a single meaning *to reply*. The stative verb *like* is used in an imperative sentence and its meaning is shifted to *click on the Like button*.

3.2.2.5 Discourse

Next point worth analysing is the discourse level. As already mentioned, the conversation comprises two separate, stylistically divergent parts. The initial one is the invitation bearing resemblance with an email by some syntactic and semantic elements described in the earlier part. At the same time, the cohesive expressions which occur here *So yeahhhhh* and *Sweet* have the oral character.

There are three different functions present in this initial post. The first one is referential, i.e. the text tries to describe exact details of place and time of the event. There are several deictic and anaphoric expressions (*as we all had AWESOME time last week, another, again, and same as last week*) to establish common ground before the invitation. After these two preparatory clauses comes a third one, finally announcing the event,

another unofficial social at the Ballroom. Event details and a time schedule are offered in the sentence *We have booked [...] stay there*.

The second function is conative, i.e. it appeals to the reader by using imperatives and direct addressing (*guys*), viz. the sentence *So ... ballroom!*. The use of the adjectives with positive evaluation (*AWESOME, Sweet*) makes the invitation more attractive.

The third function is expressive, focusing on creativity of the composer Emily, which is exemplified by the use of nonstandard and emphasised orthography (!!; *So yeahhhh*) and disjunctive structures (*ALSO. also also also also*) typical for informal language. These three functions cause that the post contains heterogenous linguistic material and has a dialogical character.

The reactions forming the second part of the communication are much shorter than the initial comment. There are three of them and the fourth one *Of course* (speaker Abhi) is the reply to the previous reaction.

The comments are formulated in a polite manner via various means. The comment of speaker Sophie is in the form of a polite request *Can I [...]?*. The other two are softened, speaker As uses the modal verb *might* the speaker Zoë chooses the construction *try and ...*, both of which serve as hedging of the promises. The last two reactions (of speakers Astrid and Zoë) have a very similar form and content and both contain the word *probably*. Their second clauses express the adversative and concessive meaning thanks to the conjunction *but* and the conjunct *though*. All of these elements which can be summarised as politeness formulae and hedges also occur in face-to-face conversation, hence this part of the communication carries a significant resemblance with a real conversation.

Next question of the discourse level is how these two heterogenous parts work together. All but one reaction respond directly to the invitation and despite the asymmetric character of the communication the comprehensibility remains untouched. The first reaction by speaker Sophie is a question answered in the following comment by speaker Abhi. Two other comments directly respond to the main invitation and although they are hedged, they provide a restricted promise. These adjacency pairs in form of utterance-response sequences copy the structure of face-to-face conversation.

Due to the complexity of its structure the invitation part seems to be more distant from a real conversation and closer to some written genres (such as the informal letter).

As with the previous sample, the discourse related issue is the distribution of emoticons and their relationship to gender. The personal wall conversation showed that there may be a possible link influencing their frequency. The local data reveal that the comments of three out of four female users contained at least one emoticon while the message of their male counterpart was emoticon-free.

The constructions expressing politeness and hedging, which were absent in Sample 1, are abundant here. The expressions such as *please* (speaker Emily), the constructions containing modal auxiliary verbs *can* (speaker Sophie), the past continuous *we were thinking of...*, the construction *hope to see*, the conditional sentence *If you are interested...* (all in Emily's comment) are motivated by politeness. This feature is common also in informal conversation (cf. Biber et al., 1999: 1047).

3.2.3 Quantitative analysis

The detailed description of two instances of English netspeak brought interesting but mixed results and pinpointed several dominant features of this electronic language variety. To verify whether these elements are a one-time occurrence or whether they are permanent characteristics of this CMC they will be observed on a larger scale of data.

First, some general statistics of the two Facebook corpora will be provided. The analysis is performed in the corpus consisting of 309 comments²³ of two distinct types. The first one, labeled as Corpus 1, contains 204 comments posted on a personal wall by one user²⁴ and her friends. The second one with 105 comments is represented by Corpus 2 and it contains the messages of two British university societies (*Erasmus society* and *Cafe on the Hill*) composed by their members. The reason of this disproportion of almost 2:1 is that Facebook groups of *Erasmus* and *Cafe on a Hill* societies were founded later than the personal profile and thus there are less data available. The whole corpus contains 25,130

²³ A comment usually contains a threaded conversation, i.e. a comment and its sub-comments

²⁴ The user agreed with using her personal wall comments for the purposes of this thesis

word tokens (Corpus 1 has 17,607 and Corpus 2 has 7,523 word tokens) and 3,619 word types.²⁵

The third element is the spoken demographically sampled part of the British National Corpus (hereafter Spoken-BNC).²⁶ The total number of word tokens available is 4,219,192 in 153 files. This part of the BNC comprises transcripts of informal conversations recorded by selected individuals in everyday situations during the 1980s and early 1990s.²⁷ This is a disadvantage as the time difference with Facebook data is twenty years. However, no other more recent corpus of such size containing spoken British English is available.

3.2.3.1 Wordlists analysis

Frequency-ranked wordlists will be used as a starting point of the quantitative part of the analysis. Table 2 lists the most frequent words in the three corpora compared.²⁸ A wordlist based on the written part of the BNC is included for comparison in the last column of the table.²⁹

Rank	Corpus 1	Corpus 2	Spoken-BNC	Written BNC
1	i	to	i	the
2	you	the	you	of
3	to	you	it	to
4	the	and	the	and
5	a	a	's	a
6	and	i	and	in
7	it	we	that	that
8	of	for	't	is

²⁵ The word-counts do not include symbols and emoticons.

²⁶ Data cited herein have been extracted from the British National Corpus, distributed by Oxford University Computing Services on behalf of the BNC Consortium. All rights in the texts cited are reserved.

²⁷ A detailed description of the corpus is available at <http://www.natcorp.ox.ac.uk/>

²⁸ The settings for the Wordlists include treating all data as lowercase. Therefore the Wordlists do not record, e.g., the difference between capitalised I and lowercase i.

²⁹ The frequency list for the written part of the BNC was generated using *BNCweb*, available at <http://bncweb.lancs.ac.uk/>, accessed in December 2013.

Rank	Corpus 1	Corpus 2	Spoken-BNC	Written BNC
9	in	it	a	for
10	that	of	to	it
11	my	on	yeah	was
12	for	be	he	on
13	is	at	they	with
14	's	will	in	as
15	with	in	of	be
16	have	is	oh	's
17	't	if	what	he
18	not	so	no	i
19	be	this	well	by
20	jane	your	on	at
21	me	have	she	are
22	just	can	we	not
23	was	all	was	from
24	but	't	there	you
25	lol	's	know	this

Table 2: The 25 most frequent words in the Facebook and the BNC corpora

If punctuation and numbers are recognised as tokens, emoticons enter the list of the most frequent words in the Facebook corpora. In Corpus 1 the 25 most frequent items on the wordlist include (with the rank/frequency given in brackets) :p (9/186), xxx (11/183), :) (15/148), and xx (25/104). In Corpus 2 there is only one emoticon among the 25 most frequent words, namely :) (6/94).

As it can be seen, there are no significant differences among the Facebook and Spoken-BNC lists and the grammatical words, such as personal pronouns, determiners and prepositions dominate at most positions. The lexical words are scarce and the distribution of the parts of speech tallies with the findings of the analysis of the Facebook samples. For instance, first two positions in personal wall comments are occupied by the personal pronouns *I* and *you*, which matches the findings from Spoken-BNC. Overall, it seems that Corpus 1 of personal wall comments is very similar to Spoken-BNC, which confirms its

oral character. Group wall comments are more reader-oriented with the top personal pronoun *you*. They are also more institutional in featuring the pronoun *we*, which reflects that the societies decide about the activities and communicate with their members as a whole.

The most frequent words of Spoken-BNC which do not occur on top positions of the Facebook corpora include the interjections *yeah*, *oh* and third person pronouns *he*, *she* and *they*. In oral communication *yeah* has a phatic function of signalling the attention, which is not necessary for an asynchronous type of CMC where the presence of both parties is not required. The third person pronouns are less frequent in the Facebook comments possibly due to other means of referring to other people, e.g. by writing their full names to tag them.

The use of the capital letter in the personal pronoun *I* is another remarkable area (not applicable for Spoken-BNC). Although both Facebook corpora favour the capitalised version, in the personal wall corpus the ratio of low-case and upper-case is approximately 1:3 whereas the group wall corpus contains 1:5 ratio. This again leads to the generalisation that the personal comments are more informal, using the principles of the economy of effort.

Using Spoken-BNC as a reference corpus for the two Facebook corpora enables creating Keyword lists presented Table 3. The first positions in the Keyword list of both Facebook corpora are occupied by the features of CMC, such as emoticons, :p , xxx, :) and :d. The corpora also contain many contracted forms of verbs, either of their negative particles or in connection with the personal pronouns *i*, *you*, *it* and *we*. In Corpus 1 highly ranked expressions include the acronym *lol* and proper names *Jane* and *Leah*.

Other words which rank lower (first 50 places) in Corpus 1 include the terms denoting concepts which did not exist at the beginning of 1990's, such as *zumba*, *facebook* or *text* (verb). There are also many words expressing positive evaluation and feelings, e.g. *love*, *lovely*, *miss*, *awesome*, *best* and *worst*.

Rank	Corpus 1 (personal wall)	Corpus 2 (group wall)
1	:p	:)
2	xxx	erasmus
3	:)	it's
4	xx	:d
5	i'm	email
6	lol	i'm
7	:d	will
8	jane	don't
9	:(can't
10	don't	you're
11	xxxx	tickets
12	it's	xx
13	you're	rutherford
14	x	:)
15	it'll	xxx
16	that's	to
17	i've	guys
18	haha	:(
19	<3	stonehenge
20	:/	trip
21	;(everyone
22	leah	we'll
23	love	we're
24	thats	bonfire
25	*_*	event

Table 3: Top 25 key words of the Facebook corpora (Spoken-BNC is the reference corpus)

The Keyword list of Corpus 2 seems to contain more ‘aboutness’ words, such as *trip*, *tickets* and *bonfire*. There are less emoticons and more focus on addressing the whole group, e.g. *guys* or *everyone*. There are also more references to places and time, e.g. *rutherford*, *stonehenge* and *friday*.

To conclude the section studying Wordlists, minor differences among the two corpora and the referential BNC corpus have emerged. The personal wall corpus tends to be more personal, informal and evaluative and shares many features with Spoken-BNC whereas the group wall corpus is more official and information-centred. Still, Wordlists of both corpora bear a significant resemblance with spoken communication.

Having observed Wordlists, the analysis will now move to the individual levels of linguistic analysis. The features treated in the previous section with the two samples will be revisited.

3.2.3.2 Graphetics and graphology

Emoticons are not easily retrievable via corpus software and their quantification is more imprecise. There are several reasons for that. First, due to their nonlinguistic nature and the form composed of the combination of ASCII characters it is difficult to formally define them. Moreover, although some of them are well-known to most users, their number is not limited and the list of conventionalised emoticons varies. Numerous ad-hoc and nonce formations of emoticons accompanying different comments are quite frequent. The freedom of choice, creativity and unruliness then cause problems in their interpretation.

However, it can be assumed that the well-known emoticons carry a clear message. The list of the most frequent ones in the Facebook corpora is presented in Table 4.³⁰

³⁰ The graphical emoticons in AntConc are represented in the textual form, they were manually converted, e.g. ☺ into :) and all emoticons were treated as lower-case

	form	meaning	Corpus 1 (total number)	Corpus 2 (total number)	examples
positive evaluation	:)	friendliness/ joy	152	97	Found my hat..... :) :) :) :) :) :) :)
	<3 [heart]	love/affection	74	1	LOVE YOU JANE!!! :D <3
	x	saying goodbye/ affection	406	45	you should stay like this. its highly entertaining!!! :D xxx
negative evaluation	:(sadness	47	10	oh noooo i'll be away this weekend :(
	;(crying	18	0	We didn't get to meet up! ;(
other, context- dependent evaluation	;))	joking/flirting/ saying hello	13	10	Hopefully see you soon for zumbaaa ;) xx
	:D	laughing	65	22	ahhh looking forward to Zumba! :D lol xx
	:P	irony/mocking	187	5	tis cause i'm one of the cool kids :P xx

Table 4: The most frequent emoticons in the Facebook corpora

Next point of the analysis of the emoticons is their position within the comments. The results show that most of them are clause-final or comment-final. Unlike in real conversation where the prosody and extralinguistic cues accompany the speech from the beginning, in CMC these markers of emotion and attitude are read at the very end of the section to which they are linked.

CMC examined in the samples indicates a lot of variation in the area of orthography. More examples are available in the corpora, such as lower-case spelling at the beginning of the sentences, with proper names and place and time references which are standardly capitalised in traditional writing genres, e.g. *zoe*, *london* or *tuesday*. Spelling errors are also numerous and they mostly concern typos, such as *beleave* or *irrtating*.

Writing the whole expressions, clauses and sentences in capitals such as in the event invitation comment *It's STORY NIGHT!* is customary in this type of CMC. Reduplicated letters and punctuation marks are often accompanying elements, for instance *Sooo... who's coming?????*. While the effort to write a message as quickly as possible seems to be the driving force for the occurrence of lower case and typing errors, the capitalisation and reduplication is conditioned by the need to put special importance or to express an emotion. All these phenomena resemble face-to-face conversation by their dynamism, playfulness and emotional colouring.

3.2.3.3 Morphology and syntax

The morphological level focuses on certain parts of speech with high frequency in both corpora, namely adjectives, verbs and interjections. The parts of speech with a grammatical function, such as determiners, prepositions and personal pronouns are commented on in the previous chapter dealing with Wordlists.

Adjectives have a restricted usage in the Facebook corpora. Higher frequency was found especially with the class of evaluative adjectives which were also mentioned in the Keyword lists section. Wordlists of the Facebook corpora and Spoken-BNC provided in Table 5 contain five most frequent adjectives from which *lovely*, *good*, *nice* and *great* all express positive evaluation.

Rank	Corpus 1	Corpus 2	Spoken-BNC
1	lovely	good	good
2	little	great	nice
3	nice	free	right
4	long	next	little
5	new	new	other

Table 5: The most frequent adjectives in the Facebook and Spoken-BNC corpora³¹

³¹ The frequency list of the adjectives of the spoken demographically sampled section of the BNC was retrieved from the BNCweb (available at <http://bncweb.lancs.ac.uk/> last accessed in December 2013), the lists for the Facebook corpora were obtained from Wordlists in AntConc. None of them are lemmatised.

Verbs are the second category to observe and the list of 25 most frequent verbs from the Facebook and Spoken-BNC corpora is provided in Table 6.³²

Rank	Corpus 1	Corpus 2	Spoken-BNC
1	be	be	be
2	have	will ('ll)	do
3	do	have	have
4	love	see	get
5	can	get	go
6	go	can	say
7	will ('ll)	do	know
8	get	go	will
9	would	would	can
10	know	come	think
11	see	hope	see
12	need	bring	come
13	think	meet	would
14	miss	know	want
15	come	like	mean
16	could	look	look
17	should	join	put
18	want	pay	take
19	make	buy	could
20	may	make	tell
21	say	want	make
22	feel	could	like
23	look	enjoy	give

³² The frequency list of the verb lemmata for the spoken demographically sampled section of the BNC was retrieved from the BNCweb (available at <http://bncweb.lancs.ac.uk/> last accessed in December 2013), the lists for the Facebook corpora were obtained from Wordlists in AntConc. Since the Facebook corpora are not lemmatised, the frequencies of the verb lemmata were obtained by adding up manually the frequencies of the individual word-forms of the verbs.

Rank	Corpus 1	Corpus 2	Spoken-BNC
24	hope	send	should
25	watch	wait	buy

Table 6: 25 most frequent verb lemmata in the Facebook corpora and in the Spoken-BNC

Three very frequent verbs *say*, *think* and *know*, all of which introduce *that*-clauses and also belong to the most common verbs in informal conversation (cf. Biber et al., 1999: 1049) will be analysed. In both Facebook corpora the lemma *say* is found in 62 hits, *think* in 67 hits and *know* in 74 hits. Although they rank high in both corpora, the majority of them are located in the corpus of personal comments. This imbalance can be connected to the fact that these verbs are very often bundled with the collocate personal pronoun *I*, expressing together a personal opinion or an attitude on various topics (see below examples 1-3).

- (1) nooo ;(I **think** that's our fault (Corpus 1)
- (2) I **know** right, i am pretty amazing! lol JK! But seriously so much to tell you when i next see you. I mean i **know** i talk alot but this is just crazy amount of stuff to share. lol It is too much xxxxxxxxxxxxxxxxxxxxxxxxxxx (Corpus 1)
- (3) That is not what I **said**! I **said** it was more entertaining than the film we were about to watch which was the 3 musketeers! (Corpus 1)

Table 7 (below) shows another verbal group with high frequency and these are dynamic verbs. The lemmas of four verbs describing a physical action *come*, *go*, *leave* and *meet* (present also in the two samples) have the following number of hits in both Facebook corpora and in Spoken-BNC:

action verbs	Corpus 1 in absolute numbers / per 10,000 words		Corpus 2 in absolute numbers / per 10,000 words		Spoken-BNC in absolute numbers / per 10,000 words	
come	45	26	37	5	14,522	34
go	94	53	17	23	34,367	81
leave	12	7	11	15	3,661	9
meet	9	5	22	29	443	1

Table 7: Selected action verbs (lemmas) in the Facebook corpora and Spoken-BNC

Several remarkable differences are visible by their comparison. As the usual topics of the societies based at universities are various activities, such as trips and socials, it was expected that the number of all four common action verbs would be higher. Meanwhile, only two of the four verbs *leave* and *meet* have higher occurrence in Corpus 2 than in Corpus 1. Two other verbs, *come* with the ratio 26:5 and *go* with the ratio 53:23 prevail in Corpus 1. The comparison with Spoken-BNC reveals the similarity of numbers with personal wall comments.

One of the problems with these verbs concerns the accuracy of the numbers. It is connected to the fact that these verbs have wider usage than only physical action. This was partially resolved by manual filtering by which other than non-actional uses of these verbs (such as being part of auxiliary or phrasal verbs) were excluded. However, this filtering was not possible in Spoken-BNC due to the extent of the data. Another problem posed the progressive *-ing* form of the verbs which is the same for converted deverbal nouns, such as *meeting*. Again, these instances were excluded from the numbers only in the Facebook corpora.

The most frequent modal verbs *can*, *could*, *should*, *would* and *will* from the Facebook corpora (cf. Table 6) can be analysed and compared to their distribution in Spoken-BNC. Four of them, which according to the description of informal conversation by Biber et al. (1999: 1049) belong to the most common ones, are presented in Table 8 below. Their use in the Facebook corpora is illustrated in examples 4-7.

modal verbs	Corpus 1 in absolute numbers / per 10,000 words		Corpus 2 in absolute numbers / per 10,000 words		Spoken-BNC in absolute numbers / per 10,000 words	
can	67	38	53	70	23,341	55
could	27	15	11	15	5,794	14
will	38	22	79	105	5,320	13
would	53	30	15	2	6,670	16

Table 8: Selected modal verbs in the Facebook corpora and Spoken-BNC

These numbers reflect both intrinsic and extrinsic use of modal verbs and no additional filtering was applied. The modal *can* has the highest frequency in Corpus 2, Spoken-BNC is in the middle and Corpus 1 has almost half size. A probable reason of this high number in group wall comments is its use for politeness. The modal verb *could* is almost equally distributed in all three corpora. *Will* has the lowest frequency in Spoken-BNC, followed by Corpus 1 and it dominates in Corpus 2. This asymmetry can be caused by the orientation of Corpus 2 to future and planning of upcoming events. Per contra, Corpus 2 rarely uses *would* which is more frequent in Spoken-BNC. It is also very popular with Corpus 1 where it is employed mostly to express unreal hypotheses.

(4) I am so excited. **Can** I wear a costume? lol just kidding. But I am excited to sing! (Corpus 2)

(5) OMG i miss erasmus stuff so much and i'm currently living in london so i was just wondering if I **could** join Abhi ?! and how much would it be without travel etc?! (Corpus 2)

(6) Alright gals I **will** see you there. (Corpus 2)

(7) Really? That **would** be annoying if you wanted to switch artists... a power play **would** be formed and you **would** have two 'highly regarded' artists slap dueling each other for your attention... :P (Corpus 1)

Another part of speech with high number of occurrences (also represented in the two samples) are interjections. After a manual search in the corpora it was found out that their usage is rather extensive. For instance, the interjection *haha* (and its longer versions, such *hahaha*) returned 75 hits, out of which only seven originate in Corpus 2. In most cases this

laughter-imitating expression is situated in the comment initial and final position, which is logical as it comes as a reaction on the content of the preceding text. Syntactically and semantically, interjections fulfil functions similar to emoticons, i.e. they are non-clausal expressions, mostly with evaluative meaning. Other interjections include *aww*, *yay*, *woohoo* (positive evaluation), *boo*, *gahh* (negative evaluation) and *ahhh*, *oh*, *hmmm* (context dependent meaning). Three examples of their usage in the Facebook corpora are listed below.

(8) **hahahahaha!** nobody really needs to be exposed to them though :P (Corpus 1)

(9) **Ahhhh** you have missed me! (Corpus 1)

(10) im pretty sure that wasnt the first time your housemates must've that you've gone mad.

boo!! (Corpus 2)

As to the syntax of the Facebook corpora, there is a tendency of non-clausal expressions (such as emoticons, interjections and acronyms) to gather in clusters, which is valid mainly for Corpus 1. Examples 11-14 show they are clause/comment-final and preceded by punctuation marks only if they signal a special function, as the exclamation in Example (14).

(11) I love you too :) <3 **xxx** (Corpus 1)

(12) Yeah, let me know when you're around again, I may have moved into mine and Dean's new flat by then :D **hehe xx** (Corpus 1)

(13) ahhh looking forward to Zumba! :D **lol xx** (Corpus 1)

(14) Get well soon Janey :) Please don't die!! <3 youuuuuu!! **xxx** (Corpus 1)

3.2.3.4 Lexicology

The word-formation is represented mainly by acronyms, clippings, initialisms and individual instances of nonce-formation.

A prototypical acronym for this type of CMC is *LOL* with 111 hits.³³ Only three hits are located in the corpus of group wall comments and it can be assumed that one of the reasons is a high degree of informality linked to this acronym. Its position within the comment structure is usually initial or final and it has a similar function as interjections and emoticons.

Apart from the extensive use of the acronym *LOL* and some others (*ASAP*, *OMG*, *BTW*, etc.), the corpora contain several instances of clipped or simplified expressions, frequently found in the language of text messaging, such as *msg* for ‘message’, *thot* for ‘thought’, *tomorro* for ‘tomorrow’ or *thx* for ‘thanks’. Using homophone expressions instead of letters, e.g. *2* for ‘to’ and *@* for ‘at’ was discovered as well. An example of nonce-formation (clipping and suffixation) is the noun *c-hillers* which represents a form of addressing the members of the society *Cafe on a Hill*.

The non-standard words, such as *gimme* for ‘give me’ or *kinda* for ‘kind of’ are inspired by connected speech typical for oral communication. All these innovations which are based on the relationship between visual and sound representation are specific to the genres of CMC and are not retrievable in face-to-face conversation. However, by their imitation of the features of speech, and highly informal character they approach it to a great extent.

Last but not least, informality is expressed also via a large group of phrasal verbs. For example, the expressions containing a verb and the adverbial particle *up* appeared 101 times, equally present in both Facebook corpora, e.g. *meet up*, *suit up* or *what's up*.

3.2.3.5 Discourse

Mean length of the comments is different for both corpora. Corpus 1 contains shorter one or two sentence comments which resemble Internet chat conversations. However, no significant utterance breaks (cf. Baron, 2010) are present here, i.e. the users do not split their messages as they are not under time or space pressure.

³³ In fact, although the standard version of this expression with the meaning ‘Laughing Out Loud’ uses capital letters, here it is produced in lower case in 89 hits.

The comments in Corpus 2, especially the conversation starters, usually consist of several sentences, ranging from 1 to 10. In most cases, the introductory comments are longer due to the amount of information they carry and the complexity of their communicative functions.

The whole sets of comments, i.e. the threads, also have a different number of comments in both corpora. The average thread in Corpus 1 contains four messages composed by three users (i.e. one user reacts twice). In some cases, the threads can become rather lengthy due to an exchange of turns between two people. In this situation the dialogue is formed by short comments resembling instant messaging.

Corpus 2 threads are usually longer, consisting of one main comment and several replies, on average there are five users contributing to one conversation. Many readers react non-linguistically by using the 'Like' button, thus expressing their agreement with the content of the comment.

4 CONCLUSION

The qualitative analysis of two samples representing Facebook conversation and the quantitative analysis of the two corpora supports the hypothesis that netspeak is a language variety displaying specific features on all linguistic levels.

On the level of graphetics and graphology there seem to be two features which represent this type of CMC. The first one is expressed via the preference of lower-cased sentence starters, typing errors and loose or missing punctuation. It can be explained by the economy of effort and time saving typical for quick exchanges in synchronous CMC genres, such as instant messaging and chat groups, where the lack of editing resembles the production of speech in informal conversation. The second one is demonstrated by the cases of nonstandard spelling, upper-cased emphasised words, reduplication of letters and exaggeration of punctuation marks. These elements can be perceived as tools substituting the prosodic features of face-to-face conversation, such as intonation and stress and as an expression of emphasis and of personal creativity.

One of the most dominant features of this type of CMC are emoticons. In spite of the fact that their repertoire varies to almost an unlimited number, there are some more conventionalised types which are recognisable to most Internet users. The analysis showed that they do not always operate on the principle of iconicity. The lack of a direct link between the form and the meaning does not hinder understanding and is compensated for by the shared knowledge of the communicants. They function as markers of emotions and attitude. Moreover, they are important indicators of the informal tone of the messages providing a relaxed atmosphere.

Morphology of the collection of Facebook texts is characterised by the specific representation of individual parts of speech. A large group is formed by interjections which resemble the prosodic features and add vivacity to the conversation. The presence of the parts of speech such as selected types of personal and possessive pronouns and evaluative adjectives is in accordance with their distribution in spoken conversation. The frequency of some verbal groups, such as *say-think-know* and of some dynamic and modal verbs is similar to their occurrence in informal conversation. Overall limited grammatical elaboration, repetitive nature and low diversity of nouns are typical of this kind of CMC.

Unlike morphology, syntax of Facebook texts displays features of both written and spoken genres. On the one hand, there are elements which match the characteristics of face-to-face conversation, e.g. elliptical and incomplete sentences. Frequent non-clausal elements (acronyms, emoticons and interjections) have a fixed position within the structure of the comments and clauses, occurring usually initially or finally. In some cases these grammatical isolates cluster together. Their function is to express attitudes, emotions and opinions. The brevity of the majority of sentences and the preference of compound over complex sentences also concur with spoken communication.

On the other hand, the syntactic development of the group wall comments (Corpus 2) which contain event invitations resembles the traditional written genres. The presentation of the topic of these comments is carefully preplanned and contains structural elements typical for formal writing. The presence of complex sentences and cohesive expressions confirms a higher degree of formality although it is mixed with some highly oral features (cf. Sample 2). This heterogeneous style reflects the presence of several communicative functions of the invitation where both the need to convey the factual information and to appeal to the readers are expressed.

Lexical choice in this CMC is dominated by informal expressions, such as colloquialisms, slang words and phrasal verbs. Hedging and politeness occur mainly in group wall comments due to more anonymous nature of the group and more formal relationships. Word-formation is focused on clipping, blending and acronyms and there are some syntactic or semantic changes for certain verbs. Innovative approach is reflected via word play with visual and phonetic representation.

The discourse level shows some notable differences in the length of comments and whole threads. First, it was found out that mean length of the comments in the corpus of personal wall is shorter than the one of the corpus of group wall comments. The messages containing event invitations not only have longer sentences but there are also more of them in one thread. The disproportion between more lengthy comments and shorter reactions can be explained by the amount of information they provide, i.e. an invitation contains more information and more communicative functions than a reaction on it.

There are some signals that gender influences the distribution of emoticons. The samples indicate that male users tend to employ fewer emoticons than female users.

However, this phenomenon is observed only in the two samples. Therefore, this question remains open for future research.

The concordance freeware AntConc used for the quantitative analysis provided some standard tools, such as generating Wordlists and Keyword lists and statistical data for the corpora. Although one of minor drawbacks of AntConc consists in its inability to reliably locate some nonlinguistic expressions, such as emoticons (e.g. their total frequency) and to display them in their original graphic form, it was possible to work with their context.

The linguistic analysis of the samples and of the whole corpus brings mixed results. The corpus of personal wall comments seems to incline to face-to-face conversation by the resemblance to some of its prosodic features, higher degree of informality and by the length of the comments. The corpus of group wall comments contains more formal and structurally rich comments resembling written genres. However, the features of oral communication are present here as well.

Finally, the results of the analysis confirm that English used by British university students on Facebook is a hybrid genre consisting of the features of both writing and speaking and of its own unique elements. The inclination towards oral features indicates that this asynchronous type of CMC realises a function similar to a face-to-face conversation and therefore provides its alternative.

5 RÉSUMÉ

Cílem této diplomové práce je podat charakteristiku jazykových specifik internetové komunikace v sociální síti Facebook se zaměřením na jazyk anglicky mluvících studentů britských univerzit.

Teoretická část začíná historií počítačem zprostředkované komunikace. Následuje vysvětlení základních pojmů internetové lingvistiky, jako je *internetový jazyk*, který definuje jeden ze zakladatelů internetové lingvistiky David Crystal pomocí synonymních výrazů, například *netspeak*, *weblish*, *elektronický diskurz* nebo zmíněná *počítačem zprostředkovaná komunikace* (anglický termín *computer-mediated communication*, ve zkratce CMC). Také se zde definuje termín *jazyková varieta*, jejíž definice je úzce spjata s tradiční jazykovou dichotomií mezi psanou a mluvenou formou. Tyto dvě formy je možné popsat pomocí jednotlivých jazykových úrovní, jmenovitě fonetické a fonologické, grafické a grafologické, gramatické, lexikální a diskurzí (srov. Davy & Crystal, 1992). Nejpoužívanějšími formami CMC jsou e-mail, chat, elektronické mailingové skupiny, diskusní fóra, rychlé zprávy (instant messaging) a sociální sítě. Tyto formy se dělí na synchronní a asynchronní, přičemž hlavním kritériem je přítomnost obou účastníků, která ovlivňuje rychlost konverzace. Srovnání charakteristických vlastností psané a mluvené formy s internetovým jazykem naznačuje, že synchronní typy internetové komunikace mají blíže k mluvenému jazyku a asynchronní k psané formě. Následuje charakteristika jazykových prvků přítomných v elektronické komunikaci pomocí zmíněných jazykových úrovní a také popis vlastností typických pro CMC (např. emotikony, akronyma a nestandardní a inovativní pravopis).

Zajímavý lingvistický pohled na elektronickou komunikaci nabízí koncept, který hodnotí CMC jako kontinuum (Kalman and Rafaeli, 2007). K podobným závěrům dochází i Baron (2009a), která vnímá CMC jako smíšenou modalitu, ale potvrzuje, že CMC není pouze směs psané a mluvené formy. Třetí pohled poskytuje Crystal, který se domnívá, že tento druh komunikace inklinuje k psanému projevu, i když existuje jako samostatná forma.

Poslední kapitola teoretické části práce se věnuje historii, definici a funkcím sociálních sítí, zejména Facebooku. Jeho systematický lingvistický popis zatím neexistuje, jedním z důvodů je novost média i jeho neustálé proměny a vývoj. Očekává se, že s větším časovým

odstupem bude přibývat jazykově zaměřených výzkumů Facebooku a ostatních sociálních sítí.

Praktická část práce začíná kapitolou o metodologii provedeného výzkumu. Nejdříve je vymezena hlavní hypotéza, která předpokládá, že netspeak jako jeden ze současných proudů vývoje angličtiny, funguje jako samostatná jazyková varieta. Toto tvrzení má být ověřeno na základě korpusu vybraných Facebookových textů obsahujících komunikaci studentů několika britských univerzit (jde o rodilé mluvčí angličtiny). Také se očekává, že výsledky potvrdí podobné rysy této variety a mluveného jazyka, konkrétně neformální konverzace. Proto analýza obsahuje i srovnání s prvky mluveného a psaného projevu. Samostatnou část metodologické kapitoly tvoří popis komunikačních nástrojů Facebooku, který osvětluje podmínky, za kterých vznikají jednotlivé jednotky komunikace.

Sbírka textů získaných z Facebooku byla rozdělena na dva subkorpusy. První pochází z osobní komunikace na Facebookové zdi vybrané studentky. Jde o neformální konverzace, a proto se očekává, že bude jejich styl podobný rozhovoru. Druhý korpus byl shromážděn z Facebookové zdi dvou zájmových univerzitních skupin (*societies*).

Za základní jednotku analýzy byla, podobně jako u konverzace, stanovena tzv. *turn construction unit* (TCU, pojem zavedený Sacksem, Schegloffem & Jeffersonem, 1974), tedy jedna ucelená promluva účastníka konverzace. V případě Facebooku nemluvíme o *turns*, ale o *příspěvcích* (*comments*).

Facebookové příspěvky obou subkorpusů jsou porovnány s popisem neformální konverzace z *Longman grammar of spoken and written English* (Biber a kol., 1999). Zvolený přístup rozčleňuje TCU na dva základní typy, tzv. větné a nevětné (*clausal* a *non-clausal*) jednotky. Nevětné části TCU obsahují například vsuvky a dále se člení na předmluvy (*prefaces*), dovětky se jmennými frázemi (*noun phrase tags*) a nezačleněné závislé věty (*unembedded dependent clauses*).

Analýza se dále dělí na dvě podsekcce obsahující kvalitativní rozbor dvou vybraných vzorků a kvantitativní rozbor, zaměřený na celý Facebookový korpus. Pro tyto účely byl zvolen konkordanční program AntConc, s jehož pomocí bylo možné analyzovat jednak data získaná z Facebooku, ale také data Britského národního korpusu obsahující texty mluvené angličtiny (demograficky diferencované), která byla využita pro porovnání jako referenční korpus.

Kvalitativní rozbor obsahuje malý vzorek osobní komunikace na Facebookové zdi vybrané studentky a vzorek pocházející z konverzace na zdi univerzitní společnosti *Erasmus society*, která je víc anonymní. Tady se ukazuje, že osobní komunikace obsahuje prvky neformální konverzace a skupinová komunikace je strukturně bohatší a tíhne k tradičním psaným žánrům.

Další kapitolu tvoří kvantitativní analýza Facebookových příspěvků, která si všímá vybraných jevů zjištěných v kvalitativním rozboru a zároveň využívá některé nástroje korpusové lingvistiky. Nejdříve jsou uvedeny statistické informace, jako např. rozsah dat, poté se sledují nejčastěji se vyskytující slova podle frekvenčního seznamu slov vygenerovaného pro všechny srovnávané korpusy. Další funkcí korpusového programu je vygenerování seznamu klíčových slov (*Keyword list*) za použití mluvené složky BNC jako referenčního korpusu. Výrazy, které se objevily jako příznačné pro Facebookový korpus osobní komunikace, byly například emotikony, akronyma a termíny vyjadřující pozitivní hodnocení a emoce. Projevila se zde menší nevýhoda BNC textů, která se týká aktuálnosti dat. Ta totiž pocházejí ze začátku 90. let 20. století, a tak se v seznamu klíčových slov objevují termíny, které popisují koncepty neexistující před 20 lety, například *zumba* nebo *facebook*. Porovnání korpusu Facebookové komunikace univerzitních spolků na pozadí mluvené složky BNC ukázalo, že v klíčových slovech převládají slova popisující obsah. Emotikony zde nejsou tak početné, důraz se klade na dynamiku celé skupiny, což se projevuje ve zvýšeném výskytu slov, jako jsou *guys* nebo *everyone*. Také se zde najde víc časově-prostorových výrazů.

Korpusové srovnání je uskutečněno na různých jazykových úrovních. Grafická a grafologická oblast zaznamenaly vysoký výskyt příspěvků s nestandardním pravopisem (např. věty začínající malým písmenem, pravopisné chyby a netradiční nebo chybějící interpunkce). Jedním z hlavních důvodů se jeví princip nejmenšího úsilí. Dále lze v elektronickém textu pozorovat záměrné využívání nestandardního pravopisu (slov s velkými písmeny, zdvojování a násobení písmen a interpunkce), tedy jevů, které pravděpodobně nahrazují prozodické vlastnosti běžné mluvené konverzace. Jednou z dominant CMC jsou emotikony, které mají takřka neomezený repertoár. Základní typy jsou natolik známé, že i zastřená ikoničnost často nebrání porozumění jejich významu. Jejich funkce se různí a působí také jako prvky, které snižují formálnost příspěvků a vnášejí do textu hravost.

Na morfosyntaktické úrovni jsou popsány nejfrekventovanější slovní druhy, jakými jsou přídavná jména, osobní a přivlastňovací zájmena a citoslovce. Velký prostor je poskytnut slovesům, analyzují se např. výskyty čtyř sloves pohybu *come, go, leave a meet* a čtyř modálních sloves *can, could, will a would*, které také potvrzují silnou tendenci jazyka na Facebooku k neformálnímu mluvenému projevu. Jelikož korpus nebyl morfologicky označován, mohou být některé tyto výsledky mírně zkreslené. Konkordanční řádky byly proto analyzovány ještě manuálně. Celkově byla pozorována menší míra gramatické komplexnosti, tendence k jednotvárnosti a nízký výskyt obecných podstatných jmen. Na druhou stranu je větná skladba Facebookových příspěvků heterogennější a vykazuje prvky jak psané, tak ústní formy. Zde je také patrný příklon korpusu skupinové komunikace k formálnějšímu stylu typickému pro psané žánry.

Lexikálně u netspeaku převládají neformální výrazy, kolokvialismy, slangová slova a frázová slovesa. Tzv. *hedging* a zdvořilostní prvky se vyskytují zejména ve sbírce textů univerzitních skupin, což může být podmíněno větší anonymitou jejich členů. Slovtvorba je zaměřena na krácení slov a jejich spojování (*clipping a blending*) a akronyma. V několika případech byla pozorována i syntaktická, případně sémantická změna u sloves. Vynalézavost a snaha o originalitu se projevuje ve slovních hříčkách.

Závěry kvalitativně-kvantitativní analýzy dat dvou různých Facebookových korpusů na všech jazykových úrovních ukazují, že netspeak existuje skutečně jako samostatná jazyková varieta. Je možné konstatovat, že tato varieta elektronické komunikace je vskutku třetím médiem, které není pouhým agregátem různých prvků psaného a mluveného projevu, ale obsahuje i jedinečné prvky nevyskytující se ve dvou jmenovaných tradičních formách. Také se potvrdila vedlejší hypotéza, že tento asynchronní typ elektronické komunikace inklinuje některými svými vlastnostmi k tradiční neformální konverzaci, a tak představuje její alternativu v elektronickém prostředí.

6 REFERENCES

- Baron, N. S. (2003) Language of the Internet. IN Ali Farghali, ed. *The Stanford Handbook for Language Engineers*. Stanford: CSLI Publications, 59-127.
- Baron, N. S. (2008) *Always on: language in an online and mobile world*. Oxford: Oxford University Press.
- Baron, N. S. (2009a) Instant Messaging by American College Students. American University Washington, DC Presented in Session 301 'Language on the Internet' American Association for the Advancement of Science Annual Meeting February 2005.
- Baron, N. S. (2009b) The Myth of Impoverished Signal: Dispelling the Spoken Language Fallacy for Emoticons in Online Communication. IN: Jane Vincent and Leopoldina Fortunati, eds., *Emotion and ICTs*. London: Peter Lang.
- Baron, N. S. (2010) Discourse structures in Instant Messaging: The case of utterance breaks. *Language@Internet*, 7, article 4. <languageatinternet.org > Accessed on September 30, 2013.
- Biber, D. et al. (1999) *Longman grammar of spoken and written English*. Harlow, England: Longman, 1037-1127.
- Boyd, D. M. and N. B. Ellison (2007) Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), article 11. <<http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html> > Accessed on September 30, 2013.
- Crystal, D. (2001) *Language and the Internet*. Cambridge: Cambridge University Press.
- Crystal, D. and D. Davy (1992) *Investigating English Style*. London: Longman.
- Dresner, E. and S. C. Herring (2010) Functions of the nonverbal in CMC: Emoticons and illocutionary force. *Communication Theory*, 20, 249-268.
- Ellison, N. B., C. Steinfield and C. Lampe (2007) The benefits of Facebook 'friends': Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication* 12 (4), article 1.
- Fortunati, L. (2005) Is Body-to-body communication still the prototype? *The Information Society* 21(1), 1-9 15.

- Herring, S. C. (1999) Interactional coherence in CMC. *Journal of Computer-Mediated Communication*, 4 (4). <<http://jcmc.indiana.edu/vol4/issue4/herring.html> > Accessed on September 30, 2013.
- Herring, S. C. (2001) Computer-mediated discourse. In Schiffrin, D., D. Tannen and H. Hamilton (Eds.), *The Handbook of Discourse Analysis* (612-634). Oxford: Blackwell Publishers. <<http://ella.slis.indiana.edu/~herring/cmd.pdf> > Accessed on September 30, 2013.
- Herring, S. C. (2010) Computer-mediated conversation: Introduction and overview. *Language@Internet*, 7, article 2. <languageatinternet.org > Accessed on September 30, 2013.
- Herring, S. C. (2012) Grammar and electronic communication. In C. Chapelle (Ed.), *Encyclopedia of applied linguistics*. Hoboken, NJ: Wiley-Blackwell. Preprint: <<http://ella.slis.indiana.edu/~herring/e-grammar.pdf> > Accessed on September 30, 2013.
- Cho, T. (2010) Linguistic Features of Electronic Mail in the Workplace: A Comparison with Memoranda. *Language@Internet*, 7, article 3. <languageatinternet.org > Accessed on September 30, 2013.
- Kalman, Y. M. and S. Rafaeli (2007) Modulating synchronicity in computer mediated communication. Paper presented at the ICA, San Francisco, CA.
- Kapidzic, S. (2010) Non-standard features of English used in teen chatrooms. Master's thesis, English Department, University of Sarajevo.
- Paolilo, J.C. (2011) 'Conversational' codeswitching on Usenet and Internet Relay Chat, *Language@Internet*, 8, article 3. <languageatinternet.org > Accessed on September 30, 2013.
- McArthur, T. and F. McArthur (1992) *The Oxford Companion to the English Language*. Oxford: Oxford University Press.
- Pérez-Sabater, C. (2012) The Linguistics of Social Networking: A Study of Writing Conventions on Facebook. <http://www.linguistik-online.de/56_12/perez-sabater.html > Accessed on September 30, 2013.
- Randall, N. (2002) Lingo online: A report on the language of the keyboard generation. Department of English, University of Waterloo and msn.ca. <<http://www.arts.uwaterloo.ca/~nrandall/LingoOnline-finalreport.pdf> > Accessed on September 30, 2013.

Sacks H., E.A. Schegloff and G. Jefferson (1974) A simplest systematics for the organization of turn-taking for conversation. IN: *Language*, 50, 696–735.

Tannen, D. (1982) *Spoken and written language: exploring orality and literacy*. Norwood, N.J.: ABLEX Pub. Corp.

SOURCES:

Facebook (2013) Personal and group wall comments. <<http://www.facebook.com>> Accessed in 2013.

(Corpus data available on CD at the Department of English Language and ELT Methodology, Faculty of Arts, Charles University in Prague)

British National Corpus (2007) Distributed by Oxford University Computing Services on behalf of the BNC Consortium. <<http://www.natcorp.ox.ac.uk> > Accessed in 2013.