Creation of Internet Relay Chat Nicknames and Their Usage in English Chatroom Discourse

Robert Ecker (Linz)

Abstract

In a chat discourse it is not always clear who is chatting with whom; automatic discourse analysis is especially problematic. It is important to identify the users' nicknames in the written discourse to find out the receivers of the chat messages. But the linguistic possibilities in nickname creation, and also of using them in the discourse, are various. To study how nicknames are created and used in the Internet Relay Chat (IRC), logs of 13 channels consisting of 8937 public chat messages and 7936 unique nicknames were analyzed in detail. The paper shows, among other things, the basic structure of IRC nicknames, of which parts-of-speech group nicknames are compounded, and which parts-of-speech of a nickname are omitted within the chat discourse. This knowledge leads to a better prediction as to whether there is a link between a current logged-in user and the examined word in discourse, which can be a shortened or creatively changed form of a nickname.

1   Introduction

Nicknames have been used since the Middle Ages and today in a computing context the word nickname is omnipresent, especially in computer-mediated communication. People use nicknames (also known as nicks) to identify themselves, e.g., in chat rooms (also known as channels), bulletin boards or social networks on platforms such as Facebook and Twitter. Nicknames play a special role in chat discourses for direct addressing; the way in which people address one another. A nick acts as a marker in the chat discourse, comparable with Sack's concept of "speaker select" (Bays 1998; Kortti 1999; Nash 2005). Basically, chats allow many-to-many conversations; at times, they contain sequences of one-to-many or one-to-one interactions. To prevent misunderstandings regarding the addressing of a message, the nick of the receiver is frequently put in front of the message, followed by a colon and a space. This is one of the basic (written and unwritten) rules of online communication; it is called netiquette. Direct addressing, also known as "addressivity" (Werry 1996) or "cross-turn reference" (Herring 1999), opens up the possibility of taking part in more conversation at the same time for the participants. But explicit direct addressing is not always used or required, for instance, for addressing a message to everybody in a channel (Mutton 2004a). Most IRC clients provide automatic text highlighting including manually set nicknames and variants. It helps us to know, in channels with a lot of traffic, who is talking to us.

The following work focuses on the IRC, originally written in 1988 by Oikarinen/Reed. IRC is one of the most frequently used chat systems in the world. It is a multi-user, multi-server and multi-channel text-based chat system for near real-time communication. There are several different independent IRC networks (e.g., QuakeNet, IRCnet, Undernet, EFnet). Each IRC network consists of a certain number of servers, which communicate over a well-defined open protocol. It was first formally documented in 1993 by RFC 1459, with revisions in RFC 2810,

RFC 2811, RFC 2812 and RFC 2813. IRC uses the Transmission Control Protocol (TCP) and optionally the Transport Layer Security (TLS). Before using IRC, a user must choose a nickname. Previous nicknames can be easily changed at any time by the IRC command "/nick". Thus, impersonating someone or stealing a nick just for fun are quite simple (Mutton 2004b).

For automatic discourse analysis of chat transcripts it is important to identify nicknames in the written chat messages to know who is chatting with whom. Below a possible chat discourse extract is illustrated:

> <Limbic_Region> Hi all
> <Goblin> hello Limbic_Region
> <RobiX> limbic: hi

The nicks of the speakers are surrounded by angle brackets, followed by a written user message (as they appear on IRC). This extract points out the following problems: First, as mentioned above, direct addressing is not always used or required. No nick occurs in the message "Hi all", although according to Rintel/Mulholland/Pittam (2001) "openings are an excellent starting point for investigating how interaction on IRC functions to instantiate and develop interpersonal relationships." Second, <Goblin> does not comply with the rules of netiquette, because the receiver <Limbic_Region> is not put in front of the message and followed by a colon. A word-by-word comparison between user list and each word of the message is necessary in order to find the receiver's nickname. Third, the shortened variant "limbic" does not correspond exactly to one of the nicks in the existing user list (<Limbic_Region>, <Goblin> and <RobiX>).

The aim of this paper is to analyze the relationship between IRC nicks, which are used by the current logged-in chatters, and the identified shortened or creatively changed forms of nicks from written messages for later automatic discourse analysis. The following contribution addresses some questions as follows:

- **Creation of IRC nicks:** What is the basic structure of IRC nicks? Which parts-of-speech (POS) do nicks consist of in detail? In which order are POS concatenated to a compounded nick? How and with which characters are they concatenated? Are there any special cases or features, which occur in the creation of nicks?
- **Usage of IRC nicks in English chatroom discourse:** How are nicknames used to address users in a chat discourse? How are nicks exactly written in discourse and which parts of them are omitted?

This paper is structured as follows: Section 2 gives an overview of related work. The research approach is described in Section 3. In the next section, the results of the linguistic analysis of IRC nicknames are presented. The results of the discourse analysis are described in Section 5. Finally, Section 6 concludes the paper.

2   Related Work

Nicknames can be viewed from various aspects like psychology, sociology or linguistics (Morgan/O'Neill/Harré 1979; Reid 1991; Bechar-Israeli 1995). Bechar-Israeli (1995) defines a nick as "a name we receive in addition to our legal name [which is] usually given to us by the people surrounding us." The term nickname "was established as a variant of the Middle English noun eke-name." (Lakaw 2006). It should be mentioned that not only human participants are hidden behind nicks, but also computer programs called chatbots (Döring 2003).

In this section, approaches to classifications of chat nicknames and analyses of chat communication are presented.

2.1   Focus on Typologies of Chat Nicknames

Basically, nicknames are proper nouns (Anderson 2007; Stommel 2007). Some further classifications of nicknames can be found, but these works focus mainly on a semantic point of view. Although Lakaw (2006) points out that the topic of a channel has an impact on the creation of IRC nicks, the following studies deal with a small number of investigated nicknames or channels.

One of the pioneering studies regarding IRC nicknames was done by Bechar-Israeli (1995). 260 nicknames from four different IRC channels were analyzed from the perspective of a content analysis. The focus of this semantic topology is about the origin of these nicks. The author identifies seven main categories: 1) people using their real name, 2) self-related names, 3) names related to medium, technology and their nature, 4) names of flora, fauna or objects, 5) play on words and sounds, 6) names related to figures in literature, films, fairytales and famous people, and 7) names related to sex and provocation.

Another typology is given by Johnová (2004). She analyzed nicks on a British chat site with four chat forums and 12 chat rooms. The author mentions that a nick can be, for instance, a single word, a whole sentence, a combination of lower- and upper-case letters; it can include numbers, non-alphabetical symbols or smileys. She asserts that it is "difficult to predict which part of the nickname will be retained and which part will be dropped" to shorten the nicks. And she adds, "Often several variations can be used … with each user choosing their own variant." Additionally, Johnová divides the most frequent types of nicknames into several categories: 1) legitimate names of the user (and their variants), 2) short characterization of the user (can include user's age, sex, location, physical or character description), 3) names of famous people or characters, and 4) animals, flowers or objects. A more detailed linguistic study about the usage of nicknames in discourse, for instance, which parts of a nickname are omitted, is missing.

This present study is also related to Stommel's approach (2007). She collected 83 nicknames from a German forum. The analysis distinguishes between names (proper nouns) and nouns (common nouns), which are sub-divided into six word types in nicknames: 1) commonly known names, 2) novel formations, 3) nouns and noun phrases, 4) adjectives, 5) verb forms, and 6) exclamations. However, the major differences to the present study are that Stommel uses a German corpus, and chunk tags (i.e. phrases) for nickname tagging. Additionally, compounded nicks (e.g., <Estrella1981>) are not further decompounded into single parts-of-speech (proper noun "Estrella" and cardinal number "1981").

2.2   Focus on Chat Communication

A comparison of discourse between IRC and spoken English or face-to-face communication is described in Hentschel (1998) and Kortti (1999). Further different linguistic perspectives in the IRC discourse are analyzed in Doell (1998), Stevenson (2000), and Blakeman (2004). Davis/Brewer (1997) commented that computer-mediated communication "has many characteristics of both speech and writing." In general, in comparison with oral speech, typing speed in text-based chats is slow (Dewitte/Hendricks 2005). In order to guarantee a fluid written chat-communication, the participants must react and act quite quickly (Rüggenberg 2007). Additionally, slow response time makes discourse partners bored. They forget or lose interest before getting an answer.

Key strategies to increase typing speed, "namely to save time to keep up with the speed of conversation, ..." or to reduce the number of keystrokes have been developed. This can be seen in Hård af Segerstad (2002) by, for instance, "features related to space, punctuation, spelling and case constitute strategies ...", "common grammatical features", expressions like

abbreviations, or smileys. A disadvantage is the prevalence of orthographic errors due to fast typing (Ringlstetter/Schulz/Mihov 2006; Tavosanis 2007). Words and phrases can help to stand out from the online crowd. Sun (2010) "pointed out that besides some common word-forming methods like derivation and compounding, abbreviation, blending are the typical methods to form Internet words." Leetspeak is another possibility to create new words. It is a form of Internet slang (also known as netspeak or chatspeak). In leetspeak, or leet for short, letters may be replaced by similar letters, numbers or special signs. For example, "leet" would become "1337" (Perea/Duñabeitia/Carreiras 2008). Döring (2003) distinguishes three main functions for using Internet slang: time-saving economy function, identity function and interpretation function. The question that arises here is whether and how the Internet language has an influence on the nickname creation or spelling of a written nick in discourse.

3   Research Approach

This work addresses the important general issue of shortened or creatively changed forms of nicks within chat discourse which cannot be linked to the original nickname. But these links are important for automatic discourse analysis to know and to understand who is chatting with whom. Therefore, it is necessary to find out how nicks are created and how they are used in discourse. This study examined logs of IRC interactions using discourse analysis, which combined qualitative perspectives (to find similarities and differences) and quantitative perspectives (to create statistics). An overview of the research approach is given in Figure 1.
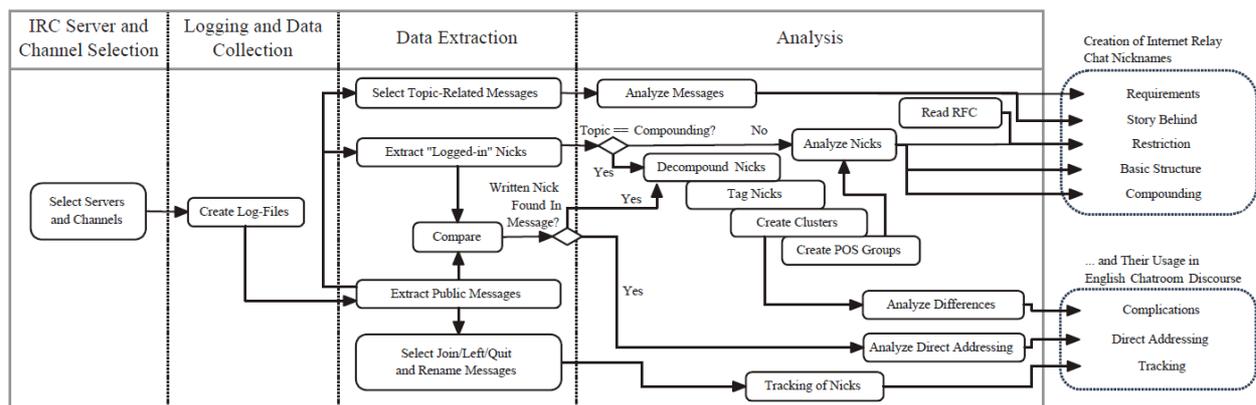
Figure 1: Overview of the Research Approach

3.1   IRC Server and Channel Selection

Gelhausen (2008) divided frequently searched chat terms into nine categories. Therefore, at least one channel per category was selected to find as many different nicknames as possible, because the topic of a channel influences the choice of nicknames (Lakaw 2006). A total of 7 different IRC networks and 13 different public IRC channels were used (see Table 1 and 2).

Category	Network	Channel(s)
cars	GameSurge	#cars
celebrities	QuakeNet	#eminem
countries, languages, cities	IrCQ-Net	#English
	QuakeNet	#england
conversation	freenode	#defocus, #freenode
	DALnet	#chat-world
	SwiftIRC	#talk
games, sports	EFnet	#soccer
love, relationship	IrCQ-Net	#Romance
music	QuakeNet	#eminem
religion	DALnet	#church
technology, Internet	freenode	##hardware, #perl

Table 1: Selected Channels for the 9 Categories

Network	Server Address
DALnet	jade.va.us.dal.net
EFnet	irc.servercentral.net
freenode	irc.freenode.net
GameSurge	irc.gamesurge.net
IrCQ-Net	irc.icq.com
QuakeNet	irc.quakenet.org
SwiftIRC	irc.swiftirc.net

Table 2: Used IRC Networks

3.2 Logging and Data Collection

Most public conversations of our selected channels were logged from June 27, 2008 to July 28, 2008. Because of logging problems, the logging of channel #eminem started on June 28. The log function of the Java program LogBot[1] was used, which remained silent and merely observed the chat. mIRC[2], a popular IRC client, was used to communicate with the participants. The user's replies about the formation of their nicknames were quite helpful. Each chat log from a channel was stored in a single file per day. Private messages (whispering) which are only seen by a particular recipient were not logged. The main language spoken in all logged channels was English.

3.3 Data Extraction

Channel	Public Messages Qual.	Quant.	System Messages Qual.	Quant.	User Messages Qual.	Quant.	Nicknames Qual.	Quant.	
#cars	9658		1398		8260		82	19	(25 - 6)
#chat-world	313216		175707		137509		39079	2041	(3038 - 997)
#church	10327		8821		1506		841	75	(101 - 26)
#defocus	136454	5389	35441	1365	101013	4024	4118	330	(516 - 186)
#eminem	2346		1380		966		96	5	(5 - 0)
#england	49139		11307		37832		1275	16	(30 - 14)
#English	700170		207826		492344		49547	1927	(2244 - 317)
#freenode	88670	3548	38625	1593	50045	1955	6744	394	(630 - 236)
##hardware	76988		16125		60863		1687	133	(220 - 87)
#perl	139559		37768		101791		3909	339	(631 - 292)
#Romance	683995		233328		450667		52201	2341	(2649 - 308)
#soccer	20949		6265		14684		451	59	(92 - 33)
#talk	172306		29795		142511		4032	257	(332 - 75)
Σ	2403777	8937	803786	2958	1599991	5979	164062	7936	(10513 - 2577)

Table 3: Logging Statistics

Table 3 shows the summaries of public messages (all logged messages; summary of system and user messages), system messages (reported by the IRC server), user messages (written by users), and nicknames of the logged-in users, which were found in the data. In summary 2403777 public logged messages (803786 by system, 1599991 messages written by users) were analyzed. For quantity analysis the log-files of two public channels (#defocus, #freenode) from July 1, 2008 were used; in summary 8937 (2958 by system, 5979 by users). Approximately one third of all public messages were system messages. 164062 nicknames (unique 150278, case-insensitive 141898) were extracted from the join/leave/quit messages, the nick-change messages, and senders of the messages. Shortened nicknames and variants were extracted by reading the log-files to make a comparison between the used nicknames in the chat discourse and the logged-in users. For quantity analysis the 13 log-files from July 1, 2008 were used (10513 nicks in summary; 7936 analyzed in detail, 7420 unique, 7326 case-insensitive). 2577 nicks (24.51%) remained unspecified, because of no user's feedback.

---

[1] http://www.jibble.org/logbot/, accessed November 2, 2011.
[2] http://www.mirc.com/, accessed November 2, 2011.

3.4 Analysis

The data were analyzed "by hand" and with some helping tools (see page 16). The steps "decompound nicks", "tag nicks", and "create clusters" (Fig. 1) are important to recognize which parts of a nickname are omitted later in the chat discourse. Thus, state-of-the-art techniques of natural language processing like POS tagging and n-grams were applied. Additionally, the step "create POS groups" allows a simpler classification of nicknames. The analyses and results are described in the next two sections.

4 Creation of IRC Nicknames

Nicknames provide a way of distinguishing between chat users. For that reason the significance of unique nicks is important. Decision making can be self-determined (e.g., requirements for the nick), other-determined (e.g., restrictions by the IRC network), or both (e.g., story behind the nick). The inventive way of nickname creation will be focused on in the following section.

4.1 Requirements for a "Perfect" Nickname

In a computing context, a nickname "creates the first impression of a user and is therefore the first condition for successful communication" (Johnová 2004). It "must be chosen with care and easy to use both for the speaker and for the listener." (Bechar-Israeli 1995). Nicknames can provide users with anonymity and a touch of freedom to obtain a new identity (Reid 1991; Gelléri 1998; Johnová 2004). The art of nickname creation lies in the picking of an attractive unique nick. The attributes "sounds good" and "short" were mentioned during the chat discourse analysis. But nicks, which look like other frequently used terms (e.g., common words, Internet slang) in the discourse, can easily be confused with them. An extract is presented below.

Log Number 1: Channel #Romance
01   *** lol has joined #Romance[3]
02   <^Zoe> he probably has cheese breath
03   <Ky|e> lol oh

In the above discourse the nick <lol> (line 1) can be confused with the common element of the Internet slang "LOL" (line 3), which means "laugh (or laughing) out loud". Similar nicks were <rofl> ("rolling on the floor laughing"), <roflmao> ("rolling on the floor laughing my ass off"), or <asl> ("age/sex/location").

4.2 The Story Behind the Chosen Nickname

When a chat nickname is chosen there is always some thought behind it (even if a nick consists of common words or is a random sequence of letters). In the best case, close relationships between users and their chosen chat identity are built up. There can be a great, interesting story behind the nickname. The following are typical of the kind of questions that can be asked:

- What does your nick mean?
- Why did you choose it?
- What is its history, background or the story behind it?

The author would like to thank the many participants on IRC who shared with him the colorful story of their nicks. Some extracts of the private messages are given below:

---

[3] Typically three asterisks at the start of the line with no punctuation around the user's nick are automatically generated system messages. They report when users join/leave/quit the channel, or change their nicknames.

Log Number 2: Private Messages

01   <ahf> it's my initials for my name
02   <lnf_> LiNux Fan
03   <nuba> Networked Ultimate Battle Android
04   <jenova> jenova is a non-playable character from final fantasy 7
05   <MrSkitZo> i have 2 personalities, one drunk one sober :)
06   <MrSkitZo> SkitZo aka Schichofrenic
07   <JermSnap> my name is jeremy, in the mountains, when I was a pro snowboarder, you get ONE
                sylable,so you can shout peoples names easier,so people called me JERM ..then I started
                breaking 1-2 snowboards a day so the guys at the factory started calling me JermSnap...
                there ya have it.
08   <Dave-O> It's my personal version of the nicjk nanme "steve-O" from jackass
09   <NightKhaos> Honestly, Chaos was likely to be taken everywhere I went, by changing it to a K i had
                a higher change of actually getting the nickname rather than NightChaos2,481,092
10   <Adys> was my first charname in wow
11   <Adys> which i typed randomly on the keyboard and made readable
12   <Rtkwe> on a qwerty keyboard rtkwe is tyler shifted one key to the left
13   <kunwon1> it's an anagram for 'unknown' with a letter missing. More importantly, it's unique enough
                that no one else uses it or a variant of it
14   <muicalc> its calcium backwards
15   <Bspec> well, i named myself after a mode in a video game called Gran Turismo 4
16   <Bspec> which in of itself contains many cars which have these certain tuned versions such as
                "M-spec" or "V-spec"
17   <Bspec> it's called "B-Spec"
18   <Rov> My original and rsn is Aeselrov but i like Rov the most
19   <Rov> Well...Aeselrov means Dunkeyass in danish...Aesel = Dunkey and Rov = Ass
20   <nim> it is the short form of nimitz, which came from a book by david webber, a treecat was named
                nimitz in the honor harrington series (not the aircraft carrier)
21   <nazgjunk> I started off with "netjunk" when I got on the internet first, found out it was used aplenty,
                so I had to come up with something new - that became nazgjunk. Nazg means ring in the
                Dark Tongue of Mordor (Tolkien), and I'm using junk to mean "junkie" or addict
22   <Kooothor> Thor is a viking god
23   <Kooothor> and Kooo is because I like the letter "K" and for some reasons I wanted to have "o" in my
                nick :)
24   <imorrOw> Im swedish and the swedish word of tomorrow is imorgon, so my nick is a mix of thoose
                words cuz im a person thats pretty good of doin things tomorrow insted of today :>
25   <aplsin> well it's short for "Apelsinmannen" (swedish for Orange man), which is the name of an urban
                legend about a man who took so much acid he thought he was an orange
26   <aplsin> i came up with "aplsin" when entering a highscore on a game that only alowed 6 letters
27   <aplsin> the L sound like "el" when you say it, so i replaced "el" by "l"

4.3   Nickname Restriction

Creativity in the creation of a nickname is limited. There are some restrictions on the choice of nicks by the IRC network (Oikarinen/Reed 1993).

4.3.1   Nickname Collision

Chatters "prefer and consistently use one nickname" (Reid 1991). But, if a server detects more than one instance of a nickname on the network, a nickname collision occurs. The nickname registration service (NickServ), which is available on a large number of IRC networks, solves this problem. It allows users to register their favorite nicks and protects them from being used by others. The registered nickname expires after a few days of inactivity (for example, 30 days) and becomes available for registration by other users.

4.3.2   Erroneous Nickname

The Latin alphabet, digits and special characters are available to create IRC nicknames. All these permitted characters were used. For example, the frequencies in Table 4 can be

compared with frequencies of the letters of the English language by Lewand (2000). In summary 82.11% of all 69686 used characters for creating 7936 nicknames were letters of the Latin alphabet. Compared with word-formation in English, digits (10.83%) and special characters (7.06%) have a much higher influence on the creation of nicknames (see Table 5 and 6).

Letter	Frequency Abs. (Lower-Case and Upper-Case)		Relative
a, A	5984	(5185 + 799)	8.59%
b, B	1307	(858 + 449)	1.88%
c, C	1694	(1241 + 453)	2.43%
d, D	1723	(1362 + 361)	2.47%
e, E	5905	(5522 + 383)	8.47%
f, F	876	(589 + 287)	1.26%
g, G	2111	(1027 + 1084)	3.03%
h, H	1696	(1397 + 299)	2.43%
i, I	3834	(3498 + 336)	5.50%
j, J	522	(275 + 247)	0.75%
k, K	1063	(861 + 202)	1.53%
l, L	3168	(2629 + 539)	4.55%
m, M	2352	(1591 + 761)	3.38%
n, N	4005	(3553 + 452)	5.75%
o, O	3321	(3028 + 293)	4.77%
p, P	946	(672 + 274)	1.36%
q, Q	103	(67 + 36)	0.15%
r, R	3553	(3071 + 482)	5.10%
s, S	3735	(3020 + 715)	5.36%
t, T	3232	(2885 + 347)	4.64%
u, U	2336	(2151 + 185)	3.35%
v, V	575	(477 + 98)	0.83%
w, W	612	(447 + 165)	0.88%
x, X	388	(279 + 109)	0.56%
y, Y	1803	(1641 + 162)	2.59%
z, Z	375	(286 + 89)	0.54%
Σ	57219	(47612 + 9607)	82.11%

Table 4: Frequencies of the Latin Alphabet

Digit	Frequency Absolute	Relative
0	695	1.00%
1	892	1.28%
2	1004	1.44%
3	879	1.26%
4	873	1.25%
5	601	0.86%
6	630	0.90%
7	656	0.94%
8	698	1.00%
9	622	0.89%
Σ	7550	10.83%

Table 5: Frequencies of Digits

Character		Frequency Absolute	Relative
`	grave accent	159	0.23%
^	caret	431	0.62%
_	underscore	3242	4.65%
\	backslash	13	0.02%
|	pipe	156	0.22%
[	left square bracket	129	0.19%
]	right square bracket	94	0.13%
{	left curly bracket	15	0.02%
}	right curly bracket	14	0.02%
-	hyphen	664	0.95%
Σ		4917	7.06%

Table 6: Frequencies of Special Characters

IRC nicks which contain characters like slashes, umlauts, punctuation marks or whitespaces are not possible. Hyphen, digit, or space are not allowed at the beginning. A space is omitted or replaced by special characters (see Table 21). Another technical limitation is that IRC cannot handle diacritics like German umlauts (e.g., "ä", "Ö", "ü") or the German ligature "ß" in nicks. The proper way is to replace them with the underlying vowel (with or without a following "e"). The apostrophe is used in English to indicate possession, or used in writing contractions. In IRC, an apostrophe is sometimes omitted, or a grave accent is used as a substitute. A mapping of non-permissible characters is shown below.

Character		Mapping	Example(s)
space		omitted	<IamTheBest>, <CharmyRoman>
		special character	<Keyboard-Cat>, <b|rd^dog>
apostrophe	possessive case	omitted	<DaddysGirl>, <Nobodys_Girl>
		grave accent	<Satan`s-Angel>, <ProudSerb`s_DaD>
	contraction	omitted	<cantyouread>, <Hiitsnick>
diacritic	German umlaut	underlying vowel with "e"	<zahnaerztin_pretty>
		underlying vowel without "e"	<bayernMunchen>
	German ligature	double "ss"	<DerGrosseMann>
	accent	letter without diacritic mark	<Cesc_Fabregas>, <deja_vu>

Table 7: Mapping of Non-Permissible Characters

40.52% of all analyzed nicks were exclusively created with letters (Table 8). A nick cannot start with a digit. Therefore, nicknames which only consist of digits are not possible.

Character Class			Frequency		Example(s)
Latin Alphabet	Digits	Special Characters	Absolute	Relative	
yes	no	no	3216	40.52%	<BhaalWK>, <Jassim>
yes	no	yes	1908	24.04%	<clock_>, <^jenessa>
yes	yes	yes	1699	21.41%	<mib_nxelq2>, <Valerio_886>
yes	yes	no	1102	13.89%	<tek1024>, <Guestguy996>
no	yes	yes	7	0.09%	<[61814]>, <|0_0|>
no	no	yes	4	0.05%	<``>, <_____>
no	yes	no	0	0.00%	
Σ			7936	100.00%	

Table 8: Character Classes of Each Nickname

4.3.3   Letter Case

In general, IRC nicks (and also channel names, or commands) are case-insensitive. It means that <AMan> is the same as <Aman>, and only one of them can be online at the same time. Nevertheless, the handling of letter case for a nickname creation is considered differently. In Table 9 the letter case of all analyzed nicks are shown in detail. In general, these nicknames are stem-based nicks, because they mainly consist of letters (see page 12). Unknown nicknames (2577 nicks) were excluded. Even some chatters used different conventions for the capitalization of their own nicks (for example, <^^ARNOLD> and <^^arnold>).

Letter Case		Frequency		Examples
		Abs.	Rel.	
lower-case		3390	42.72%	<_what_is_love>, <corvette>
upper-case		360	4.54%	<ALASKA>, <^JENN^>
mixed-case	only first letter to upper-case	303	3.82%	<`Maxi_vacationmode>
	first letter of each word to upper-case	3093	38.97%	<AMuslimGirl>, <Walt>
	alternate each letter	93	1.17%	<AcTiVaTe>, <BaD_575>
	alternate each word	30	0.38%	<AXL_roses>, <notNULL>
	random	647	8.15%	<Dr-knoK>, <JEsus>
undefined	nick only consists of number(s)	7	0.09%	<`4_8_15_16_23_42>, <[61814]>
	non stem-based nick	9	0.11%	<[-_-]>, <XxxxX>, <``>
	mixed-based nick	4	0.05%	<fun8]>, <Springfield_XD>
Σ		7936	100.00%	

Table 9: Letter Case

4.3.4   Maximum Nickname Length (NICKLEN)

According to RFC 1459, the maximum nickname length that a client can use is 9, but this is actually determined by the server. The minimum length is one character. The length of the nickname restricts the creation. Depending on the maximum length, long nicks are shortened to make them suitable (see Table 10). Frequently used conventional abbreviations are, e.g., "m" (male, man), "f" (female), and "gf" (girlfriend). Dropping letters is a quick and simple way to shorten nicks. Especially the readability when vowels are dropped often remains intact.

Category	Example(s)
abbreviation	<Youngprof-male>, <Mr-Destructive>
drop letter(s)	<SxySnglMan>
cut off the end	<benJIman'on'holi>, <|414RequestTooLo>
arbitrary shortening	<NkinOnHevnsDoor>

Table 10: Variants of Shortening

The NICKLEN parameter depended on the respective IRC networks. They were 9 (EFnet), 15 (IrCQ-Net, QuakeNet), 16 (freenode), and 30 (DALnet, GameSurge, SwiftIRC). The average nick length of all channels was 8.78 characters (see Table 11). <DrPraetor>, a chatter at #perl mentioned, that "this network [freenode] has criminally low nick length limits". He added, "On other networks I can fit much more interesting nicknames". This statement is all the more astonishing because only 186 users (2.34%) took the opportunity of using the maximum possible length.

Channel	Nickname Length			Nick with Its Length			
	Max.	Count	Average	Minimum		Maximum	
#cars	30	0	9.00	4	<fork>	18	<Cactus-Jack|Server>
#chat-world	30	1	9.09	3	<era>	30	<ThisNickNerverWillBeRegistered>
#church	30	0	8.68	2	<``>	19	<AzuzephreCommunity->
#defocus	16	7	8.23	3	<AHA>	16	<MannyTheMolecule>
#eminem	15	0	6.80	1	<Q>	12	<MrSkitZo^AFK>
#england	15	0	7.38	1	<Q>	15	<sandreinalove_b>
#English	15	54	8.67	2	<no>	15	<SilVeR_Sh[a]DoW>
#freenode	16	7	8.04	3	<Inf>	16	<LordGreystoke422>
##hardware	16	1	7.92	3	<row>	16	<black_Nightmare_>
#perl	16	4	7.83	3	<Kev>	16	<master_of_master>
#Romance	15	98	8.98	2	<hi>	15	<Gener[a]lPublic>
#soccer	9	14	6.86	3	<NEC>	9	<JingleBel>
#talk	30	0	9.47	3	<Zap>	29	<City|Sleep|ILY_Kacey|Call|Me|>

Table 11: Average Length of Nicknames per Channel

4.3.5 Inappropriate Nicknames

Inappropriate nicknames such as abusive or swear words will be removed by operators. If a user on the auto kick list (AKICK) attempts to join the channel, a channel service bot (ChanServ) will automatically kick and ban the participant from the channel.

4.4 Basic Structure of IRC Nicknames

The results of the analysis show that nicknames can be divided into stem-based (99.84%), non stem-based (0.11%), and mixed-based nicks (0.05%). Note that some of the nicknames have special signs in front of them. It indicates that these users have individual rights (and duties) in a channel. For example, an at-sign in front of their names indicates that they are channel operators. They have the power to kick and ban people. Chatters with a plus have voice privileges and can talk in moderated channels ("normal" users cannot). Further user mode signs are "%" (user is a half-op on the current channel), "&" (user is an admin), and "~" (user is founder/owner of the current channel). These special signs are not part of the nick.

4.4.1 Stem-Based Nickname

Stem-based nicks are the most frequently used IRC nicks. They can be divided into the following parts: stem, status, clan, concatenation and decoration. But not each part of this basic structure needs to exist for a nick. An overview of the basic structure with possible positions is shown in Table 12. The position of the stem has the value 0. A negative (positive) value of the position is used to address the part before (after) the stem.

Part / Position	-5	-4	-3	-2	-1	0	1	2	3	4	5
clan	no	optional	no	no	no	no	no	no	no	no	no
stem	no	no	no	no	no	optional	no	no	no	no	no
status	no	no	no	no	no	no	no	no	no	optional	no
decoration	optional	optional	no	no	optional	optional	optional	no	no	optional	optional
concatenation	no	optional	no	optional	no	optional	no	optional	no	optional	no

Table 12: Basic Structure of Stem-Based Nicknames with Possible Positions

For example, the hypothetical nick <^_^[p]Germ{a}n_boy-15^_^|Away> is subdivided into the following parts, illustrated in Table 13.

Part \ Position	-5	-4	0	1	2	4
clan		p				
stem			Germ a n boy 15			
status						Away
decoration	^_^	[ ]	{ }	^_^		
concatenation			_ -			

Table 13: The Parts of the Nick <^_^[p]Germ{a}n_boy-15^_^|Away>

In almost all cases the main part stem was fixed and was not only optional. But the stem can be empty, for example, the user <blizzard-> went away and renamed nick into <_away>. Additionally, only a handful of users created nicks with a different basic structure, because the order of the parts stem and status were swapped (see Table 14). Further examples are <brb-susu-time> (renamed from <susu-sipper>), <afkmum> (<animum>), and <[Away]Ollie> (<Ollie`>).

Part \ Position	-4	-2	0
status	BRB		
stem			Ghost
concatenation		_	

Table 14: The Parts of the Nick <BRB_Ghost>

The 10 most frequent structure templates are shown in Table 15. They cover 98.54%. A further 31 templates cover the remaining 1.46%.

-1 Decor.	0 Stem	Decor.	Conc.	1 Decor.	2 Conc.	4 Status	Abs.	Rel.	Example(s)
no	yes	no	yes	no	no	no	4810	60.71%	<jason_k>, <JayJay>
no	yes	no	no	no	no	no	2153	27.17%	<women>, <sniper>
no	yes	no	no	yes	no	no	303	3.82%	<Infinito->, <carl_>
no	yes	no	yes	yes	no	no	143	1.80%	<no_way_>, <MrBig->
yes	yes	no	no	yes	no	no	85	1.07%	<{rani}>, <_Mike_>
yes	yes	no	yes	no	no	no	81	1.02%	<[_m36>, <_miss_egypt>
yes	yes	no	yes	yes	no	no	68	0.86%	<^cube3^>, <[Jon-Hall]>
yes	yes	no	no	no	no	no	67	0.85%	<_fanmeile>, <_Toni>
no	yes	no	no	no	yes	yes	51	0.64%	<Valentino`Away>
no	yes	yes	no	no	no	no	46	0.58%	<fri^da>, <rapha^ela>

Table 15: Top 10 of Basic Structure Templates

The **stem** is the main part of this type. It consists of one or more letters or digits, which can form words (<yellow>), phrases (<The_Dark_Dragon>), or sentences (<IamTheBest>). The average number of parts is 1.86. Note that these parts include words and numbers. Due to the small number, sentences are rare. An overview is given in Table 16.

Part(s) Count	Absolute	Relative	Example(s)
1	2646	33.40%	<black>, <Gentle_>, <|cute|>, <good>
2	4039	50.98%	<One-Kiss>, <Sugar_875>, <TexasGirl>
3	1004	12.67%	<art_m_45_>, <Call_Of_Booty>, <oneforall>
4	187	2.36%	<Male_With_Cam1>, <within-the-dark-mind>
5	39	0.49%	<JIM_UK_MSNonCAM>, <aB0Y4aGal>
6	6	0.08%	<the_man_who_sold_the_world>
7	2	0.03%	<i-am-a-girl-for-a-reason>
Σ	7923	100.00%	

Table 16: Stem: Part(s) Count

Characters before the stem can stand for some sort of game **clan** (electronic sports) or other organization. A clan is a group of users that play games together over the Internet or on local

area networks. IRC - especially the IRC network QuakeNet - is very popular among players of many different games. Therefore, the number of clan tags, which show the membership of a specific clan is significantly higher than in other networks. A clan tag is mostly a shortened form of words or phrases (see Table 17).

Example	Explanation
<[p]SandMan>	"[p]" is a distinctive mark for a Counter-Strike[4] team. "p" means prominent.
<[rYs]ToScA->	"[rYs]" is an Unreal Tournament 99 clan. "rYs" means "resurrected Yoga slaughterers".
<LLˆspyhunter>	"LL" (short for "LatterligLett") is a Counter-Strike clan. It means "ridiculously easy".

Table 17: Clan

It is common politeness that users change their nickname with the command "/nick" to indicate their current **status**. Thus, everyone can see at a glance that the user is currently not available for a while and cannot read messages right now (see also page 22). For example, statuses can be abbreviations, adjectives, or nouns. They are primarily single words. The most frequent status is named "away". A list of statuses (with or without additional information) is shown in Table 18.

Category		Status	Nick
		Example(s)	
unspecified	action	AFK (away from keyboard), away, BNC (bouncer), BUSY, Gone, IDLE, off, Offline, out	<|FlorianˆoFF>
	time, date	bbiab (be back in a bit), BBL (be back later), BRB (be right back), BBS (be back soon)	<[PowerZ|BRB]>
specified	action	Bath, cook, dinner, eating, FishinG, food, Guitar, meditating, phone, Shower, Sleeping, STUDY, training, work, zzz	<ˋjordan|Eating>
	location	AtWork, Bath-ROOm, Bed, Doctor, Home, Market, pooltime, Pub, School, store, Toilet	<Samurai|Doctor>
	time, date	in2-5, till18th	<Guru|BRB-in2-5>
	feeling, emotion	Sick, Depressed, Sad, Headache, Enraged, ILY_Kacey|Call|Me|	<Smurf|Sad>

Table 18: Status

The main parts (stem, clan, and status) can be **decorated** in front of, within, or behind. In particular special characters (<}{Muffin}{>) beautify nicknames, but also letters, especially the letter x (<XxmelixX>). Decoration can highlight letters, words, or parts of words. Chatters use most of the time the same look and style of their nick for individualization. Examples for text decorated stems are given in Table 19.

Category		Example(s)
in front of		<|[wayne]|>, <ˆˋAnne>, <ˋrocky>
within	split stem	<raphaˆela>, <memori_es>, <Serˋena>
	highlight (letter)	<Sa[i]nT>, <gH[ˆ0ˆ]ST>, <[T]I[G]E[R]>, <D_R_A_C_U_L_A>
	highlight (word)	<{I}{Walk}{Alone}>, <hot{man}>
	highlight (part)	<{G}EN{TLE}MAN>
behind		<iceman_->, <Coach__>, <magnetsˆˆ>

Table 19: Decoration of a Stem

An overview of the top 5 decorations for each position is shown in Table 20. Mostly stems are decorated in front of and behind them.

---

[4] A wide-spread first-person shooter game, played online and usually in teams.

Position	Character	Frequency Abs.	Rel.	Position	Character	Frequency Abs.	Rel.
-5 (in front of clan)	[	7	77.78%	1 (behind stem)	_	191	30.08%
	ˇ	1	11.11%		^	88	13.86%
	|	1	11.11%		-	54	8.50%
-4 (within clan)		0	100.00%		`	52	8.19%
-3 (behind clan)	]	7	100.00%		_	43	6.77%
-1 (in front of stem)	^	87	27.27%	3 (in front of status)	[	8	53.33%
		83	26.02%		{	3	20.00%
	[	28	8.78%		`	2	13.33%
	^^	21	6.58%		|	2	13.33%
	`	21	6.58%	4 (within status)		0	100.00%
0 (within stem)	_	28	26.92%	5 (behind status)	]	8	42.11%
	-	21	20.19%		}	4	21.05%
	[	17	16.35%			2	10.53%
	^	11	10.58%		_xx	2	10.53%
	`	8	7.69%		|	2	10.53%

Table 20: Decoration (Top 5 of Each Position)

Additionally, the main parts (stem, clan, status), and parts of them can be **concatenated** with different characters. In general, IRC nicks with spaces are not allowed (see restriction, page 9/10). Connecting two POS without spaces is the most common way. An overview is given in Table 21.

Position	Char	Frequency Abs.	Rel.	Position	Char	Frequency Abs.	Rel.
-4 (within clan)		0	100.00%		-	9	9.68%
-2 (between clan and stem)	|	1	100.00%		^	9	9.68%
0 (within stem)	space	3558	52.28%		_	9	9.68%
	_	2675	39.30%	4 (within status)	space	8	44.44%
	-	435	6.39%		|	5	27.78%
	^	51	0.75%		-	2	11.11%
	`	35	0.51%		^	2	11.11%
2 (between stem and status)	|	50	53.76%		_	1	5.56%
	ˇ	15	16.13%				

Table 21: Concatenation (Top 5 of Each Position)

4.4.2   Non Stem-Based Nickname

They mainly consist of special characters to create emoticons (a blend of emotion and icon), festoons, or further ASCII art objects. Eastern-style emoticons can be read without turning the head to the left. Similar looking emoticons are bixies (BIXies), which were used by Byte Information Exchange (BIX). Examples are given in Table 22.

Category			Examples
ASCII art	emoticon		<d˘‿˘b>, <d[o_o]b>, <(*_*)>, <[-_-]>
	object	festoon	<o-Oo_0_oO-o>, <O-OO-O-o-o-o-o_o>, <XxxxX>
		else	<˘_——–-˘>, <|_b>, <˘oˆ>[5]
special character			<{}>, <|>, <\\\\>

Table 22: Non Stem-Based Nicknames

4.4.3   Mixed-Based Nickname

As the name suggests, mixed-based nicks are a combination of stem based and non stem-based types. For example, the nick <Springfield_XD> is made up of the stem "Springfield" and the smiley face emoticon "XD", which expresses laughing very heartily. Further examples are given in Table 23. Note that adding a status to an emoticon is another possibility to create mixed-based nicks.

---

[5] Helicopter, cup, bat.

Category	Examples
mixed-based	<fun8]>, <danielita_xD>

Table 23: Mixed-Based Nicknames

4.5   Compounding of Nicknames

This subsection deals mainly with the compounding of stem-based nicknames, especially with the main part stem and its styling. Apart from these the name of the clan is normally strictly predefined. A more detailed research on the part status was not carried out, because it is mostly made up of one single word.

Some nicknames look similar to each other. This happens if they have the similar idea of creation or through adapting nicks from others. The five most used templates with identical stem, decoration and concatenation to create nicknames are shown in Table 24. The basic idea of these templates is that random numbers were added at the end. Copying nicknames (<zhang>) and adding numbers (<zhang2008>) is a simple way to vary nicknames.

Template	Frequency		Examples
	Absolute	Relative	
"Guest_"<number>	375	3.57%	<Guest_161>, <Guest_236>
"Guest"<number>	180	1.71%	<Guest22410>, <Guest34312>
"Guest_"<number>"_"<number>	49	0.47%	<Guest_109_601>, <Guest_978_828>
"BRAZILGIRL_"<number>	15	0.14%	<BRAZILGIRL_294>, <BRAZILGIRL_841>
"Unknown"<number>	15	0.14%	<Unknown28115>, <Unknown63013>

Table 24: The Five Most Used Templates to Create Nicknames

There are numerous classical (which are used to create, e.g., new English words) and non-traditional mechanisms of constructing new creative IRC nicknames (respectively stems).

4.5.1   Creating a Stem

To find out in detail which parts-of-speech a nick consists of, the whole nick - especially the stem - must be decompounded into single POS and tagged. This is crucial to our understanding of answering such questions as the following.

- Which parts-of-speech do nicks consist of in detail?
- In which order are POS concatenated to a compounded nick?
- Which parts of a nick are omitted in discourse?

Decompounding of nicknames: All 10513 nicks were decompounded into single POS with *NickDecompounder,* which the author has written in Java. Figure 2 shows a simplified activity diagram of nickname decompounding.
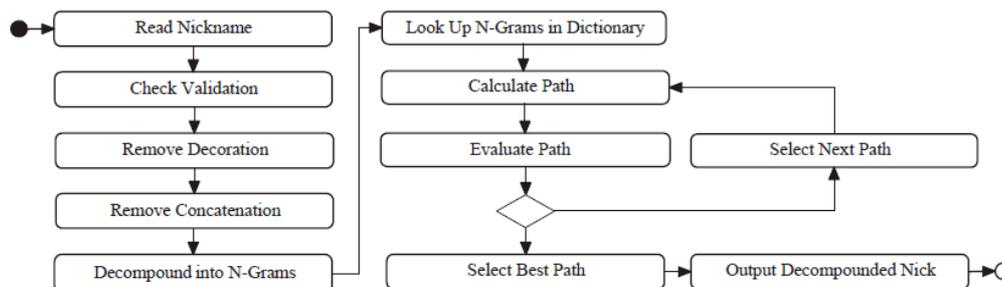
Figure 2: Decompounding of Nicknames with NickDecompounder

A nick with length *n* has $\sum_{i=1}^{n}$ n-grams and 2^{n-1} possibilities (paths) to split the nick. In the following example, the nick <your_heart> is decompounded with the *NickDecompounder*.

- Read nickname: The nickname <your_heart> is read in from a text file.
- Check validation: The string is a valid IRC nickname.
- Remove decoration: No decoration is found.
- Remove concatenation: After removing the concatenation "_" the new string is "yourheart" (with length 9).
- Decompound into n-grams: 45 n-grams are created ("y", "o", ..., "r", "yo", ..., "yourhear", "ourheart", "yourheart").
- Look up n-grams in dictionary: Nine n-grams ("a", "art", "ear", "he", "hear", "heart", "our", "you", "your") exist in the common dictionary, which does not include jargon and slang terms.
- Calculate path: In summary, 256 paths are possible.
- Evaluate path: For example, one path of the examined nickname consists of the n-grams (words) "art", "heart" and "you". The n-grams "you" and "heart" fit into *<yourheart>* at once, but "art" and "heart" overlap. Therefore, "······art", "····heart", "you······", "you···art" and "you·heart" are possible. In this case "you·heart" is the best, because 8 positions of the whole nick are covered by the n-grams "you" and "heart" (88.89% of the total).
- Select best path and output decompounded nick: Finally, the best decompounding of the nick <yourheart> is into "your heart", which covers all nine characters (100%). This result is better than "your he art", which needs more n-grams.

Tagging of nicknames: After all decompounded nicks were manually checked and marked with a full stop at the end ("your heart ."), automatic tagging was executed with the "Stanford Log-linear Part-Of-Speech Tagger" (Toutanova/Manning 2000) using the "Penn Treebank Tag-set" (Marcus/Santorini/Marcinkiewicz 1994). Tagging can be complex because some words represent more than only one POS (e.g., "heart" can be a noun or verb). Incorrectly tagged words were manually corrected. The tags were clustered into coarse-grained syntactic categories (see Table 25).

Cluster POS Tag	Penn Treebank POS Tags
[JJ]	JJ, JJR, JJS
[NN]	NN, NNS, NNP, NNPS
[PRP]	PRP, PRP$
[RB]	RB, RBR, RBS
[VB]	MD, VB, VBD, VBG, VBN, VBP, VBZ
[WP]	WDT, WP, WP$, WRB

Table 25: Cluster of Penn Treebank POS Tags

Not every unknown word as part of a nick refers to a foreign word. They can consist of smileys, non-words (<zrttrtr>), or pseudowords (<Adys>). Additionally, jargon, slang and non-traditional mechanisms for the creation of stems make tagging difficult. The table below shows the four new POS tags, which have been added.

New POS Tag		Explanation
IW	(illegal word)	a) non-word: not pronounceable nor meaningful
		b) pseudoword: pronounceable but not meaningful
MB	(mixed-based nick)	see page 15
NSB	(non stem-based nick)	see page 15
UW	(unknown word or nick)	word or nick of unknown origin

Table 26: New POS Tags

There is no need to use the original POS tag LS (list item marker), "which includes letters and numerals when they are used to identify items in a list." (Marcus/Santorini/Marcinkiewicz 1994). The original POS tag SYM (symbol) is divided into "mathematical, scientific and technical symbols" (now subpart of NSB) and "expressions that aren't words of English" (subpart of IW). The classification of the rest remain equal (e.g., CC, CD, UH). Examples are shown in Table 27.

Tag	Examples	Tag	Example(s)
CC	*and* (<youandmee>), *n* (<bonnie-n-clyde>)	PDT	*such* (<suchaniceday>)
CD	*47* (<^Man47USA>), *25* (<boy25spain>)	POS	*`s* (<Satan`s-Angel>)
DT	*The* (<ManOfTheYear>), *A* (<[[[[A-Man4U>)	PRP	*You* (<You_248>)
EX	*there* (<there_was_nothing_to_lose>)	RB	*n0t* (<l1k3_n0t>), *Just* (<JustAGuy>)
FW	*schokokeks* (<schokokeks_3388>)	RP	*up* (<ce_never_give_up>)
IN	*for* (<oneforall>), *with* (<Always-with-u>)	TO	*To* (<BoredToTears>), *2* (<talk2me->)
IW	<dfgdatlrwfr>, <gghdfhgdfh>	UH	*hehe* (<hehe-[boy]>), <okee>, <wau>
JJ	*sweetest* (<sweetest_lady>), *Bad* (<[-BadBoy>)	UW	*gt li v* <carla18_gt_li_v25>
MB	<fun8]>, <danielita_xD>	VB	*call* (<call_me>), *s* (<hesthere>)
NN	*yasmine* (<yasmine_72>), *eagle* (<the—eagle>)	WP	*what* (<whatsyourname>)
NSB	<|{0_0}|>, <[--]>, <_____>		

Table 27: Tagging Nicknames

Sequences of the same tag were merged together, as in the example below:

Nick	**Penn Treebank Tag-set**	**Cluster**	**POS group**
<MariaPia>	NNP NNP	→ [NN NN] →	(NN)

Table 28 shows that the top 10 of frequently used POS groups cover 89.59% of all 7936 nicks with user's feedback (excluding 2577 nicks without feedback). The POS group (NN) is the most used. (NN) is, e.g., a first or last name, a name related to a town, country, or film.

POS group	Frequency		Examples
	Abs.	Rel.	
(NN)	3348	42.19%	<women>, <monaliza>
(NN CD)	1972	24.85%	<FTorres9>, <User6>
(JJ NN)	907	11.43%	<slimjim>, <^badboy>
(JJ NN CD)	198	2.49%	<coollady2008>, <GoodGuy_446_36>
(JJ)	188	2.37%	<|cute|>, <crzy>
(FW)	145	1.83%	<je_suis_belle>, <HaKuNa_MaTaTa>
(JJ CD)	136	1.71%	<handsome_8286>, <Crazy_51>
(NN JJ)	94	1.18%	<fransisca_cute>, <girl-cool>
(NN IN NN)	73	0.92%	<ianinAmsterdam>, <Sabina-in-syd>
(DT NN)	49	0.62%	<TheChicken>, <_A-Girl>

Table 28: Top 10 of POS Groups

Morphological processes to create stems: As already mentioned in (Crystal 2004; Beißwenger/Storrer 2008), the Internet language uses smileys, jargon (e.g., technical term <Bspec>), slang (<macnoob>; a noob is a person who is new or inexperienced in a subject), abbreviations, and contractions (<Lets^chat>), which also come across within the stem creation. Another possibility is alternative spelling, which changes some letters in a word which is pronounced in a similar manner (e.g., <BigDawg> is an alternative spelling of "big dog"). A further interesting variant is the spelling with the help of homophones. Homophones are words with the same pronunciation, but different spellings and meanings (e.g., <hairypotter> instead of "Harry Potter"). Various mechanisms of word-formation and inflection are used to create IRC nicks. The most common word-formation mechanisms are

back-formation, blending, clipping, compounding, conversion, derivation, and neologism (e.g., loanword, onomatopoeia) (Bussmann 1996; Plag 2003; Trask 2007). There is no clear-cut classification. All these morphological processes are important for nickname creation. Examples are shown in Table 29.

Category	Example	Explanation
abbreviation	\<ProfLee\>	"prof" is an abbreviation for "professor"
acronym	\<Radar\>	acronym for "radio detection and ranging"
back-formation	\<EDIT\>	removing suffix "or" from noun "editor"
blending	\<pulsar\>	combining of the two words "pulse" and "quasar"
borrowing	\<zebra_591\>	adopted word from the Bantu language
clipping	\<Fridge\>	reduction of the word "refrigerator"
coinage	\<Aspirin_720\>	invention of a totally new term,
compounding	\<EarthQuaqe\>	stringing together words "earth" and "quake"
conversion	\<green\>	noun (referring to a putting-green in golf) is derived from the adjective
derivation	\<simply_boyish\>	adding suffix "ish" to the noun "boy"
inflection	\<badboys\>	adding inflectional plural affix "s" to the noun "boy"
initialism	\<FBI\>	initialism for "Federal Bureau of Investigation"
loanword	\<zeitgeist_\>	the spirit of the times or age
onomatopoeia	\<meow\>	sound uttered by cats

Table 29: Classical Mechanisms for the Creation of Stems

A popular strategy is the using of self-related nicks, which disclose personal information to the other users. This has already been pointed out by (Bechar-Israeli 1995). A classification of nicknames which include personal information are shown in Table 30. In contrast, non-self-related (\<sky\>) and unknown nicks (\<cr_yg\>) have a high degree of anonymity which means "that the nickname used by a participant does not reveal any information about the user's on-line identity." (Lakaw 2006).

Category		Examples
name	first, last, initials, nick	\<john117\>, \<Wachert\>, \<chb\>, \<Ollie\>
age	years, year of birth	\<Dana_22\>, \<ROGER1975\>
location	country of origin, residence	\<From_Russia_\>, \<Rick_London\>
	country code	\<zaggy-nl\>, \<young_PL\>
sexuality	gender identity	\<ukgirl\>, \<FEMALE\>, \<[-_English_Man\>
	orientation, preferences	\<alicia25bi\>, \<lesbian24\>, \<foot-fetishman\>
interpersonal relationship	marital status, family	\<[Single_Boy]\>, \<grandmothera\>
religion		\<christianman\>, \<IslamicGirl\>
career	education, job	\<HighSchoolGirl\>, \<geologist\>
physical appearance	aging, appearance, clothing	\<young_303\>, \<TallGuy\>, \<GuyinBlueJeans\>
attitude	positive or negative view	\<Hate-My-Self\>, \<_Life_is_Beautiful_\>
behavior	character	\<Intelligent_guy\>, \<nicegirl\>
emotions	feeling, wish, interjection	\<feelgood\>, \<StillWannaBeInBali\>, \<haha\>
interest, activity	hobby	\<^sportyGirl^\>, \<ArtLove\>, \<mr_computer08\>
	reason to be online	\<letstlkboutsex\>, \<here4fun\>
favorites	club, actor	\<ChelseaFC_Girl\>, \<[[[Leonardo_Di_Caprio]]]\>

Table 30: Personal Information

Nevertheless, it may occur that the new created nick is still not unique. Some creative non-traditional morphological processes change this state and help to adopt nicknames to distinguish between chatters with the same names. Deliberate eccentric strategies to make a nickname unique are, for example, replacing a letter to stand out from other similar variants, or shorten the nickname.

Category		Example(s)
add	number	\<Guest17447\>
	element of obscurity	\<adamx\>
	variable (e.g., for ages)	\<RobiX\>
drop	vowel (consonant writing)	\<sxyfml\>
	consonant	\<lngblknhrd\>
replace	letter	\<NightKhaos\>, \<^dRaGuLa^\>
reduplicate	letter	\<^LoooVeeeR^\>, \<MMEEGGAA\>
	part	\<brarraveheart\>
backward		\<muicalc\>
anagram		\<kunwon1\>
swap	syllable	\<ConSeannery\>
shift		\<Rtkwe\>

Table 31: Non-Traditional Morphological Processes

Universal Leet (L337, L33T, 1337) Converter[6] is a small by-product of this examination. This script is web-based and written in PHP. It translates text-to-leet and leet-to-text. Leetspeak is a creative form of Internet slang, which can also be utilized for nickname creation. The converter has already implemented different levels of letter mappings. The examples below show how letters, words, or parts of them are represented in leetspeak. Both encoding and decoding are not bijective.

Letter Mapping			Example(s)	(Part of) Word M.			Example
a	↔	4	<Cr4zy-Legend>, <YuN4_Ch4N|	are	↔	r	<UrBabe>
b	↔	3, |3	<ˆCho03y_girlˆ>, <|3ug>	-ate	↔	8	<i_h8_j0ck5>
e	↔	3	<Fir3blad3>, <Pix3l>	-ate-	↔	8	<sk8r_Grl>
g	↔	6	<an6eL>	for	↔	4	<aB0Y4aGal>
h	↔	|-|	<|-|0T_Guy>	for-	↔	4	<friend4ev>
i	↔	1, |	<Damn_G1RL>, <v|s|ble>	one	↔	1	<Just1OfTheGuys>
k	↔	|{,]{	<Sm0|{e>, <|{ilroy>	-one	↔	1	<som1forme>
l	↔	1, |, |_	<fireba11>, <Breath|ess>, <DazZ|_3R>	to	↔	2	<love2chatman>
n	↔	|\\\|	<|\\\|ala>	you	↔	u	<FuNK-U->
o	↔	0	<T0tal>, <v0id>	you-	↔	u	<UrBabe>
r	↔	|2	<|2ebeL_girL>				
s	↔	5, c, z	<f00li5h>, <D0c5i5>, <pRinc3ZZ>				
t	↔	1	<AfterDea1h>				

Table 32: Letter and (Part of) Word Mapping

Several types of orthographic errors, which can be intentional or unintentional, are found in the IRC. These may be due to either typographical errors (errors caused by pressing the wrong keys), or spelling errors (errors due to insufficient language competence). A misspelled word can be a correct spelling of another word but can also lead to a new nickname.

Category	Examples
wrong	<sexyandpritty>, <canadiansximan>
missing	<anyting4u>, <luckysot>
extra inserted	<Horney_Man>, <wildflopwer>
transposed	<Ronadl>, <shanghaigilr>
misspelled	<beautyful>, <crazi_boy>

Table 33: Orthography

One or more of these morphological processes are arbitrarily combinable. There are theoretically (and maybe practically) no boundaries.

4.5.2 Styling

Adding a decoration or concatenation alters not merely the appearance. Furthermore, it can make an already used nickname unique without changing the meaning. Examples for nickname styling with decoration and concatenation are shown in Table 34. Both options can also be combined.

Decoration	Concatenation
Gentleman_	[[webcam-for-u
Gentleman__	[[webcam-for_u
Gentleman___	[[webcam_for_u
ˆGentlemanˆ	[[web-cam-for-u

Table 34: Different Styling Possibilities

5   Usage of IRC Nicknames in English Chatroom Discourse

Once connected to an IRC network via an IRC client program (like mIRC), every user has its own unique nickname within the network. After joining a channel, IRC presents a page

---

[6] http://www.robertecker.com/hp/research/leet-converter.php, accessed November 8, 2011.

containing the room's conversation, and a list of logged-in users. These are human beings, but also software programs (chatbots). While adding chat messages, text lines continuously scroll up.

This chat discourse analysis gives an overview about how nicknames are written in discourse and how they are used to address chatters. Different variations of direct addressing and inexact spelling make automatic detection of nicknames and clear referring difficult.

5.1   Direct Addressing

Direct addressing occurs when chatters insert nicknames of other users for addressing (Nash 2005). It prevents discourse confusion. 14.87% of all 8937 investigated public chat messages included at least one nickname.

If direct addressing is necessary a message should start with the receiver's nickname, followed by a colon and a space. 77.13% of all direct addressing starts with a nickname and 58.77% looks similar to the first example in Table 35. Other observed styles at the beginning were dropping the colon or replacing it with other characters. Additionally, nicknames are found within or at the end of messages. This makes the detection of nicknames more complex.

| Category | Frequency | | Example |
	Absolute	Relative	
colon with space (chatiquette)	781	58.77%	<CBG> GPT: good point.
comma with space	100	7.52%	<sarkar112> MooCow, hi :P
dropped colon with space	79	5.94%	<Wigyan> xteddy is a bot?
"–>" with space	24	1.81%	<B_O_F_H> GPT–> but yes it is
semicolon with space	8	0.60%	<eyecue> miranda; how do you figure?
else (e.g., within a message)	337	25.36%	<dmb> hey Starnestommy :D
	1329	100.00%	

Table 35: Single Direct Addressing

It is important to note that not every word followed by a colon at the beginning of a message is a nickname. For example, this word can be part of a quotation, note, enumeration, or definition.

Log Number 3
```
01      <wigyanpy> quotes: " CBG: Male? Female? Martian?"
02      <ei> btw: iron man is just a name
03      <tonny_m> First: How are you?
04      <Telek> CFM: Cubic Feet per Minute.
```

Multiple direct addressing has also been sighted within a message. The recipient names were listed separated by different characters (with or without a space behind), as shown in Table 36 below.

Category		Example
comma and space		<sarkar112> hi MoiraA, Jacco, Kagome
	transposed	<tom100> hello ,00Jafo ,00Sir
comma without space		<academy> rindolf,vincent: thanks
semicolon without space		<Nadine_635> HI SHADOW;DARIO
slash without space		<pigskin_chaser> Starnestommy/hrist: cool!
ampersand and space	one ampersand	<Daniel0> jpds & nalioth: doesn't work
	two ampersands	<pkrumins> hello Zoffix && mauke
conjunction word "and" with space		<Miranda> hi theblue and KangGuru

Table 36: Multiple Direct Addressing

Sometimes there is no link between the nickname used in discourse and the original logged-in one. For example, the problem occurs, when 1) the chatters use another nick in the meantime (e.g., by changing nickname; line 1, 2), 2) they do not use their regular nicks (line 3 - 10), 3)

the input of the message including a nick was in a wrong channel (line 11, 12), 4) the user left the channel or quit the chat (line 13 - 15). Conversely, some words in discourse refer to names of logged-in users who just happen to fit (see example with <lol> on page 8).

Log Number 4

01	*** worried is now known as darkmyst
02	<computer_user> @worried, your lol key is stuck
03	*** Guest_585 has joined #English
04	<MichaelFromH3ll> hey mithos are you here?
05	<Guest_585> how r u micharl
06	<MichaelFromH3ll> I'm fine guest
07	<MichaelFromH3ll> why don't you change your nick mithos
08	<MichaelFromH3ll> call yourself mithos
09	<MichaelFromH3ll> I like that name
10	*** Guest_585 is now known as mithos
11	<notion> sfoster: Mibbit in "Dojo Toolbox". Release date? ;-)
12	<notion> Sorry wrong channel ...
13	<Kagome> well gotta go sleep and then i will go bye bye
14	*** Kagome has left #defocus
15	<miranda> Kagome: bye e-mail..drat...

Direct addressing often arises a) after joining the channel (greeting), b) before leaving the channel or quitting the network (farewell), and c) when expressing thanks. In many cases, nicknames occur after special signal words (like "hello" in "<Werdna> hello RichiH"). Examples of signal words are shown in Table 37.

Category	Examples
greeting	bonjour, hello, hey, hi, hiya, how are you, what's up
farewell	adios, aloha, bbl, bye, ciao, cu, cya, farewell, g2g (got to go), goodbye, goodnight, greetings, night, peace out, see you, ttfn (ta ta for now), ttyl (talk to you later)
expressions of thanks	cheers, merci, thank you, thanks, thx, tnx, ty (thank you)

Table 37: Signal Words

The exceptions are signal words which are used alone, in combination with punctuation (e.g., "hi!"), messages to everybody ("hi there", "bye then") or named groups, e.g., in the positive sense of friends or partners ("hi bro").

Category	Examples
everybody	all, room, channel, world, people, everyone, now, there, then, smileys like ":)"
friend/partner, rival/enemy	babe, bro(ther), chicken, dudes, folk, friend, geek, guy, loser, mate, mom, partner, sis(ter), stranger

Table 38: Words after Signal Words which Mostly are not Nicknames

5.2   Tracking of Nicknames

Chatters are potential discourse partners as long as they remain in the channels. They can be receivers of written chat messages. Therefore, it is necessary to observe users and their nicknames, which join, leave or quit the channels, or change nicks. In addition to simple nickname comparing, a smarter way is comparing the IRC hostmask. The user addresses ("hostmask") on IRC networks typically consist of three parts; a nickname, the ident (user id) and either hostname or IP address. For example, a hostmask looks like "Nickname!username@host.name" ("!" and "@" are separator characters). Commands like "/who" or "/whois" inform people about the specified nick (Charalabidis 2000). The log below shows these changes of one specific user, which is mainly called <Dysaniak>.

Log Number 5: Channel #defocus

01	[28.06.2008 01:20] *** Dysaniak has joined #defocus
02	[28.06.2008 02:06] *** Dysaniak has left #defocus
03	[01.07.2008 04:16] *** Dysaniak has joined #defocus
04	[01.07.2008 06:01] *** Dysaniak is now known as Inf
05	[01.07.2008 06:01] *** Inf is now known as Dysaniak
06	[01.07.2008 06:02] *** Dysaniak is now known as GP1
07	[01.07.2008 06:02] *** GP1 is now known as Inf
08	[01.07.2008 06:04] *** Inf is now known as Dysaniak
09	[01.07.2008 06:08] *** Dysaniak is now known as DeadBaby
10	[01.07.2008 06:09] *** DeadBaby is now known as A
11	[01.07.2008 06:09] *** A is now known as Aman
12	[01.07.2008 06:09] *** Aman is now known as Skeleton
13	[01.07.2008 06:11] *** Skeleton is now known as NaughtyChild
14	[01.07.2008 06:11] *** NaughtyChild is now known as Dysaniak
15	[01.07.2008 06:24] *** Dysaniak is now known as Paraelectri1
16	[01.07.2008 06:24] *** Paraelectri1 is now known as meninslack
17	[01.07.2008 06:33] *** meninslack is now known as Dysaniak
18	[01.07.2008 08:45] *** Dysaniak has left #defocus
19	[06.07.2008 07:36] *** Dysaniak has joined #defocus
20	[06.07.2008 08:27] *** Dysaniak Quit ("Leaving.")

<Dysaniak> seems to be the user's regular nick, because a) it was used most frequently, and for the longest periods, b) he always logged in with <Dysaniak> (most IRC clients include a setting to automatically set nickname at startup). Note that he flooded the channel with his nick-changes (line 4 - 14), but he was not banned. Two kinds of visualization of the above log are shown in Table 39. In the left figure (network diagram) nicknames are represented by knots, nick-changes by sequentially numbered links. Additionally, in the right figure timeline and knots with different states are visualized (white: change nickname, green: join channel, red: leave channel, black: quit/disconnect from the server). Further possibilities of statistical analysis are about, for example, number of written messages, words, characters, addressees, words per line, last online time, active or inactive status.

Knot	Nickname	Visualization
A	<Dysaniak>	
B	<Inf>	
C	<GP1>	
D	<DeadBaby>	
E	<A>	
F	<Aman>	
G	<Skeleton>	
H	<NaughtyChild>	
I	<Paraelectri1>	
J	<meninslack>	
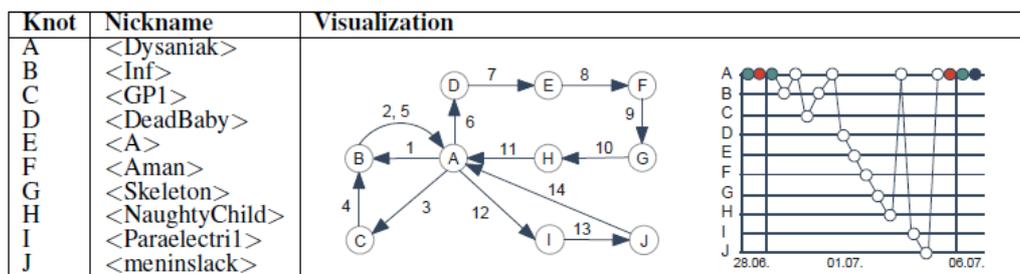
Table 39: Visualization of Nick-Changes

231 nick-changes ("is now known as") were found and analyzed. The results in Table 40 show how nicks were renamed whilst chatting (Top 5). In most cases, the stem was changed into a completely new one. A common reason to change the current nickname is when the user currently cannot pay attention to IRC. Therefore, the nickname is renamed into a new one including an additional argument (status). 13.85% were especially used for changing the status.

Stem	Decoration (Behind)	Concatenator (Between Stem/Status)	Status	Frequency Abs.	Frequency Rel.	Nickname (Example) Old	Nickname (Example) New
new	-	-	-	137	59.31%	<Diluvium>	<daemon>
-	remove	-	-	29	12.55%	<guru_>	<guru>
-	-	add	add	15	6.49%	<Qst>	<Qst_away>
-	-	remove	remove	9	3.90%	<Qst_away>	<Qst>
add part	-	-	-	8	3.46%	<brown_ca>	<brown_cat>

Table 40: How Nicks were Renamed Whilst Chatting (Top 5)

A better way of changing nicknames is using the command "/away", because the current nickname is not modified. The advantages are that 1) the user still remains reachable, 2) no public notification of nick-changing is necessary, and 3) nobody can "steal" the nick, because it is not changed. Now, if someone uses "/whois" or writes a private message, a description including the away message is shown. Command "/away" without any additional argument will remove the away message.

Mutton's *PieSpy Social Network Bot* monitors a set of IRC channels and visualizes social networks on IRC. It uses direct addressing of users, monitoring nick changes, temporal proximity, and temporal density to infer relationships between pairs of users (Mutton 2004a). But linguistic play and creativity are not taken into account during discourse.

5.3   Complications in the Detection of Nicknames whilst Chatting

Saving keystrokes to reduce time and effort play an important role whilst chatting. But this is not always the case; some variant forms of linguistic playing with nicks sometimes result in even more keystrokes than the original nickname or require more effort. All these underlying processes, including punctuation and orthographic errors, influence and complicate the detection of nicknames in discourse.

5.3.1   Punctuation

IRC Nicknames cannot contain certain punctuation characters like a comma, colon, question and exclamation mark, ellipses, period, or round bracket. But they can be surrounded within a message by any punctuation; various letters are found in front of the nick (e.g., punctuation marks at the beginning of a quotation), or behind it (e.g., question marks for interrogation, periods for the end of sentences). Especially, if a space is forgotten to separate words. Using punctuation marks for self-correction or applause, mostly unique to online chat, is another important feature of written language. The misspelled word is replaced in the next message with the word marked by an asterisk. Furthermore, one plus means applause, the more pluses, the louder and warmer is the applause for this user. Below in Table 41, examples of using nicknames in connection with punctuation are shown.

Category		Nick (Original)	Example (Discourse)
in general	interjection, exclamation	\<javanon\>	"JAVANON!!!!!"
	interrogation	\<runatrain\>	"runatrain?"
	possessive case	\<Starnestommy\>	"Starnestommy's cat has 5 legs"
	word divider (no space)	\<Jesus\>	"thank you freefull and jesus:D"
chat-specific	self-correction	\<cal\>	"cal*" after "cak: go eat Cybergeek2021"
	applause	\<Zoffix\>	"Zoffix++"

Table 41: Punctuation

5.3.2   Orthographic Errors

Orthographic errors can especially occur while quickly typing a nick which is similar to the creation of IRC nicknames. A mistaken written letter for another similar-looking letter can be considered as an unintentional error. They are ambiguous for human readability (e.g., 0 ⇔ O, 1 ⇔ l, 8 ⇔ B, q ⇔ g, 2 ⇔ Z, ` ⇔ ', m ⇔ rn). Therefore, some nicks were written wrongly. This mostly depends on the chosen font. The same problem happens "when handwritten forms are scanned and optical character recognition (OCR) is applied" (Christen 2006). Nevertheless, several errors seem to be obviously intentional. They look like a provocation. The additional input of spaces complicates the nickname detection because separated words are created. Examples of orthographic errors due to typing are shown in Table 42.

Category		Nick (Original)	Nick (Discourse)
inadvertence	mistyped (wrong)	\<jonsmith1982\>	"jonsmith1984"
	mistyped (missing)	\<sarkar112\>	"akar112"
	mistyped (extra inserted)	\<wahnfrieden\>	"wahnfrienden"
	mistyped (transposed)	\<sheep23591\>	"sheep23951"
	misread (similar-looking)	\<gavin_1\>	"gavin_l"
intention	provocation	\<lady08\>	"Lazy08"

Table 42: Orthographic Errors Due to Typing

Users make several text normalization decisions to correct created nicks for the use in discourse. Correcting misspelled words, dropping of reduplicated letters, or simply writing out the nick in correct English are some of these procedures. Some nicknames written in leetspeak were fully, partially or incorrectly reconverted into a "normal" text.

Category		Nick (Original)	Nick (Discourse)
mistyping	correct a mistake	\<bloodoby\>	"bloodboy"
reduplication	drop reduplicated letter	\<SeXXXyyy\>	"sexy"
	drop reduplicated part	\<nickck\>	"nick"
leetspeak	fully reconverted	\<WORLDW1D3\>	"worldwide"
	partially reconverted	\<f00li5h\>	"f00lish"
	incorrectly reconverted	\<|Stea|thBabe|\>	"steathbabei" (instead of "StealthBabe")
correct English	marking of possessive case	\<maripiasteacher\>	"maripia's teacher"

Table 43: Text Normalization

5.3.3 Saving Keystrokes, Time, and Effort

91.65% of all written nicks in discourse referred exactly to the logged-in ones. One reason for this high number of correctly written nicknames is the nick autocomplete feature. Some IRC clients provide this functionality which completes the rest of the nickname after pressing the first letters of the nick and a special key (e.g., the tabulator key).

However shortened nicknames and variants were also found in the chat discourse. Shortening or omitting parts of the nickname are two important strategies to save keystrokes. Examples are given in Table 44. Especially the length of the dropped substring seems to be arbitrary. The problem thereby is that a new semantic of the remaining word is possible. Interestingly, short nicks have even been shortened again. This can lead to confusion. Therefore, many chatters only react if their nicks are exactly written in discourse. \<_Tom\>, a user at channel #talk, confirmed this impression: "I only respond when somebody says _Tom directly. Because there's many toms on this network".

Category		Nick (Original)	Nick (Discourse)
shortening	acronym, initialism	\<HellDragon\>	"HD"
	abbreviation	\<GuCCiGuRL\>	"guc", "GuCC"
omitting POS	clan	\<[P]SandMan\>	"sandman"
	status	\<mouz^away\>	"mouz"
	decoration	\<|Soujiro|\>	"soujiro"
	concatenation	\<B_O_F_H\>[7]	"bofh"

Table 44: Shortening or Omitting Parts of the Nickname

The comparison between the original chatter's nickname and the used variant in the discourse show us which part of a nickname has been omitted in detail. Not only clan, status, decoration, or concatenation can be omitted (see Table 44); also parts of a stem (see Table 45). Note that sequences of clusters were not merged together. The below mapping is not bijective.

---

[7] Acronym for "bastard operator from hell".

Clustered POS from Nick ...		Example	
in Discourse	Which is Logged in (Original)	Nick (Discourse)	Nick (Original)
[CD]	[CD RB VB], [NN CD]	"404"	<|404NotFound|>
[DT]	[DT CD]	"The"	<The_868>
[FW]	[FW CD]	"sanctus"	<sanctus2099>
[JJ]	[JJ CD], [JJ NN], [JJ PP], [NN JJ NN]	"latina"	<sexylatina>
[JJ NN]	[JJ NN CD], [JJ NN CD NN], [JJ NN IN NN]	"black beauty"	<blackbeauty_720>
[NN]	[DT NN], [JJ NN], [NN CD], [NN CD NN], [NN IN NN], [NN JJ NN], [NN NN], [NN NN CD]	"Dragon"	<ThE_DrAgOn>
[NN NN]	[NN CD NN], [NN NN CD], [NN NN NN]	"engel girl"	<Engel08girl>
[RB]	[RB TO NN]	"Back"	<Back2Basics>
[VB]	[VB IN PP NN)]	"dream"	<Dreamofmebab>

Table 45: Omitted POS of Nickname Whilst Chatting

A prediction of which part-of-speech will be dropped is not easy to make; but the tendencies are as follows.

- Omitted parts are mostly cardinal numbers.
- Adjectives or/and nouns often remain.
- Nicknames are shortened after the first one or two POS.

A key strategy to increase typing speed and minimize response time is the handling and matching of lower and upper-case letters of a written nick which does not always correspond exactly to the original user's nickname. This can be a consequence of typing quickly, when nicks or the whole messages are written in lower-case letters. On the other hand, messages written entirely in upper-case letters are considered to be shouting.

5.3.4   Creative Linguistic Playground

Chatters play with the original nickname; regardless of saving keystrokes, time, or effort. Reduplicated letters signal stressed syllables. They are used to express emotions whilst chatting. Other creative methods in using nicknames in context are by adding decorations, or writing backwards. All these methods are well known from nickname creation. Not only the creator of the nickname wants to stand out, but also the chatter who uses it (see Table 46).

Category	Nick (Original)	Nick (Discourse)
reduplication	<Rita>	"Riiiitaaaaaa"
decoration	<adaptr>	"[adaptr]"
backwards writing	<GanjaMan>	"naMajnaG"

Table 46: Creativity

Messages become harder to understand and analyze, if addressees are based on the user's real name (Ex. 1), diminutive forms are used (Ex. 2), background information is necessary (Ex. 3 - 5), or linguistic playing occurs with the nick (Ex. 6).

Log Number 6

01	<fallingfromyou> hiya Michael
	<Necromant> hi jen
02	<Chewbacca_360> me?
	<t3hbowie> Yeah, you, baccy.
03	<OliverWoods> hay guys
	<Chewbacca_360> hello harrypotter's friend
04	<mercury> I'm chinese, from ShangHai,China
	<eileen_> hi chinese guy
05	<MiSsUnDeRwEaR> HOWS U Tom
	<Tom> hugs missdoubleyummyunderwear
06	<GothicAngel> MyBabe :))
	* UrBabe kisses GothicAngel on the cheek[8]

6   Conclusion

This paper presents an empirical study of nicknames in IRC, with quantitative and qualitative observations. The aim was to analyze the link between written nicknames in discourse, and current logged-in users. A series of examples were provided that can help researchers and practitioners to improve the quality of methods for automatically processing IRC chat, especially nickname detection in discourse. The results show that the linguistic possibilities in the creation of nicknames are various. However, already the top 10 of frequently used POS groups, and the top 10 of basic structure templates cover most of the investigated nicks. In particular, creativity and saving keystrokes complicate the detection of nicknames in discourse. They make exact string matching problematic. The surprising fact is that about 90% of all written nicks are written exactly like the logged-in ones. This value is probably much lower in other chats without an autocomplete feature. Additionally, it could be illustrated which parts of a nick were often omitted within the chat discourse. This study helps to predict whether words in discourse match the original nicknames. A one-to-one string matching (excluding focus on letter case), and a string comparison of the first two POS groups are mostly sufficient. But for a completely automatic discourse analysis it is important to identify all nicks (and other links like pronouns) in the written chat messages to find out more about the discourse structure (Holmer 2008), thread detection (Shen et al. 2006), or who is chatting with whom.

Limitations of this study include the fact that private messages were not and could not be part of the study. These could be different (as regards the creation and especially the use of nicknames) from those in the public IRC channels.

There have also been several suggestions on how to design a nickname creator so that the generated nicks look different and look as if they are created by a human being. These suggestions can be applied to other social media like Facebook and Twitter because IRC nicknames are often found on Facebook and Twitter (and vice versa). Such a nickname generator has been developed and is already being used in practice. But this nickname generator is not part of this paper.

References

Anderson, John M. (2007): *The Grammar of Names*. New York: Oxford University Press.

Bays, Hillary (1998): "Framing and face in Internet exchanges: A socio-cognitive approach". In: *Linguistik online* 1, 1/98.

Bechar-Israeli, Haya (1995): "From 'Bonehead' to 'cLoNehEAd': Nicknames, Play and Identity on Internet Relay Chat". In: *Journal of Computer-Mediated Communication* 1/2.

---

[8] The IRC command "/me" performs an "action" on the current channel. It shows up one asterisk at the start of message with no punctuation around the user's nick.

Beißwenger, Michael/Storrer, Angelika (2008): "Corpora of Computer-Mediated Communication". In: Lüdeling, Anke/Kytö, Merja (eds.): *Corpus Linguistics*. *An International Handbook*. Berlin, Mouton de Gruyter: 292-308. *(= Handbücher zur Sprach- und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science 1)*.

Blakeman, Adam (2004): *An investigation of the language of Internet chat rooms*. Dissertation. LAMEL, Lancaster University.

Bussmann, Hadumod (1996): *Routledge Dictionary of Language and Linguistics*. London: Routledge.

Charalabidis, Alex (2000): *The Book of IRC – The Ultimate Guide to Internet Relay Chat*. San Francisco: No Starch Press.

Christen, Peter (2006): "A Comparison of Personal Name Matching: Techniques and Practical Issues". In: *Proceedings of the Sixth IEEE International Conference on Data Mining - Workshops*. Washington, DC, USA: IEEE Computer Society: 290–294.

Crystal, David (2004): *Language and the Internet*. Cambridge University Press.

Davis, Boyd H./Brewer, Jeutonne P. (1997): *Electronic Discourse: Linguistic Individuals in Virtual Space*. New York: State University of New York Press.

Dewitte, Siegfried/Hendricks, Hendrik (2005): "Qualitative market research online. Easier said than typed." In: *Research Center of Marketing*. K.U.Leuven: 1–20.

Doell, Wernfrid (1998): *The Internet Relay Chat (IRC): Linguistic Perspectives*. http://www.robertecker.com/hp/research/publication/Doell2005.zip, accessed November 11, 2011.

Döring, Nicola (2003): *Sozialpsychologie des Internet. Die Bedeutung des Internet für Kommunikationsprozesse, Identitäten, soziale Beziehungen und Gruppen*. Göttingen: Hogrefe.

Gelhausen, Andreas (2008): *IRC statistics*. http://irc.netsplit.de/, accessed November 2, 2011.

Gelléri, Péter (1998): *The IRC Vernacular: A Linguistic Study of Internet Relay Chat*. MA Thesis.

Hård af Segerstad, Ylva (2002): *Use and Adaptation of Written Language to the Conditions of Computer-Mediated Communication*. Doctoral Dissertation. Sweden: Göteborg University.

Hentschel, Elke (1998): "Communication on IRC". *Linguistik online* 1, 1/98.

Herring, Susan C. (1999): "Interactional Coherence in CMC". In: *Proceedings of the Thirty-Second Annual Hawaii International Conference on System Sciences*. Vol. 2. HICSS '99. Washington, DC, USA: IEEE Computer Society.

Holmer, Torsten (2008): "Discourse Structure Analysis of Chat Communication". In: *Language@Internet* 5.1.

Johnová, Markéta (2004): "The Language of Chat". In: *PHILOLOGICA.NET: An Online Journal of Modern Philology*.

Kortti, Heikki (1999): *On some similarities between discourse in the Internet Relay Chat and the conventions of spoken English*. http://www.robertecker.com/hp/research/publication/Kortti1999.zip, accessed November 11, 2011.

Lakaw, Alexander (2006): *Hiding behind nicknames. A linguistic study of anonymity in IRC chatrooms*. Student thesis. Växjö University. http://lnu.diva-portal.org/smash/record.jsf?pid=diva2:206935, accessed November 11, 2011.

Lewand, Robert Edward (2000): *Cryptological Mathematics*. Washington: The Mathematical Association of America.

Marcus, Mitchell P./Santorini, Beatrice/Marcinkiewicz, Mary A. (1994): "Building a Large Annotated Corpus of English: The Penn Treebank". In: *Computational Linguistics* 19.

Morgan, Jane/O'Neill, Christopher/Harré, Rom (1979): *Nicknames: Their Origins and Social Consequences*. London/Boston/Henley: Routledge & Kegan Paul. (=*Social Worlds of Childhood*).

Mutton, Paul (2004a): "Inferring and Visualizing Social Networks on Internet Relay Chat". In: *Eighth International Conference on Information Visualization (IV04)*. IEEE, 9 total.

Mutton, Paul (2004b): *IRC Hacks – 100 Industrial-Strength Tips & Tools. First*. Beijing: O'Reilly.

Nash, Carlos M. (2005): "Cohesion and Reference in English Chatroom Discourse". In: *Hawaii International Conference on System Sciences* 4. http://www.computer.org/portal/web/csdl/doi/10.1109/HICSS.2005.143, accessed November 8, 2011.

Oikarinen, Jarkko/Reed, Darren (1993): *RFC 1459: Internet Relay Chat Protocol*.

Perea, Manuel/Duñabeitia, Jon Andoni/Carreiras, Manuel (2008): "R34D1NG W0RD5 W1TH NUMB3R5". In: *Journal of Experimental Psychology: Human Perception and Performance* 34/1: 237–241.

Plag, Ingo (2003): *Word-Formation in English. Cambridge Textbooks in Linguistics*. New York: Cambridge University Press.

Reid, Elizabeth M. (1991): *Electropolis: Communication and Community On Internet Relay Chat*. Honours Dissertation. University of Melbourne.

Ringlstetter, Christoph/Schulz, Klaus U./Mihov, Stoyan (2006): "Orthographic Errors in Web Pages: Toward Cleaner Web Corpora". In: *Computational Linguistics* 32.3: 295–340.

Rintel, E. Sean/Mulholland, Joan/Pittam, Jeffery (2001): "First Things First: Internet Relay Chat Openings". In: *Journal of Computer-Mediated Communication* 6/3.

Rüggenberg, Sabine (2007): *So nah und doch so fern. Soziale Präsenz und Vertrauen in der computervermittelten Kommunikation*. Dissertation. Germany: University of Cologne.

Shen, Dou/Yang, Qiang/Sun, Jian-Tao/Chen, Zheng (2006): "Thread Detection in Dynamic Text Message Streams". In: *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '06. Seattle, Washington, USA: ACM, pp. 35–42.

Stevenson, Jon (2000): *Language Data Investigation. The Language of Internet Chat Rooms*. Originally published on http://www.demo.inty.net/Internet%20Relay%20Chat.pdf.

Stommel, Wyke (2007): "Mein Nick bin ich! Nicknames in a German Forum on Eating Disorders". In: *Journal of Computer-Mediated Communication* 13/1.

Sun, Hong-mei (2010): "A Study of the Features of Internet English from the Linguistic Perspective". In: *Studies in Literature and Language* 1/7: 98–103.

Tavosanis, Mirko (2007): "A Causal Classification of Orthography Errors in Web Texts". In: *IJCAI-2007: Workshop on Analytics for Noisy Unstructured Text Data*.

Toutanova, Kristina/Manning, Christopher D. (2000): "Enriching the knowledge sources used in a maximum entropy part-of-speech tagger". In: *Proceedings of the 2000 Joint SIGDAT conference on EMNLP/VLC*. Hong Kong: Association for Computational Linguistics.

Trask, Robert Lawrence (2007): *Language and Linguistics – The Key Concepts*. London: Routledge.

Werry, C. (1996): "Linguistic and interactional features of Internet Relay Chat". In: Herring, Susan C. (ed.): *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives*. Amsterdam: John Benjamins.