Read Writ (2016) 29:1225–1243
DOI 10.1007/s11145-016-9634-z

CrossMark

Is instant messaging the same in every language? A basque perspective

Jasone Cenoz[1,2] · **Garbiñe Bereziartua**[1,3]

Published online: 14 March 2016
© Springer Science+Business Media Dordrecht 2016

Abstract This study focuses on computer mediated communication (CMC) in instant messaging using the Basque language in a context where exposure to English is very limited outside the classroom. This context provides an opportunity to analyze the universality of linguistic features identified in CMC in English. The corpus consists of 54 naturalistic dyadic conversations between Basque secondary school students, using the medium of instant messaging. Thirty-four of those students then took part in six focus group discussions so as to obtain information about their perception of the linguistic features used in instant messaging. The results indicate that the linguistic features used in CMC are in general terms similar in English and Basque with some exceptions which are related to the specific linguistic features of the Basque language.

Keywords Computer mediated communication (CMC) · Textisms · Instant messaging · Basque language

Introduction

Computer mediated communication (CMC) refers to the exchange of messages via a computer connection. It has been defined as "communication that takes place between human beings via the instrumentality of computers" (Herring, 1996, p. 1).

✉ Jasone Cenoz
jasone.cenoz@ehu.eus

Garbiñe Bereziartua
garbibere@gmail.com; garbine.bereziartua@ehu.eus

[1] Faculty of Philosophy and Education, UPV/EHU, University of the Basque Country, Tolosa Hiribidea, 70, 20018 Donostia-San Sebastián, Spain

[2] Faculty of Education, Philosophy, and Anthropology, Tolosa Hiribidea, 70, 20018 Donostia-San Sebastián, Spain

[3] Bilboko Irakasleen Unibertsitate Eskola, Sarriena Auzoa, 48940 Leioa, Bizkaia, Spain

Nowadays, CMC is a cover term that not only refers to computer-based communication but also to communication that takes place through mobile phones. Other terms have also been used such as ‘cyberspeak’, ‘internet language’ or ‘netspeak’ though CMC is the most widespread. Herring, Stein &Virtanen (2013) discuss the term CMC and argue that even if mobile phones can be considered a type of computer, it may be more difficult to consider voice calls as CMC. They discuss other possible terms but conclude that CMC, which is broadly used, is still a useful term.

There have been some very important developments in CMC over the last 20 years (see for example Herring, 1996; Herring et al., 2013; Thurlow, Lengel & Tomic, 2004). According to the International Telecommunication Union (2015), there were 3.2 billion people using the internet by the end of 2015 and more than 7 billion mobile phone subscriptions. According to Baker and White (2010, p. 1591), teenagers “represent the largest and fastest growing demographic sector using the Internet”. Indeed, it is between the ages of 13 and 19 that young people access the internet and communicate with each other using computers and mobile phones more than other age group (Herring & Kapidzic, 2015; Lenhart, Purcell, Smith & Zickuhr 2010). One of the most common uses of CMC is interpersonal communication. CMC is replacing telephone conversations among adolescents but it is also extensively used by children and adults for personal and professional purposes.

According to Tagliamonte and Denis (2008), there are two parameters that can be useful to classify different types of CMC messages: the number of recipients of the message and whether the message is synchronous or asynchronous (see also Crystal 2001, p. 151). A message can be dyadic, adopting the form of one-to-one dialogue, or multiparticipant when it is aimed at a group of people. When messages are synchronous, an instant or near-instant reply is expected. Herring (2010) explains that in the case of synchronous messages there is less opportunity to edit them and they look more like turns in a face-to-face conversation. Synchronous messages also tend to be short but they need to be relevant to those preceding. This means that there must be some coherence between the exchanges as is the case with face-to-face conversations.

Another characteristic of CMC is that it is dynamic. This is due to advances in technology and the different characteristics of new hardware and software. For example, there have been significant developments in instant messaging on mobile devices in recent years. At the same time, language also changes and as Baron (2003, p. 22) says, the description of CMC “tends to be at once a diffuse and a moving target”.

Most studies on the characteristics of the language used in CMC among teenagers have focused mainly on English. This article looks at CMC in instant messaging in a minority language, Basque, in order to see its similarities and differences with linguistic features identified in other languages. It also includes reflections made by participants about their choices. In the following sections we summarize the linguistic features of CMC and the context of the Basque language, and then we report on a study conducted in a Basque-speaking town before finally comparing the results to those of studies conducted in other contexts.

Linguistic features of CMC

CMC has some special characteristics that are reflected in linguistic features. Although the language used in CMC is written, it shares many characteristics of oral language. It is obvious that there are important differences between traditional writing and CMC. Herring (2010) explains that users experience CMC in similar ways to spoken conversation and use verbs such as "talk", "say" and "hear" when referring to CMC. However, there are also marked differences between oral language and CMC. Crystal (2001), who calls CMC language "netspeak", considers that a clear difference is the lack of simultaneous feedback in CMC until the message is sent. In comparison, speakers in an oral face-to-face conversation can get feedback from their interlocutors and see if what they say has been understood, liked or disliked. Indeed, participants in face-to-face conversations can provide feedback simultaneously and they can also interrupt their interlocutors. Crystal considers that another basic difference between CMC and oral language is the lack of prosody and paralanguage. CMC lacks intonation, facial expressions, gestures and body language to express attitudes and opinions. Crystal (2001, p. 47) concludes that CMC "has far more properties linking it to writing than to speech".

However, there are others who consider CMC to be closer to oral than to written language. For example, Herring (2010) explains that there are a lot of similarities between CMC and informal spoken conversations highlighting the idea that CMC has many of the same social functions as spoken conversations. In fact, CMC is somewhere between oral and written language yet with some specific characteristics that are not shared by either of them. Baron (2004) analyzed a corpus of 23 CMC conversations among American college students with a total of 11,718 words and she reported characteristics of both spoken and written language. Participants took multiple turns to close the conversation as can happen in face-to-face oral interaction but made attempts to correct their own typing errors as is done when editing a written text.

The language used in CMC messages has been studied for its specific features. The general term *textisms* is widely used to refer to these features (see, for example, De Jonge & Kemp, 2012; Grace, Kemp, Martin & Parrila, 2014). Some of these features are associated with the technical characteristics of CMC and its limitations compared to face-to-face interaction. One of these limitations is the need for messages to be short because of the size of screens, keypads and limitations in the number of characters. Another characteristic is the need to show emotions by using some resources that are different from those used in oral language. The use of textisms compensates for some of these limitations but it goes further than that. It can also be a way to develop group membership, particularly among teenagers (Baron, 2003; Kemp, Wood & Waldron, 2014; Prieto-Arranz, Juan-Garau & Jacob, 2013). Teenagers choose to use some common textisms and can even create others because CMC is a joint activity carried out with friends outside school.

Studies on language use in CMC have developed different taxonomies of textisms for CMC in English (see for example, De Jonge & Kemp, 2012; Grace et al., 2014; Varnhagen et al., 2010). When comparing these taxonomies it can be

seen that some pay attention to form and others also to the function of textisms. For example, Grace et al. (2014) propose a categorization of the difference between textisms and standard written language based on form. It includes 15 categories which are characterized by the way the text has been adapted. Kemp, Wood & Waldron (2014) also focus on form when distinguishing four broad categories: missing punctuation, missing capitalization, word and grammatical errors and unconventional punctuation.

Varnhagen et al. (2010) combine formal and functional criteria and specify three broad categories: short cuts, pragmatic devices and errors. Short cuts and errors focus on form comparing textism and standard language while pragmatic devices look at the function of the device used. In this article we consider the three broad categories proposed by Varnhagen et al. (2010) including some of the specific categories proposed by Grace et al. (2014) as well as Kemp et al. (2014). Therefore, we will consider the following categories: (1) short cuts and abbreviations; (2) pragmatics and emotional resources; and (3) errors.

1. ‘Short cuts and abbreviations’ can include the following:

   - *Word combinations* When multiple words are contracted into a single, phonetically spelled word: *wanna* for *want to*, *gonna* for *going to.*
   - *Shortenings* When words are shortened by removing one or more phonemes or morphemes: *Tue* for *Tuesday*, *tmrw* for *tomorrow, u* for *you*, *pls* for *please, wont* for *won't*, *havin* for *having*.
   - *Acronyms or initialisms* When words are formed from the initial letters or words in a series of words: *brb* for *be right back*, *omg* for *Oh my God.*
   - *Homophones or phonetic abbreviations* When a word or part of a word is substituted with an alphabetic name *b* for *be*, *wat* for *what.*

Grace et al. (2014, p. 861) distinguish shortenings (*Tue* for *Tuesday*), contractions (*pls* for *please*), omitted apostrophes (*wont* for *won't*) and G-clippings (*havin* for *having*) as separate categories but we consider them all to be short cuts because in all cases the word is shortened. Some of the categories identified by Grace et al. (2014), such as omitted apostrophes and G-clippings, can work very well for English but not for other languages because these specific linguistic structures are not present in all languages. Varnhagen et al. (2010) consider the category ‘alphabet letter’ (*u* for *you*) which we would classify as shortening because it is similar to the reduction from seven letters to three in the case of a word like *Tuesday* but it could be considered a homophone or phonetic as well. Varnhagen et al. (2010) have also incorporated the category ‘insider word’ for slang words but these are not included here because they are not short cuts or abbreviations.

Short cuts or abbreviations serve an important function in that they save space, time and effort. As Baron (2003, p. 21) points out abbreviations and acronyms are not new and have long been used in handwritten manuscripts and print. Nowadays, abbreviations are widely used in CMC.

2. ‘Pragmatic and emotional resources’ can be the following:

    - Pragmatic lengthening refers to extra letters or extra words used to mirror the intonation of the oral language: *hellooo* for *hello or love love* for *love.*
    - Upper case or extra capitals when upper case letters are used to represent the words said at a higher volume: *WHAT* for *What.*
    - Emotion words such as *hahaha* (laughter or the sound of laughter).
    - Emotion punctuation refers to the extraneous use of punctuation for emphasis (e.g., !!!!) and emoticons (e.g., :-) for a smiling face).

This category includes resources that bring CMC closer to oral language and compensate for the limitations of short written messages to express communicative intent, attitudes and emotion. For example, the lengthening of a specific word can show emphasis that in oral language can be expressed by intonation and facial expressions. An emoticon with a smile can soften a message by showing a friendly attitude. Emoticons could be included in this category but they will not be considered in this study. There is variation in the way different taxonomies categorize pragmatic and emotional resources. For example, Varnhagen et al. (2010) include *lol* (*laughing out loud*) as an emotion acronym while it can also be considered a short cut or abbreviation. Emotion punctuation is the term used by Varnhagen et al. (2010) yet Grace et al. (2014) refers to it as ‘extra punctuation’ and Kemp et al. (2014) ‘unconventional punctuation’.

3. The third category, ‘Errors’, refers to the following:

    - *Typografical carzy* for *crazy*
    - *Non-standard spellings progect* for *project*. This category also includes omitted capitals: *bob* for *Bob*.

There is some overlap between this category and the others. For example, Androutsopoulos (2000), distinguishes six types of non-standard spellings by pointing to more specific categories. Two of the categories he proposes ‘phonetic spelling’ and ‘shortenings’ are already included in ‘short cuts and abbreviations’. Similarly, the category ‘prosodic spellings’ is included in ‘pragmatic and emotional resources’. The other three types of non-standard spelling identified by Androutsopoulos (2000) are ‘colloquial spellings’, ‘regiolectal spellings’ and ‘interlingual spellings’ which could all be considered as ‘non-standard spellings in the categorization above. Androutsopoulos uses this taxonomy in the study of non-standard spelling in media texts in Germany but his work is also highly relevant for CMC in instant messaging. An important point is that when non-standard spelling is adopted in a collective way it can express social and cultural identity. Ling & Yttri (2002) highlight how the specific use of linguistic features and slang demarcates the boundaries between teenagers and their parents as well as strengthening group membership.

Apart from the description of the different types of textisms that can be used in CMC, another important consideration is the extent to which these resources are used.

Thurlow and Brown (2003) analyzed 544 messages provided by 135 Welsh university students and found textisms in almost 20 % of the total words. Baron (2004) analyzed 23 conversations among American university students and abbreviations were 0.3 % of the corpus and acronyms 0.8 %. Ling and Baron (2007) analyzed 191 text messages from 22 female American university students. They only reported about 5 % textisms in their corpus. Tagliamonte and Denis (2008) collected data from 71 individuals aged between 15 and 20 in Canada and reported that 2.4 % of the corpus were textisms. De Jonge and Kemp (2012) asked 53 Australian university students and 52 high school students to translate messages from standard English to the way they would use these messages in CMC and found that textisms accounted for 13–16 % of the words in the messages. However, the study conducted by De Jonge and Kemp (2012) was not based on spontaneous instant messaging but on translations made by participants from standard English to the way they would use the same utterances in CMC. In a more recent study, Drouin and Driver (2014) asked 183 American university students to transcribe five personal messages, The percentage of textisms for the whole corpus was 24 % of the total of 7161 words. Taken together, these studies show the diversity of results but also that textisms do not reach 25 % of the words in CMC.

CMC language has been considered impoverished because it is constrained by the physical conditions of the medium (Herring, Stein & Virtanen, 2013). As these physical conditions are the same or very similar all over the world, the characteristics of CMC language would by implication be similar in different languages, provided they are determined by technology. It is also possible that textisms are not only due to the limitations of screens, keypads or number of characters. According to Herring et al. (2013: 9), new technology “enables new kinds of participation, new kinds of fragmentation, and new ways of co-constructing meaning that transcend traditional notions of conversation, narrative, exposition and so forth”.

Most research has looked at the way the English language is used in CMC while research on CMC in other languages is more limited (see for example, Androutsopoulos, 2015; Danet & Herring, 2007; Dyers & Davids, 2015). The findings of research on languages other than English are not many but provide valuable information. For example, Danet and Herring (2007) discuss CMC in several languages other than English and explain some characteristics of CMC in Japanese. An interesting finding is the importance of playful performance in Japanese and the frequent use of emoticons. Androutsopoulos (2015) conducted a study among seven Greek students in Germany. It was an ethnographic study based on online observation in Facebook. The Facebook contributions shared a repertoire that included German, Greek, English and some elements of other languages. Students used both the Roman and Greek scripts in Greek. Androutsopoulos (2015) not only reported examples of mixing and switching but also abbreviations and shortening. Dyers and Davids (2015) reported a study conducted at Western Cape University in South Africa. Participants were 315 undergraduate students who were asked to submit five sent and five received SMSs as well as the context for each message. The analysis of the messages indicated that there were blends of English with Afrikaans, isiXhosa and Setswana. It was interesting to see that there were characteristics of texting in Afrikaans and isiXhosa but not in Setswana.

Herring (2010) and Herring et al. (2013) highlight the need to carry out more studies in languages other than English so as to see if the characteristics of CMC described for English are similar to those found in other languages. The study reported in this article contributes to filling this gap by looking at the linguistic features of a minority language, Basque, in CMC among teenagers.

The basque language

Basque is a non-Indo-European language of unknown origin which is classified as "vulnerable" according to UNESCO Atlas of the World's Languages in Danger (Moseley, 2010). Basque, a highly inflective ergative language, is spoken on both sides of the Pyrenees extending along the coast of the Gulf of Biscay, in Spain and France. Basque is the main language of instruction in the Basque Autonomous Community (BAC) where Spanish is the majority language (Gorter, Zenotz, Etxague & Cenoz, 2014). In this community English is taught as a third language. In the BAC about 32 % of the population (aged 16 and over) is bilingual in Basque and Spanish, another 17.4 % can understand Basque and is fluent in Spanish and 50.6 % is monolingual Spanish (Basque Government, 2012).

The town where the study was carried out, Azpeitia, is located in an area where Basque is widely used for everyday communication. Azpeitia is a town of 14,540 inhabitants (Eustat, 2011). For most inhabitants Basque is their first language (76.5 %), followed by those with Spanish as a first language (18.1 %). Early bilinguals (Basque-Spanish) are 3.9 % of the population and 2.7 % speak other languages as first languages.

Basque is widely used in Azpeitia. According to a recent study based on observation, Basque is used in public spaces in the town by 91 % of the children and young people (Soziolinguistika Klusterra, 2012). A regional variety of Basque is used for oral communication. This variety has some differences when compared to the standardized language used at school. The following two examples can illustrate this point:

Example 1

Standard Basque: *Joan egin behar dut*
Regional variety: *Jun inber det*
Translation: *I have to go*

Example 2

Standard Basque: *Ez dakit liburua non dagoen*
Regional variety: *Eztakit liburue nun daon*
Translation: *I don't know where the book is*

Besides phonetic changes in the way the words are pronounced (e.g. *joan* vs. *jun*; *dut* vs. *det*; *liburua* vs. *liburue*) the variety of Basque used in Azpeitia has some shortenings (*daon* for *dagoen*) and combinations (*inber* for *egin behar*). Amonarriz (2008), who reported the use of some textisms among Basque university students,

suggested that perhaps there were fewer additional shortenings among speakers of Basque regional varieties than in standard Basque because the shortenings and combinations already existed in the regional variety.

Exposure to English in the context of this study is, in most cases, limited to the study of English as a school subject and some additional English language classes in private language schools.

Mobile phones are widely available to teenagers in Azpeitia. Even though there is no specific data about their use in this town, the Spanish National Institute of Statistics (Instituto Nacional de Estadística, 2014) reported that over 95 % of teenagers aged 14–15 use computers in Spain and 85.6 % have their own mobile phone at age 14 and 90.3 % at age 15. The Basque Country is one of the communities with the highest internet use in Spain and these figures are likely to be higher among teenagers in Azpeitia (Instituto Nacional de Estadística, 2014).

The study

A study based on a language that is linguistically different from English, and in a context where there is very limited exposure to English outside school, can provide valuable information about the universality of the linguistic features of CMC. However, the extended use of English as a language of international communication could mean that teenagers may be exposed indirectly to textisms in English and use some of them in their own language. This study contributes to the analysis of CMC based on a naturalistic corpus collected from real conversations (see also Tagliamonte and Denis, 2008; Varnhagen et al., 2010). It aims at analyzing the characteristics of CMC in Basque among teenagers and the perception these teenagers have of their own use of textisms. Specifically, we address the following research questions:

1. Are the linguistic features of CMC among teenagers communicating in Basque similar to those described for English CMC? This research question aims to compare linguistic features described for textisms in English to those found in the Basque corpus collected for this study.
2. How do Basque teenagers perceive their own use of CMC? This research question takes participants' own views into account and looks at the reasons Basque teenagers give to explain the use of textisms. These reasons can potentially be similar or different from those described in the case of English.

Methodology

Participants

Participants were 299 students in the 3rd and 4th years of secondary education in the town of Azpeitia. The students were between 14 and 16 years old with a mean age of 14.79. The distribution of gender was 47.8 % male and 52.2 % female. The

participants consisted of the whole population of 14 to 16-year-olds in Azpeitia and were students at one of the three schools in the town. All the students had Basque as the language of instruction and they studied Spanish and English as school subjects.

The participants had a complex linguistic repertoire. They could speak standard Basque and the variety of Basque used in Azpeitia, they were fluent in Spanish and they were learning English. Ninety eight per cent of the participants had acquired Basque before they were 5 years old. Basque was the basis of their everyday communication in all contexts (inner thoughts 88 %; among siblings 84.4 %; at school 95.6 %; among friends 94 %; in social networking sites 91.6 %). Participants used the standardized variety of Basque in the school context, as a language of instruction, but interpersonal communication among themselves or in the town took place in the regional variety of Basque. For instance, 91 % of the participants said that they used the regional variety when communicating in social networks and 100 % of the CMC corpus collected for this research study was written in the regional variety.

Instruments

The instruments used to obtain the data were a background questionnaire, social media conversations and focus group discussions:

Background questionnaire This was used to obtain information about the knowledge and use of language in different contexts (family, friends, school, technology).

Social media conversations Participants were asked if they would volunteer to supply naturalistic conversations on Tuenti. This is a social networking site, similar to Facebook, which is very popular among Spanish and Basque adolescents. A total corpus of 54 dyadic conversations, in which teenagers interacted with each other, were obtained. These conversations were provided by some of the 199 participants but it was not possible to identify them because it was agreed that the conversations had to be anonymized from the very beginning. Some students decided to give the conversations directly to the researcher but others gave them to the class president. The 54 conversations were given either digitally or in print. The corpus had 7411 words and the conversations ranged between 23 and 433 words, the average length being 137.2 (SD 94.4).

Focus group discussions Six focus group discussions with students were carried out, one group per school and course. The size of the group for each of the discussions was between four and eight students. A total of 34 participants took part, 19 girls and 15 boys. Participants in the focus group discussions were chosen by the class president and other representatives of students from each class. The students participating in the Focus group discussions were not necessarily the ones who had provided the anonymized on-line conversations. The duration of these discussions was between 52 and 85 min, the average being 68 min.

Procedure

The students, families and schools were informed about the purpose of the research study and gave their consent. The textisms observed in each conversation were categorized according to the three categories given above: short cuts and abbreviations, pragmatic and emotional resources, and errors, as well as the specific sub-categories for English described above. As previously mentioned, the taxonomy used in this study is based on that proposed by Varnhagen et al. (2010) but takes into account other taxonomies as well (e.g. Grace et al., 2014; Kemp et al., 2014).

The focus group discussions were organized with participants from the different schools. The discussions were held out of the school context and informed consent was obtained from all individual participants and their parents. The discussions were transcribed and analyzed for content, paying specific attention to the reasons participants had to use textisms.

Results

The first research question aims to study the characteristics of CMC among teenagers communicating in Basque so as to analyze the similarities and differences with those of CMC in English. In order to answer this question, we identified the linguistic features of the 54 dyadic conversations provided by participants and compared them to the categories in the taxonomy based on Varnhagen et al. (2010), Grace et al. (2014) and Kemp et al. (2014), described above. We then looked at the frequency of these features. The CMC features identified in our Basque language corpus compared to those in English can be seen in Table 1.

The results in Table 1 indicate that the taxonomy used for English also works for Basque. In fact, there are linguistic features for the three main categories: short cuts and abbreviations, pragmatic and emotional resources, and errors.

1. *Short cuts and abbreviations* Within the short cuts category, there were several types of linguistic features in the Basque corpus that were similar to those reported for English:
    - *Word combination* The combination of two or more words into one can be found in Basque CMC *eztakit* for *ez dakit* (I don't know)*; junber* for *joan behar* (I have to go). Combining words is a distinctive feature of the regional varieties of the Basque language such as the one used in Azpeitia, the town where the corpus for this research study was collected. Since the conversations are written in the Azpeitian variety of the Basque language, these combinations are very common. However, they do not necessarily have to be a characteristic of CMC but of the regional variety used in the town.
    - *Shortenings* Vowels are dropped in some words in the same way as has been reported for English: *nd* for *ondo* (good), *bstla* for *bestela (otherwise).*

Table 1 Features of CMC in Basque and English

CMC features	English	Basque
1. Short cuts and abbreviations		
Word combination	*wanna* for *want to*	*eztakit* for *ez dakit* (I don't know)
Shortenings	*tmrw* for *tomorrow*	*nd* for *ondo* (good)
		t for *eta* (and)
Acronym	*brb* for *be right back*	*zmz* for *zer moduz zaude* (How are you?)
Homophone spelling	*b* for *be*	*wapa* for *guapa* (pretty)
Omission of "h"		*emen* for *hemen* (here)
2. Pragmatic and emotional resources		
Pragmatic lengthening	*Hellooo*	*Kaixooo* (hello)
	love love	
Upper case	*WHAT* for *What*	*ZER for Zer* (What?)
Emotion words	*Hahahah*	*Jajajajaj*
Emotion punctuation	*!!!!*	!!!!
3. Errors		
Typographical error	*carzy for crazy*	*iksui* for *ikusi* (to see)
Misspelling	*progect for project bob for Bob*	*azpaldi* for *aspaldi* (long ago)

- *Acronyms* Both in English and Basque, the initial letters are used to form the acronyms as can be seen in *zmz* for *zer moduz zaude* (How are you?).
- Homophone spellings. Some cases of homophone spellings were found in the data such as *wapa* for *guapa* (pretty) where the initial "gu" has been replaced by "w". "Guapa" is a word in Spanish but it is common to use some specific Spanish words when communicating in Basque.
- Omission of "h". The omission of "h" in CMC has also been reported for English (Tagg, 2012) and can be found in words like *wat* for *what* which is considered to be a homophone spelling. This feature has been dealt with separately here because of its special characteristics and the very high frequency in our data. Unlike English, Basque is a transparent language with phoneme-to-spelling correspondence. The "h" is pronounced in the varieties of Basque used in the Northern Basque Country, in France but not in Spain where it is mute.

2. *Pragmatic and emotional resources* The resources found in our corpus were similar to those reported for English:

- *Pragmatic lengthening* This pragmatic/emotional resource is used both in English and Basque in a very similar way: *Hellooo* for *hello* in English and *kaixoooo* for *kaixo* (hello) in Basque.
- *Upper case* This pragmatic/emotional resource is also used in both languages to give emphasis as if it were being spoken at a higher volume in oral language: *WHAT* for *What* or *ZER* for *Zer* (what?). The pragmatic/emotional force of this resource can be seen more clearly in the following example: *EZ ETSI!* for *Ez etsi!* (Don't give up).

- *Emotion words* This resource is also used both in English and Basque in a very similar way: *Hahahah* in English *Jajajajaj* in Basque. The difference in spelling is due to the different pronunciation of this emotion word in English and Basque.
- *Emotion punctuation* Both languages use this pragmatic/emotional resource (e.g., !!!!).

3. They can be found in English and Basque but there are some differences:

- *Typographical errors* These happen in both languages because they are related to ability in the use of the keyboard and the lack of editing. Some examples in Basque are *iksui* for *ikusi* (to see) *or huandi* for *haundi* (big).
- *Misspelling* This type of error is less common in Basque because of the phoneme-to-spelling correspondence. Apart from the "h", which we consider to be a short cut or abbreviation rather than a misspelling, some confusion between the letters "s" and "z" can be found: *hotzein* for *hotsein* (call). We are aware of the fact that the avoidance of capital letters (*'bob' for 'Bob')* can be an intentional technique associated with social identity but it was not common in our corpus.

Besides evaluating the types of linguistic features used in CMC by Basque teenagers, it is also important to see how often these features are used. Table 2 shows the percentages of each of the features identified in the corpus.

We can see that the most frequent characteristic among short cuts and abbreviations is the omission of the "h" (64.1 %). The strategy for economising on letters is to write the language as it sounds, instead of writing it respecting the orthographic rules. The most commonly used strategy among those which aim to imitate oral language is the emotion punctuation (26.8 %), while pragmatic lengthening is a strategy used almost as frequently (25.9 %). Word combinations are also quite common (11.3 %). Other textisms such as shortenings (5.0 %) and the use of emotion words (5.3 %) are not as common. The same can also be said for the remaining linguistic features studied.

Table 2 Percentages of CMC features among Basque teenagers

Short cuts or abbreviations	Word combination	11.3
	Shortenings	5.0
	Acronym	0.7
	Homophone spelling	0.2
	Omission of "h"	64.1
Pragmatic and emotional resources	Pragmatic lengthening	25.9
	Upper case	0.5
	Emotion words	5.3
	Emotion punctuation	26.8
Errors	Typographical error	1.5
	Misspelling	0.3

The second research question aims at analyzing the way Basque teenagers perceive their own use of CMC. Specifically, the analysis looks at the reasons for the use of textisms as discussed in the focus group discussions. The following categories for justifying their textisms were identified: CMC as different from school language, a need for shortening and abbreviations, becoming a habit, changes over time, oral language, belonging to the group, gender differences:

1. *CMC versus school language* Basque is the language of instruction at school, so the participants in the study are taught every subject (except Spanish language and English language) through this medium. The type of Basque used at school is “Euskara Batua” or standardized Basque. Participants use the Azpeitian variety of Basque in instant messaging and not standardized Basque. When asked about the reasons for this choice they clearly indicate that they choose to use the variety that is different from standard Basque used in the academic context, as can be seen in Extract 1.

   Extract 1

   Researcher *Pixket honei buruz geyo zeoze esatie ba al dakezute? Ze nik garbi ikusi detena da euskalkiye ibiltzezutela erabat.*
   (Could you say something else about this topic? Because I have seen that you always use the regional variety?)

   Student A *bai, formala egiye da igual eskoliekin lotze deula. Eta ordun nola eskolan ite deun ezteu nahi ezer jakitie ya horrekin.*
   (Yes, we link the standardized variety to school. So, as it is the language used at school, we don’t want to use it in our private lives)

2. *Need for shortenings and abbreviations* Participants believe that they need to use shortenings and abbreviations when using the computer or the mobile phone and that this is something associated with CMC. This may be because it is easier or more convenient as seems to be implied by Student B. Student C and D also confirm the idea of this need and how writing is different in CMC.

   Extract 2

   Student B *eske ordenadoran idazteakun edo mobilin mensaje bat bialtzezunin, ahal dezun motxena.*
   (You use the computer or the mobile phone so when you send a message, as short as possible)

   Researcher *bokala kendu…*
   (without vowels)

   Student C *bai, edo beste era batea idaztezu*
   (yes you write in a different way)

Student D *edo "h" gabe. Igual zuk, normalin zuk "h"ik eztezu idazten ordenadorin-da hola*
(or without "h", you don't write the "h" when using the computer)

3. *Becoming a habit* The idea that some types of texting can become a habit can already be seen with regard to shortenings in extract 2. However, this is even more obvious when students E and F discuss the lengthening of vowels as in *Kaixooo* (hellooo) in extract 3. They have got used to doing that and so they (go on doing) continue with this practice, though they do not seem to be completely aware of the way this habit was acquired.

Extract 3

Researcher *eta zeatiken itezute hori?*
(and why do you do that?)
Student E *ya ohitu in naiz, ni behintzet. Hasiko nitzen eunen baten da ya…*
(we got used to it, at least me. I started sometime and..)
Student F *eske ya jun itea bi aldiz zapaltzie teklie… jaja*
(it just goes like that, you press the key twice…jaja)

4. *Changes over time* Another interesting finding was that acronyms such as *zmz* for *zer moduz* (How are you?), similar to *lol* (*laughing out loud*) in English were used more when they had just started socializing via social media, but later they are only used a formula to start the conversation. There seems to have been a development in the use of these shortenings or abbreviations as can be seen in extract 4:

Extract 4

Student G *[…] txikitan… […] messengerrakin haste zeanin da, igual, "zmz" igual "zemouz" galdetzeko, baño oin ya ez, nik behintzat ez*
(When we were children, when you start using messenger, you may use "zmz" in order to say "how are you", but not now, at least I don't
Student H *ez* (I don't)
Researcher *ya haundiyek zeate ta*
(You are older now)
Student I *eske hori hasierako emoziyue da. Hasieran bat hastea ta, hola idazteik ordenadoran… baño gero ya pasau itezu.*
(That's the emotion of a beginner. You start writing that way on the computer… but as time goes by you stop writing that way)

5. *Oral language* Participants seem to be aware that CMC is a special way of writing and have some understanding of how it shares characteristics with oral language. It seems that at least for student L in extract 5, CMC is a way of writing that mimics oral communication and which seems to cause shortenings and lengthenings. It appears that there is a need to use pragmatic and emotional resources to compensate for the functions of intonation and gestures.

Extract 5

Student J *eta letrak luzau're…*
(and we lengthen the letters as well)
Researcher *eta letrak luzau, hori bestie*
(lengthening the letters, that's another thing)
Student K *eske "epa" bat hola bakarrik oso bordie sonau leike*
(if you just say "epa" it would be rude)
Student L *nola hitzeite deun bezela nahi deun idatzi, ba atea ite zaizu luzio eo motzo*
(as we want to write as we talk, you write the word in a longer or shorter way)

6. *Belonging to the group* The idea of membership seems to be important as we can see in the case of Student K in extract 5. To use just *epa* the colloquial Basque form for "*Hi*" would be rude and therefore something else, such as the lengthening of the vowels, is needed so that it is acceptable to other participants. This idea of caring about the reactions of others is also reflected in the use of emotion words to end the conversation as is explained in extract 6. According to student M, to end with "a kiss" is a sign of intimacy among the members of the group who share the conversation. This interpretation is confirmed by student N who uses the shortening *mx1* for *muxu bat.*

Extract 6

Student M *"muxu bat" hori tipikue da, hori konfiantzie bezela da*
("a kiss" is typical, it shows intimacy)
Student N *bai. "mx1*
(yes. "mx1")

7. *Gender differences* The use of goodbye kisses is widespread and has become a habit but there are differences between boys and girls. Girls use this emotional way to say goodbye but boys do not do so in conversations with other boys.

Extract 7

Student O *… bukaeran ohitura bezela "aio mx"*
(at the end as a habit "bye kiss")
Researcher *ohitura bezela, ez?*
(as a habit, am I right?
Student O *bai. Lagunai're bai "mxx" edo "mx"*
(yes. With friends "mxx" or "mx")
Researcher *eta zuek, mutilek'e bai hori? holako afektu erakuste holako zeaik?*
(And you boys as well? Do you also show affection like that?)
Student P *mutilen artin ez*
(not among boys)

Discussion

The study reported here analyzes the linguistic features of CMC in Basque instant messaging among teenagers and takes into account participants' views as they discuss their use of textisms. The study compares the linguistic features used in CMC in Basque with those reported for English in other contexts. It tries to find out if linguistic features of CMC are similar or different in the case of a minority language which is the main language of communication and the language of instruction in a context with very limited exposure to English.

The analysis of the corpus indicates that CMC in Basque shares many characteristics with those reported for English (De Jonge & Kemp, 2012; Grace et al., 2014; Ling & Baron, 2007; Varnhagen et al. 2010). In fact, the taxonomy based on Varnhagen et al. (2010) and its categories and subcategories can also be used for the Basque language. Basque shows the same underlying features that have been categorized as short cuts or abbreviations, pragmatic and emotional resources and errors. The specific subcategories such as word combination or emotion punctuation proposed by Varnhagen et al. (2010) have also been identified in Basque. The reasons for these similarities may be related to different factors. Instant messaging requires quick reactions and the use of keyboards has some limitations, so it can be expected that CMC in any language will have short cuts and abbreviations. There is a limited number of possibilities to make messages shorter and they are similar in both English and Basque so it is not remarkable that they are used in both languages. The data from the focus group discussions also shows how Basque participants consider it necessary for messages to be short. Time pressure and the technical characteristics of CMC may also explain some spelling errors.

Similarities in the use of pragmatic and emotional resources are more difficult to explain. Basque teenagers report that the lengthening of vowels is used to approximate oral language, while the use of emotion words serves to express intimacy within the group. These resources have been reported in other studies (De Jonge & Kemp, 2012; Grace et al., 2014; Varnhagen et al., 2010). Why do speakers of Basque use the same linguistic features (e.g. extra capitals, extra letters, emotion punctuation) for the same purposes as those reported in CMC in English? Why do they develop the use of these specific resources and not others? One possibility could be that there is some type of influence from English to Basque. Participants do not use English in their daily lives apart from in language classes at school but some type of indirect contact cannot be excluded. They may have seen some English textisms in a movie even if those movies are dubbed. It is also possible that the participants in this study have come into contact with other teenagers who have more exposure to English. These possibilities cannot be excluded, yet it is intriguing that even some trends in the use of acronyms are similar in different languages and contexts. The study by Tagliamonte and Denis (2008) shows that "the use of *lol* declines systematically according to age, with the younger individuals using it the most" in a study about teenagers in Canada. Similarly, one of the focus group discussions among Basque adolescents shows that *zmz* (how are you?) is seen as something adopted by younger individuals when they start using instant messaging.

In both contexts, teenagers outgrew some of the forms they once used because they considered them too childish. It is interesting that in both cases they are very widely used acronyms.

The similarities are related to the way teenagers develop their own social identity and they want to establish clear boundaries between themselves and their parents or their schools. As Tagg (2012, p. 19) points out: "*Respellings in texting can carry social meaning: they create a sense of spoken informality and intimacy, they signal deviance from the norm and they create an illusion at least of brevity*". The development of social identity in adolescence can share many characteristics in different societies. One of these characteristics is gender identity and it is interesting to see that in one of the focus group discussions, female participants acknowledge the use of more emotion words than male participants. This difference between male and female participants in CMC has also been reported for other contexts (Herring & Kapidzic, 2015; Kapidzic & Herring, 2011).

There are, however, some specific elements identified in the Basque corpus which are different from those reported for English. These differences are related to the frequency of the linguistic features, not to the type and can be explained by the characteristics of the Basque language and the use of a specific variety. For example, combinations of words are among the most common features in Basque, while they are not as common in other studies (Varnhagen et al., 2010). As we have already seen, this is explained by the fact that our participants use a regional variety of Basque which is used orally and has plenty of combinations. The use of regional varieties in writing has also been reported in other contexts (Tagg, 2012).

There are few examples of spelling errors and homophone spelling in Basque except in the case of the "h". Basque is a transparent language with phoneme-to-spelling correspondence and this may explain why there are fewer differences between standard writing and CMC in Basque than in English.

In sum, this study shows that there are more similarities than differences in the use of CMC in instant messaging in Basque when compared to English. The significance of this study lies in the naturalistic corpus collected from real conversations in Basque together with the focus group discussions that allow us to listen to teenagers' views on their own texting. In this study we add evidence about the universality of textisms in a context where exposure to English is very limited. The study also confirms that the use of textisms is not only related to the technological characteristics of instant messaging but also to linguistic characteristics and social identities (see also Grace et al., 2014). It is the combination of these factors that can explain the similarities and differences in instant messaging in teenage communication.

Acknowledgments We would like to acknowledge the funding by the Spanish Ministry of Economy and Competitiveness EDU2012-32191, EDU2015-63967-R and the Basque Department of Education, Research and Universities IT-362-10 (UFI 11/54).

References

Amonarriz, K. (2008). Gazte hizkerak - hizkera gazteak. *BAT Soziolinguistika Aldizkaria, 68*, 167–183.

Androutsopoulos, J. K. (2000). Non-standard spellings in media texts: The case of German fanzines. *Journal of Sociolinguistics, 4*, 514–533.

Androutsopoulos, J. (2015). Networked multilingualism: Some language practices on Facebook and their implications. *International Journal of Bilingualism, 19*(2), 185–205. doi:10.1177/1367006913489198.

Baker, R. K., & White, K. M. (2010). Predicting adolescents' use of social networking sites from an extended theory of planned behaviour perspective. *Computers in Human Behavior, 26*, 1591–1597.

Baron, N. S. (2003). Language of the internet. In A. Farghali (Ed.), *The Stanford handbook for language engineers* (pp. 59–127). Stanford: CSLI Publications.

Baron, N. S. (2004). See you online: Gender issues in college student use of instant messaging. *Journal of Language and Social Psychology, 23*, 397–423.

Basque Government. (2012). *Fifth sociolinguistic survey: Basque autonomous community, Navarre and Iparralde*. Donostia-San Sebastián: Basque Government, Ministry of Culture. http://www.euskara.euskadi.net/r59-738/en/contenidos/informacion/sociolinguistic_research2011/en_2011/2011.html. Accessed 22 April 2015.

Crystal, D. (2001). *Language and the internet*. Cambridge: Cambridge University Press.

Danet, B., & Herring, S. C. (2007). Multilingualism on the Internet. In M. Hellinger & A. Pauwels (Eds.), *Language and communication: Diversity and change. Handbook of applied linguistics* (vol. 9, pp. 553–592). Berlin: Mouton de Gruyter. http://ella.slis.indiana.edu/~herring/danet.herring.hb.pdf.

De Jonge, S., & Kemp, N. (2012). Text-message abbreviations and language skills in high school and university students. *Journal of Research in Reading, 35*, 49–68. doi:10.1111/j.1467-9817.2010.01466.x.

Drouin & Driver. (2014). Texting, textese and literacy abilities: A naturalistic study. *Journal of Research in Reading, 37*, 250–267. doi:10.1111/j.1467-9817.2012.01532.x.

Dyers, C., & Davids, G. (2015). Southern African linguistics and applied language studies post-modern 'languagers': The effects of texting by university students on three South African languages. *Southern African Linguistics and Applied Language Studies, 33*(1), 21–30. doi:10.2989/16073614.2014.999994.

EUSTAT. (2011). *Azpeitiko datu estatistikoak*. http://eu.eustat.es/ci_ci/municipal/datos_estadisticos/azpeitia_e.html#axzz2lypOPGls. Accessed 18 May 2015.

Gorter, D., Zenotz, V., Etxague, X., & Cenoz, J. (2014). Multilingualism and European minority languages: the case of Basque. In D. Gorter, V. Zenotz, & J. Cenoz (Eds.), *Minority languages and multilingual education: Bridging the local and the global* (pp. 278–301). Berlin: Springer.

Grace, A., Kemp, N., Martin, F. H., & Parrila, R. (2014). Undergraduates' text messaging language and literacy skills. *Reading and Writing, 27*, 855–873. doi:10.1007/s11145-013-9471-2.

Herring, S. C. (1996). Introduction. In S. C. Herring (Ed.), *Computer-mediated communication: Linguistic, social and cross-cultural perspectives* (pp. 1–10). Amsterdam: Benjamins.

Herring, S. C. (2010). Computer-mediated conversation: Introduction and overview. *Language@internet, 7*. http://www.languageatinternet.org/articles/2011/Herring Accessed 21 May 2015.

Herring, S. C., & Kapidzic, S. (2015). Teens, gender, and self-presentation in social media. In J. D. Wright (Ed.), *International encyclopedia of social and behavioral sciences* (2nd ed., pp. 146–152). Oxford: Elsevier.

Herring, S. C., Stein, D., & Virtanen, T. (2013). Introduction to the pragmatics of computer-mediated communication. In S. C. Herring, D. Stein, & T. Virtanen (Eds.), *Handbook of pragmatics of computer-mediated communication* (pp. 3–31). Berlin: Mouton.

Instituto Nacional de Estadística. (2014). *Encuesta sobre Equipamiento y Uso de Tecnologías de Información y Comunicación en los Hogares*. http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t25/p450/base_2011/a2014/&file=pcaxis. Accessed 25 July 2015.

International Telecommunication Union. (2015). *ICT. Facts and figures. The world in 2015*. http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2015.pdf. Accessed 25 July 2015.

Kapidzic, S., & Herring, S. C. (2011). Gender, communication, and self-presentation in teen chatrooms revisited: Have patterns changed? *Journal of Computer-Mediated Communication, 17*, 39–59.

Kemp, N., Wood, C., & Waldron, S. (2014). Do I know its wrong: Children's and adults' use of unconventional grammar in text messaging. *Reading and Writing, 27*, 1585–1602. doi:10.1007/s11145-014-9508-1.

Lenhart, A., Purcell, K., Smith, A., & Zickuhr, K., (2010). *Social media and mobile Internet use among teens and young adults*. Pew Internet and American Life Project. http://pewinternet.org/~/media//Files/Reports/2010/PIP_Social_Media_and_Young_Adults_Report_Final_with_toplines.pdf. Accessed 17 March 2012.

Ling, R., & Baron, N. (2007). Text messaging and IM: Linguistic comparison of American college data. *Journal of Language and Social Psychology, 26*, 291–298. doi:10.1177/0261927X06303480.

Ling, R., & Yttri, B. (2002). Hyper-coordination via mobile phones in Norway. In J. E. Katz & M. Aakhus (Eds.), *Perpetual contact: Mobile communication, private talk, public performance* (pp. 139–169). Cambridge, UK: Cambridge University Press.

Moseley, C. (Ed.) (2010). *UNESCO interactive Atlas of the world's languages in danger*. http://www.unesco.org/languages-atlas/es/atlasmap.html. Accessed 05 September 2015.

Prieto-Arranz, J. I., Juan-Garau, M., & Jacob, K. L. (2013). Reimagining cultural identity: Transcultural and translingual communication in virtual third-space environments. *Language, Culture and Curriculum, 26*, 19–35. doi:10.1080/07908318.2012.

Soziolinguistika Klusterra. (2012). *Hizkuntzen kale-erabileraren VI neurketa*. Andoain: Soziolinguistika Klusterra.

Tagg, C. (2012). *Discourse of text messaging*. London: Continuum.

Tagliamonte, S. A., & Denis, D. (2008). Linguistic ruin? LOL! Instant messaging and teen language. *American Speech, 83*, 3–34.

Thurlow, C., & Brown, A. (2003). Generation Txt? The sociolinguistics of young people's text-messaging. *Discourse Analysis Online, 1*. Retrieved from http://extra.shu.ac.uk/daol/articles/v1/n1/a3/thurlow2002003-t.html

Thurlow, C., Lengel, L., & Tomic, A. (2004). *Computer mediated communication: Social interaction and the internet*. London: Sage.

Varnhagen, C. K., MacFall, G. P., Pugh, N., Routledge, L., Sumida-MacDonald, H., & Kwong, T. E. (2010). LoL: New language and spelling in instant messaging. *Reading and Writing, 23*, 719–733.