



Universidad de Valladolid

FACULTAD de FILOSOFÍA Y LETRAS
DEPARTAMENTO de FILOLOGÍA INGLESA
Grado en Estudios Ingleses

TRABAJO DE FIN DE GRADO

From texting to Internet Language. Analysis of
contemporary language evolution. The case of
YouTube

Daniel Marrón Fernández de Velasco

Tutor: Laura Filardo Llamas

2014-2015

ABSTRACT

Thousands of years have passed and language continues to evolve. New technologies have massively contributed to the expansion, both in terms of usage and quantity, of Internet Language communication that has as its main purpose the economy of the language. Taking as a starting point text-messaging and its main representative features, this essay will try to explain how Internet Language is changing the way we communicate using new technologies. It will deal with YouTube comments, and we will extract non-standard English vocabulary and classify the words in different categories to study their internal functionality. Hypotheses on phonetic causes or even historical explanations will be covered in this essay to add a possible reasoning of why users, consciously or unconsciously, create certain abbreviations or variations of words. Also, a discussion on the differences between uses in 2008 and in 2015 will be presented for future hypotheses.

Keywords: Internet Language, phonetics, historical explanation, abbreviations, creativity, language economy

Han pasado miles de años y el lenguaje sigue evolucionando. Las nuevas tecnologías han contribuido masivamente a la expansión, tanto en uso como en cantidad, del lenguaje de internet que tiene como objetivo principal la economía del lenguaje. Tomando como base los mensajes de texto vía teléfono móvil y teniendo en cuenta sus características, este trabajo tratará de explicar cómo el lenguaje de internet está cambiando la manera de comunicarse usando las nuevas tecnologías. El material de trabajo consistirá en comentarios de YouTube, a través de los cuales extraeremos el vocabulario no reglado del inglés y clasificaremos las palabras en distintas categorías para estudiar su funcionalidad. Se tratarán, asimismo, hipótesis en relación a las causas fonéticas y explicaciones históricas añadiendo de este modo un posible razonamiento al por qué los usuarios crean determinadas abreviaturas o variantes de palabras consciente o inconscientemente. También se establecerá una diferenciación entre los usos identificados en 2008 y 2015 para futuras hipótesis.

Palabras clave: Lenguaje de Internet, fonética, explicación histórica, abreviaturas, creatividad, economía del lenguaje

INTRODUCTION

In this essay we can see a connection between the contents covered in it and the courses developed in these four years of English Studies. As we have dealt with phonetic causes of non-standard English vocabulary, the most relevant courses have been English Instrumental Language I and English Phonetics and Phonology. Another important course has been Fundamentals on English Language History as part of a historical explanation to the evolution of certain words. Finally, English Language: Applied Linguistics II has helped us with ideas for the creation of a corpus.

With these courses in mind, we have tried to provide an analysis on different changes in the evolution of language. Language has experienced different changes from its creation, which dates from around 80,000 years ago in an oral form and 5,500 in a written form. This means that writing is a tool based on how people speak changing it into a deliberate and slow means of communication (McWhorter 2013). Both communicative systems –written and oral– have changed in different periods of history depending on the purposes and contexts within which they are used.

Nowadays, the English language is considered the most relevant and global one (Crystal 2005). As it happens with other languages, English is experiencing a process of change influenced by new technologies. Computers, mobile phones, the Internet, social media; they are all contributing to increasing the use of non-standard English vocabulary. Specially, text-messaging has played an important role due to its internal characteristics. When text-messaging first appeared, there were several constraints for the user such as a limitation on the number of characters in a single message, the keyboard system that implied a time-consuming way of writing, and the cost of each message. Users had to experiment with language to avoid or to minimize those limitations by creating abbreviations or variations of words, which were easily accessible in a small keyboard.

More recently, the evolution of technology has made computers and mobile phones closer to each other in terms of usage. For example, the QWERTY system has facilitated writing by increasing speed and key accessibility. Also, the creation of applications such as 'Whatsapp' or 'Viber' has made instant messaging free and there is no longer a character limit. Finally, by having Internet on the mobile phones, we have access to some of the same

platforms that can be accessed in a computer. For these reasons, we could say that text-messaging is gradually changing into Internet Language (IL), which we understand as the language uploaded or sent via the Internet with a standard keyboard or a QWERTY system.

Text-messaging and IL have contributed to the creation of a new way of writing in which oral communication coalesces with written communication. Mobile phones and the Internet have provided a tool for a type of communication in which formality is not a requirement and in which instant messages have broken the barrier between sender and receiver in terms of time-lapse. Standard and on-line written conversations, which were slow and followed English grammar and structure, are gradually changing into person-to-person-like conversations, which are fluid and break grammatical and structural rules. To this, we should add the accommodation of language to this technology and the necessity to avoid superfluous letters in order to transform the disadvantages of key accessibility and character limit into advantages. This means that pronunciation has become one of the main sources users have relied on in order to adapt how they say words to how they write them. In the words of John McWhorter (2013) "texting [and IL are] developing [their] own kind of grammar and conventions."

This essay aims to study IL by analyzing non-standard English vocabulary in a corpus compiled in 2015. In order to do this, we will explore previous analyses on the subject trying to update the reasons of use and the internal characteristics of text-messaging. We will reflect on phonetics applied to IL to study the connection between what has been written, how it is normally written and how it is pronounced. Also, we will try to establish a relation between words and their historical evolution to identify other possible reasons for the use of nonstandard words in IL.

METHOD

In 2008, linguist David Crystal published a book called *Txtng: The gr8 db8*. In it, he deals with the concerns, characteristics, and reasons that can explain how we use language with new technologies. Specifically, he writes about text-messaging. In chapter 3, Crystal (2008: 35-62) analyzes what he calls distinctive features that explain how texts are written. These distinctive features are divided into six main categories: logograms (1), pictograms

(2), initialisms (3), omitted letters (4), nonstandard spellings (5), and shortenings (6). This essay will use these categories when analyzing the words that form the corpus we have created to analyze IL.

Logograms (1) are words, syllables or noises –associated with actions–, which are represented by letters, numbers, or symbols. In order to analyze this type of category, the focus should be narrowed to the pronunciation of the letter, number or symbol instead of the writing. For instance, the word 'be' is represented by the letter and the word 'at' by the symbol '@' (Crystal 2008: 37). However, logograms are more than words represented by one letter, number or symbol. They can be the combination of more than one letter, number, or symbol; or they can also be the combination of a letter, number, or symbol plus a syllable or part of a syllable. Some examples are the words 'before' as 'b4' –combination of a letter and a number–; 'atoms' as '@oms' –combination of a symbol and part of a syllable–; 'today' as '2day' –combination of a number and a syllable–; and the action of sleeping as 'zzz' –repetition of the same letter (Crystal 2008: 38).

Pictograms (2) refer to the representation of objects, concepts, or faces, which stand for emotions or characters, among other things. The focus in this type of category is in the shape of what is written. Crystal (Crystal 2008: 38), provides some examples, such as @(- - -` - - -` - - -` - - -` representing a rose, :-@ representing a face screaming, or (*o*) representing a surprised face. Pictograms present similarities with logograms as they can use the same symbols. However, they are different and they can be distinguished from one another. For example "[t]he @ in the word @om and the @ in the emoticon of 'screaming' [...] perform completely different functions. In the first case, you listen to the sound; in the second case, you look at its shape" (Crystal 2008: 39).

Initialisms (3) consist on reducing words to the first letter of each word; that is, generating acronyms. They can appear both in uppercase or lowercase. There are several combinations in which initialisms can appear: single words, compound words, phrases, elliptical sentences, whole sentences, and exostulations. Some examples are: N for no, GF for girlfriend, FTF for face to face, DK for don't know, SWDYT for so what do you think?, or YYSSW for yeah, yeah, sure, sure, whatever! (Crystal 2008: 42).

Omitted letters (4) are shortened words written without one or more letters in the middle or at the end of those words. Normally, the omission is produced by avoiding the use of a vowel, yet the consonants may be also omitted in this process. Also, one word can be written in different ways by omitting different letters. An example would be 'pleasd', 'plsed', 'plsd', etc. There are no fixed rules, which provides this category with a big amount of variations. In Crystal's book we can find different examples of omitted letters, such as the word 'please' and its shortened version 'plsed', where two vowels have been omitted. In the word 'coming' the final letter <g> may be elided –comin. Likewise, in the word 'getting' we can omit the final <g>, the <i> and reduce the double consonant –gtn. Not common but also possible is the omission of a consonant at the beginning of a word when it is not pronounced, as we can see in the word 'write' –rite (Crystal 2008: 46).

Nonstandard spellings (5) derive from the misspelling of a word. This can be consciously or unconsciously done. According to Crystal (2008: 48), nonstandard spelling is a conscious device since writers would not use new technologies if they did not know how to read and write. Some examples are the words 'should' –shud–, 'school' –skool–, 'what' –wot–, 'going to' –gonna–, or 'been' –bin, bn (Crystal 2008: 48-49). In these nonstandard spellings we can see how, for example, in the word 'school' 'ch' is changed to 'k' because, in this case, the pronunciation is the same: /k/. In the word 'what' the letter 'h' is omitted since it is not pronounced, and 'a' is changed to 'o' because of the similarity in their pronunciation –/wat/ for 'what' and /wɒt/ for 'wot'.¹

Finally, shortenings (6) are abbreviations produced by dropping one meaningful element. Normally, the omitted element is found at the end of the word but it can also occur at the beginning. For instance, 'day' is occasionally omitted in the days of the week –mon or sat. The endings of some months are also omitted –jan or dec. Other examples are the words 'university' –uni–, 'doctor' –doc–, 'approximately' –approx–, or 'telephone' –phone–, where the beginning of the word is not written (Crystal 2008: 51).

Using these categories, we will be able to see if they are applicable to the study of IL. Even if on the Internet, there are several platforms we can use to communicate with others such as forums, chats, or YouTube, we had to consider a number of issues before

¹ All transcriptions are extracted from *Longman's dictionary* (1995) using American English transcriptions

deciding to collect our data from YouTube. On the one hand, forums usually request users to write following grammatical rules. Also, not all forums are easily accessible: a user may need an invitation, or to register or participate in order to access more information. On the other hand, chats do not record the information, so previous conversations cannot be accessed. On the contrary, YouTube stores all the comments and it is easily accessible. This makes YouTube a tool used by all kind of users covering different ages, nationalities, levels of education, etc.; this decreases possible biased results.

In order to collect the data for the analysis presented in this essay, we established some criteria. The first criterion was to use the most viewed videos of YouTube. To access them, we searched on Google "most viewed youtube videos" and we found a list given by YouTube. Departing from this list (Table 1), we tried to distance ourselves from selecting videos that could help us with our research, while at the same time we thought that, since these videos are the most viewed ones, through them, we could cover all type of users' profiles. The second criterion was to use English YouTube comments no matter what the mother tongue of the users was. The third criterion was to narrow the year of the comments to 2015 because of the difficulty to access previous comments. YouTube lets the user load a few comments at a time and after a certain number of loaded comments, in two different videos, it did not let us load more comments. Also, as we can see in Table 1, the most viewed video, which was uploaded in 2012, has more than five million comments. Trying to access comments of 2012 would have been too time-consuming and it would exceed the scope of this essay. The fourth criterion consisted on selecting only YouTube comments that contained at least one word with one of the categories described above per comment. Only in the first month of 2015, each of these videos had more than one thousand comments, including those with and without distinctive features. Creating and analyzing a corpus with thousands of comments would need a longer period of time and would also lie outside the scope and requirements of this essay. The last criterion was to narrow the length of the comments to a maximum of 160 characters to facilitate the analysis and to make a parallelism between the length of text-messaging and YouTube comments.

	Video	Views	Comments
1	PSY - Gangnam style	+2 billion	+5 million
2	Justin Bieber - Baby ft. Ludacris	+1 billion	+6 million
3	LMFAO - Party Rock Anthem ft. Lauren Bennet, GoonRock	+800 million	+600 thousand
4	Charlie bit my finger - again !	+800 million	+600 thousand
5	Eminem - Love The Way You Lie ft. Rihanna	+800 million	+600 thousand
6	Jennifer Lopez - On The Floor ft. Pitbull	+800 million	+400 thousand
7	Katy Perry - Dark Horse (Official) ft. Juicy J	+800 million	+300 thousand
8	Shakira - Waka Waka (This Time for Africa) (The Official 2010 FIFA World Cup™ Song)	+800 million	+700 thousand

Table 1: Most viewed videos according to YouTube from February 2nd, 2015 to February 4th, 2015

ANALYSIS

In the analysis section, we have tried to follow Crystal's (2008) classification of the most common processes that can be found in text-messaging and apply it to the study of IL. The findings can be summarized as follows. The analysis shows a tendency to use IL with an average of a 20%. These results have been obtained after analyzing 4510 words taken from 521 YouTube comments, as summarized in Table 2. This shows that the analyzed comments, all of which present at least one category, contain an average of 8.6 words and 1.7 distinctive features per comment.

	Video	Comments	Words	Distinctive features	Percentage of use
1	PSY - Gangnam style	63	546	122	22.34%
2	Justin Bieber - Baby ft. Ludacris	70	604	139	23.01%
3	LMFAO - Party Rock Anthem ft. Lauren Bennet, GoonRock	86	719	136	18.91%
4	Charlie bit my finger - again !	52	453	79	17.43%
5	Eminem - Love The Way You Lie ft. Rihanna	80	668	132	19.76%
6	Jennifer Lopez - On The Floor ft. Pitbull	30	232	46	19.82%
7	Katy Perry - Dark Horse (Official) ft. Juicy J	59	599	94	15.69%
8	Shakira - Waka Waka (This Time for Africa) (The Official 2010 FIFA World Cup™ Song)	81	689	144	20.89%
		521	4510	892	19.77%

Table 2: General statistics

By analyzing the corpus, we have been able to extract all the categories and classify them (Table 3) to be able to see which one is the most commonly used, with which percentage they are used, how many times they appear, and the relation between them in

terms of usage. We have also been able to see how and why these categories are created and used. As a general rule, the different categories seem to share one main reason for using them: language economy; that is the omission of those parts that could be considered as superfluous.

Initialism	Logogram	Nonstandard spelling	Omitted letter	Pictogram	Shortening	Others
257	227	191	82	81	51	3
28.81%	25.44%	21.41%	9.19%	9.08%	5.71%	0.33%

Table 3: Categories

We have found that initialism is the most significant category. This has a correlation with language economy, as initialism is the category with the highest number of omitted letters. It is important to mention that there is not a lot of variation amongst the examples, which could be narrowed to four repeated uses: 'lol' instead of 'laughing out loud' (example 1.8)², 'omg' instead of 'oh my god' (example 1.11), 'wtf' instead of 'what the fuck' (example 1.22), and 'lmao' instead of 'laughing my ass off' (example 1.61). Some of these examples, as well as other initialisms, may present genuine novelties. A genuine novelty is "a form of language play - the desire to 'up the ante' and outdo what has been done before." (Crystal 2008: 53) Our corpus includes 'omfg', which means 'oh my fucking god' (example 3.32) and its the genuine novelty of 'omg'. These four features and their novelties represent a 76% of the total findings. One of the possible reasons that may explain this percentage derives from the globalization caused by the Internet and the communication between languages that it offers through chats, forums, YouTube, and other sites. For example, in YouTube comment "Idk why people dislikes this vídeo :/" (example 4.47), there seems to be a native Spanish speaker writing in English the initialism 'idk' -I don't know. The native language of this user can be inferred from the accent in the word "vídeo." On the Internet, other languages are borrowing English initialisms –among other abbreviations– to represent the same idea as the original, such as 'lol', which is found in several languages (Crystal 2008). Moreover, as English is considered a global language, other languages use not only abbreviations but also complete words that are shorter than the ones in the source language. Using Crystal's words (2008: 129-130), "I have seen *now* in place of *maintenant* in French

² All examples are included in the corpus, which can be seen in the CD attached to this essay

– which is certainly a time-saver! [...]. As most parts of the world are bilingual, I would expect code-mixing to develop to be a major characteristic of texting", and therefore of IL.

The majority of the initialisms that we have found are characterized by their three-letter format, as we can see in 'lol'. However, there are examples of initialisms with more or less words, and we have even found one-word initialisms. Within four-letter initialism, we have found 'idek' meaning 'I don't even know' (example 7.15). A five-letter initialism would be 'idgaf', which means 'I don't give a fuck' (example 8.38). Less than three letters can be found in two-word initialism 'af' for 'as fuck' (example 5.9) or the one-word initialism 'b' for 'billion' (example 1.47).

After analyzing each example, we may conclude that initialisms do not seem to follow an arbitrary pattern. With the exception of two examples, which will be discussed below, initialisms, in this corpus, always have a phonetic cause. If we consider each word in isolation and not as connected speech, all the words used begin with a primary stress. This gives prominence to the first letter of each word, which will represent the distinctive feature. It seems that users have established as an unconscious condition for creating initialisms that each word contains a primary stress in the first syllable. The reason why the words need to be treated in isolation is because connected speech alters the pronunciation pattern of words to follow the English pattern of stressed-unstressed-stressed syllables. Connected speech may also alter the nominal or alternative version of vowels in non-functional words changing them into schwas and thus losing their prominence. The nominal version of a vowel refers to a long vowel and the alternative version refers to a short vowel. Also, the nominal version coincides with the name of each vowel in the alphabet, as for example the vowel <a> pronounced in English as /eɪ/ (Cámara 2010: 27). An example of a word that changes depending on whether it appears in isolation or in connected speech is 'and', which has two transcriptions: /ænd/, containing the alternative version of the vowel <a>, and /ənd/, with the unstressed version.

Both of the exceptions refer to the same initialism which has two meanings in our corpus: 'bc' as because and 'bc' as Before Christ. The first one is an initialism of a single word with the following characteristic: the primary stress of this word is in the second syllable and thus it does not follow the previously mentioned hypothesis. The first syllable

is unstressed but with the peculiarity of adding a /ɪ/ instead of a schwa (/bɪ'kɔz/). As discussed below, this characteristic will be repeated in several cases. The second exception does not follow the same rule, as it is the only initialism in the corpus prior to the new technological era: specifically the phrase 'Before Christ' dates from the 1660s and its abbreviation from 1823³.

To make sure that the hypothesis of the primary stress in the first syllable could be right, we have also analyzed all the initialisms provided by Crystal (2008). The results have confirmed this pattern of primary stress in the first syllable. Moreover, by analyzing those examples that do not follow this rule, we have found a similarity with the first use mentioned for 'bc' (because). Words used in initialisms with a primary stress in the second syllable, when the previous letter is an <e> or an <o>, can be used in initialisms if they are not transcribed as a schwa (Table 4). However, there are exceptions as for example the initialisms 'cid' (Crystal 2008: 191) or 'imo' (Crystal 2008: 193). They stand for 'consider it done' and 'in my opinion' respectively. The words 'consider' and 'opinion' break the /ɪ/ rule as <con> and <o> are unstressed syllables transcribed as schwas –/kən'sɪd·ər/, /ə'pɪn·jən/.

Initialism	Meaning	Analyzed word	Transcription
aisb	as I said before	before	/bɪ'fɔr/
asl	age, sex, location	location	/ləʊ'keɪ·ʃən/
bion	believe it or not	believe	/bɪ'li:v/
dur	do you remember	remember	/rɪ'mem·bər/

Table 4: examples from Crystal's corpus

All in all, and using statistics both from the corpus created for this essay and the corpus provided by Crystal, we can conclude that initialisms with primary stress in the first syllable are used around 87% of the times. Initialisms with primary stress in the second syllable, when the previous letter is a <e> or a <o>, are used around 8% of the times. Finally, initialisms that do not follow any of the rules mentioned above are used around 5% of the times. These statistics are extracted from a total of 109 different initialisms.

Maybe the most remarkable example that we have found with initialisms is 'kek' (7.23). To know the meaning of this initialism, as well as other examples, we have used the

³ All etymological explanations are extracted from the *EtymoOnline Dictionary* available at www.etymonline.com

*Urban Dictionary*⁴. It is not considered a highly academically source but since IL is evolving and changing faster than standard dictionaries, we had no chance but to use it because it is highly updated. According to the *Urban Dictionary* 'kek' refers to a word used in the video game *World of Warcraft*. In it, there are two factions: the Horde and the Alliance and they use different languages so that they cannot be understood by the other faction. However, there exist certain patterns in the languages, and when a horde writes 'lol', an alliance reads 'kek'. Here, there is another example of language play and how writers try to update, change, or outdo what has been done before.

Although initialisms show certain characteristics differentiating them from the rest of the categories, logograms share some similarities with initialisms. The use of logograms is closely related to the one of initialisms with only a 3% difference of use. In total, these two categories represent more than 50% of all the IL examples analyzed in this essay. The reason, again, seems to be language economy as all words are reduced to one letter, one number, two letters, or one letter and one number. As it happened with initialisms, with few examples we can cover the majority of variants. These examples are reduced to two letters in three combinations: 'u', 'r', and 'ur'. The repetition of these logograms represents a 74% of the total findings in the corpus.

The phonetic cause is also noticeable in the case of logograms. In the examples, we can observe substitutions of a letter for a word (1.10) or part of a word (7.30) by its phonetic representation, which usually coincides with the pronunciation of a letter, a number, or a symbol. More details of this phenomenon can be seen in example 1.10 ("[I]ike if u just came here to see the amount of views") where the letter 'u' appears in isolation substituting the personal pronoun 'you'. To explain the reason why this happens, we have to see their transcription. The letter 'u' is transcribed as /ju/ and the word 'you' is also transcribed as /ju/.

In example 7.30 ("[h]ey all u fucks, she admitted selling her soul because her career wasn't going well b4. It's illuminati. Now shut the fuck up"), the letter 'b' appears with the number '4'. This combination substitutes the word 'before'. Again, a transcription shows the reason for its use: the letter 'b' is transcribed as /bi/, the number 4 is transcribed as /fɔr/, and

⁴ The Urban Dictionary is available at www.urbandictionary.com

the word before is transcribed as /br'fɔr/. Notice that there is a slight change in the pronunciation of the 'b' when it is combined with the second syllable. To understand this process, it is necessary to mention that /i/ only appears in open syllables that do not precede a consonant (reality: /ri'æɪ·ɪ·ti/) (Cámara 2010: 22).

Even if the phonetic cause may be timesaving, it may also create ambiguity in the writing. Table 5 shows ambiguous logograms to illustrate this. Nonetheless, "context, along with the place of the word in the sentence, invariably makes it clear whether a texter means *we'd* or *wed*, *can't* or *cant*, *we'll* or *well*, and so on." (Crystal 2008: 18-19). For instance, in the sentence "[I]ike if ur watching in 2015 and hating [...]" (example 1.17), the analysis of the syntactic structure solves the ambiguity. The word 'watching', treated outside this sentence, may function as a subject in the NP or as a verb in the VP. 'Watching' is preceded by 'ur', which means that 'watching' cannot function as the subject. Finally, we need a subject in nominative case (being 'you' and not 'your') and we need to complete the present continuous tense of the verb: 'you are/ you're watching in 2015' and not 'your watching in 2015'. Another example is the number '2' and the two words that it may refer to: 'two' and 'too'. In YouTube comment "looks like you had 2 much starbucks" (example 8.46) we need to know the meaning of the number. The surroundings of the word avoid the ambiguity: the verb 'had' permits both interpretations; however the word 'much' narrows the interpretation to the word 'too' since 'you had two much' is not an English structure.

Logogram	Meaning	Examples
Ur	<ul style="list-style-type: none"> You are Your 	<ul style="list-style-type: none"> Like if ur watching in 2015 and hating this just as much as you used to in 2012 (example 1.17) napalm you wasted ur time typing that (example 3.21)
2	<ul style="list-style-type: none"> Two Too 	<ul style="list-style-type: none"> Over 2 fucking billion views.... LOL almost the whole world have seen it (example 1.48) Looks like you had 2 much starbucks (example 8.46)

Table 5: Ambiguity in logograms

The two categories already analyzed share a particularity: Even if they are the most commonly used ones, they present a low amount of different words with the features needed to be initialisms or logograms. These categories are divided into less than 50 variants with 484 repetitions. Unlike these, nonstandard spelling is a category with around 70 variants and 191 repetitions. While the most frequently-used initialism and logogram are repeated 99 and 120 times respectively, the most repeated nonstandard spelling appears

only 13 times. These numbers make a contrast as nonstandard spelling is repeated only around a 4% less than logograms and 7% less than initialisms. The reason why this happens seems clear. Initialisms divide each sentence, clause, or word into the first letter of each word or syllable (in the case of some single words), logograms divide each word into its minimum phonetic representation. For this reason, both distinctive features present limitations when reducing words; that is, one sentence, clause, or word has only one representation. Something different happens with nonstandard spelling. One word may have several representations, which may be used for different reasons, as we can see in Table 6.

Standard spelling	Nonstandard spelling
You are	Your; youre; ure
Because	Cuz; 'cuz; bcuz; cos; coz; cz
Please	Plz; pliz

Table 6: different representations of one single example

One of the main reasons justifying the use of nonstandard spelling is the need to abbreviate a word. To exemplify this, we could mention the word 'thanks' –with the nonstandard spelling 'thx' (example 3.68)– or the word 'please' –with the nonstandard spelling 'plz' (example 3.84). As seen in these examples, users change part of a syllable for a letter representing the sound of that part of a syllable. But, they also abbreviate a word by omitting the apostrophe both in the subject+verb contraction and the Saxon genitive. It is important to remember that nonstandard spelling may be consciously or unconsciously done. According to Wray (1995), 72% of undergraduate English students make mistakes in the use of the apostrophe. For this reason, apostrophes may be the most difficult example to analyze in terms of the will of the writers to abbreviate. In contrast, Crystal (2008: 18) considered that a study carried out by Thurlow and Borwn (2003) "should give some reassurance to members of the Apostrophe Protection Society". Thurlow and Brown analyzed 544 messages from which there were a total of 192 instances of the apostrophe (one out of three messages contained one). As the total number of samples is similar to the ones in our corpus, we have decided to see if Thurlow and Brown's findings with text-messaging could be extrapolated to IL. The results have been that the apostrophe in the subject+verb contraction and the Saxon genitive are used less than 100 times (almost one

out of five comments contains an apostrophe). Hence, we see a difference between both studies even if the reasons cannot be easily identified. In an attempt to clarify them, we have also analyzed the number of times that the apostrophe is omitted in our corpus being this number 60 times. The conclusion may be the following: As the number of times in which the apostrophe is used is higher than the number of omitted ones, it seems that the omission responds more to an abbreviated form, than to an unconscious misspelling. To this, we should add that a YouTube comment might contain at the same time constructions with apostrophes and constructions without them, as seen in example 8.7 ("whats ur problem?!?! -.- don't just talk cause they didn't want Ebola in the first place so u have no right to talk that way about Africans okay? -.-"). In this example, we can see how the user omits the apostrophe in "whats" (what's) but later this same user uses "don't" and "didn't".

In terms of the phonetic cause, we have found three different ways to reduce a word. The first one consists on writing according to the pronunciation of the words. This is the case of the apostrophe in examples such as 'its' (3.25) or 'whats' (8.7). As the apostrophe is not pronounced, it is omitted. However, an apostrophe may have a schwa pronunciation. Borrowing Sklar's (1976: 175) words, "[i]n the genitive of singular nouns ending in a fricative or an affricate, the apostrophe in some sense "stands in" for a pronounced /ə/ [...] as in *church's*". Apart from the omission of the apostrophe, we may include in this first way of reducing a word other examples such as the words 'could' –with the nonstandard spelling 'cud' (6.27)–, the word 'fuck' –with the nonstandard spelling 'fack' (1.2)–, or the words 'he is' –with the nonstandard spelling 'hez' (2.64). If we analyze the transcription of the words 'could', 'fuck' and 'he is' –/kʊd/, /fʌk/, and /hi+/ɪz/–, we can understand the processes that the writers have followed to connect writing and pronunciation.

The second way of reducing a word consists on writing a word according to the way in which it is often said instead of the standard pronunciation. Examples of this could be the nonstandard spellings 'gonna' (5.8), 'lemme' (4.22), or 'wanna' (5.31). These examples stand for 'going to', 'let me', and 'want to' respectively.

Finally, we can reduce a word by establishing a connection between writing and the closest sound of a letter in order to make the word shorter. To illustrate this explanation, we have the word 'that' and the nonstandard spelling 'dat' (see example 1.13). In this example

we have a voiced dental fricative (/ð/) and a voiced alveolar plosive (/d/) (Ripman 1931). As both have a 'd' sound and 'd' is shorter than 'th', both writings become interchangeable in IL.

We have also found a very unique example in terms of evolution: luv (see example 3.39). According to the *Urban Dictionary*, it is "[a] casual way of saying you really like someone without [...] saying I love you." However, the phonetic cause is highly noticeable and possibly made without the writers realizing the process. The word 'love' is transcribed as /lʌv/. The nonstandard spelling 'luv' is also pronounced as /lʌv/ if we follow the formula of <u> + C + // –being C a consonant and // the end of the word (Cámara 2010: 191). Notice how writers drop the final 'e' in 'luv' since <u> + C + <e> produces a /u:/ sound (Cámara 2010: 193) and thus it changes the original pronunciation. We could also give a conscious reason since the word 'love' is an exception in terms of pronunciation. According to Enrique Cámara (2010: 150) the <o> + C + <e> formula has a /oʊ/ pronunciation. Users may be expressing a logical writing for a /lʌv/ pronunciation. But, this example is also very unique because of its evolution throughout history. The word 'love' comes from Old English 'lufian' and Middle English 'luf' and 'lufe' (Fernández 1982: 647; 666). We can see different processes happening in this word: the first one is the change from <f> to <v> as Middle English phonology included voiced fricatives within consonants and not just as allophones as it had happened in Old English (Millward 1996: 147). The reason why it changed into a voiced fricative is because it is in an intervocalic position and it is affected by the voiced sound produced with the vowels. The second process, and the one that interests us, is the change of the <o>. In the 13th century, the vowel <u> changed into an <o> when it was followed by a nasal or a <v>, among other combinations (Fernández 1982: 123). If we now look at the Middle English word 'lufe' we could deduce the changes: the inclusion of the voiced fricative <v> joined with the intervocalic position of <f> and the change from <u> to <o> because of the <v> consonant made the word evolve into 'love'. What seems really interesting is how the original word had a <u>, then a <o>, and more recently the <u> has re-appeared in the nonstandard spelling 'luv'. Also, a final <v> may derive in a devoiced /v/ making it similar to 'lufe'.

One of the most interesting aspects of nonstandard spelling is the substitution of some endings for the letter <a> (Table 7). This seems to have the same function as 'gonna' or 'wanna'; that is, to make speech more fluid. It is also interesting that the examples provided in Table 7 follow two main patterns: functional word+preposition constructions and words ending in -er. Among these examples, we can find a word with a historical background: 'nigga' (3.18). 'Nigga' dates from 1925 and it was used to represent the pronunciation of Southern-American English.

Standard spelling	Nonstandard spelling
You	Ya
The	Da
Hell of a	Hella
Want to	Wanna
The fuck	Dafuck
Ever	Evah
Fucker	Fucka
Going to	Gonna
Kind of	Kinda
My	Ma
Mother	Motha
Motherfucker	Mothafucka
Motherfucking	Mothafukin
Nigger	Niga
Nigger	Nigga
Whatever	Whateva

Table 7: nonstandard spellings ending in -a

Taking in mind the use of the final <a>, we have formulated a hypothesis about the nonstandard spelling 'bruh' (example 3.33). It is another way to say 'bro', which is the shortening of 'brother'. This word follows the pattern of -er endings changed into a <a>. However, what interested us was its nonstandard spelling. Our hypothesis was that if 'brother' could change into 'brotha', then 'bro' could change into 'bruh', following the same pattern of the <a> sound. The reason why there is a <u> instead of an <a>, as seen in all the other examples exposed in Table 7, is because there is an attempt to avoid ambiguity since 'bra' is an English word with a complete different meaning. Moreover, the <h> plays an important role because without it, and following the formula <u> + //, the pronunciation would be /u:/ (Cámara 2010: 191). With the <h> the formula changes to <u> + C + // –as mentioned above– and the pronunciation would be transformed into /ʌ/.

As it happened between initialisms and logograms –reducing words to their minimum representation–, nonstandard spellings also share some similar traits with the following category: omitted letters. Both categories omit letters, particularly vowels, in some words. Also, specific consonants may be omitted in both cases. This characteristic will be discussed later. The difference between them is that nonstandard spelling usually changes one letter for another. To illustrate the similarity, which consists on the omission of letters, and difference, which consists in changing a letter for other letter(s), we could analyze the word "fucking" and the different writings that appear in our corpus. There are five variants divided into these two categories: fuckin (example 5.19), fokin (example 7.23), fcking (example 3.10), fking (example 7.25), and foking (example 8.48). The first one is an example of omitted letters as it is written with the standard spelling but with the letter <g> omitted. The second one belongs to nonstandard spelling because, even though the letter <g> is also omitted, there is a <o> instead of a <u>. In the third and fourth examples, the letter <u> and the letters <u> and <c> are respectively omitted. In the last case, which is a nonstandard spelling, letter <c> is omitted and the letter <u> is replaced with <o>.

Omitted letters answer to a phonetic cause and different types of omissions can be found: end vowels (1), middle vowels (2), end consonants (3), and middle consonants (4). End vowels (1) are omitted if the transcription of a word has no vowel in that position. The word 'people', for example, appears in the corpus as 'ppl' (example 1.38) and the transcription is /'pipəl/. Since there is no vowel in the final position when it is pronounced, the vowel is omitted. Something different happens if the pronunciation implies an end vowel. For instance, the words 'seriously' or 'really' maintain the final <y> with an /i/ sound: srsly (example 1.36) and rlly (example 3.17). Oxytone words, which are those whose primary stress is in the last syllable with a V+C+e formula where V is a vowel and C is a consonant omit the <e> in their pronunciation (Cámara 2010). Examples with this characteristic are the words 'hate' -hat (2.64)-, 'have' -hav (8.58)-, or 'love' -lv (5.44). The omission of the final -e also has a historical explanation. During the 11th century, end vowels were reduced to <e> and during the Middle English period the final -e started to disappear in the pronunciation yet not in the writing (Fernández 1982: 149). Also final -e

was used to indicate that a vowel was long if it appeared in an open syllable (Görlach 1993: 47). For these reasons, final -e serves to track the historical evolution of words. Nonetheless, the omission of the end vowel <e> can produce an alteration in the pronunciation. Moreover, it can produce an ambiguity that needs to be resolved by the context. This is the case of the above mentioned hate -hat- where the formula <a> + C + <e> implies /eɪ/ (Cámara 2010: 46) but the formula <a> + C + // implies /æ/ (Cámara 2010: 40). Also, 'hat' creates ambiguity as it means a complete different thing from 'hate'.

Middle vowels (2) are omitted indistinctly of whether the transcription of the vowels include their nominal version, their alternative version, or a /ə/. This is due to the fact that consonants generate a vowel sound by themselves ('b': /bi/) and context helps readers to identify the word. In the analyzed corpus, we can find examples with the words 'friends' –frnds (1.20)– or 'watching' –wtching (1.53). Within omitted middle vowels, we can also include the reduction of a diphthong into a pure vowel if the pronunciation of a word does not imply a diphthong. This is the case of the word 'meant'. Some irregular verbs, whose present tense contains the diphthong <ea>, make their past tense with an /e/ pronunciation instead of the /i:/ pronunciation characteristic of this diphthong (Cámara 2010). For this reason, the analyzed word appears as 'ment' (example 2.67). The letter <a> is dispensable in the pronunciation but it is helpful to trace the evolution of the word in terms of pronunciation. The verb 'mean' comes from Old English 'mænan' where <æ> was a long vowel pronounced as /ɛ:/ which during Early Modern English evolved into /i:/. In the Middle English period the <t> was added to form the past and the vowel was thus shortened (Millward 1996).

End consonants (3) may be omitted according to three criteria. The first criterion refers to the omission of a consonant that is not reflected in the pronunciation. Thus, in the word 'back' the letter <k> is omitted –bac (example 1.63)– since both <c> and <k> can be pronounced as /k/. For the same reason, in the word 'damn' –dam (example 4.35)– the letter <n> is omitted as shown in its transcription: /dæm/. The <n> can be omitted if it is a final <n> preceded by a <m> as for example 'hymn' or 'solemn' (Fernández 1982: 196). The second criterion refers to the omission of a consonant whose pronunciation is altered by the coda to which it belongs. This is the case of the -ing coda, repeated in the corpus with

different words. The pronunciation of /ŋ/ in some contexts makes the letter <g> dispensable. Some examples are fuckin (5.19), doin (5.69), dancin (3.38), or freakin (3.67). In some cases, writers substitute the letter <g> by an apostrophe that indicates the writer's awareness of this omission, as seen in freakin' (example 3.73) or makin' (example 5.67). Some American English dialects change /ŋ/ for /n/ if the -ng segment appears in final position of an unstressed syllable; this process is known as "g-dropping" (Wolfram and Schilling-Estes 1998: 72). The third and last criterion refers to the omission of a final consonant to follow the rules of connected speech. In this corpus, we have found two variants of the /t/ sound depending on its surroundings. In YouTube comment "wher is my bible? i wan to purge dis demon in the video" (example 4.27) the word 'want' appears with an omitted <t>, which can be omitted because it is between consonants (Ripman 1931). The second variant is seen in YouTube comment "HOW DAR U HAT ON JB!! HEZ DA BEZZZD U JUS JELEOUS[...]" (example 2.64). The word 'just' appears without the <t> because it is followed by <j>, creating a coalescence process which results in a change of the pronunciation (<t>+<j>=/tʃ/) (Ripman 1931).

Middle consonants (4) are also omitted following three similar criteria. As in the previous case, consonants can be omitted if they are not reflected in the pronunciation. For instance, the letter <c> is omitted in the word 'dick' –dik (example 8.5). Notice the similarity and difference with the above mentioned 'bac'. The omitted consonant changes yet the pronunciation does not vary (/k/). The second criterion refers to the omission of a consonant in a double-consonant structure. Here, we have found examples such as tennis –tenis (8.54)–, better –bter (8.5)– or faggot –fgt (2.8). The last criterion is related to the omission of consonants that, by being in a cluster, change their pronunciation. The cluster 'sh' in the word 'shut' is an example of this, as it becomes 'sut' (example 8.3), which results in a change from an affricate consonant into a fricative consonant.

Within omitted end consonants, we may include the inflectional use of final <s> as plural and third person singular marker. These cases do not follow any of the criteria discussed above. It could be even argued that third-person-singular <s> omission may rather derive from a grammatical error than from a conscious omission. To prove this, we have analyzed the number of times the third person singular marker appears. The results


show more than thirty examples of third person singular marker. Only in two occasions is this marker omitted: with the words 'play' (example 2.7) and 'start' (example 4.45). In the latter ("Omg its so cute how he talks and the baby start laughing when he said it hurt.lol") two markers appears: one grammatically correct (talks) and one ungrammatical (start). We can then conclude that these omissions seem to respond more to a conscious or even unconscious omission than to a grammatical error

The last two categories differ from initialisms, logograms, nonstandard spellings, and omitted letters in the lack of a phonetic cause, as there seems to be no connection between writing and pronunciation. Pictograms respond to the desire of changing written communication so that it is more informal, direct, and personal. This way, writers have no need of further explanation if they are being, for example, sarcastic. To achieve this, writers in this corpus use three types of pictograms. The first, and most commonly used, type is faces. They try to convey the emotions that writers would have if they were saying out loud what they write or read. Faces can be divided into two subtypes: premade faces and custom faces. Premade faces are those pictograms given by the software. Custom faces are those written by the user and that do not change their appearance on the screen. After analyzing them, we have found that custom faces (59 times) are more frequently used than premade faces (19 times). It seems surprising since custom faces need more keystrokes than premade faces. However, 'language play' also takes part in this process, as premade faces cannot cover all necessities writers may have. Table 8 shows examples of language play given by Crystal (2008: 74). In our corpus, we have found some examples of language play with the pictograms XD (1.34), xD (6.30), or x'D (5.63). Our interpretation of these is the following: laughing, laughing more (being the mouth 'D' bigger than the rest 'x'), and laughing even more (being 'a tear in the eye).

Pictogram	Meaning
:~)	happy
:~))	very happy
:~)))	very very happy
:~))))	ecstatic

Table 8: language play with pictograms (based on Crystal 2008: 74)

The second type of pictograms is the use of symbols representing words. The reasons for this are both language play and shortening. This is seen in the heart symbol,

which can be identified with the feeling of love. In YouTube comment "Do you ♥ me?" (example 2.60) the symbol substitutes the word 'love'. Notice the difference with logograms. The heart could be considered as a logogram if, instead of love, the symbol substituted the word 'heart'. The third type of pictograms is gestures. By using them, writers try to be more explicit, as with  (example 1.19). This pictogram may have two interpretations depending on the context: it shows a hand with the second finger up or it shows a penis.

Finally, shortenings respond to the desire of language economy. This category is the less representative one, with only 51 repetitions. The reason for this seems to be the difficulty to select words that could be understood once they are shortened. We have found four categories of shortenings: well-known shortenings (1), context shortenings (2), phonetic shortenings (3), and cultural shortenings (4). Well-known shortenings (1) include those that are also used outside IL. An example would be uni as a shortening of university (6.30) or mil instead of million (1.15). Context shortenings (2) are those words used in YouTube and which are related to this platform, yet they are not exclusive of it. These words are repeated several times making them prone to abbreviation. Examples like 'subscriber' as 'sub' (4.25) or 'favorite' as 'fav' (7.45) could be included within this category, even if they are not exclusive of YouTube or IL. Within phonetic shortenings (3) we have only found one example in the word 'though' as 'tho' (8.18). Its transcription $-/\dot{\theta}ou/$ shows the lack of the 'gh'. Also, the <u> is omitted because oxytone words ending in <o> are pronounced as the diphthong $/ou/$ (Cámara 2010) making the <u> unnecessary in writing. Back to the -gh digraph, it is interesting to see how, even though it is not pronounced, we still write it to track the historical evolution of this word. In Old English the /h/ sound was spelled as <h>. During the Middle English period the 'gh' appeared in substitution of the <h> grapheme. An example would be 'riht' and 'right' (Millward 1996: 161). Later, during Early Modern English period, 'gh' in final position lost its pronunciation ('although') or changed its sound into a /f/ ('tough'). Also, the loss of /h/ made preceding short vowels longer, hence explaining the existence of present day English diphthong (Millward 1996: 251). Finally, cultural shortenings (4) include the word 'brother' and its shortening 'bro' (example 8.39), which dates from the 1660s and was used as a colloquial word.

All in all, we have seen how abbreviations or variations of words are still in use creating new examples and following some patterns as can be language economy, language play, phonetic causes, or even historical explanations.

DISCUSSION

While analyzing the corpus, we have encountered some difficulties that should be taken into account for deeper or further research. The first difficulty when extracting the data was the 'asd' combination (example 1.58). According to the *Urban Dictionary*, 'asd' means nothing and it is used mainly to spam in video games. The reason why players use those three letters may be because video games normally use the letters w, a, s, and d to move. For this reason, the hand of the player is always in that part of the keyboard and the 'asd' combination seems the easiest one to press. The difficulty comes when cataloguing this structure since it seems that none of the categories described throughout this essay can include this structure within their defining scope. If we analyze the examples mentioned above, we can always see a logical connection between that example and the full word or structure it stands for. For instance, by looking at the same time to 'lol' and 'laughing out loud', we can establish without further knowledge that there exists a logical connection between both of them since the user is reducing the words to their initial letters. The same happens with 'fone' and 'phone' (Crystal 2008: 48), where we know that <f> and <ph> have the same pronunciation (/f/). This is not what happens with 'asd' and its meaning 'nothing'; there is no connection between them. The problems associated to the 'asd' example prove the necessity to create new defining categories as IL is expanding and language play is continually evolving.

The second difficulty derives from the use of the YouTube platform itself. As different languages coalesce at the same time, we had to filter the comments where English was the only used language. However, this filter may not be perfect in the case of those comments that do not have a context for two reasons: First, because we may not know the language; and second, because one structure may be used in different languages and for different purposes. For example, we can mention the case of the word 'cwel' (2.61), which appears in isolation and by reading it, it could be included in English initialisms or English

omitted letters, thus proving how faulty the English only filter may be. 'cwe'l' is used in Polish slang, according to the *Urban Dictionary*. Likewise, we could mention again the 'asd' structure, which in Italy can also mean 'laughing out loud', again according to the *Urban Dictionary*. As we do not have a context, because these words appear in isolation, we cannot provide a reliable conclusion about their use.

The third and last difficulty is related to the lack of information provided by Crystal regarding the reason(s) why certain abbreviations are catalogued in a category or another. For instance, all the examples in nonstandard spelling (Crystal 2008: 48-49) seem to be considered as such because at least one letter does not appear in the standard spelling (Table 9). An exception is the word 'should' and its nonstandard spelling 'shud'. The nonstandard spelling in this example does not show different letters from those in the standard spelling. The only process that it suffers is the omission of the letters <o> and <l> because they are not pronounced. In omitted letters, we have another modal verb 'would' with its abbreviated form 'wd' (Crystal 2008: 27). The only apparent difference between these two examples is that in 'would' two vowels are omitted instead of only one, and yet the two abbreviations are included in different categories. We have tried to see if this division could be a consequence of the evolution of the words, but, they are very similar in dates and evolution: 'Should' comes from 'sceolde', which is the past tense of 'sceal' (shall) and it dates from around 1200; and 'would' comes from 'wolde', which is the past tense of 'willan' (will) and it dates from around 1300.

Nonstandard spelling	Standard spelling
fone	phone
skool	school
sum	some
thanx	thanks
wot	what
ya	you
dat	that
wiv	with
wenja	when do you

Table 9: examples of nonstandard spelling provided by Crystal (2008)

In an attempt to study how language is evolving as a consequence of the use of new technologies, we have tried to see similarities and differences between 2008, using Crystal's corpus, and 2015, using the corpus compiled for this essay. Since we do not have

information about how many texts were collected for Crystal's findings, we can neither give the percentage of the abbreviations in relation with the total number of words nor the number of times that each abbreviation is repeated. Nonetheless, we are able to provide the total number of different abbreviations. Crystal's corpus consists on 224 different abbreviations while our corpus includes 213 different abbreviations. We can see here how the 7 year-lapse does not seem to have an effect on the amount of abbreviations used, yet we still lack information regarding Crystal's corpus. What we have found while analyzing the two corpora is that language used in new technologies is not consistent in its use since only 28 abbreviations appear in both corpora, as seen in Table 10. Also, there are three abbreviations with the same writing but different meaning. These examples are 'b' as 'be' according to Crystal and 'billion' in our corpus; 'f' as 'friend' according to Crystal and 'fuck' in our corpus; and 'z' as 'said' according to Crystal and 'is' in corpus. For this reason, we could conclude that IL is not fixed and users prioritize language play and commodity rather than consistency. Thus, it seems that this is not an issue related to literacy.

2	cos	h8	lol	omg	thx	wtf
4	cuz	idk	luv	pls	u	y
b4	doin	imo	m8	ppl	w8	ya
btw	gf	lmao	n	r	wot	yr

Table 10: shared examples in the two analyzed corpora

As seen in the analysis section, we have included some possible reasons for using certain words according to their historical evolution. This may justify or explain why they are used even though users probably cannot trace the historical origin of those words. Crystal, again in his book *Txtng: The gr8 db8*, does not show the historical evolution of language, but he provides examples of how all the categories here explained are not new and have been used throughout history. Abbreviations have been used for centuries and so what it is currently considered IL or text-messaging could be seen in reality as the normal path that languages follow to evolve, survive, and play, and not as an issue to be concerned about. Speakers are adapting language to their new technological needs, but the processes they are following have been long in used.

Logograms, for example, present a similarity with ancient Egyptian writing. They both use what is called rebus, i.e. "a message which [...] consists entirely of pictures that are

used to represent the sounds of words, rather than the objects they refer to." (Crystal 2008: 39) Crystal provides an example with the word 'beast' written in English with the picture of a bee plus the letters <st>. The connection between the word and the image is the pronunciation; the word 'bee' –/bi/– and the beginning of the word 'beast' –/bi:st/– have the same pronunciation. Crystal also posits that rebuses are also found in the art and literature of Europe by authors as Leonardo da Vinci, Ben Jonson, William Camden, or Lewis Carroll, who used rebuses in some of his letters. Logograms, for this reason, are "part of the European ludic linguistic tradition" (Crystal 2008: 41).

Initialisms were and are widely used and the reasons seem to be the same: to abbreviate not only a word but also several words. Following a chronological order, Crystal gives some examples of this category used in English such as 'pm' meaning '*post meridiem*' and used since 1666, 'RIP' meaning 'rest in peace' and used since the nineteenth century, 'NBG' meaning 'no bloody good' and used in the first years of the twentieth century, or 'SNAFU' meaning 'situation normal, all fouled/fucked up' and used in the middle years of the twentieth century. Some initialisms are nowadays so deep-rooted that we are not able to identify them as such and we are "unable to say what the letters originally stood for, as in the case of *laser* ('light amplification by the stimulated emission of radiation')." (Crystal 2008: 44).

Some past examples of omitted letters are still very common nowadays as 'Mr' – Mister– and 'Mrs' –Mistress–, or the abbreviations referring to weights or measures such as 'cm' standing for centimeter or 'kg' standing for kilogram. Crystal (2008: 47) also includes in this group some abbreviations used in text-messaging that are, in turn included in Partridge's *A Dictionary of Abbreviations* (1942). Within these examples, we find 'aftn' referring to 'afternoon', 'gd' referring to 'good', 'togr' referring to 'together', or 'btwn' referring to 'between'. There are other examples that seem to have been lost such as 'cml' instead of 'commercial', 'clk' instead of 'clerk', or 'rly' instead of 'railway', which could be currently confused with 'really'.

Some nonstandard spellings have the peculiarity of being included in the *Oxford English Dictionary* (Crystal 2008: 49). Here, we can find examples such as 'cos' –included in 1828–, 'luv' –included in 1898–, or 'ya' –included in 1941. Other nonstandard spellings

are not included in the dictionary although they have been known for more than a century. Some of these words are 'gonna' instead of 'going to', 'thru' instead of 'through', or 'wiv' instead of 'with'. Moreover, writers have been using this type of words to represent dialects in literature. Among these writers, we find Charles Dickens, Mark Twain, D. H. Lawrence, or even more recently J.K. Rowling.

Finally, some shortenings created in the past are very common nowadays; for instance the words 'exam' instead of 'examination', 'vet' instead of veterinarian, 'fridge' instead of 'refrigerator', or even 'bus' instead of 'omnibus'. Apart from these, in 1711, there were some shortenings exposed to criticism including 'rep', which means 'reputation', or 'pos', which means 'positive'. Crystal posits within this category the following: "[i]t is not a criticism of texting to say that one cannot understand [a shortening because] any informal letter between friends would present similar difficulties of interpretation to outsiders." (Crystal 2008: 52)

After analyzing Crystal's examples used outside text-messaging and IL, and making a comparison between these examples and the corpus analyzed in this essay, we can see how some of them are still in use. Table 11 shows the words and the abbreviations that appear in both sources. Hence, the data seems to state that text-messaging and IL should not be exposed to criticism in the use of abbreviations without having in consideration previous data.

Category	Meaning
u	you
r	are
b	/bi/
4	/fɔr/
wot	what
cos	because
luv	love
thanx	thanks
ya	you
gonna	going to
wanna	want to

Table 11: Shared examples between corpus and Crystal's (2008) historical examples

It should be also pointed out that some of the examples exposed here present phonetic suprasegmental features, which could be defined as the changes that a segment or segments experiment in terms of pitch, stress, and quantity (Lass 1976: 225). An example is

the 'OMG' initialism (1.1), differentiating it from 'omg' (1.11). Though, again, initialisms could be written in capital letters or not without any concrete reason, it is true that capitalized letters give intensity and can be even considered as shouting (Crystal 2008: 124). Another example of suprasegmental feature is 'o....m....g' (7.54) where the sequence of dots make the letters or words longer adding, for example, a surprise or incredulity feeling. Finally, we could mention 'lollll' (example 5.20) where the user writes repeatedly the letter <l> simulating the lengthening of the /l/ sound. With this process, the user tries to emphasize that something is funnier than if it had been written with one single <l>. These suprasegmental features should be taken into account for further research since they change the purpose of the speakers in terms of abbreviation. The number of features in our corpus is extended and their analysis would lie outside the scope of this essay.

CONCLUSION

In the current essay we have tried to provide an explanation and an analysis to the evolution of language as a consequence of new technologies. Using a classification of nonstandard English words provided by David Crystal in his book *Txtng: The gr8 db8* (2008), we have been able to analyze YouTube comments in 2015. The classification divides each nonstandard English word into different categories depending on their internal features. In this essay, these features have responded to language economy; a desire of creativity; a connection between oral communication and written communication; and a connection between changes made in the past and in the present.

While analyzing YouTube comments, we have found several problems using Crystal's classification. Some IL words cannot be classified within any of the categories provided in his book; therefore, there seems to be a necessity to adapt these classifications to new contexts. Also, there seems to be a necessity to specify or redefine the describing features of each category. This necessity comes from the contrast of some words provided by Crystal and the use of several categories in one single word.

The main reason for using nonstandard English words still seem to be language economy even if it is written with a standard keyboard. All categories share the feature of reducing words to avoid keystrokes and to increase speed. The omission can go from a

single letter to all letters except one while allowing the user to infer the meaning of the word in isolation or in context. What initially became a necessity, it is now a habit characterized by its inconsistency and expansion in terms of variability. Considering Crystal's corpus and our corpus as general language corpora, we may think that the vocabulary used is similar, and what we have seen is that almost 200 abbreviations or variations of words used in 2008 are used correctly in 2015. This has a correlation with language play, as it is not only a matter of reducing a word but also of being unique when reducing it. Creativity plays an important role, as users may want to express their writing and adapt it to their style. It seems that users are aware of aesthetics trying to convey a new way of writing. This creativity may overlap the idea of illiteracy as the cause or effect of text-messaging and IL. Also, the amount of standard English vocabulary understates the idea of user's illiteracy. Users seem to be aware of the difficulty of reading sentences full of abbreviations and they try to balance comprehensibility and language play.

With these abbreviations and language play, users are probably making a logical connection between what it is said and how it is written. McWhorter (2013) says that the beginning of writing was a reflection of how people spoke. In the Early Modern English period, there was an attempt to simplify spelling and to try to make it closer to phonetics (Millward 1996: 232). Also, in this period, schools made spelling and pronunciation be closer to each other (Görlach 1993: 12). This is what happens nowadays with IL; users are defending a spelling where the superfluous is omitted and the pronunciation is reflected. We can see this process through hypotheses on transcriptions and historical explanations. IL seems to be thus just following a previously-established trend within language evolution.

To all this, we should add the necessity to do further research regarding two ideas: First, the problems of Crystal's classification should be solved in order to give what we think is a more accurate way of describing nonstandard English words. Their defining features should also be updated. Second, suprasegmental features should be taken into account because they are changing the path of abbreviations. While text-messaging abbreviates and gives importance to what it is written, IL abbreviates words and lengthens letters giving importance to how it is written. We are thus seeing an evolution towards face-to-face communication in which we include pronunciation, intonation, and even emotions.

As a final remark we could say that the use of abbreviations with new technologies is the normal path of language evolution tending towards a logical connection between written and oral communication. It can be seen with the creation of oral language and its representation in written form, the Early Modern English period, etc., or the phonetic and historical explanations here exposed. Language seems to be like a circle, and we could establish a parallelism between this idea and Yeats' words in his poem *The Second Coming* (1996: 187) saying that language is "[t]urning and turning in the widening gyre".

WORKS CITED

- Baugh, Albert & Cable, Thomas. *A History of the English Language*. 5th ed. London: Routledge, 2005. Print
- Cámara, Enrique. *La Vocal Inglesa: Correspondencias Grafo-Fonémicas*. 3rd ed. Valladolid: Universidad de Valladolid, 2010. Print
- Crystal, David. *La Revolución del Lenguaje*. Madrid: Alianza Editorial, 2005. Print
- Crystal, David. *Txtng: The gr8 db8*. New York, Oxford University Press, 2008. Print
- Fernández, Francisco. *Historia de la Lengua Inglesa*. Madrid: Editorial Gredos, 1982. Print
- Finneran, Richard, ed. *The Collected Poems of W. B. Yeats*. New York: Simon & Schuster, 1996. Web. 20 May, 2015 <http://goo.gl/ORiNMx>
- Görlach, Manfred. *Introduction to Early Modern English*. 2nd ed. Cambridge: Cambridge University Press, 1993. Print
- Harper, Douglas. *Online Etymology Dictionary*. 2001. Web. 7 April 2015 <http://goo.gl/M60vxU>
- Lass, Norman J. *Contemporary Issues in Experimental Phonetics*. New York: Academic Press, 1976. Print
- McWhorter, John. "Is Texting Killing the English Language?" *Time*. 2013. Web. 26 April, 2015 <http://goo.gl/Kpk0xU>
- Millward, C. M. *A Biography of the English Language*. Boston: Wadsworth, 1996. Print
- Partridge, Eric. *A Dictionary of Abbreviations*. London: George Allen & Unwin, 1942. Print
- Ripman, Walter. *English Phonetics: A Revised Version of "the Sounds of Spoken English"*. London: J.M. Dent & sons, 1931. Print
- Sklar, Elizabeth S. "The Possessive Apostrophe: The Development and Decline of a Crooked Mark." *College English* (Oct., 1976), 175-183. Print
- Summers, Della. *Longman Dictionary of Contemporary English*. 3rd ed. Essex: Longman Group, 1995. Print
- Thurlow, Crispin & Brown, Alex 2003. "Generation Txt? The sociolinguistics of young people's text-messaging." *Discourse Analysis Online*. 2003. Web 2 April, 2015 <http://goo.gl/qaXKRC>
- Urban Dictionary*. Urban Dictionary LLC, 1999. Web. 2 April, 2015 <http://goo.gl/4Zc5nG>
- Wolfram, Walt & Schilling-Estes, Natalie. *American English: Dialects and Variation*. Oxford, Blackwell Publishers, 1998. Print
- Wray, Alison. "The Occurrence of 'Occurance' (and 'Alot' of Other Things 'Aswell'): Patterns of errors in undergraduate English." *Language and Education*. Southampton: Longdunn Press, 1995. 94-106. Print

ANALYZED SOURCES

- EminemVEVO. Eminem - Love The Way You Lie ft. Rihanna. Online video clip. *YouTube*. YouTube, 2010. Web 3 February, 2015 <http://goo.gl/HDqhXI>

HDCYT. Charlie bit my finger - again !. Online video clip. *YouTube*. YouTube, 2007. Web 2 February, 2015
<http://goo.gl/sxmw15>

JenniferLopezVEVO. Jennifer Lopez - On The Floor ft. Pitbull. Online video clip. *YouTube*. YouTube, 2011.
Web 3 February, 2015 <http://goo.gl/3eUrPc>

JustinBieberVEVO. Justin Bieber - Baby ft. Ludacris. Online video clip. *YouTube*. YouTube, 2010. Web 1
February, 2015 <http://goo.gl/Z7JkD4>

KatyPerryVEVO. Katy Perry - Dark Horse (Official) ft. Juicy J. Online video clip. *YouTube*. YouTube, 2014.
Web 4 February, 2015 <http://goo.gl/TQW8s5>

LMFAOVEVO. LMFAO - Party Rock Anthem ft. Lauren Bennet, GoonRock. Online video clip. *YouTube*.
YouTube, 2011. Web 2 February, 2015 <http://goo.gl/rwXCMU>

MyTop100Videos. Most Viewed Videos of All Time. *YouTube*. YouTube, 2012. Web 1 February, 2015
<http://goo.gl/5UkNov>

officialpsy. PSY - Gangnam style. Online video clip. *YouTube*. YouTube, 2012. Web 1 February, 2015
<http://goo.gl/pPpd96>

shakiraVEVO. Shakira - Waka Waka (This Time for Africa) (The Official 2010 FIFA World Cup™ Song).
Online video clip. *YouTube*. YouTube, 2010. Web 4 February, 2015 <http://goo.gl/iQsWe4>