



HAL
open science

INTERNET REPRESENTATIONS OF DIALECTAL ENGLISH

Laura Gabrielle Goudet

► **To cite this version:**

Laura Gabrielle Goudet. INTERNET REPRESENTATIONS OF DIALECTAL ENGLISH. CORELA - COgnition, REprésentation, LAngage, 2018. hal-01847025

HAL Id: hal-01847025

<https://hal.science/hal-01847025v1>

Submitted on 23 Jul 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INTERNET REPRESENTATIONS OF DIALECTAL ENGLISH

Laura Gabrielle Goudet
Université de Rouen
ERAC (EA7338)

Abstract:

This paper presents an account of how alternative spellings found online can be linked to phonetic and phonological roots, especially in settings where dialects are encouraged. These spellings are part of Netspeak uses, and dialectal spellings are influenced by official, lexicographical sources and a personal desire to spell words as they are pronounced. Consonants and vowels are affected by these choices, and this paper will mainly discuss consonantal changes and the limits of vocalic alternations.

The first research question focuses on how non-traditional spellings aim at showing an alteration of the same type of sound, irrespective of the dialect, to see if there are common underlying mechanisms between very different dialects of English. Exposure to the Internet culture, and forum-specific uses may account for neography. Limits reached by these respellings (homophony, semantic ambiguity) will also be discussed.

The corpus of this study is composed of written productions found in African American and Scottish online communities, *Black Planet* and *Scotster*, to examine the parallel development of alternative spellings in two dialects that contravene the rhoticity of the variety of English spoken in their country.

The wider ecological¹ context of writing online is the first focal point: Netspeak influences alternative written practices, and censorship is a peculiarity of this environment. The typology of alternative spellings will be defined: they are not limited to simple deletion or doubling of letters, and their difference with spelling mistakes will be specified.

The stance of linguistic studies on spelling and lexical differences with standard English will shed light on the official stances regarding these two dialects, and which rules alternative forms found on the forums are supposed to follow. While Scots spelling is seen with a little more prescriptivism, works on African American Vernacular describe forms and sanitize them (there is very little profanity in dictionaries).

Bearing in mind these specificities and the wider online context, consonantal substitutions will be tackled, both generally, and then more specifically when they are applied to grammatical words. Vowel changes and limits created by homophony will be the point of the last subsection.

Keywords: African American Vernacular, dialect, graphophonemics, Scots, spelling

¹ In the sense of “digital ecology”, i.e. the wide interconnection and network formed by all the elements interacting online.

Mots-clefs : dialecte, graphie, phonographématique, Scots, vernaculaire afro-américain

Introduction

English spelling online greatly differs from accepted orthography. There are no regulating bodies policing utterances, so website members tend to express themselves freely, with adaptations of the written norms of English. The only limit to what can be said is framed within Terms of Service, supposedly acknowledged by platform users when they first sign in. Censorship can also be exercised by scripts surveying any or all messages posted; the censored words being replaced by alternative versions. This, in turn, pushes the users towards non-traditional spellings which are not listed by these automated replacement tools: “computer-based censorship cannot keep up with simple substitutions of letters, let alone community-based insults which do not belong to any standardized list of profanities” (Goudet, 2013).

The corpus is made of forum threads from two sites aimed at gathering people around a common ethnic or geographical background: *Black Planet*² (hereinafter referred to as BP), an African American online community with more than 25 million members; and *Scotster*³, a Scottish-oriented social networking site with 5,000 users. The members are about the same age (35–54 years old for BP, 45–54 years old for *Scotster*). The threads selected for this study contain more than 10 answers in order to observe how people interact, in the most popular sections of the forum. The selected message corpora are about the same size, 25,000 words each.

These native English speakers master spoken and written English and the sociolects they resort to, at least orally. As Internet users, they are exposed to shorthand practices both on new communication media, and online. These are encompassed by the wide concept of Netspeak (Crystal 2001). They are also exposed to website-specific uses and vocabulary, such as the “timeline” (for *Facebook*'s news feed). On *Black Planet*, statuses belong to the “what’s new?” section, whereas *Scotster* uses the expression “the Pulse”.

On regional interest forums which gather people around a common geographical or ethnic identity, these common uses are also intertwined with minority languages or dialects.

I will first draw a quick overview of some practices found online, regarding spelling and graphophonemic equivalences. Netspeak can account for some unorthodox graphic forms of the corpus, and can be viewed as another dialectal layer. The type of alternative spellings, which can bear the same graphic forms as spelling mistakes, will then be examined, before moving on to the role censorship plays in online exchanges. Official views on these two varieties, especially by lexicographers and academics who study these dialects will be discussed in the second part: they differ from one another, and also from what the members of the two websites use. Last, graphophonemic replacements of consonants, in a general way and more specifically in grammatical words will be studied before moving to the limits of vowel substitutions and problems of homophony.

1. Typology of alternative spellings online

² Available at the address <<http://www.blackplanet.com/>>. Last access: Jan. 2, 2016.

³ Available at the address <<http://www.scottishsocial.com/>>. Last access: Jan. 2, 2016.

Alternative spellings are often motivated by the oral nature of the transformations applied to written discourse. There is a mental equivalence between dialects and graphic representation of pronunciation: “Orthography is a system designed for readers who know the language [...] Such readers can produce the correct phonetic forms, given the orthographic representation and the surface structure, by means of the rules that they employ in producing and interpreting speech” (Chomsky & Halle, 1968, p. 49). I will first evoke Netspeak, especially the spellings commonly used on the Internet that are inspired by pronunciation, before examining the very close relation between alternative graphic forms and spelling mistakes. The role of censorship is also paramount in some graphic choices, and will be the subject of the third section about spellings online.

1.1. Netspeak

Online exchanges are essentially free from control, and practices become standard as they are being used by members of the community. Netspeak is a broad concept, encompassing “the words, idioms, and peculiarities of spelling and grammar that are characteristic of online documents and communications” (McFedries, 1995). Some of these peculiarities are adaptations of written forms in order to be closer to their pronunciation: “c’mon” (BP) is a shortened version of “come on”, and compresses the two words into a spelling showing how it is pronounced (/kə'mɑ:n/). The unstressed vowel is replaced with an apostrophe, one of the most common strategies used to symbolize a reduced sound in non-orthographic spellings. The importance of orality is also palpable in the proportion of interjections (<huh>, <hmp>, SC and BP).

The form < cuz > (*‘because’* BP and SC) attests the common apheresis of this conjunction. Grammatical words tend to be shortened or affected by alternative spellings more often than lexical words, because of their relative frequency in discourse and their structuring goal within sentences. The alterations can be either generally shared online (in Netspeak) or belong to a dialect⁴. Many of these alternative forms are highly symbolic: the choice of the representation <u> for “you” is another example of a graphophonemic equivalence being selected by new technology users, on the Internet and in text messages alike (Crystal 2008). This type of spelling does not preserve the length of the original orthography, but its frequency online⁵ is high (one example of <u> for every 97 <you>. This ratio is much higher than the form < cuz > for <because>, which is only 1:295). The reduced size of the corpus allows for human verification of all these ambiguous, shortened forms. Such concise spellings are problematic for automated analyses: the form <u> could only be the single letter and not a full word, or an initial.

Discourse online resorts to alternative spellings as a way to transform written messages into a more expressive and unchecked hybrid, especially through its frequent uses of oral forms. These spellings may be exceptional or widely accepted online. One of the biggest issue when dealing with these is the fact that they may take the same form as simple mistakes. This will be the main point of the next section.

⁴ These dialectal changes will be examined in sections two and three of this paper.

⁵ 165,264 tokens on the corpus of global Web-based English (GloWbE), compared to 16,129,419 results for the form <you>.

1.2. Unorthographic spellings

Deciding whether a spelling that does not follow the standard orthography is an alternative spelling or a spelling mistake is a tricky task. Misspellings are involuntary and irregular and revolve around either spelling or typing errors for native English speakers. Uncertain spellings often rely on the incorrect selection of a grapheme to symbolize a sound. The spelling <gurdle> for “girdle” (BP) is repeated by a user in one of the threads, and is a common spelling mistake, both on the forum and online⁶. It is caused by the choice of <ur> to represent /ɜː/ or /ɜː/ (lexical set of NURSE) which can be written <earC>, <yrC>, <ir> or <urC>. Instead of selecting the orthography, the users favour another phonologically logical representation, even if the correct spelling was found on the thread. Typing mistakes often revolve around one grapheme only: additions, omissions, substitutions of a single letter for another are regularly observed, as well as metatheses. They are the consequence of improper proofreading, such as <contintue> or <voilence>.

The example of <gurdle> (‘*girdle*’) can be compared to the form <gurl>, the alternative spelling of the very common word “girl”. Contrary to <gurdle>, <gurl> is frequent online, and has a ratio of one occurrence in every 154 use of “girl”⁷, and is even the name of a website aimed at teenage girls. It also has several entries in *Urban Dictionary*: it is semantically slightly different from the word it stems from, and this leads to the conservation of this new form. However, <gurdle> does not differ semantically from <girdle>, appears less frequently, with a ratio of 1:516, and Google’s algorithm suggests the correct spelling when searching for <gurdle> (message “did you mean: girdle” between the search box and the results). As such, we can safely assume <gurdle> is a spelling mistake, with the same graphophonemics explanation as <gurl>, but lacking any semantic or pronunciation contrast to the orthographic form.

The substitution of one grapheme for another is a common strategy used by Internet users to add an extra semantic layer to their discourse. Some are obvious, such as <Amerikkka> (an offensive blend of “America” and the initials of “Ku Klux Klan”, referring to racist America), and act as portmanteau words where the second element is forced into the first as an infix. Others are motivated by phonological realization, such as the form <pix> (‘*pic[ture]s*’). The apocope triggers the substitution of the last consonant of the syllable by an <x>. This grapheme is more uncommon in English than <c>, and it acts as a replacement for <cs>, both combinations representing the pronunciation /ˈpɪks/.

Alternative spellings use the same mechanisms as misspellings, and consequently, may be graphically identical. Spelling mistakes show an inadequacy between the preferred standard and the written form, all the while showing how accurate graphophonemic instincts are for native English-speakers. Apart from impressing a new semantic layer onto a word, or trying to nuance its actual meaning, alternative spellings may originate from platform-specific censorship. This will be the main point of the next subsection of this paper.

⁶ There are 310 tokens in the COCA, which contains an online section, and 915 in the GloWbE.

⁷ Results from a Google search on January, the 11th, 2016:12,700,000 tokens of “gurl” and 1,960,000,000 occurrences of “girl”.

1.3. Effects of censorship

On the two forums under scrutiny, censorship is not exerted the same way. Users of *Scotster* tend to censor their messages themselves. They either symbolize taboo words with the initial and the number of asterisks corresponding to the remaining number of letters (“F*** sake”), or paraphrase what they could write (“preceded by a swear word.”[sic]). That kind of censorship is endogenous, as there is no moderation when a user resorts to profanity: the word “shit” (a hapax on SC) is not altered by the social network.

On BP, an automated script neutralizes offensive strings of characters, converting them into the sequence <%#&@\$! >, found 62 times in the corpus. Thus, the word “fucking” becomes <%#&@\$! ing> (sic). In turn, users of this forum have adapted their discourse to circumvent these replacements. Some may have no phonological underlying motivation, such as the deletion of the first letter of the incriminated word. The form <igger> (*nigger*, hapax) changes the word for the censoring script but not for the reader, as this taboo word is often evoked on the site. This spelling has no homograph, and the context of a censoring website accounts for this written form. Other highly recognizable words can also undergo substitutions of letters for logograms that bear a strong iconic resemblance with the missing letter, such as <@ss> (6 tokens), or <b!tch> (87 examples in the corpus).

Substitutions of letters for others with the same or a close phonic value are also commonly used. The graphemes <i> and <y> have the same pronunciation in a closed syllable context. This accounts for the transformation of <shit> and <bitch> into <shyt> (80 occurrences) and <bytch> (112 examples). When no actual homophone exists, near-homophonic characters can symbolize the missing letters: hence <azz> (71 tokens on BP) replaces a consonant having a [-son] realization by its [+son] counterpart. The form <asz> is also found 59 times in the corpus, but not the spelling <azs>, as it would certainly evoke the morpheme <-s>, added to a non-existent word.

Alternative spellings of insults on BP are more numerous than the replacement implemented by the website. This suggests literacy on the part of the users, who know how to adapt their discourse to the constraints of the network. On both platforms, endogenous censorship exists, whether it is a change in the standard word, or a euphemistic phrasing, such as <N—r> and <N word> (both occurring in the same message). Stylistic choices are highly personal, but most of the changes on swear words occur on vowels, which are more prone to be replaced by another character than consonants. Another important factor is the conservation of the number of letters in a word, in order to recognize it easily, and have it fill the same space the orthography would.

After this account of Internet-triggered alterations of spelling, the second section of this paper I will examine how dialects are considered, before seeing how speakers use their dialect in what they write.

2. Dialectal peculiarities of spelling

Within these communities, attitudes towards the variety they speak are diverse. A centripetal movement draws users of the same website to adopt the group's accepted codes and literacy. The speakers' interest for metalinguistic discourse and the views in education (either acknowledging the existence of AAV, or actively educating pupils about the spelling freedom of Scots), as well as the type of discourse held in dictionaries will be compared for the two linguistic communities, in order to see how ethnocentrism⁸ affects them.

2.1. Lexicographical stances on African American dialect

African American dictionaries tend to have a descriptivist approach (Kearse, 2006, Smitherman, 2000). They are either composed by academics, or by speakers of African American Vernacular (AAV), such as Kearse, who compiled an "urban slanguage" (sic) dictionary as he was serving time in jail. Smitherman said she observed all kinds of African American people and used their production, as well as cultural productions such as songs, radio programs, or even letters (Smitherman, 2000). The proportion of standard English words is significant, although their meaning may have been altered and accepted alternative spellings appear as head words. The phatic term of address "homey" is defined by Kearse as "males who are from the same city, state or neighborhood as you" and as "a person from one's neighborhood. A Black person" by Smitherman. This form does not exist in the corpus (although this form, as well as <homie> are attested in the forum), but the term "homegirl" appears once in the corpus, and 18 times only in the forum. There is a discrepancy between <homey>, the spelling adopted by the dictionary (a homograph with an adjective meaning "cosy and comfortable"), which is only used 45 times in the whole forum, and the spelling adopted by African American members of BP, <homie>, which is three times more frequent in discussion threads (141 occurrences). Both spellings try to emulate the pronunciation /'houmi/, but the dictionary does not reflect actual, contemporary uses. The only listed definition is "comfortable or familiar like home" (Merriam-Webster, 2011), not Kearse's, although that meaning is represented online and in slang dictionaries. The graphic representation of the word is problematic as well. The spellings <homy> and <homey> are accepted for the standard English adjective, and the second spelling is also correct for the AAV word, but mainly in dictionaries, as internet users prefer <homie>.

There is also a greater proportion of idioms and full phrases in AAV dictionaries, such as "I'm not talkin' too fast, you're listenin' too slow" (Kearse, 2006), than for Scots. These are rarely, if ever, attested in the corpus. On *Black Planet*, there is no result on the forum for queries containing the words 'talk/speak slang, vernacular' pertaining to using African American Vernacular. Certain terms such as <overstand>, defined as "referring to profound knowledge and insight, over and beyond that of mere 'understanding'" (Smitherman, 2000) and African American identity are closely interrelated. These never call for an explanation in the corpus, and only occur in conversation about ethnic matters. The reanalysis of the morphological structure of the lexeme accounts for the semantic change. It is not always the case, as for the word <supremacy>, a discreet play between "supremacy" and "man" (two tokens in the corpus). It

⁸ In the anthropological sense, where language ties together society and the language spoken, modelling the speakers' view on the world: "Language as symbols of beliefs and values communicates the perceptions of a society. These symbols will, in varying degrees, exercise control over the individual and perception" (Seidner, 1982).

appears in the corpus as a way to talk about white supremacists, and is both graphically and phonetically close to the original spelling. This example evokes the form <Amerikkka> (two instances in the corpus), where a homophonic portmanteau word superimposes two semes.

Contemporary African American spellings are in most cases word plays and expressions that are described by scholars and speakers alike, in dictionaries. Creativity is paramount in these respellings, hence the need to document them for lexicographers, a parallel to internet users' want to create new spellings. The next section will show how Scottish views on language differ.

2.2. Scottish lexicography and prescriptivism

Compared to AAV lexicographers, most Scots dictionaries reflect a more prescriptivist approach to the dialect. As Scots is used in schools in Scotland, the general stance on education is to let the pupils choose themselves how they would like to spell Scottish words, within carefully defined educational moments. Some dictionaries have almost no explanation as to how data was gathered, and no field works are evoked, not even the reason why these dictionaries are compiled (Warrack, 2000). Guidelines and style sheets for Scots, such as Warrack's introduction to his dictionary or the Makar's Club recommendations which are still seen as today's standard, have common features, such as advocating against the use of "parochial apostrophes". These are perceived as demoting Scots, as they point at considering Scots spelling as an adaptation of the English orthography. The spelling <caw> is an example of accepted representation for "call", but <ca'> is prohibited (Makar's Club, 1947).

The importance of historical details about the evolution of the dialect may suggest that the main goal of Scots dictionaries is to read older texts. The same may apply to the *Dictionary of the Scots Language*, split between the *Dictionary of the Older Scottish Tongue* (Scots before 1700) and the *Scottish National Dictionary* (SND), which contains words of Modern Scots, and a corpus of texts up to 2005. It is also the only dictionary with a name translated into Scots, "Dictionar o the Scots Leid". Eagle's *Online Scots Dictionary* (OSD) and the SND are the only ones to give phonetic transcriptions, with regional variations. The entry "bird" in the OSD lists /bɪrd/, /bɛrd/ and /bʌrd/, but the SND only suggests the last one, along with /bø:rd/ and /bu:rd/. This example shows how imprecise these transcriptions are, seeing they refer to several varieties of the same dialect. As the spelling of the word is not fixed, both dictionaries propose several spellings, such as <burd>, <bord> and <bird>.

Definitions in the dictionaries are exclusively written in standard English, implicitly turning them into Scots-English dictionaries: as a recommendation goes in the OSD, "British English spelling is used!". Hence, standard spelling is a means used to unify different pronunciations of words, and is all the more important that literacy in English is taught using its orthography exclusively.

On *Scotster*, members frequently refer to Scots or Gaelic. The subsection of the forum called "Scottish Languages" contains 289 threads, devoted to asking users to translate messages in Gaelic, or test their knowledge of the two languages. "Parochial apostrophes" are common in certain words, as the form <o'> "of" (37 occurrences) shows. This corpus demonstrates the

discrepancy between what prescriptivist sources tend to advocate, and how people perceive their pronunciation.

Academic stances regarding both dialects differ in the way these dialects are documented, as well as the way their speakers envision them on the forums. These dialectal lexemes are used online and nonstandard graphic representations are needed to suggest their pronunciation. Data collected from the corpus sheds light on how dialectal pronunciations are incorporated in non-standardized spellings.

3. Spellings

This part will tackle dialectal spellings which are triggered by specific pronunciation features in Scots and AAV. First, consonantal differences between standard English and alternative written forms will be discussed, after which a more precise analysis of consonantal changes within grammatical words will take place. Lastly, I will look at vocalic substitutions and the problems they raise.

3.1. Consonantal changes

Figure 1 sums up some of the most frequent consonantal changes, whether these are grammatical or lexical words. There are more alternations in the Scottish corpus than in the African American one, and both corpora exhibit fairly similar proportions of consonantal changes when looking at manners of articulation. The exception is the treatment of fricatives, mainly because of the tendency to drop <v> in “over”, “have” (spelled <hae>), “give” (<gies>, ‘gives’), or replace it with another letter, such as <w> (<ower>, ‘over’) in Scots.

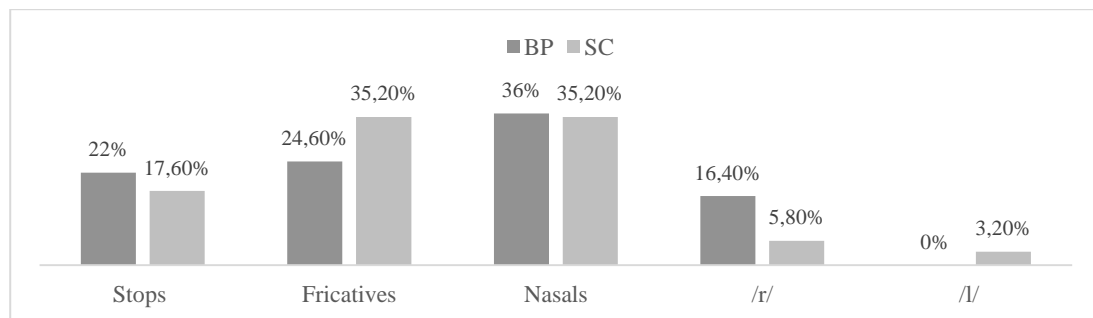


Figure 1: Consonantal changes according to articulation

A tendency towards the elimination of <d> and <t> in syllabic codas, which may not be replaced by any other grapheme is common in Scots. The form <da> for “dad” is used by two members of the forum, and it is not always paralleled by <ma> for “mother”. It is documented in the OSD, whereas the DSL does not have this form, probably because of its vernacular nature. In the graphic representation <thouless> (also spelled <thowless> in the DSL), the deletion of the letter corresponds to the pronunciation /^hθʌlʌləs/ recorded in the OSD. As for African American, there is no example of the dropping of /d/, but the spelling <lil’> is a shortening hinting at a pronunciation closer to /^hlɪl/, or even /^hli/. It is heavily documented in African American (Rickford, 2000).

Phenomenon	Example (SC)	Example (BP)
/d/ → ø, /ʔ/	<da> ('dad')	—
/t/ → ø, /ʔ/	<thouless> ('thoughtless')	<lil(>' ('little')
/g/ → /j/	<forjet> ('forget')	—
/ŋ/ → /n/	<daein> ('doing'), <havn> ('having')	<lookin> ('looking'), <fallin> ('falling')
/l/ → ø	<ca(> ('call'), <aw> ('all')	—
/r/ → ø	—	<fatha> ('father')

Table 1: Examples of the consonantal changes suggesting a dialectal pronunciation

The approximation of /g/ into a /j/ in the form <forjet> is implied through the use of <y>, and the pronunciation /fər'jet/ is attested in dictionaries. No user of SC chose another written form, which may prove that none of them knew, or wanted to hint at the Older Scots spelling <forjet>, using a close graphic equivalent for the yogh, for instance <3>⁹.

Both dialects feature consonant cluster reduction, especially in final position, and the devoicelessening of the last consonant (/ŋ/) is common in the corpora. It is a frequent phenomenon in AAV (Clark, 2014), and in informal speech. This change only occurs when <ng> is in final position, and is especially used when it is part of the functional morpheme <-ing>.

Scots is the only dialect where the approximant /l/ is dropped when it is right before the phoneme /ɑ:/. The letters representing /l/ in standard English (<ll>) are either substituted with a parochial apostrophe, or a <w> that symbolizes the pronunciation /ɑ:/ of the sequence <aw>. Scots recommendations involve changing <all> cluster into <aw>.

The sequence <aw> serves an entirely different purpose in AAV, where it often underlines a stylistic exaggeration following the dropping of <r> in postvocalic position, as it is a non-rhotic variety of English. It is the case in the form <lawd> ('Lord'), and may be a stylistic exaggeration, that is a pronunciation closer to /'laud/. In both cases, adding the letter <w> in a postvocalic environment represents either lengthening of the sound, or a diphthongization.

Other graphic substitutions of a postvocalic /r/ are used, such as <-a> for the standard English <-er> in forms like <holla> ('holler') or <betta> ('better'). Voicing of the unstressed syllabic /r/ is common in AAV (Mufwene 1998), and spellings such as <ho> ('whore') are frequent, both on the forum (27 occurrences) and online¹⁰. The semantic shift, from "prostitute" to "woman", may account for its popularity on the website. Due to its extremely derogatory nature, its standard English counterpart is censored on BP. Scots being a rhotic variety, there are very few alterations for postvocalic <r>, and final <-r>, and <-er> sequences are conserved.

⁹ A commonly used strategy for Arabic speakers is to use digits to replace Arabic characters. Thus <3> often replaces <ع> (Younes & Souissi, 2014, p. 71).

¹⁰ 1.090.000.000. hits for <ho>, against 60.200.000 results for "whore" on Google, on January 5, 2016. Because of the shortness of the alternative form, it is impossible to know precisely in what proportion <ho> is a substitution for "whore", and not an interjection like Santa's traditional "ho ho ho".

After this quick overview of the results for alternative representation of consonants in general, I will focus on how grammatical words are spelled differently in both communities. As they are more frequently used by forum members, these exhibit greater changes.

3.2. Grammatical words

When it comes to grammatical words, both SC and BP exhibit similar tendencies towards the elision or the change of certain sounds, reflected in spelling, although these do not have the same historical causes. Grammatical lemmas are more affected by spelling changes, because they are more common than lexical words. Only alternative spellings with the same underlying phonological explanation will be shown in table 2¹¹.

Phenomenon	Examples (SC)	Examples (BP)
<t#> → Ø	<disnae> ('does not')	<tryna> ('trying to')
/ð, θ/# → /z, t/	<wiz>, <wi> ('with')	<wit> ('with')
#/ð/ → /d/	—	<da> ('the')
/v/ → ø	<oer> ('over'), <hingoer> ('hangover')	<shoulđa> ('should have')
/f/ → ø	<o, eh, ae> ('of') <o't> (('of it'))	<lotta> ('lot of'), <o> ('of')
/r/ → ø	<fae> ('from')	<yo> ('your'), <fow> ('for')

Table 2: Most common alternative forms of grammatical words

African American forms such as <tryna>, <shoulđa>, <lotta> are snowclones (Pullum, 2004), as they are all made with the same model, that is <gonna> and <wanna>, where a verb is fused with the following element, whether it be a preposition, for <tryna>, or an auxiliary, as for <shoulđa>. The exception is the form <lotta>, where it is a quantifier which blends with the following preposition. All forms result in transforming the end of the word as <-a> (pronounced /ə/). Both <shoulđa> and <lotta> show the elision of a final fricative consonant (/v/ and /f/, respectively). They are the only words where the last element bears a consonant in the coda position, contrary to the other examples, which all end with “to”. “Of” is pronounced /v/ unless there is an assimilation, and is also attested in the form <o> (for both AAV and Scots). A hypothetical devoicing of /v/ into /f/, dropped afterwards, may explain the pronunciation /ʃudə/ for <shoulđa>.

Other known models can trigger alternative spellings. The negative adverb “not” in Scots, <nae>, is more often used at the end of an auxiliary (<disnae>, <arenae>, <huvnae>), and is also used when it does not follow any grammatical rule, as in the extract “yurr *gonnae* send it tae yer pals”. Here, the user seems to add an extra <e> to <gonna>, even if it is not negative. The ending <-ae> is common in grammatical words, hence its use here, where “going to” replaces the modal “will”. The form <nae> stands for both “not” and “no” in Scots, and is used widely on the forum (31 occurrences in the corpus).

Both linguistic groups converge when it comes to modifying the manner of articulation of /θ/ or /ð/, used in “with”. The standard American pronunciation of this conjunction is /wɪθ/, and it is also more common in Scotland than in the rest of the United Kingdom, where /wɪð/ is preferred

¹¹ These phenomena are taken into account in table 1 and complement them.

(Wells, 2008). Stopping (<wit> BP), deletion (<wi>, SC) or alveolarization (<wiz>, SC) of the final consonant may happen. These characteristics are documented for both dialects (Corbett et al., 2003, Mufwene, 1998). Stopping of the initial /ð/ is exceptional in the corpus from BP, with only two occurrences, written by two different members of the forum. This typical feature of AAV seems to mainly be linked to stylistic choices, as the users who resort to the graphic form <da> use <the> in the rest of their messages. This spelling only appears at the end of posts, in aggressive conversations: “I say fuk da tea party”, or “link da fuk up” (*post a link about*). This could hint at progressive assimilation of place, since the sequence /kð/ in “link the” and “fuck the” could be produced as /kt/, where /t/ is represented by <d>.

Both corpora show a tendency towards dropping the letter <v> (and of the corresponding sound /v/), especially in grammatical words, whether at the onset or the coda position of syllables. The Scots spelling <oer> can be realized as /'ʌu[ə]r, 'ɔ:ɹ, 'o:r, 'u:r/ according to the OSD and the DSL, and the Scottish propensity to omit /v/ in words such as “give”, spelt <gie> in the corpus is attested in other sources (Corbett, 2003). The very frequent dropping of /v/ explains the difference in proportion between Scottish and African American spellings in figure 1.

The preposition “from” is spelled ‘in different ways by members of both communities. The graphic form <for> for “from” (AA) is an example of a final consonant deletion, a phenomenon nasals are especially vulnerable to in AAV (Mufwene, 1998), probably followed by a metathesis, creating the form <for>. Some users resort to the spelling <fow> to replace “for”, and avoid ambiguities between the two prepositions. The Scottish form <fae> is historically attested and pronounced /fre:/. It is represented by the spellings <fra> and <fae>. The second spelling shows the reduction of the initial consonant cluster, and is pronounced /fe:/.

R-lessness (Rickford 2000) accounts for the spelling <yo> of “your”, which is probably phonologically closer to /'jʊv/ or /'jʊə/. The same mechanism explains the graphic representation of “for” as <fow>, as postvocalic /r/ tends to be voiced in AAV (Mufwene 1998). The letter <w> is used to show possible diphthongization (/foə/ or /foʊ/), as was the case for <lawd>, discussed in 3.1.

Prepositions and grammatical words are commonly written in Scots on Scotster, as table 3 suggests. The more frequent a word, the more alternative spellings it will have in the corpus. However, the most frequent word, “and”, is spelled in a standard way more often than the other three examples, because English speakers are used to writing it orthographically, hence limiting the number of substitutions. This is not the case for “with”, very often represented by <wi>¹², as in “Am wi you” (*I am with you*). This shorter equivalent is easy to understand in context, and is common in Scottish texts, since it is the only form used in the dictionary entry in the DSL.

This particular focus on prepositions can be explained by the greater number of Scottish people who want to practice or learn Scots: the frequency of these grammatical words immediately give a text a Scottish flavour.

¹² 23 occurrences in the corpus, which is quite a large proportion in the corpus.

Standard form	BNC rank (Frequency)	Alternative spellings (SC)	Ratio Alternative:Standard
And	3	<i>N; ane; an</i>	1:11
From	20	<i>Fae; fra</i>	1:3.5
Over	135	<i>Oer; ower</i>	1:5.5
With	14	<i>Wiz; wi</i>	1:2.3

Table 3: Comparison of uses of standard and alternative spelling of prepositions (Scotster)

One of the messages about Scots is metalinguistic, and the user tends to use as many alternative spellings as possible:

1. “Now adays people fae the next village speak a different twang, they hav aw got there ane different words. ay noticed that the further out you go fae a town or city, you still get the old local dialects.” (*Nowadays people from the next village speak a different twang, they all have their own different words. I’ve noticed that when you get far from a town or city, you still hear old local dialects.*)

In example 1, none of the alternative spellings affect nouns, and most function words (except “out”) bear a Scottish form. The word “twang” actually has a different meaning in Scots¹³, and its standard English definition is to be understood here.

Function words in general tend to be more altered in Scots than in AA. Vowel changes may be more problematic, as there are only 6 graphic characters to represent the 24 lexical sets of English, Scots and AAV. The focus of the next part will be possible homophonic confusions.

3.3. Homophony and limits of vocalic substitutions

Vowels tend to be affected by these alternative spellings, but homophonic confusion tends to arise more readily than with consonantal replacements. I will restrict the discussion to issues of potential homophony in this section, and how vocalic substitutions can create equivocal interpretations.

Some graphic alterations are easy to understand, such as <thang> for “thing” in AAV, as the lowering of /i/ to /æ~ɑ/ is part of the Northern Cities Chain Shift (Labov, 1991), and is a frequently used substitute for the original word¹⁴.

The first type of problems that can arise is between speakers of different dialects, especially with specific lexemes. As this does not happen on BP, I will only focus on Scottish productions, with example 2, about the lexeme “midden”:

2. “My word fae they dae is midden. When christine Aquilaira [C.Aguilera] was in Scotland someone called here a “radge wee midden,” [*crazy little slattern*] she thought it meant beautiful young girl.”

¹³ The DSL defines it as “a sudden, sharp pain”.

¹⁴ 51,500,000 hits on Google (Jan. 14, 2016). This form was made famous by Dr. Dre and Snoop Doggy Dogg’s hit song “Nuthin’ but a G Thang” (sic) in 1992.

This extract shows the chasm between a word in Scots and a very close English phonic neighbor, “maiden”. The graphic form of <midden> (/ˈmɪdən/) is attested in all Scots dictionaries, and it is defined as “1. a dunghill. [...] 3. a dirty, slovenly woman” (Warrack, 2000). Even speakers of the dialect struggled to understand, as they only knew the metaphorical use of the first two definitions (“I have been told that a midden is when your house is in a mess”), and the singer evoked in ex. 2 thought it was the English word, as pronounced by a Scottish speaker, with a monophthongization of /eɪ/.

Substituting a vocalic grapheme for another can cause misunderstandings, as there is no phonetic input or context to guide the other users. In example 3a, a user proposes a Scots spelling for the word “woods”, <wids>. As he does not provide any context, various English translations such as “weds” and “would” are posited. The skeletal consonantal structure is too short, so the instinct of the other speakers is to assume <i> stands for either the fronting of /ʊ/ or /u:/, or the raising of /e/ into a pronunciation closer to /ɪ~i/.

- 3a. Heres anither wan "Wids" (*here's another one "woods"*)
- 3b. it's either- Woods, or would?? Am I right?
- 3c. I thought it was 'weds'

These evaluations of graphophonemic replacements are fairly accurate, and show how speakers envision the differences between their dialect and standard English. The pronunciation of this word, spelled <wuid> (DSL) or <wid> (OSD) is indeed /wɪd/ in dictionaries, but the non-conservation of the number of letters in the alternative spelling was responsible for this confusion. That is the reason why altering vowels is often more likely to cause ambiguities than replacing consonants in non-standard spellings.

Conclusion

Language online is specific to the media, and Netspeak elements pervade both forums, as there is no official regulating body assessing the grammatical or orthographical correctness of messages. This constitutes a fertile ground for the creation of alternative spellings, which may bear the same forms as spelling mistakes. These creative manipulations are not always consistent, as stylistic changes can occur within a community, or even within a user’s productions. Stylistic choices may also be influenced by the presence of censorship, or another avatar of regulation of written productions.

Censorship is traditionally endocentric, and the first letter of the taboo word may bear its symbolic weight, as in “the F-bomb” or “the [#C] word”. On BP, censorship is exocentric and exerted automatically by the platform, after the users have composed and posted their message. To circumvent the effects of censorship, users tend to alter vowels with a close equivalent, and the initial letter, bearing an important symbolic weight, is very rarely replaced or dropped.

Official recommendations are usually not followed by members of both communities, and are outdated by this kind of corpus. African American innovations are more varied (at least 34% of all messages contain an alternative token), but dictionaries and lexicographical works do not reflect how people chose to express themselves on the website. There is no real difference between African American morphemes and those used in standard English. The non-rhotic nature of AAV accounts for words traditionally ending in <-er> being turned into <-a>. However, this is

not as frequent in the corpus as one could expect, and only appears on three words. The corpus also contains very little voluntary metatheses, such as <aks> (5 examples in the corpus) or <ax> (31 tokens on the whole forum) for “ask”, although these are often cited as AAV features. Many alternative spellings occur on profanity, as BP censors these words.

The examples from *Scotster* contain many more alternations: the corpus is stylized, and the wish to preserve the dialect explains the diversity of forms for any given word. Multiple dialects account for the various forms found in the corpus, but there is a high proportion of hapax legomena in the corpus. Even if members of this community are older than those on BP, 43% of messages contain at least one alternative token. Using Scots is a factor of cohesion although spelling recommendations are not always followed in this corpus either. The extra Scottish letter, yogh (<ɣ>) is replaced by its closer phonological neighbor <j>, which is not its closest graphic counterpart: the representation of the pronunciation is more important than respect for older spellings. Scots users display a greater stylistic flexibility, but spelling changes are mainly focused on grammatical words, accounting for the larger proportion of messages containing an alternative form.

Alteration of consonants have no correlation, either with the frequency of letters in the corpus or with speech (Crystal, 2003), and the develarization of <-ing>, symbolized by the spelling <in> is a very widespread phenomenon in English. It is the most common alternative in the corpora. Identical written representations may be used in both dialects, even if their underlying phonological motivations diverge. This suggests the importance of English spelling rules when it comes to creating a new written form for a word. These can have a historical reason as well, as is the case for the multiple forms of prepositions in Scots.

As for vowels, they are often affected by these changes, whether purely cosmetic (<!> or <y> for <i>), or a proof of diphthongization. Misunderstandings happen more frequently in short words, as the small number of letters representing vocalic sounds does not allow for fine-tuned written depiction of sounds. In such equivocal situations, English graphophonemic rules can be used to interpret the vocalic sound, but without any phonetic confirmation, these ambiguities remain.

Dialectal spellings are a composite mix between various written and graphophonemic codes in order to denote pronunciation, express oneself freely within the limits imposed by the platform while striving to be intelligible for others.

References:

- Davies, Mark. (2008-) *The Corpus of Contemporary American English (COCA): 520 million words, 1990-present*. Available online at <http://corpus.byu.edu/coca/>. Last access: January 22, 2016.
- Davies, Mark. (2013) *Corpus of Global Web-Based English (GLOWBE): 1.9 billion words from speakers in 20 countries*. Available online at <http://corpus.byu.edu/glowbe/>. Last access: January 2, 2016.

Dictionary of the Scots Language. 2004. *Scottish Language Dictionaries Ltd*. <http://www.dsl.ac.uk/entry/snd/burd_n1> Last access: January 10, 2016,
Eagle, A (2002—). *Online Scots Dictionary*. Online resource. URL: <http://scots-online.org/dictionary/scots_english.asp>. Last access: January 15, 2016.
Merriam-Webster.com. (2011), Retrieved Jan 8, 2016, from <https://www.merriam-webster.com/dictionary/>
Urban Dictionary (2003-). Available online at <http://www.urbandictionary.com/>. Last access: January 30, 2016.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Cambridge, Mass: MIT Press.
Clark, J. T. (2014). *Negotiating Elite Talk: Language, Race, Class and Identity Among African American High Schoolers*. Routledge.
Corbett, J., McClure, J. D., & Stuart-Smith, J. (2003). *The Edinburgh companion to Scots*. Edinburgh: Edinburgh University Press.
Crystal, D. (2001). *Language and the internet*. Cambridge: Cambridge University Press.
Crystal, D. (2003). *The Cambridge Encyclopedia of the English Language*. Cambridge, U.K: Cambridge University Press.
Crystal, D. (2008). *Txtng: The Gr8 Db8*. Oxford: Oxford University Press.
Deorowicz, S, and M G. Ciura. (2005) "Correcting Spelling Errors by Modeling Their Causes." *International Journal of Applied Mathematics and Computer Science*. 15: 275–286.
Education Scotland. (2012). *Scots spelling—Writing in Scots—Knowledge of Language*. URL: <<http://www.educationscotland.gov.uk/knowledgeoflanguage/scots/writingin Scots/scotsspelling/index.asp>> Last access: January 15 2016.
Goudet, L. G. (2013). “Alternative spelling and censorship: the treatment of profanities in virtual communities” in *Aspects of Linguistic Impoliteness* [Cambridge Scholars Publishing].
Kearse, R. (2006). *Street Talk: Da Official Guide to Hip-hop & Urban Slangue*. Fort Lee, NJ: Barricade Books.
Labov, W. (1991). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
McFedries, P. (1995). *The word spy*. <<http://www.wordspy.com/>> Last access: January 25, 2016.
Mufwene, S. (1998). *African American English: Structure, History, And Use*. London: Routledge.
Pullum, G. K. (2004). “Snowclones: lexicographical dating to the second”. *Language Log*. <http://itre.cis.upenn.edu/~myl/language-log/archives/000350.html>. Last access: January 25, 2016.
Rickford, J. R. (2000). *Spoken soul: the story of Black English*. New York: Wiley.
Seidner, S. S. (1982). “Ethnicity, language, and power from a psycholinguistic perspective”. *Proceedings of the International Symposium Scientific Committee: Centre de Recherche sur le Plurilinguisme*.
Smitherman, G. (2000). *Black Talk: Words and Phrases from the Hood to the Amen Corner*. Boston: Houghton Mifflin.
Warrack, A. (2000). *The Scots Dialect Dictionary*. New Lanark, Scotland: Waverley Books.
Wells, J. C. (2008). *Longman pronunciation dictionary*. Harlow: Pearson-Longman.