



# Individual Sensitivity to Change in the Lingua Franca Use of English

Irene Taipale<sup>1\*</sup> and Mikko Laitinen<sup>1,2</sup>

<sup>1</sup>Department of Foreign Languages and Translation Studies, University of Eastern Finland, Joensuu, Finland, <sup>2</sup>Center for Data Intensive Sciences and Applications, Linnaeus University, Växjö, Sweden

The study of ongoing change in English typically focuses on studying evidence from codified varieties of the language. Recent corpus studies show, however, that advanced non-native users of English may display heightened sensitivity to features undergoing frequency shifts similar to that experienced by native speakers. In addition, most studies aiming to detect patterns of linguistic regularity utilize large data sets that attempt to minimize the presence of the individual. In this study, we focus on change in ELF and place non-native individuals at the center of attention. Our empirical section examines how aggregated features that are currently undergoing change in codified varieties of English vary in the repertoires of ELF users of Twitter. To carry out this task, this study utilizes geo-tagged tweets retrieved from the Nordic Tweet Stream. The data obtained from this real-time monitor corpus are freely available for research and re-use at <https://cs.uef.fi/nts/>. For the analysis itself, we selected the idiolects of 150 individual users who actively tweet in English from geographically varying locations in Finland. As American English predominates with several patterns of linguistic change in codified varieties of English, a simplified dichotomy between American and British features is utilized as a conceptual tool for inspecting variation. The idiolects are analyzed from the perspective of spelling and lexico-grammatical and morphological variation, such as V + *-ing* | V + infinitive (e.g. *start doing* | *start to do*) and expanded predicates (e.g. *take a look* | *have a look*). The quantitative observations show that, particularly in the case of grammatical features, ELF speakers appear to have generally adhered to ongoing linguistic change.

**Keywords:** individual variation, idiolect, ongoing change, English as a lingua franca, second language, sensitivity, americanization

## OPEN ACCESS

### Edited by:

Cristina Suarez-Gomez,  
University of the Balearic Islands,  
Spain

### Reviewed by:

Al Ryanne Gatcho,  
International Islamic University  
Malaysia, Malaysia  
Ali Zangoei,  
University of Gonabad, Iran

### \*Correspondence:

Irene Taipale  
[irene.taipale@uef.fi](mailto:irene.taipale@uef.fi)

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Communication

**Received:** 06 July 2021

**Accepted:** 08 December 2021

**Published:** 27 January 2022

### Citation:

Taipale I and Laitinen M (2022)  
Individual Sensitivity to Change in the  
Lingua Franca Use of English.  
Front. Commun. 6:737017.  
doi: 10.3389/fcomm.2021.737017

## INTRODUCTION

This study focuses on ongoing frequency shifts in English as observed in advanced lingua franca use of English (ELF). The study of ongoing shifts has mainly focused on evidence from codified and standard varieties of English and has, in consequence, concentrated on the role played by native speakers as the driving force of change. However, digitalization has brought about a situation in which a very large number of people nowadays use English as a common resource in naturalistic settings in social media (Leetaru et al., 2013; Gonçalves et al., 2018). As an illustration, on some platforms the share of English social media posts in the Nordic region covers roughly 30–40% of all the available data (Coats 2019). This has created a unique sociolinguistic setting, which calls for novel empirical approaches in studying the role of advanced non-native users of English. What is more, it also opens up theoretical approaches to rethink the role played by multilingual individuals in language change.

As an illustration of some of the new theoretical openings, recent sociolinguistic evidence suggests that advanced ELF usage in social media is embedded in social settings in a way that has been understudied previously and calls for further investigation of the role played by non-native individuals. In a computational sociolinguistic investigation, Laitinen and Lundberg (2020) use metadata from tens of thousands of social media accounts from the Nordic region, notably involving a dataset that has been automatically pruned to exclude public figures and celebrities. Their study focuses on the size and quality of the social networks of multilingual individuals using Twitter. The results show that the social networks of those who prefer English over national/domestic languages tend to be substantially different from those of other groups. Their networks are larger and comprise a substantially greater number of weak ties, suggesting that the social embedding of advanced ELF use results in sociolinguistic conditions that promote change. According to a widely accepted view in sociolinguistics, weak tie conditions in which individuals are predominantly linked through occasional and insignificant ties (such as acquaintances) tend to support innovation diffusion (Milroy and Milroy 1978; Milroy and Milroy, 1985; Milroy, 1987; Lippi-Green 1989). The underlying reason for the strength of the weak ties is that adopting new ideas is always socially risky, while there is a smaller risk of disruptive outcomes in loose-knit settings than in close-knit surroundings.

Despite its being an observation taken from a single study, this finding shifts the focus to the level of individuals and calls for further research into the role played by multilingual and multicompetent non-native individuals in language change. Sociolinguistic embedding makes ELF a unique testbed for observing ongoing frequency shifts. Weak ties and the propensity for change in general could explain why advanced ELF users may display heightened sensitivity to features that are undergoing frequency shifts (Laitinen 2016). Corpus data suggest that the use of ELF does not follow the pattern observed elsewhere whereby post-colonial second language varieties in many contexts prove to be more conservative than native varieties (Collins 2009). Rather, the quantitative patterns observed in ELF are more complex, with ELF often trailing the leading variety; it may also be that recent frequency shifts are further polarized by ELF users.

We operationalize ongoing frequency shifts through Americanization, which in previous studies has been considered to be one of the main forces of change in present-day English (Leech et al., 2009; Baker, 2017). In turn, this points to processes whereby those forms that are more frequently found in American English (AmE) tend to spread to other varieties of English, and closely overlaps with other processes such as colloquialization and informalization. Partly because of this, it is also difficult to determine whether these trends of change can be rightfully described as Americanization, or if these changes simultaneously take place in many Englishes independently of one another. The dichotomy perceived between American and British features is naturally a simplification, and national preferences refer primarily to quantitative tendencies observable in large datasets rather than strict opposites. In addition, the putative distinction also suggests that American

English is a single entity, which is a partial misconception since it is subject to change and variation itself, as shown, for example, by Grieve (2016). The term ought therefore to be regarded simply as a conceptual tool that facilitates accessing ongoing frequency shifts and observing them in ELF data. Americanization makes it possible to pool a set of aggregate linguistic variables, thus enabling researchers to distance themselves from excessive use of randomly selected individual variables, concomitantly increasing the empirical validity of their observations.

In addition to shifting the focus to the individual level, this investigation extends to two less frequently studied areas of non-native speaker sensitivity to change. Firstly, it uses geo-tagged tweets that enable zooming in to the individual level with precision. Our observations are drawn from the *Nordic Tweet Stream*, which is a real-time monitor corpus that is freely available for research and re-use at <https://cs.uef.fi/nts/> (Laitinen et al., 2018). As we show in the following section, some previous studies of Americanization have used textual data from Twitter as their material. For instance, the global survey of Americanization in Englishes around the world by Gonçalves et al. (2018: 8) made use of a large collection (c. 31mi) of tweets, but despite this they show how “the real problem is the lack of data in certain cells” from countries in which English is not the main language. The NTS has been collected to fix the small data problem in such ELF settings. Secondly, most studies aiming at detecting patterns of linguistic regularity utilize datasets that either attempt to minimize the presence of the individual or rely on a very small number of individuals. This study places advanced non-native individuals at the center of attention as we investigate the idiolects of 150 individual users who actively tweet in English from geographically varying locations in Finland.

This article aims to provide an answer to the following research questions:

- 1) To what extent and in the usage of which linguistic variables are individual, advanced Finnish ELF users sensitive to an Americanization trend detected in codified varieties of English?
- 2) Do individuals from geographically varying locations and varying large online social networks exhibit different patterns of sensitivity?

The structure of this article consists of the following. **Section 2** introduces relevant previous literature and lacunae in the study of non-native English. **Section 3** details our material and methods, and **Section 4** discusses our observations. Finally, **Section 5** discusses the implications of our findings and presents thoughts on topic areas for future sociolinguistic study of advanced non-native Englishes.

## BACKGROUND AND PREVIOUS STUDIES

Frequency shifts connected to Americanization have become increasingly prevalent in recent decades (Leech et al., 2009: 252–259; Baker 2017). In Mair (2013) system of world Englishes, he suggests that it is appropriate to label AmE as the “hub variety,” a variety that is relevant to speakers of all other varieties. Keys to understanding this hierarchical system of

Englishes, as proposed by Mair, can be found in social factors, and he argues that the fact that linguistic forms preferred in AmE have spread to other varieties stems from the demographic weight of the US and a range of institutional factors prevailing in the country. These factors include the political, economic and military pre-eminence of the United States.

The phenomenon of Americanization has been observed in the study of lexis and grammar in native metropolitan varieties of English (Depraetere, 2003; Leech et al., 2009; Rohdenburg and Schlüter, 2009a; Baker, 2017). In addition, various post-colonial varieties have been investigated, including Americanisms in Nigerian English phonology and lexis (Awonusi 1994), the study of spelling and lexis in Kenyan, Singaporean, and Trinidad and Tobagoan Englishes (Hänsel and Deuber 2013). A range of studies has focused on empirical evidence concerning Philippine English syntax (Schneider 2011; Collins et al., 2014a; Collins et al., 2014b; Collins 2015), while the spelling and lexis of this variety may be found in Fuchs (2017).

In terms of data, the majority of past studies rely primarily on stratified corpus data that represent standard and edited language use, but some novel approaches also use societal big data from freely-available sources, such as the APIs of various social media applications. As an illustration, a global survey of English by Gonçalves et al. (2018) makes use of geo-located tweets written in English, thus expanding the empirical basis towards more naturalistic language use from across the world and improving the empirical validity in the study of non-native production. Their results show the strong presence of American English variants across the globe, except in countries where colonial “British influence has been strongest” (2018: 12). In addition, the data show that in non-native settings in the expanding circle contexts “American orthography and vocabulary dominate” (2018: 9–10).

Past approaches that focus on Americanization across varieties have focused on linguistic analyses, while in the process they have downplayed the sociolinguistic aspects. We argue that being sensitive to change ultimately makes the quantitative study of all non-native users a variationist sociolinguistic endeavor. According to Weinreich et al. (1968), the most essential aspect in the study of linguistic change resides in orderly heterogeneity, in which variation is not random but systematic. To understand this systematicity and to study change in social context involves a range of levels of analysis. Its actuation involves charting for the emergence of innovative linguistic items, while change can be constrained by a range of systemic factors. Transition focuses on investigating how change proceeds from one stage to another, from the individual to the community. The question of embedding examines how change is embedded in the linguistic and social structures in language use, while social evaluation takes into account how linguistic forms are assessed by users. The core is to approach forms as the linguistic variables of writers’ choices between alternative ways of expressing the same meaning. By applying a variationist approach to our data, we are able to look into how frequency shifts might be embedded in the social structures and how they are evaluated.

The variationist approach offers us a systematic apparatus for investigating variation and change in ELF (see Laitinen and

Lundberg 2020 for a study of embedding). Despite the obvious benefits, variationist approaches have not yet been widely used in the study of ELF, as micro-level investigations have predominantly focused on the cognitive processing capabilities of individuals (Vetchinnikova 2017; Mauranen 2018). The role played by an individual has been more extensively highlighted recently by Vetchinnikova and Hiltunen (2020), who argue for the need to observe variability in ELF on the individual level. Their study sets out to investigate the extent to which variability in advanced ELF use stems from individual variation. Observation of ten individuals in online ELF environments reveals that the individual and the communal levels are different and the communal level ought to be “seen as emergent from the individual” and as being qualitatively different from it (also Vetchinnikova 2017).

In addition to social embedding, sociolinguistic evaluation is closely linked to the setting from which we have drawn our data. Our primary data consist of English data from Finland, where English is used as an additional language alongside the national languages (Finnish and Swedish). It goes without saying that ELF in such contexts is not a focused variety with a clearly defined speech community or a widely-agreed set of norms and common structural basis. However, language users are faced with a choice related to the norms of what the preferred language variant is. Large-scale survey data reflecting attitudes in this setting suggest that American and British varieties are generally considered prestigious. In Leppänen et al. (2011: 70–74), the respondents were requested to assess which varieties of English they found most appealing. The findings show that people in urban areas prefer British English, whereas those living in the countryside prefer American English. The division is clear since the more highly educated respondents – in general, those living in urban areas and older informants – prefer BrE, whereas the respondents living in the countryside and young informants in general prefer AmE (Leppänen et al. 2011: 73).

The following section details a data source that provides researchers with access to naturalistic advanced non-native use of English on one social media platform. The dataset contains geo-location information which makes it possible to access non-native language use from multilingual populations in various regions, thus opening up new empirical avenues in the study of advanced ELF use.

## MATERIAL AND METHODS

Three sets of data from geographically distinct areas were compiled using *the Nordic Tweet Stream* (NTS). The data consist of tweets from the time period of June 1, 2016 to April 30, 2017. The first set of data aims at representing the largest cities in Finland and includes tweets from Helsinki, Tampere, Oulu, Turku, and Jyväskylä. The second set represents medium-sized localities, while the third group contains data from the countryside and small towns. It consists of small amounts of data from several different locations in all parts of the country. Each regional set of data consists of the idiolects of 50 users, resulting in altogether 150 idiolects for analysis.

The sorting was, for the most part, automated. However, some manual checking was executed, as the API crawler falsely identified many English tweets as instances of some other language. In addition, spam, tweets automatically generated by third-party applications, and extensive amounts of news and forecast reports written in a repetitive form were removed from the data through manual inspection.

According to recent estimates, geo-tagged tweets offer a good overview of local language ecologies despite representing only a small fraction of the material (see, for example, Lamanna et al., 2018). However, to reduce the possible effects of traveling, additional information was used to fine-tune the data. The NTS collects various kinds of metadata ranging from the time of the tweet being sent to the information that the users have written about themselves on their profile. The tweets were sorted by discarding those that were sent from a user who has clearly set their hometown set to something other than the three areas mentioned earlier. For example, the tweet was removed in cases where a tweet written in the countryside was sent by a user who has inserted Helsinki as their hometown.

The information available about the most active users was also inspected qualitatively to see if there were further signs of their nationality and place of residence. Accounts that appeared to belong to people who were clearly from countries other than Finland were also removed from the analysis. The tweets of the 50 most active users in each region have been selected for this final examination.

Subsequently, the final data set was syntactically annotated using the GATE Twitter part-of-speech tagger (Derczynski et al., 2013) designed specifically for tagging tweets. The tagger utilizes the Penn Treebank part-of-speech tags. It is approximated to achieve 91% accuracy, which we consider to be acceptable for our research questions in this text category. The final set of data consists of a total of 34,830 tweets and 546,542 tokens.

With regard to methods, we focus on variables in the spelling and in the lexico-grammar. Spelling differences among AmE and BrE are relatively stable and may be known to laymen, whereas empirical evidence suggests that most of the grammatical variables are undergoing change, which we assume the ELF users are unlikely to be aware of. Comparing non-native usage of these kinds of variables offers a unique perspective into the study of intraindividual variation. The analysis examines a total of seven spelling variables in which AmE and BrE tendencies differ significantly. The variables come from a total of 60 distinct types or word stems, which were searched for with the aid of WordSmith Tools 7. The full list of the individual words is presented in **Appendix 1**.

A methodological issue concerns the differences between American and British English. A fundamental problem ingrained in many publications addressing the topic is that they merely state that certain spelling conventions and lexical items set BrE and AmE apart, but without empirical evidence to support these claims (Trudgill and Hannah 2008). The variables selected have been chosen on the basis of a number of empirical studies concerning the differences between AmE and BrE and how they have adapted to change (Mair 2006; Vosberg 2006; Hundt 2009; Leech et al., 2009; Rohdenburg and Schlüter 2009b;

Tieken-Boon van Ostade 2009: 38; Tottie 2009; Ishikawa 2011; Baker 2017).

When the variables of this study are compared to a similar study by Gonçalves et al. (2018), some differences emerge. For example, *-ize* | *-ise* is omitted from Gonçalves et al. (2018). Other scholars such as Baker (2017: 35) are also wary of considering *-ize* | *-ise* as a true spelling difference between AmE and BrE, as it has been speculated to be a by-product of word-processing software. Regardless of the origins of this prevalence of *-ise* in BrE, it seems that *-ise* is nevertheless regularly used in BrE (Tieken-Boon van Ostade 2009: 38; Ishikawa 2011), which is why it is included in this study. In addition, Gonçalves et al. (2018) examine lexical variation and exclude grammatical variables from their analysis. The present study does not include lexical variation as it is prone to many pitfalls, the most challenging of them being that many lexical items known to separate AmE and BrE have slightly different meanings, which means that they are not in true co-variation with one another. As an illustration, *biscuit* is often cited as the BrE equivalent of the American *cookie*, but *biscuit* is also used in AmE to refer to a different type of pastry (Trudgill and Hannah 2008: 90; Murphy 2018: 72). In their study, Gonçalves et al. (2018) include *cookie* | *biscuit* and some other lexical items whose semantic equivalence is debatable (see, for example, Baker 2017: 127 on *truck* | *lorry*). In addition, Baker's (2017: 149) empirical findings show that lexical variation between the two varieties is rather low.

The set of aggregated spelling variables and examples of them are assembled in **Table 1**. The spelling variables selected consist of *-or* | *-our*, *-er* | *-re*, geminate consonants, *-ense* | *-ence*, *-yze* | *-yse* and some instances of *-ize* | *-ise* (see Appendix for a complete list of the inspected variants). In addition, some individual words have also been inspected.

The raw findings were manually pruned to exclude false positives. The example in (1) illustrates a case that was excluded from the final results. The writer displays mixed use of the AmE *gray* and BrE *grey* at the level of a single tweet, which is an accompanying note for a sepia-toned photo of a hamburger.

1) Burger in 50 Shades of Gray #Burger #50shadesgrey #Joytrip #Tampere #Koskipuisto [LC30]

One can only speculate whether this choice is conscious or not. The writer may have made a deliberate choice of using both the AmE variant *gray* in their main tweet but also the correct name of the erotic novel franchise in question, which adheres to the British norm in their hashtag. However, the hashtag is most likely not written in an attempt to entice fans of *50 Shades of Grey*, because the tweet is merely a pun and its theme is not the franchise per se. The hashtag has also been written without the preposition *of*, so even with the BrE *grey*, the formation is not the name of the novel in its totality.

Each illustration is replicated in the exact form that it appears in the original dataset, apart from the attached URLs and emoticons. Each user is given a random number based on the location of the tweet in an abbreviated form: CS is short for countryside, MST for medium-sized towns and LC for large cities. More detailed metadata, such as the exact location, has been

**TABLE 1 |** The examined spelling variables.

Variable	AmE example	BrE example
or   -our	color	colour
-er   -re	center	centre
geminate consonants -l-   -ll- and -p-   -pp	traveled; skillful; worshipped	travelled; skillful; worshipped
-ense   -ence in nouns	license	licence
-yze   -yse and some cases of -ize   -ise	analyze	analyse
	realize	realise
individual word spellings	toward, gray, math	towards, grey, maths

**TABLE 2 |** The grammatical variables used in the study.

Variable	AmE example	BrE example
gerund complement   to-infinitive	start doing	start to do
preterite and past participle forms of verbs such as <i>burn + get</i> and <i>have gotten/have got</i>	burned; have gotten	burnt; have got
downtoners	kind of, kinda	sort of, sorta
expanded predicates	take a look	have a look

excluded because the data is not analyzed at the level of individual cities, but with the privacy of the users in mind as well.

Most of the grammatical variables that are included in this study consist of lexico-grammatical items. They exist on the border between grammar and lexis, such as the variation between gerunds and to-infinitives in the complementation of catenative verbs that express the onset or continuation of activities or states of being (*start*, *begin* and *continue*). Other lexico-grammatical items include expanded predicates (e.g., *take | have + a look* etc.), the downtoners *kind of* and *sort of* and their reduced forms *kinda | sorta*. Some variables are morphological, such as two forms of irregular verb inflections (*burned | burnt*), in which the AmE variant is the more advanced. However, in the case of *gotten | got*, the opposite applies, which illustrates how conceptualizing AmE as the more advanced variety is a simplification as well (see Hundt 2009 for a detailed inspection of the formation of *gotten* in AmE). **Table 2** presents the chosen lexico-grammatical and morphological features.

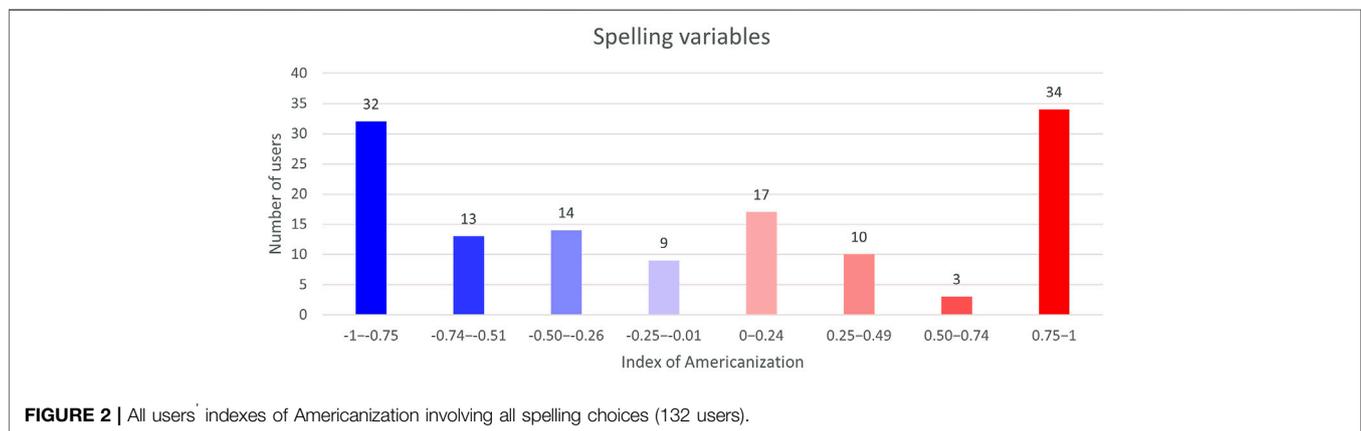
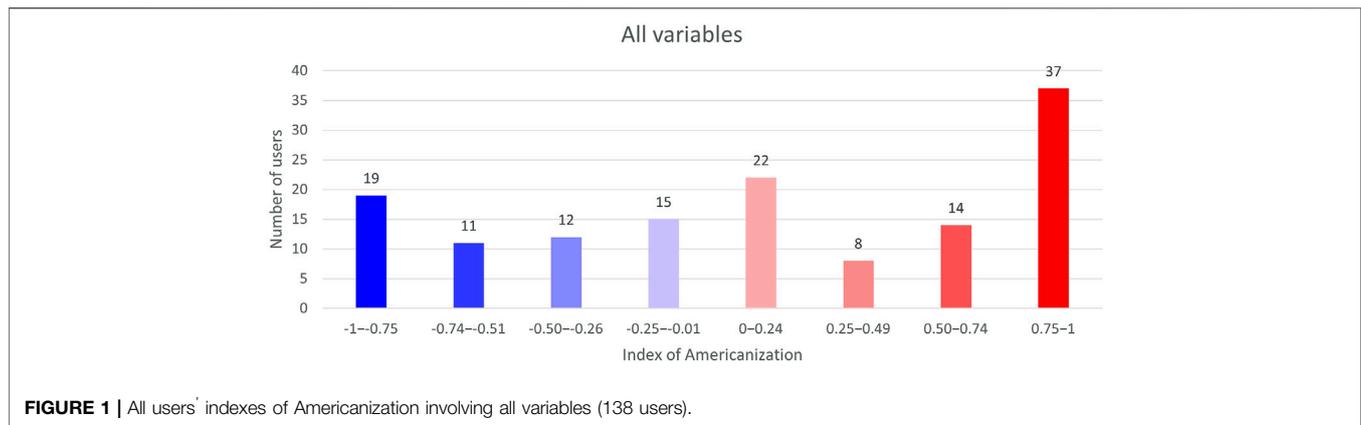
As with the spelling, each word or word stem was searched individually. This involved manually checking the concordance lines to ensure that the instances represent relevant language uses. Firstly, spellings that are known to have separate meanings were excluded. For instance, spellings that refer to measuring devices, such as *barometer* and *thermometer*, were excluded from the analysis of *-er | -re*, as the spelling is the same in both AmE and BrE. In addition, many proper nouns, such as usernames, were excluded. Hashtags are included in the analysis for the most part: in cases such as *#favorite*, the hashtag does not seem to link the tweet to any specific online phenomenon. Hashtags that are most likely related to a specific chain of tweets or a phenomenon, such as *#cannabislicense*, or hashtags that are otherwise written in a fixed way, such as *#votelabour*, would have been excluded, but these kinds of hashtags were not detected in the manual analysis.

Ensuring that the hits for grammatical queries were in fact instances of the grammatical variation in question required

equally careful manual pruning. For example, only some meanings of *have got* are in co-variation with *have gotten* (e.g., Biber et al., 1999: 398–400; Algeo 2006: 14), and syntactic annotation does not account for these shades of semantics. In fact, most of the initial hits were instances that referred to possession or necessity and obligation, which are not expressed via *gotten*. These kinds of false positives were excluded via a manual inspection. Similarly, search queries for expanded predicates generate thousands of false positives, as the syntactic annotator does not differentiate between expanded predicates and other verb + noun constructions, such as *take a photo*. In consequence, it was decided to search for a selection of common expanded predicates. The object nouns selected are *bath*, *break*, *look*, *nap* and *shower*. In the case of downtoners, instances where the downtoner precedes a noun phrase, as in (2), were excluded. This is because only grammaticalized uses of downtoners as estimators (3) appear to hold regional contrasts (Algeo 2006; Rohdenburg and Schlüter 2009b):

- 2) We're asking for innovations, but what kind of innovations? Pokemon games? A marvelous innovation #basicincome is not implemented, Why? [LC45]
- 3) I'm kind of concerned about that dog's wellbeing. XD Maybe it's time to change the owner? Based on that scream :P [MST11]

To acquire quantifiable results, we utilize an index of Americanization, adopted from Gonçalves et al. (2018). This index determines the share of variants associated in the data with American English. It also ensures comparability between variables, which vary in token frequency and provide concrete numbers to analyse the results with. This has been used for the analysis in each of its levels, ranging from individual users' preferences to normalized total data sets. It has been calculated using the following formula:



$$AmE\ index = \frac{(AmE\ variants - BrE\ variants)}{(AmE\ variants + BrE\ variants)}$$

Using this index, the results range from -1 to 1: an index of 1 would signify that instances of a certain variety feature conform solely to the American norm. Conversely, -1 would mean that only British variants are used. An index of 0.5 can be considered to be the threshold of highly Americanized usage.

The results are visualized with a corresponding color continuum, where an index of 1 is represented with red, -1 with blue and 0 with white; in other words, the stronger the index of Americanization, the darker the shade of red.

## RESULTS

### Overall Results Concerning Spelling and Lexico-grammar and Morphology

Of the 150 users' idiolects that were investigated, 138 provided relevant hits. **Figure 1** provides an overview of all users' total indexes of Americanization, meaning that all of the instances of both spelling variables and lexico-grammatical and morphological variables were examined together:

The figure displays a peak at the Americanized end of the scale: a total of 37 users appear to operate vigorously with American variants. However, a total of 57 users' total indexes land

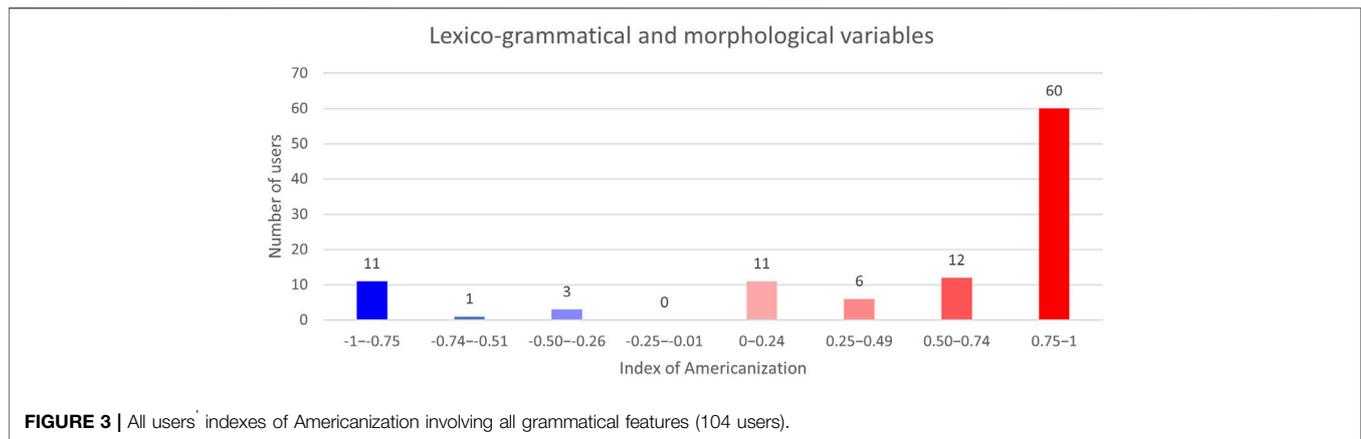
somewhere between -0.5 and 0.49, meaning that approximately 41% of all users steadily alternate between AmE and BrE forms.

**Figure 2** illustrates the distribution when only the spelling variables are examined. When this restriction is applied, the users' preferences appear to be rather polarized.

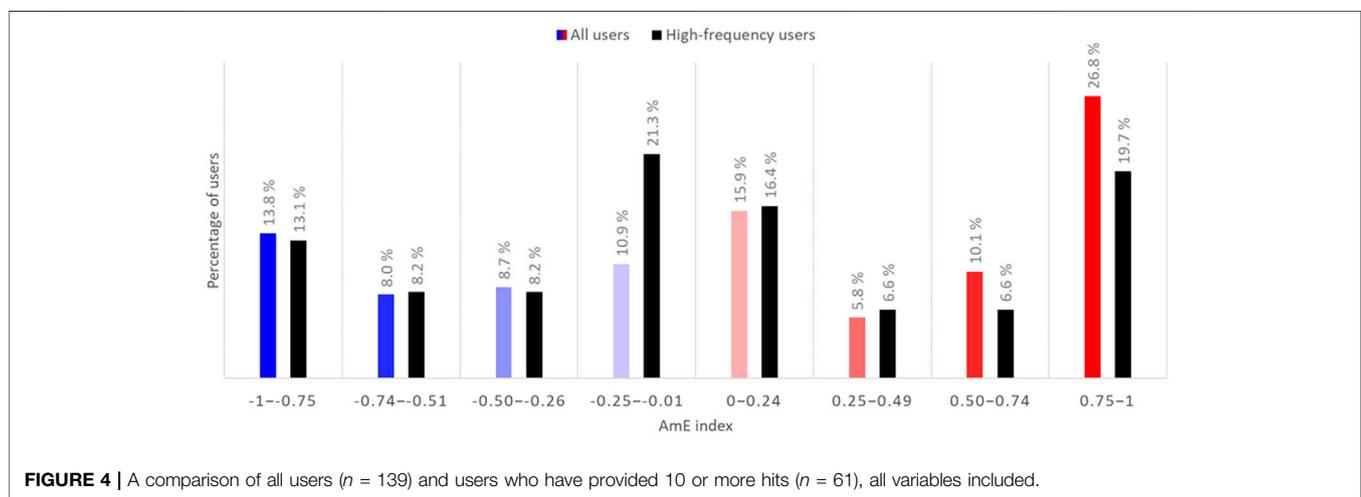
In fact, the number of users that exhibit polarised usage is roughly the same at both ends of the spectrum. There is also, however, an approximately equal portion of users hovering between the spelling norms of both varieties.

The results involving lexico-grammatical and morphological variation exhibit a clear trend in their use of advanced American variants. A total of 60 users (equal to 58% of the users examined) have adopted the AmE variants extensively, as shown in **Figure 3**.

There are only a few exceptions, since 11 users appear to consistently favour the British norms. From the perspective of sensitivity, the grammatical variants that are prone to linguistic change are the most relevant. Namely, at least the vast majority of users are laymen who are most likely unaware of lexico-grammatical and morphological tendencies observed in present-day Englishes. The users are more likely to recognize regional spelling contrasts that are, for instance, taught in school, and thus they may consciously choose the more pleasing variant regardless of which variant they have been exposed to more frequently.



**FIGURE 3 |** All users' indexes of Americanization involving all grammatical features (104 users).



**FIGURE 4 |** A comparison of all users ( $n = 139$ ) and users who have provided 10 or more hits ( $n = 61$ ), all variables included.

However, the most polarized results must not be accepted *de facto*. Namely, many of the users who exhibit an extremely polarized usage of variants are those who have provided relevant data in lesser quantities. **Figure 4** illustrates how the repertoires of the so-called high-frequency users', i.e., those who have produced a minimum of 10 instances for inspection, differ from the data as a whole.

The two groups exhibit, for the most part, similar patterns of variation. However, mixed use of AmE and BrE variants is more common for those users who have provided 10 or more relevant linguistic occurrences. A total of 21.3% of high-frequency users' AmE indexes extend from -0.25 to -0.01, whereas when all users are examined the percentage is only 10.9. In addition, the groups that display the most Americanized use of variables differ in size, with the high-frequency users' group being 7.1% smaller.

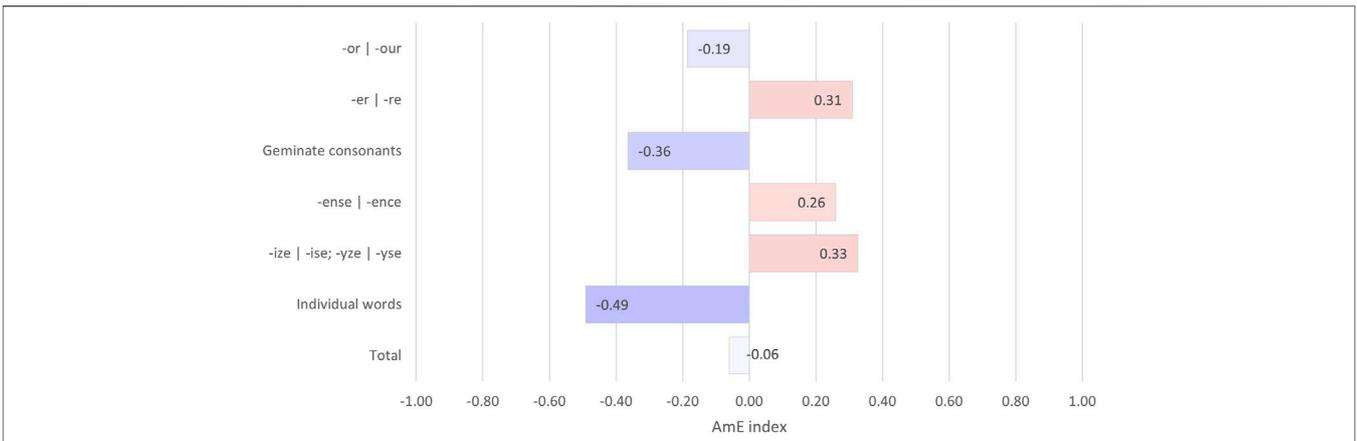
### Individual Features

When the results at the level of individual variables are inspected, several diverging patterns emerge. As a first step towards analyzing how much individual variation there is in the use of each variable, the AmE indexes of the occurrence of each variable were calculated for each user. The means of these indexes are

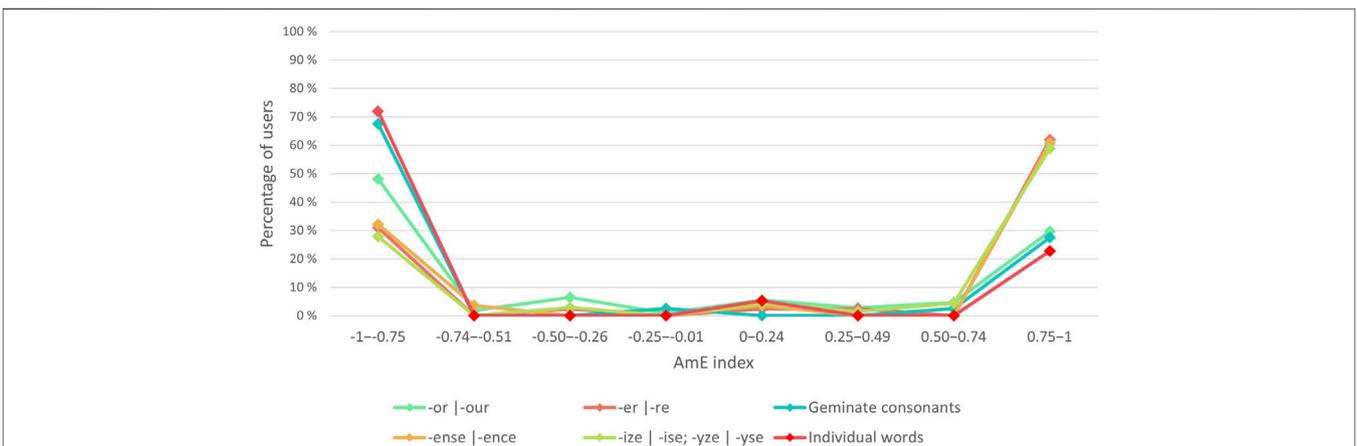
presented **Figure 5** and **Figure 6**. **Figure 5** presents the results involving the individual spelling variables. The results have been calculated based on a total of 582 relevant instances.

As a group, individual ELF speakers appear to use the inspected spelling variables in a mixed way, the mean being close to zero (-0.06). There are also significant differences in the use of individual variables: individual BrE spellings such as *grey* and *towards* are more common than their AmE variant, whereas forms such as *-er* and *-ize* lean more toward the AmE side. However, none of the spelling variables exceeds the threshold of Americanized usage (0.5); instead, the spelling variables that are closest to appearing in a form consistent with either variety are individual words that adhere to British norms with greater frequency. The polarity of this result may be explained by the fact that the variable group in question has the least individual variables. Groups of variables with roughly equal numbers of total instances (geminate consonants; *-ense* | *-ence*) appear to be used in an even more mixed way. This may indicate that some words may have a more fixed spelling than others.

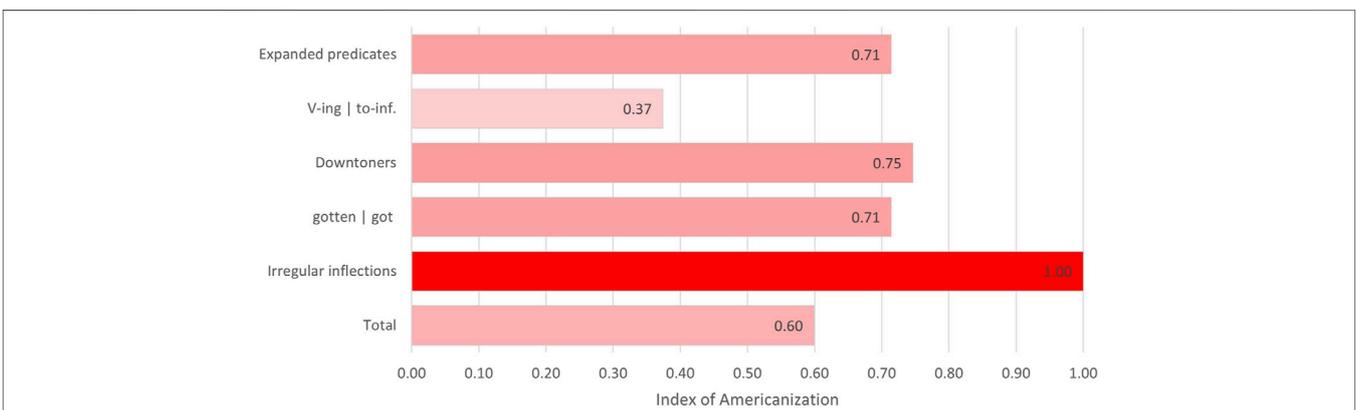
However, when the users are divided into groups based on their AmE indexes and hence are examined as individuals rather



**FIGURE 5 |** Individual variation in the use of spelling variables presented as AmE index averages.



**FIGURE 6 |** The distribution of users' AmE indexes at the level of individual spelling variables.



**FIGURE 7 |** Individual variation in the usage of morphological and lexico-grammatical variables presented as AmE index averages.

than as a group, converging patterns appear. Namely, it appears that individual users opt for either AmE or BrE variants consistently, as **Figure 6** demonstrates:

The figure shows how the majority of users are positioned at either one of the polarized ends of the spectrum, whereas mixed usage inside an individual group of variables is rare. Users often

simultaneously use AmE variants in the case of one group of words and the BrE variants in the case of another variable. The most polarized spelling variables are individual words that more than 70% of users write according to the BrE standard, while geminate consonants come second. Spelling forms that are predominantly written in the AmE style include *-er*, *-ense* and *-ize/ -yze*. The variation between *-or* and *-our* appears to divide users the most, as roughly half of the users strongly prefer *-our*, while approximately a third of users mostly opt for *-or*. For example, the single user MST28 in (4), (5), and 6) exemplifies a typical conjoined usage of spelling features from both metropolitan varieties (AmE *-er* and *-or* and BrE *towards*):

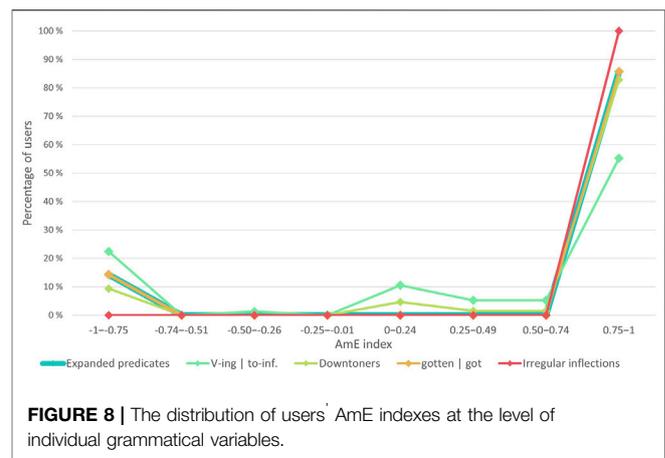
- 4) @user Off center, of course. This is such an undebatable question [MST28]
- 5) @user I like the EU version more, looks more colorful to me His character is fucking weird but charming. I'm weird. [MST28]
- 6) I feel like I'm the only person that has walked on earth who doesn't like Zelda #FlippedLearning #AI and #Robotics? Or will the pedagogic pendulum swing back towards the #lecture? [MST28]

A shift of perspective from the traditional, large-scale group and their averages to the level of individual variation provides a clear difference in outcome: what appears at the level of all users as mixed, inconsistent usage of a certain linguistic variable is in fact a result of polarities emerging from both extremities of the scale. In other words, as a larger population, Finnish ELF users use spelling variants in a mixed way, but as individuals they are usually consistent at the level of individual variables. Of course, as was noted earlier, some of the observed polarities are the result of a lack of large quantities of data.

A similar comparison of group tendencies and variation at the level of the individual was carried out in the case of lexico-grammatical and morphological variables. The averages concerning these variables are presented in **Figure 7**. The results are based on a total of 360 relevant hits.

In the case of lexico-grammatical and morphological variables, the results are more consistently Americanized than in the case of spelling, even at the level of averages. Only the variation between gerundial and infinitive complements falls below an AmE index of 0.5. Irregular inflections appear to be the most Americanized of the variables; however, a plausible reason for this result is the scarce amount of data available for this variable (a total of 19 instances), decreasing the significance of the result.

As in the observation of spelling variables above, users can be divided into groups based on their AmE indexes at the level of variables to see how consistent the individual users truly are. **Figure 8** shows that most users opt for advanced, American variants. Mixed usage of any variable is very rare, but 11% of the users have a tendency to mix *to*-infinitives and gerunds, and an even more minuscule group (5%) of users sometimes uses both *kind of/kinda* and *sort of/sorta*. In the case of verb complements, this kind of mixed usage is what could be anticipated, as the use gerunds has not been observed to completely replace *to*-infinitives (Leech et al., 2009: 195, Mair 2006: 128–130).



**FIGURE 8 |** The distribution of users' AmE indexes at the level of individual grammatical variables.

While AmE lexico-grammatical and morphological variants are used overwhelmingly, rather than BrE forms, there is room for individual variation. This is particularly evident in the usage of gerunds and *to*-infinitives, which are the least Americanized of the variables. Examples 7) and 8) show how gerunds and *to*-infinitives can appear in co-variation at the level of an idiolect. The user in question uses both the gerund and the *to*-infinitive to convey what appears to be exactly the same meaning with the same verb *hate*:

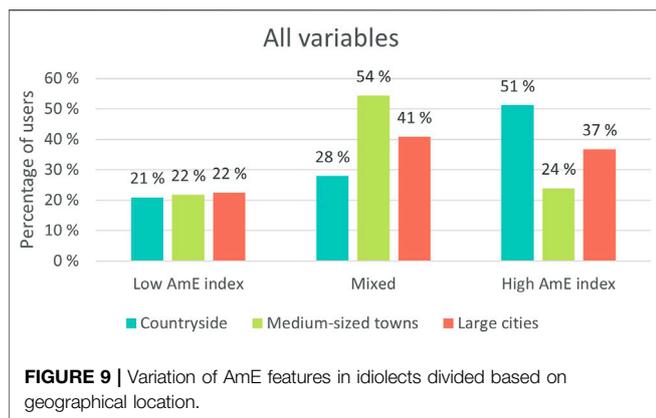
- 7) i lit ('literally') can't live my life without constantly being afraid that people start to hate me out of nowhere [LC32]
- 8) Every time when i see a goodlooking person I immediately start hating myself i can't help it akbenrkgnf [LC32]

What makes the example even more relevant is that *hate* is a stative verb, which at least according to traditional grammars usually pairs up with infinitives (e.g., Mair 2003). While there would be room for more qualitative analysis of syntactic and semantic restrictions that might govern the users' choices, examples 7) and 8) show that the users' choices are at least not fully explainable via these kinds of constraints.

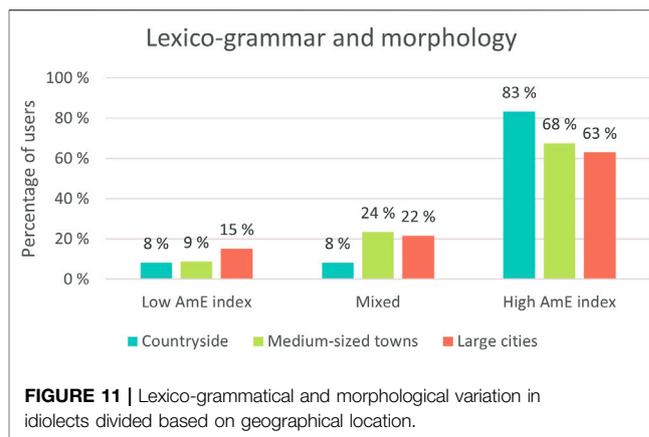
## Regional Variation

This section makes use of geotagging in the material and examines possible regional differences in the users' AmE indexes. To make the results more tangible, the 142 users were divided into three groups based on the level of Americanization of their idiolects. The first group, which represents those with a low AmE index, consists of individuals whose AmE index ranges from -1 to -0.51. The second group, which represents those whose repertoires are mixed, contains users with indexes between -0.5 and 0.49. The third group consists of users whose repertoires exceed the threshold of Americanized usage (0.5) and who thus may be most sensitive to ongoing change. Firstly, when all of the variables are inspected relative to the region the results take the form visualized in **Figure 9**.

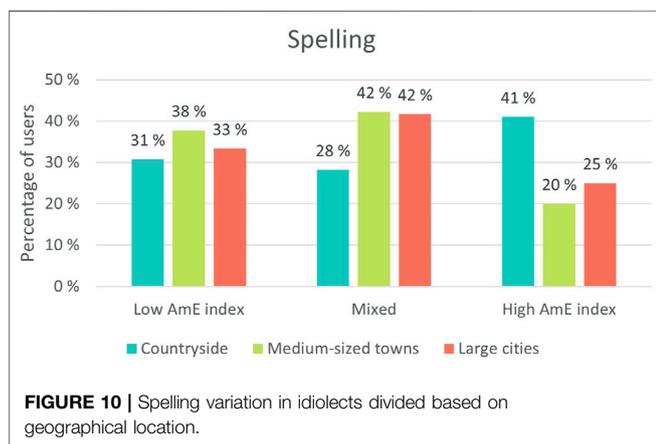
As **Figure 11** illustrates, the proportion of users with a low AmE index is almost equal in every region. However, in the case of users who mix AmE and BrE variants, differences



**FIGURE 9** | Variation of AmE features in idiolects divided based on geographical location.



**FIGURE 11** | Lexico-grammatical and morphological variation in idiolects divided based on geographical location.



**FIGURE 10** | Spelling variation in idiolects divided based on geographical location.

emerge. Over half (54%) of the users from medium-sized towns are prone to mixing the variants. Almost a third (28%) of users from the countryside belong to this group, and an even larger number (41%) of users from large cities also hover between the AmE and BrE variants. In the case of users with a high AmE index, users from the countryside make use of American forms the most, as 51% of them have a high AmE index. Thus, it appears that this small selection of country-dwellers may be most sensitive to AmE influences. Individuals from the medium-sized towns data set, on the other hand, clearly juggle between AmE and BrE variants the most (54%), with large city users coming second (41%). The share of users whose AmE indexes are low is almost equal in each regional group.

When spelling variables are inspected by separating them from the lexico-grammatical and morphological variables, the quantitative patterns shown in **Figure 10** emerge:

As in the overall results, non-native English users from the countryside appear to be the most prone to using AmE variants, as 41% of them have high AmE indexes, whereas the approximately 42% of individuals from both medium-sized towns and large cities mix these variants. In the case of medium-sized towns and large cities, the number of users who tend toward the BrE side is also higher than the number of those whose spelling tendencies are highly Americanized.

Labov’s (1966: 499) well-known distinction between “change from above” and “change from below” may be seen in the users’ repertoires. In Labov’s dichotomy, change is categorized according to whether change takes place above or below the boundaries of conscious awareness. In conjunction with our corpus results, Finnish people living in the countryside reportedly find American English more appealing than those living in urban areas, whereas Finns from large cities prefer BrE (Leppänen et al. 2011: 70–74). In addition, BrE is quite obviously considered a prestigious variety, and the inspected users are most likely aware of the status of BrE. It may be that people living in large cities find the prestigious British variety more compatible with their identities than country-dwellers do, and it may also be that this association is reflected in their spelling choices that may play a part in the users’ construction of their social media personae. Based on these findings, it can thus be speculated, on the one hand, that via their spelling choices users may wish to make use of existing associations and collective beliefs concerning AmE and BrE and the people whose native languages these varieties are.

On the other hand, in the case of grammatical change, which most likely takes place below conscious awareness, the Americanization trend is evident in each region. Some regional differences do, however, emerge. The share of users from countryside with a high AmE index surpasses that of other regional groups, as seen in **Figure 11**:

Users from large cities, on the other hand, appear to use American lexico-grammatical and morphological forms the least. Hence, not only do country-dwellers appear to be most sensitive to American influences, but large city users appear to be the least sensitive. Thus, at least from the perspective of the raw numbers, it appears that advanced ELF users from the countryside are the most sensitive to Americanization. However, this result is not statistically significant, as the seemingly towering difference in proportion is a result of there being a relatively small number of users to compare: 24 in the countryside, 36 in medium-sized towns and 46 in large cities. Thus, the usage of American lexico-grammar and morphology appears to be a phenomenon that unifies the regions: users appear to have picked up AmE lexico-grammar and morphology regardless of their location. It would have also been counterintuitive that countryside users’ repertoires

**TABLE 3** | Network sizes across geographical sub-corpora.

	Median of twitter followers	Median of twitter friends	Index of strong ties
Countryside	408	378.5	0.92
Medium-sized towns	474	424.5	0.90
Large cities	717	690.5	0.96

**TABLE 4** | Network sizes and users' AmE indexes.

	Median of twitter followers	Median of twitter friends	Index of strong ties
High AmE index	448	317	0.71
Mixed	447	482	1.07
Low AmE index	759	813	1.07

were significantly more advanced. Furthermore, Eisenstein et al. (2014) observe how new words that are coined in Twitter spread in a regionally tied manner from city to city, despite the possibility of social media transcending these kinds of geographical boundaries.

In addition, the social network theory predicts that weak and uniplex networks are the most fruitful platforms for change, which would make users with the largest number of friends and followers most susceptible to being agents of change. **Table 3** shows that, with regard to both friends and followers, users from countryside locations have smaller networks than those from medium-sized towns and large cities. The index of strong ties indicates the relative proportion of the users' contacts who are friends compared to followers: Twitter friendship requires reciprocity and thus indicates that the users share a stronger bond. As can be seen from **Table 3**, the countryside users' social networks, while smaller, contain fewer strong ties than the networks of users from the other areas. This kind of looseness could help to explain why users from the countryside may be more sensitive to change than the rest of the users.

Similar patterns emerge when network sizes are examined in comparison with the users' AmE indexes, as shown in **Table 4**. Contrary to what would have been expected, it appears that the lower the AmE index the larger the users' networks are. However, if instead of pure size the users' networks are approached via the index of strong ties, those with the highest AmE indexes clearly have more loose networks (0.71 vs. 1.07).

All in all, it appears that the countryside users differ from the rest of the data by exhibiting more sensitivity to Americanization than the groups of users. While this is a surprising result, it may be explained by the social network theory, which predicts that those with loose networks are prone to acting as agents of change: while their networks are smaller than the networks of people living in cities, the networks are weaker in nature. However, the results should be accepted with caution. Namely, the countryside data is the weakest in the sense that there are fewer idiolects to work with and fewer hits. Thus, a considerable portion of the individuals may only appear to use AmE variants in a polarized way. The reality may be that their polarized AmE index is simply

a result of them providing only one relevant hit, and more data might reveal inconsistency. Hence, more data is needed to verify this possible tendency, but these results can thus be seen as an interesting starting point for future research with greater amounts of data.

## DISCUSSION AND CONCLUSION

Our study, while preliminary, takes on understudied areas of linguistics that are related to large-scale changes in the English language continuum. We hope that this study has illustrated the potential that the study of ELF idiolects possesses to reveal in-depth patterns of change and variation, as the results are strikingly different when individuals are examined rather than the group as a whole. Our main results show that individuals exhibit considerable variation in their spelling choices and lexicogrammatical and morphological tendencies. In addition, the countryside data differs from the rest of the data: it appears that Finnish users of English from non-urban areas are perhaps most sensitive to Americanization.

With regard to our first research question, we can observe that individuals exhibit considerable variation in their respective usage of different linguistic variables. Overall, spelling tendencies appear to be mixed, whereas lexicogrammatical and morphological variation clearly tends toward AmE. Our observations were of patterns similar to those found by Gonçalves et al. (2018), but our study provides more detail on how both idiolects and individual users' usage of various spelling features varies. When analyzed as a large group, the users appear to mix AmE and BrE variants and the overall index of Americanization is neutral. However, when inspected at the more sophisticated level of idiolects and individual features, it seems that users are in fact rather polarised in their choices between spelling variants. They may also be consistent in their usage of individual spelling variables (such as *-or* vs. *-our*) but some individuals may use the AmE variant in the one spelling category and the BrE variant in another. Thus, our findings complement the macro-level observations of Gonçalves et al. (2018). By zooming in to the level of individual speakers and individual linguistic variables, we can

distinguish the role played by individual variation in the overall mixed usage of AmE and BrE variants.

With respect to our second research question, the observations show that users from the countryside may be more sensitive to change, and this may be explained by the nature of the users' social networks. By the very nature of their audience, Twitter users always communicate in a way that enables a very loose network or the general public to see their tweets. In addition, with regard to the size of the users' networks consisting of friends and followers, the users can be said to operate in loosely-knit networks. The larger the municipality the users tweet from, the larger their networks are; however, users in the countryside have the smallest number of strong ties. According to the main hypotheses of the social network theory, which suggest that diffusion takes place more readily in loose social networks, users from the largest cities were expected to be the most prone to sensitivity. However, as the proportion of strong ties is smaller in the countryside-dwellers' online social networks, countryside users can be positioned as more likely to exhibit sensitivity related to the principles of the social network theory. Overall, our findings do not confirm the hypothesis presented in the social network theory but nor do they necessarily challenge it, as the results concerning regional differences in lexico-grammatical and morphological variation are not statistically significant. Instead, it appears that users are sensitive to ongoing grammatical change in roughly equal proportions in every region.

When both spelling and grammar are included in the analysis, statistically relevant differences do, however, emerge. Thus, our findings offer empirical support for the idea that Finnish individuals' repertoires may in fact comply with their observed preferences, particularly at the level of what may be conscious decision-making. However, as noted by Baker (2017: 52), another factor that may contribute to spelling choices is computer software: for example, Microsoft Word does not make it easy for its users to deviate from the spelling conventions of the set language variety. It is unlikely that text editors are used when writing tweets, but smartphones, for example, may also favour a certain variety feature, which is something that must be taken into consideration when drawing conclusions about the results of this study. In addition, in the case of *-ize* | *-ise*, the occasional preference for *-ize* may sometimes be influenced by the preference and influence of a prestigious United Kingdom publisher, such as OUP.

All in all, our results reinforce the observation that AmE is in many respects ahead of other Englishes in the case of grammatical change and perhaps even leads these changes (cf. Leech et al., 2009; Baker 2017), and Mair's (2013) conceptualisation of Englishes can be backed up by more empirical evidence. In fact, some ELF speakers appear to be even more sensitive to change than native speakers, as lexico-grammatical and morphological choices adhere to the American forms to a strikingly high degree. Grammar may also be the layer that is more prone to unconscious choices and thus reflects language change and sensitivity more thoroughly than spelling, which users may be aware of and with regard to which they may make conscious stylistic choices. However, one should also note that present-day usage in Leech et al. (2009) is represented with corpora from 1991/2, which is almost equally distant from the actual

time at which the data used in this study was written as its point of comparison, ergo corpora from the 1960s. One might expect that the pattern of change observed by Leech et al. has intensified during the 2000s. In addition, due to the absence of large amounts of data from individual users, the results of this study are in need of verification.

By making use of one of the first non-native English data sources that enable the inspection of intranational regional variation, we offer new and unique perspectives for the study of ongoing change in English. Digital microblogging data is increasingly used to complement what traditional methods of linguistic enquiry have previously brought to the field. For example, Huang et al. (2016) study American regional dialects in Twitter, and by using solid empirical evidence, their study both confirms and enriches previous understandings of US dialects. At the same time, their approach challenges the traditional ideal informant, i.e., the non-mobile, old, rural male (NORM), since Twitter data is known to over-represent the younger generations (e.g., Cramton et al., 2013; Longley et al., 2015). Utilizing data that have been produced by in many respects the polar opposite of the NORM is particularly fitting for our study, which attempts to trace ongoing change in English among a population of non-native social media users.

Regarding future research, the present study offers many potential paths for future inquiry. Firstly, the themes of the study could be approached via larger and even more sophisticated sets of data. In addition, we hope that this study could function as an inspiration for the study of other linguistic mechanisms of non-native English that were not explored in this study. This area of study would benefit from the examination of a more comprehensive set of linguistic variables by including, for example, lexical items. In addition, using up-to-date reference corpora from the native varieties would elevate the empirical reliability of future endeavors in the field. This kind of quantitative approach to idiolects is unique. In addition, the study leaves room for more qualitatively oriented viewpoints. The study of individual variation would benefit from an analysis that would include the semantic and syntactic factors that may govern the users' choices between different variants. In addition, the users' social networks could be inspected in a more fine-tuned way, such as focusing on the level of interaction between certain individuals and analyzing variation and diffusion at the level of entire networks by inspecting online social networks of varying size and type. The most fruitful path of future inquiry would naturally be one where the macro and the micro level complement one another.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://cs.uef.fi/nts/>

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## REFERENCES

- Algeo, J. (2006). *British or American English? A Handbook of Word and Grammar Patterns*. Cambridge: Cambridge University Press.
- Awonusi, V. O. (1994). The Americanization of Nigerian English. *World Engl.* 13 (1), 75–82. doi:10.1111/j.1467-971x.1994.tb00284.x
- Baker, P. (2017). *American and British English: Divided by a Common Language?* Cambridge: Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E. (1999). *The Longman Grammar of Spoken and Written English*. London: Longman.
- Coats, S. (2019). Language Choice and Gender in a Nordic Social media Corpus. *Nord. J. Linguist.* 42 (1), 31–55. doi:10.1017/s0332586519000039
- Collins, P., Borlongan, A. M., Lim, J.-H., and Yao, X. (2014a). “The Subjunctive Mood in Philippine English,” in *Contact, Variation, and Change in the History of English*. Editors S. Pfenninger, O. Timofeeva, A.-C. Gardner, A. Honkapohja, M. Hundt, and D. Schreier (Amsterdam: Benjamins), 259–280. doi:10.1075/scls.159.13col
- Collins, P., Borlongan, A. M., and Yao, X. (2014b). Modality in Philippine English. *J. English Linguist.* 42, 68–88. doi:10.1177/0075424213511462
- Collins, P. (2009). Modals and Quasi-Modals in World Englishes. *World Engl.* 28 (3), 281–292. doi:10.1111/j.1467-971x.2009.01593.x
- Collins, P. (2015). “Recent Diachronic Change in the Progressive in Philippine English,” in *Grammatical Change in English World-wide*. Editor P. Collins (Amsterdam: John Benjamins), 271–296. doi:10.1075/scl.67.12col
- Cramton, J., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., and Wilson, M. (2013). Beyond the Geotag: Situating ‘big Data’ and Leveraging the Potential of the Geoweb. *Cartogr. Geogr. Inf. Sci.* 40 (2), 130–139. doi:10.1080/15230406.2013.777137
- Depraetere, I. (2003). On Verbal concord with Collective Nouns in British English. *English Lang. Linguist.* 7 (1), 85–127. doi:10.1017/s1360674303211047
- Derczynski, L., RitterClarke, A. S., and Bontcheva, K. (2013). “Twitter Part-Of-Speech Tagging for All: Overcoming Sparse and Noisy Data,” in Proc. Int. Conf. Recent Adv. Nat. Lang. Process ACL, Hissar, Bulgaria, September 7–13, 2013.
- Eisenstein, J., O’Connor, B., Smith, N. A., and Xing, E. P. (2014). Diffusion of Lexical Change in Social Media. *PLoS ONE* 9 (11), e113114. doi:10.1371/journal.pone.0113114
- Fuchs, R. (2017). The Americanisation of Philippine English: Recent Diachronic Change in Spelling and Lexis. *Philipp. ESL J.* 19, 60–83.
- Gonçalves, B., Loureiro-Porto, L., Ramasco, J. J., and Sánchez, D. (2018). Mapping the Americanization of English in Space and Time. *PLoS ONE* 13 (5), e0197741. doi:10.1371/journal.pone.0197741
- Grieve, J. (2016). *Regional Variation in Written American English*. Cambridge: Cambridge University Press.
- Hänsel, E. C., and Deuber, D. (2013). Globalization, Postcolonial Englishes, and the English Language Press in Kenya, Singapore, and Trinidad and Tobago. *World Engl.* 32 (3), 338–357. doi:10.1111/weng.12035
- Huang, Y., Guo, D., Kasakoff, A., and Grieve, J. (2016). Understanding US regional linguistic variation with Twitter data analysis. *Comput. Environ.* 59, 244–255. doi:10.1016/j.compenvurbsys.2015.12.003
- Hundt, M. (2009). “Colonial Lag, Colonial Innovation or Simply Language Change,” in *One Language, Two Grammars? Differences between British and American English*. Editors G. Rohdenburg and J. Schlüter (Cambridge: Cambridge University Press), 13–37.
- Ishikawa, S. (2011). Duality in the Spelling of English Verb Suffixes -ize and -ise: A Corpus-Based Study. *IPEDR* 26, 390–396.
- Labov, V. (1966). *The Social Stratification of English in New York City*. New York: Harcourt. Brace.
- Laitinen, M., and Lundberg, J. (2020). “ELF, Language Change, and Social Networks,” in *Language Change: The Impact of English as a Lingua Franca*. Editors A. Mauranen and S. Vetchinnikova (Cambridge: Cambridge University Press), 179–204. doi:10.1017/9781108675000.011
- Laitinen, M., Lundberg, J., Levin, M., and Martins, R. (2018). “The Nordic Tweet Stream: A Dynamic Real-Time Monitor Corpus of Big and Rich Language Data,” in Proc. DHN 2018, Helsinki, Finland, 7 March 2018 through 9 March 2018, 349–362.
- Laitinen, M. (2016). “Ongoing Changes in English Modals: On the Developments in ELF,” in *New Approaches in English Linguistics: Building Bridges*. Editors O. Timofeeva, S. Chevalier, A.-C. Gardner, and A. Honkapohja (Amsterdam: John Benjamins), 175–196.
- Lamanna, F., Lenormand, M., Salas-Olmedo, M. H., Romanillos, G., Gonçalves, B., and Ramasco, J. J. (2018). Immigrant Community Integration in World Cities. *PLoS One* 13 (3), e0191612. doi:10.1371/journal.pone.0191612
- Leech, G., Hundt, M., Mair, C., and Smith, N. (2009). *Change in Contemporary English: A Grammatical Study*. Cambridge: Cambridge University Press.
- Leetaru, K., Wang, S., Cao, G., Padmanabhan, A., and Shook, E. (2013). *Mapping the Global Twitter Heartbeat: The Geography of Twitter*, 18. First Monday. doi:10.5210/fm.v18i5.4366
- Leppänen, S., Pitkänen-Huhta, A., Nikula, T., Kytölä, S., Törmäkangas, T., Nissinen, K., et al. (2011). National survey on the English language in Finland: Uses, meanings and attitudes. *Studies in Variation, Contacts and Change in English* 5. Available at: <http://www.helsinki.fi/variang/series/volumes/05/>.
- Lippi-Green, R. L. (1989). Social Network Integration and Language Change in Progress in a Rural alpine Village. *Lang. Soc.* 18, 213–234. doi:10.1017/s0047404500013476
- Longley, P. A., Adnan, M., and Lansley, G. (2015). The Geotemporal Demographics of Twitter Usage. *Environ. Plan. A.* 47 (2), 465–484. doi:10.1068/a130122p
- Mair, C. (2003). “Gerundial Complements after Begin and Start: Grammatical and Sociolinguistic Factors, and How They Work against Each Other,” in *Determinants of Grammatical Variation in English*. Editors H. Rohdenburg and B. Mondorf (Berlin: Mouton de Gruyter), 329–345.
- Mair, C. (2006). *Twentieth-Century English: History, Variation and Standardization*. Cambridge: Cambridge University Press.
- Mair, C. (2013). The World System of Englishes: Accounting for the Transnational Importance of Mobile and Mediated Vernaculars. *Engl. World-wide* 34 (3), 253–278. doi:10.1075/eww.34.3.01mai
- Mauranen, A. (2018). “Conceptualising ELF,” in *The Handbook of English as a Lingua Franca*. Editors J. Jenkins, W. Baker, and M. Dewey (London: Routledge), 7–24.
- Milroy, J., and Milroy, L. (1978). “Belfast: Change and Variation in an Urban Vernacular,” in *Sociolinguistic Patterns in British English*. Editor P. Trudgill (London: Edward Arnold), 19–36.
- Milroy, J., and Milroy, L. (1985). Linguistic Change, Social Network and Speaker Innovation. *J. Ling.* 21, 339–384. doi:10.1017/s0022226700010306
- Milroy, L. (1987). *Language Change and Social Networks*. 2nd edition. Oxford: Blackwell.
- Murphy, L. (2018). *The Prodigal Tongue: The Love-Hate Relationship between American and British English*. New York: Penguin Books.
- G. Rohdenburg and J. Schlüter (Editors) (2009a). *One Language, Two Grammars? Differences between British and American English* (Cambridge: Cambridge University Press).
- Rohdenburg, G., and Schlüter, J. (2009b). “New Departures,” in *One Language, Two Grammars? Differences between British and American English*. Editors G. Rohdenburg and J. Schlüter (Cambridge: Cambridge University Press), 364–423.
- Schneider, E. W. (2011). “The Subjunctive in Philippine English: An Updated Assessment,” in *Studies of Philippine English: Exploring the Philippine Component of the International Corpus of English*. Editor M. Bautista (Manila: Anvil), 159–173.
- Tieken-Boon van Ostade, I. M. (2009). *An Introduction to Late Modern English*. Edinburgh: Edinburgh University Press.
- Tottie, G. (2009). “How Different Are American and British English Grammar? and How Are They Different,” in *One Language, Two Grammars? Differences between British and American English*. Editors G. Rohdenburg and J. Schlüter (Cambridge: Cambridge University Press), 341–363.
- Trudgill, P., and Hannah, J. (2008). *International English: A Guide to Varieties of Standard English*. London: Routledge.
- Vetchinnikova, S., and Hiltunen, T. (2020). “ELF and Language Change at the Individual Level,” in *Language Change: The Impact of English as a Lingua Franca*. Editors A. Mauranen and S. Vetchinnikova (Cambridge: Cambridge University Press), 205–233. doi:10.1017/9781108675000.012
- Vetchinnikova, S. (2017). “On the Relationship between the Cognitive and the Communal: a Complex Systems Perspective,” in *Changing English: Global and Local Perspectives*. Editors M. Filppula, J. Klemola, A. Mauranen, and

- S. Vetchinnikova (Berlin: Mouton de Gruyter), 277–310. doi:10.1515/9783110429657-015
- Vosberg, U. (2006). “The Great Complement Shift. Extra-semantic Factors Determining the Evolution of Sentential Complement Variants in Modern English,” in *English and American Studies in German*, 19–22.
- Weinreich, U., Labov, W., and Herzog, M. (1968). “Empirical Foundations for a Theory of Language Change,” in *Directions for Historical Linguistics*. Editors W. P. Lehmann and Y. Malkiel (Austin: University of Texas Press), 95–188.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2022 Taipale and Laitinen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

## APPENDIX 1 THE LIST OF SPELLING ITEMS (BARE FORM) USED FOR THE QUERIES.

Variable	AmE variants	BrE variants
or   -our	ardor, armor, behavior, color, colorful, endeavor, favor, flavor, harbor, honor, humor, labor, neighbor, odor, rumor, savor, tumor, vapor, vigor	ardour, armour, behaviour, colour, colourful, endeavour, favour, flavour, harbour, honour, humour, labour, neighbour, odour, rumour, savour, tumour, vapour, vigour
-er   -re	caliber, center, fiber, luster, maneuver, manoeuver, meager, meter, somber, specter, theater	calibre, centre, fibre, lustre, manoeuvre, manœuvre, manoeuvre, meagre, metre, sombre, spectre, theatre
Geminate consonants	canceled, canceling, cruelest, fulfillment, labeled, labeling, leveled, leveling, marvelous, modeled, modeling, traveled, traveling, traveler, panelist, skillful worshiped, worshipping	cancelled, cancelling, cruellest, fulfilment, labelled, labelling, levelled, levelling, marvellous, modelled, modelling, panellist, skilful, travelled, travelling, traveller, worshipped, worshipping
-ense   -ence in nouns	defense, license, offense	defence, licence, offence
-yze   -yse	analyze, analyzing, catalyze, catalyzing, dialyze, electrolyze, hydrolyze, paralyze, paralyzing	analyse, analysing, catalyse, catalysing, dialyse, electrolyse, hydrolyse, paralyse, paralysing
-ize   -ise	apologize, apologizing, criticize, criticizing, organize, organizing, realize, realizing, realization, recognize, recognizing	apologise, apologising, criticise, criticising, organise, organising, realise, realising, realisation, recognise, recognising
Individual word spellings	gray, math, toward	grey, maths, towards