Contrastive Focus Capitalization: Nonstandard Usages of Capital Letters in Web-based English and their Capital-I Implications

Josh Linden
Georgetown University
jsl169@georgetown.edu

Like most languages using the Roman alphabet, English has an upper- and lowercase form of each letter and several interconnected patterns governing their use. This paper explores the ways those patterns are changing in the age of the Internet and proposes a novel usage of sentence-internal capitalization called Contrastive Focus Capitalization (CFC). CFC mainly targets nouns and conveys a number of meanings related to legitimacy and givenness as well as drawing attention to the most prototypical or salient meaning as the intended one. This phenomenon is explored via analysis of a 2.2 million-word sample of GloWbE, the Corpus of Global Web-based English, consisting mainly of blog posts made by English speakers around the world. The related but distinct practice of capitalizing common nouns as if they were proper nouns is also discussed. It is found that the latter is more common, but both are used especially in American English. Observations are made about the scope and connotations of these forms of nonstandard capitalization and parallels are drawn to other, less orthography-dependent structures with similar meanings. These findings are then considered in the broader context of Internet-based language with the goal of examining the relationship between spoken language and written language in the Digital Age.

1. Introduction

It has long been typical that changes in written language are predicated on related changes in spoken language. Novel forms tend to appear first aloud and then, one they become established, make their way into writing. The current paper argues that this pattern is not due entirely to the nature of speaking and writing but also in part to the fact that novel forms tend to develop within informal, social communication. This type of communication has been almost exclusively spoken in the past, but we now see a large amount of it happening digitally via computer-mediated communication (CMC). This seems to be leading to innovations originally and organically emerging within written forms of the world's languages in a way that, if not entirely new, is at least more common and widespread

than ever before. If we as linguists wish to understand this phenomenon, we must look at novel forms that emerge organically in written language.

Many such forms have been the focus of extensive academic study. Perhaps the best example is the phenomenon of emoji. Their creation was intentional, as they were designed by a team of developers and incorporated into Unicode to be made available for use online. However, the conventions surrounding their usage and meaning are an excellent example of organic development taking place in written language. See the next section and Na'aman (2017) for further discussion, but for now it suffices to point to them as an example of the kind of development posited above. The current paper explores use of capitalization online including the frequent blurring of the line between proper and common nouns, and proposes another example of an organic, novel usage of written language with its own features and meaning independent of any spoken analog, called Contrastive Focus Capitalization (CFC). CFC is defined as the capitalization of the first letter of a word other than a proper noun sentence-internally to convey one of several closely related meanings: legitimacy, salient nature and/or the intent for an established or understood meaning. Consider for instance example (27), presented below, which describes discontinuation of ferry service in parts of the Middle East as "another casualty of 'The Situation,'" which can be understood to have a different meaning from something that would be described as "the situation" without capitalization. This phenomenon demonstrates that written language, when used as extensively and informally as is today, is subject to evolution not directly predicated on shifts in spoken language. CFC is explored via analysis of a 2.2 million word sample of informal, Web-based written English from English-speaking countries around the world. A corpus of 55,835 English-language SMS messages is also considered. It is found that CFC is most common on nouns and multiword noun phrases, but likely not limited to them entirely based on preliminary evidence. The precise meaning is explored in various contexts and comparisons are made to spoken ways of conveying similar meanings. Critically, CFC does not appear to represent any particular element of spoken language. It is even sometimes explicitly evoked aloud. An example of this comes from NPR's radio show *Fresh Air*: when describing a fictional character becoming an art student under Nazi rule, host Terry Gross says, "the only thing you're allowed to paint is, like, uplifting things in service of, like, The People with capital letters" (Gross 2019). This kind of spoken reference to the written form implies that it has its own nuance divorced entirely from any peculiarity of spoken language. Still, methods to encode this meaning are certainly not absent from spoken English.

The current paper draws on the framework of Ghomeshi et. al. (2004), as

that paper's Contrastive Focus Reduplication (CFR) conveys meanings very similar to CFC (and in fact was the inspiration for the name of and choice to capitalize Contrastive Focus Capitalization). See section 5 for a discussion of the similarities and differences in the scope and meaning of CFC and CFR. While the exploration of the semantics of CFC presented here follows the model of Ghomeshi, this paper differs from that one in ways to do with the written nature of CFC and the broader implications of that written nature. It aims to define, explore and discuss the meaning of nonstandard capitalization in written discourse, in particular the kind of nonstandard capitalization called Contrastive Focus Capitalization. It then compares its usage with some similar forms both spoken and written and looks to draw some basic conclusions about how written language evolves in online spaces.

2. Background

Ghomeshi et. al. (2004) presents the concept of Contrastive Focus Reduplication. Examples of the phenomenon given in that paper include "I'll make the tuna salad, and you make the SALAD–salad," and "Oh, we're not LIVING-TOGETHER–living-together." The authors "characterize this effect as denoting the prototypical instance of the reduplicated lexical expression." This kind of reduplication in colloquial English narrows the possible interpretations of the reduplicated material to the most "contextually salient readings." While the form is quite different, the meaning is very similar to that of CFC. Much of that paper aims to define the scope of grammatical uses of CFR. It can target words of many grammatical categories, but only contentful words. Nouns, verbs and adjectives are most commonly targeted, but reduplication of prepositions is possible when they are contentful. Idioms behave differently than syntactically identical non-idiomatic phrases, and it's nontrivial to predict how much material is reduplicated as there is variation in which elements of an idiomatic phrase are always, never, or optionally duplicated. In some cases the reduplication extends to an idiomatic direct object but in others it does not. The discussion of the meaning of CFR (as opposed to the scope) will be most applicable to the current paper, but CFC and CFR can be compared in both those domains. Ghomeshi concludes that CFR draws attention to other possible interpretations of the reduplicated material in order to specify that they are not the intended reading. Instead, CFR encodes the most situationally salient, obvious or prototypical interpretation. The forms of CFR and CFC are very distinct, and although comparisons of their scope proved informative it is the overlap in meaning that makes Ghomeshi's work so relevant here.

As for the form of CFC, it shares much more in common with some other well-attested phenomena found in CMC. One such occurrence, alluded to above, is that of emoji. Emoji are not themselves written language in the conventional sense, in that they often do not represent specific sounds, words or ideas from spoken language. They do so on occasion (Na'aman 2017), but they more commonly supplement the existing symbols with suprasegmental information like intonation, facial expression and gesture. It may be possible to attribute their swift rise in popularity to their ability to fill this very specific need: formal written language on the whole is less in need of ways to mark tone, but social communication must include sarcasm, emphasis, tonal variation, etc. in order to be fully able to play the role that spoken language does within a social circle. Na'aman (2017) looks at the use of emoji on Twitter, dividing examples into two broad categories. Use of emoji either directly represents spoken language "as content and function words," or encodes paralinguistic information "as multimodal affective markers." Examples of the former include "The 🔑 to success is 🍕," where the key and pizza emoji each take the place of that content word, and "I 🍩 like you," where the doughnut emoji stands in for the function words "do not." The latter is exemplified by "Let my work disrespect me one more time… 🙃," where the upside-down face emoji indicates an attitude but no specific English word or words. According to that paper, it was easy to find criteria to distinguish between those two kinds of uses but harder to find more specific meaningful classifications. Even though emoji are a well-studied phenomenon relative to how new they are, some of the variations in their applications are less well understood. Even so, they serve as a useful point of comparison when considering other Web-based developments in written language such as CFC.

Another feature of Web-based writing that has received some attention in the literature is the use of nonstandard spellings. Eisenstein (2015) argues that such forms mimic speech in some unexpected ways. For instance, coda deletion in social media posts tends to follow spoken patterns within a given community, exhibiting the same dependencies on part of speech and phonological context. English speakers are likely to write "sayin" or "jus" without a final g or t on Twitter according to patterns that closely mirror tendencies to do the same aloud. It is important to remember that, despite the possible counterexamples presented in this paper, writing following speech rather than vice versa is still the more common occurrence. Eisenstein's finding that there is a strong relationship between CMC and spoken language in the conventional direction ought to be encouraging to anyone looking to make comparisons between CMC and

more traditional forms of writing as this analysis does. Similarly, Kalman (2014) finds that letter repetition in emails usually emulates phoneme lengthening in speech.

Rodríguez-Hidalgo (2017) discusses some of the other ways paralinguistic information can be conveyed online, especially when discussing emotions and offering emotional support. That study examined some of the psychological implications of text-based paralinguistic cues in blog posts, including for example typing the word *hug* to stand in for offering a hug as one would in person. The study found that patterns of use and mirroring, meaning adopting the paralinguistic cues of another writer when addressing that writer, matched observed face-to-face human interactions in some robust and unexpected ways. This suggests that such paralinguistic cues have at least some degree of psychological reality to them, again lending credibility to the idea that other written forms like capitalization should be given similar consideration.

Another such domain of CMC features is punctuation, which is often observed to be used differently than it is in other forms of written communication. Gunraj (2016) discusses the meaning of periods in text messages: in a controlled experiment, texts ending in periods were judged to be less sincere than otherwise identical texts without periods. This effect was specific to text messages; when the same messages were presented as handwritten notes the effect disappeared. For example, when presented as part of a text message conversation, the response "Sure." was judged to be less sincere than the response "Sure" without a period. This difference was not found when identical conversations were presented in handwritten form. This is evidence that what is standard practice in formal written English has developed an independent meaning in this kind of informal context. Danet (2014) also discusses the use of punctuation, as well as many other features of CMC in English. According to Danet, "Writing a message all in capitals is generally understood as shouting, though marking a single word in all capitals just emphasizes it." She also observes that abbreviations like *brb* (for "be right back") and *ttyl* (for "talk to you later") are less common than is typically perceived, and that "this generation largely transferred writing habits learned in school-based literacy to IM." Danet goes so far as to claim that not all CMC is even viewed as writing by those participating in it. She continues: "Because online textual communication is dynamic, interactive, and ephemeral, we often experience it as speech-like. Yet, compared to speech, it is also attenuated, since the non-verbal and paralinguistic cues that accompany speech are missing." As an element of CMC, CFC shares various qualities with each of the forms discussed here and one goal of the current paper is

to place CFC within the framework of other, better studied features of online communication.

Having looked at an overview of the literature on computer-mediated communication in English, we turn now to the other key element that comprises CFC: capitalization. Perhaps the most important background in this realm is Lee's (2016) account of the history of capitalization in English:

> What we can glean from historical accounts is that around the early 11th century in England, capitalization was used frequently 'but not always' for initial words of a sentence as it is today (Irvine, 2006, p. 52). By the beginning of the 18th century, authors would often use capitalization mid-sentence to stress particular words, and some printers would even capitalize all nouns, which may appear to modern readers as giving the text 'an unduly formal character' (Corrie, 2006, p. 243). Evidence suggests that during this time, the rules of capitalization were determined on an ad hoc basis, subject to the vagaries of local house guides or the predilections of individual writers and editors. It is apparent that by the early 19th century, however, the conventions of capitalization had more or less stabilized to resemble those we tend to encounter today. (Lee 2016)

This characterization emphasizes the recent and haphazard nature of modern capitalization standards in English writing and supports Lee's conclusion that "capitalization is by no means a timeless or even natural feature of English." The paper offers critical historical and colonial context for any discussion of modern capitalization: deviation from capitalization standards is not (and has never been) uncommon, and the ability to deviate from them for intentional effect depends on cultural and linguistic privilege. Some users are able to challenge those conventions, but speakers of less mainstream varieties of English and speakers for whom English is not their first language "are chronically relegated to a continual process of having to legitimize their 'errors' as conscious, deliberate, effective, or appropriate." Lee illustrates this idea via a political poster created by a native speaker of Korean, which lacks capitalization entirely. The poster includes several unusual capitalization choices which can be read as deliberate or not. They are widely seen as erroneous, and Lee proposes that it would be impossible for that person to create such a poster with access to the full range of capitalization techniques available to a native speaker. This context is especially relevant when proposing a novel form of intentional deviation from established convention. Not everyone has equal access to such forms, and this may help to explain CFC's relative popularity in the United States compared to the other

countries represented in the available data (which is discussed further in section 6).

As a final piece of context, Maia (2018) conducted a series of controlled eye-tracking experiments looking for effects of text-based manipulations on readers. Maia found evidence that typographic emphasis such as italics and capitalization affects readers' eye movement. Previous research had conceptualized the effects of emphasis as acting on memory and retention: an emphasized word or phrase would be parsed as normal but would be more likely to be recalled later. Instead, Maia takes the approach that it conveys additional information and his experiments offer strong support for the ability of text-based information channels like italics and capitalization to encode semantic material. Maia's paper is not about the use of capitalization in discourse, but rather uses capitalization as one of several manipulated variables. His results are promising for any proponent of CFC, as they demonstrate that it is possible for such a form to encode novel information as opposed to merely focusing attention and increasing recall of the information otherwise encoded. Maia makes "A data-driven case […] that typographic emphasis is a visible carrier of content that serves a linguistic function in the text." Manipulations of these variables on noun phrases produced the strongest results, which closely reflects the findings presented below. Maia also introduces several other concepts and terms which are useful in the context of this paper, largely surrounding the ideas of focus and emphasis. Languages have many tools and structures to direct what Maia calls "discourse focus;" consider for instance the sentence "It was Darcy who wrote Jane a letter" (Maia 2018). Each element of a sentence is not attended to equally, and speakers can use forms like these (as opposed to the more typical "Darcy wrote Jane a letter") to direct focus. These forms interact with the idea of givenness, which is roughly the extent to which a referent is understood to have been mentioned or considered before. The more given a referent, the less informatively it will be referred to. Maia's ideas and results serve as important background, offering some psychological impetus for the real-world variation in capitalization usage presented here.

3. Methods

This analysis examined two corpora of computer-mediated English searching for examples of nonstandard capitalizations, collected them and looked for commonalities and patterns. The collected data was not a statistically representative sample. These methods were chosen because this is intended as an exploration of uses of nonstandard capitalization rather than to make claims about distributions or average cases.

Limitations are discussed in greater detail in section 6. The most relevant dataset used was a 1% sample of GloWbE, the Corpus of Global Web-based English. That 1% sample consists of approximately 2.2 million words of text posted online by speakers in 20 English-speaking countries around the world, about 60% of which are in the form of blog posts. The data contained many examples of nonstandard capitalization, a selection of which are presented in the next section. For additional context, the NUS SMS Corpus also proved useful. This corpus of 55,835 English-language SMS messages was even less formal in register than the GloWbE texts. The final analysis includes only examples from GloWbE, but the NUS SMS Corpus provided important contrast when considering informal, social communication (as opposed to text intended for wider publication in some form). Capitalization in the text messages was fairly erratic and on the whole seemed less meaningful than in GloWbE, where it was less likely to have been a product of the text input method. Where capital letters were used they tended to be applied to full words and phrases, apparently to represent some of the meanings discussed above and by Danet (2014). Nonetheless, the opportunity to compare the two corpora informed decisions about what did and didn't carry meaning in GloWbE.

Data from the two corpora were downloaded into two separate files and searched for examples of nonstandard capitalization. To aid in this process, sentence-internal capitalizations were highlighted using a find-and-replace algorithm searching for spaces followed by capital letters. Before running the automatic highlighting program, several replacements intended to eliminate noise were made. Spaces were deleted where they appeared after periods, brackets and several other kinds of punctuation to avoid highlighting sentence- or section-initial capitals. Finally, each capital I surrounded by spaces was replaced with a distinctive but meaningless string ("I") to preserve the data but avoid highlighting each instance of the first-person singular subject pronoun. This process highlighted 159,343 capitalizations. The result was then scanned by the researcher for highlighted letters which were not names, places, titles or initialisms and useful examples of the phenomena discussed in this paper were copied into another file. A total of 82 examples were collected, mostly from the sections originating in Australia and the United States. This was a subjective process, and the goal of exploring the data rather than representing it statistically was foremost. Not every instance of nonstandard capitalization was discovered, and more attention was paid to sections of the data with a higher density of relevant capital letters. Finding every example was a less important goal than finding enough examples to observe patterns in the use of capitalization. In this large a dataset and without more sophisticated text-searching software, the most feasible way to gain a good understanding of the proposed phenomena

was to scan rather than read. While it is impossible to know with certainty how representative of the full data set the collected examples are, there was sufficient consistency in the types of forms to create useful categories. The collected data was then considered in an attempt to draw conclusions about the usage of capitalization in written discourse online.

4. Findings and Results

Findings are presented in two subsections which are brought together for one unified discussion in section 5. Section 4.1 examines the very common practice of blurring the distinction between proper and common nouns in the data. This pattern does have a contrastive component, but it stands apart from CFC as defined above and is therefore considered separately. Section 4.2 considers only the rarer but more contentful examples of proper CFC as defined above, where the capitalized word or words are relatively clearly not proper nouns. Lee (2016) deals heavily with the concept of intentionality and specifically discusses some of the pitfalls present in assuming speakers' intent in academic analyses; that issue will not be thoroughly explored here, but this analysis takes care not to rely too heavily on assumptions about intent. The distinction being made is more based on apparent meaning, and it's not a rigid one. The inclusion of a specific example (or even of a structure or pattern more broadly) in one category doesn't necessarily preclude its belonging to the other as well. In other words, this is not a complementary distribution by any strict definition. Rather, it is intended to separate out the primary focus of this paper for discussion while also affording due consideration of a related and much more common phenomenon.

4.1. The Proper-Common Distinction in Nouns

In standard written English only proper nouns are to be capitalized sentence-internally. The question addressed in this subsection is, what is considered a proper noun in online discourse? Lee (2016) discusses the arbitrary and recent nature of the standardization of proper nouns, and this claim is heavily supported by the current data. There is some agreement among speakers about classes of nouns that are or are not proper. For instance, place names, people's names, months and days of the week are widely considered proper nouns and are usually capitalized. It is interesting to note that this also extends to screen names and pseudonyms used online, which are usually either capitalized or rendered in CamelCase (in which each word is capitalized but no spaces are included). Very generally nouns that refer to a specific person, place or thing are

considered proper and accordingly capitalized. Common nouns are less specific, refer to a category of people, places or things and are not capitalized. However, this distinction is not precise and is often followed only roughly even in more formal contexts; there is disagreement over what to capitalize at every level of formality (Lee 2016). These disagreements are more prevalent in less formal contexts such as the texts in GloWbE, and even more so in the NUS SMS Corpus. This results in speakers making extremely contextual and individual decisions about what should be considered a proper noun. These decisions appear anchored in the ideas speakers have about the properties that proper nouns tend to share: they are of an established or institutional nature and/or refer to a specific referent. Most nonstandard sentence-internal capitalizations of the first letter of one or more words were on this border between proper and common nouns. Some examples follow:

(1) The Government has proposed changes to the level of support it will provide
(2) When pondering this on Twitter, some Tweeps kindly raised a second objection
(3) […] a transformation which will keep us competitive with the big, smart First World economies

In (1), "Government" does not refer to some unspecified government nor to the idea of governments in general, but rather to the government of Australia specifically. Similarly, (2) and (3) illustrate the specificity of referent common to many examples of nonstandard capitalization found in the data. In (2), we see the word "Tweeps" (usually meaning a person's followers on Twitter) capitalized. Note the use of "some," often considered an indefinite article in English; the Tweeps raising an objection are an indefinite subset of a specific and rigidly defined group of people following a given user on Twitter, which is arguably encoded in the capitalization pattern of "some Tweeps." There is an element of parallelism here, and it is not clear if that word is seen as a proper noun in itself or simply as a portmanteau of "Twitter" and "peeps," maintaining the original capital T. In either case, this highlights the less-well-defined distinction between proper and common nouns in informal CMC as opposed to more formal writing. The most common thread of nonstandard capitalization of nouns seems to be an assertion that the noun in question has more in common with a proper noun than it otherwise might.

The scope of this form of capitalization is limited to nouns and noun phrases; other syntactic categories cannot be compared to or considered to be proper nouns in this way. Typically the entire noun phrase is capitalized. This is especially true of compound nouns:

(4) I don't rule out the possibility of legislating a Carbon Pollution Reduction Scheme, a market-based mechanism […]

(5) We have one client that sells Breast Cancer Annual Diaries through her newsletter

Almost without exception, the decision to capitalize a compound noun extends to each word in the compound regardless of whether those words are nouns individually. The same is sometimes – but not always – true of noun phrases formed by other processes like adjunction:

(6) For a free consultation or advice on building an Effective Website, please contact us

(7) Her ineffectual Government has hocked the nation to the tune of $100 billion

As seen here, when the aspect of the noun phrase being legitimized is the adjective, capitalization can extend to it. In other cases, it may not.

As mentioned above, capitalizing a common noun sentence-internally in this way seems to have several related meanings all stemming from an apparent assertion that in this instance or usage, the noun should be read as a proper noun with all the many implications that has. Similar to a definite article, this implies that the noun used represents *the* understood, established, typical one of its class. It is common for this usage to convey a sense of legitimacy, as illustrated here:

(8) […] alternative therapies such as Aromatherapy have also proved beneficial for some

(9) We have definitely become a nation of Cooks. Did you know that around 52% of Australians eat their main meal at home seven days a week?

Furthermore, this form of capitalization can imply an association with various kinds of societal institutions such as the following.
Government and politics:

(10) […] the Carbon Tax will solve Global Warming by itself
(11) If your divorce application is successful, the Court will grant a divorce order
(12) They are tied up in the Party
(13) We have the most experienced Cabinet since 1975

Clubs and special-interest groups:

(14) […] planning a trip somewhere for some surf and Yoga
(15) When Members purchase a Welcome Pack in cinema […]

The military:

(16) […] where members of the Afghan National Security Forces kill Coalition soldiers
(17) However, Improvised Explosive Devices or IEDs continue to be the most deadly risk in Afghanistan
(18) My foray into Aid and Development in Post-Tsunami Sri Lanka was an emotional, moral reaction to the suffering and devastation around me

Human demographic groups:

(19) Look at Black Africa
(20) Queer Hero strategy: successful people from the various academic disciplines who were homosexuals will have their sexuality identified in class/curriculum materials
(21) I know many Black Americans who have gone on a Sojourn to Africa
(22) […] the vision I have to achieve equality for Deaf people
(23) If people think there is even a trace of a chance of this leading to criticism of Gay people, it will die.

It is also common for this usage to reflect the existence of a previously established, specific meaning (even if that meaning does not exemplify the associations given above):

(24) As part of this rewrite, the Avoid Common Pitfalls section was added

The unifying factor of these trends is that they tend to be elements frequently shared by nouns that are more widely agreed to be proper and that are therefore more universally capitalized. They often have a specific referent, an established meaning and a lack of ambiguity. It is useful to compare with words that can function as proper nouns or common nouns in standard English as case-studies in how their meanings change. For example, Earth is a planet while earth is a kind of dirt. Mother is the speaker's mother, whereas mother is a familial role. Delta is an airline, but delta is a geographical feature (and either is a Greek letter). The usage attested in this subsection, which makes up the majority of nonstandard capitalizations in the data, asserts that the noun being capitalized not only

has a second meaning akin to Earth, Mother and Delta but that that second (more specific, understood or established) meaning is the one intended.

Again there certainly is a contrastive element of this pattern, but it is not Contrastive Focus Capitalization in the way envisioned and defined in section 1. Nonetheless, any data-driven analysis of the current sample would be incomplete without mention of such a frequent occurrence. Ultimately this is not a paper about the definition of a proper noun, but it is worth noting that informal, web-based discourse does not appear to validate the commonly taught (or, for that matter, any one singular) definition. An assertion that a given noun is proper includes a wide range of connotations based on the criteria developed over the course of the Late Middle English and Early Modern English periods and widely circulated and taught since then. This usage is not necessarily novel, and has much in common with inconsistent capitalizations going back well before the Digital Age. Words like g*overnment* and *party* have existed on this proper-common border for some time, but the frequency of usages like this in the GloWbE texts suggests that the practice is becoming even more prominent in Web-based English than it has been historically.

4.2. Contrastive Focus Capitalization

Let us now turn to the rarer but very informative cases of sentence-internal capitalization which does not appear to simply assert that the words being capitalized are in some sense proper nouns. Again the selection criteria for this type of capitalization were not rigidly defined, but the process was heavily influenced by Ghomeshi (2004). The researcher looked for examples that appeared to be playing with the construct of capitalization, aware of the fact that the capitalization was marked and using that for an effect. This included most of the examples of sentence-internal capitalization that appeared not to be due to typographical errors and not to be a proper noun by way of usage. Even acknowledging the potential for sampling bias, it was striking how similar the meanings of most of the capitalizations in this category were to Ghomeshi's findings on the meanings of CFR. Ghomeshi characterizes that form's meaning in a way that applies almost equally well to CFC: "The semantic effect of this construction is to focus the denotation of the reduplicated element on a more sharply delimited, more specialized, range." This subsection presents some selected examples and discusses them in isolation from the other usages of capitalization discovered in the data.

Consider first these two instances of capitalization by the same author:

(25)   […] we folk in the Fat Acceptance movement were right […]
(26)   When I informed BuddingStarlet that there was no longer an Obesity Epidemic, she said […]

These usages are excellent examples of the similarities in meaning between CFC and CFR and of the differences between CFC and other forms of nonstandard capitalization. Either might be considered an assertion that the noun being capitalized is a proper noun, but the meaning and syntax don't fully support that conclusion. Notice the lowercase rendering of the word "movement" in (25). The entire compound noun, "Fat Acceptance movement," is not capitalized. An application of the findings presented in the previous subsection would strongly predict a capital M. There are no examples of compound nouns with some content words capitalized and others not in the previous section's data, and this mixed example is not easily explained by considering any part of "Fat Acceptance movement" to be a proper noun in this instance. Thus an analysis using the framework of that subsection fails here; the Fat Acceptance *movement* is the established, understood and organizational element of this compound, yet each word is not capitalized. Considering the compound as nested also fails, as the component compound "Fat Acceptance" is not a specific referent. (Consider that "the Fat Acceptance" is less grammatical than "some Fat Acceptance" when used in isolation from "movement.") This contrast with lowercase "movement" offers a sense of the motivation for distinguishing this usage from that presented in the previous subsection: these two examples have a common syntax and meaning largely absent from subsection 4.1. They draw attention to the givenness of the capitalized material while legitimizing it and excluding alternative, less salient interpretations. According to this writer there may or may not be an obesity epidemic somewhere in the world, but the familiar, widely-discussed Obesity Epidemic is no more. Let us compare three new sentences based on (26):

i. I informed her that there was no longer an epidemic.
ii. I informed her that there was no longer an Epidemic.
iii.I informed her that there was no longer an epidemic-epidemic.

Each of these is modified from the original data (presented faithfully above) to illustrate a contrast while conforming to Ghomeshi's prosaic claims about the scope of CFR; following Ghomeshi's conclusions, reduplication of "obesity epidemic" would be of questionable length and would likely be considered ungrammatical by at least some speakers. ii. is the analog of the actual data observed, and it shares certain elements with each of those surrounding it. The implied rhythm of ii. is more like that of i., though not identical. Capitalization alters sentence-level stress and

gives the third syllable of "Epidemic" the unambiguous primary stress of the sentence. This is a possible interpretation of i. as well, but it is the only possible interpretation of ii. A Maia (2018)-like eye-tracking experiment could help validate these claims. Still, the meaning of ii. is likely more similar to that of iii. than to that of i. even if more study is needed to confirm that that perception is widespread.

There are many more examples of capitalization used in a similar way, as here:

(27)   Crossing borders in the Middle East is hardly straightforward, and the ferries that used to go from Greece and Cyprus stopped years ago, another casualty of 'The Situation'

The final two words appear to refer to the situation in the Middle East in a way that is paradoxically very broad but also very specific. It's broad in that it refers to the entirety of the complex geo-political situation in that part of the world, yet it's specific in that it is precise and clear about which situation is meant. The author seems to decline to provide more detail not out of lack of willingness to do so, but out of lack of necessity. (This is not a claim about the intention of the author as much as one about the apparent meaning, which again could be verified experimentally with a fairly straightforward design.) The implication seems to be that all parties involved are aware of the situation. It's not any situation, but *that* situation, The Situation. Notice also the reinforcement of this idea with single quotation marks. Similarly, we have:

(28)   […] the guitar designer must take responsibility for every detail of the Design

This occurs in the context of an application of Dieter Rams's *10 Principles of Good Design* to the design of guitars. Contrast with another capitalization by the same author:

(29)   This is another Principle that guitar designers and builders compromise on a regular basis

This is a clear example of the difference in meaning between CFC and other forms of capitalization. (29) capitalizes "Principle" likely as a reference to the title of Rams's work, but it also conveys a sense of givenness as discussed by Maia and is a good example of the overlap between the two kinds of capitalization discussed in this section. In (28) on the other hand, "Design" cannot be reasonably interpreted to refer to the work by Rams. Instead, it conveys a Ghomeshi-like sense of salience

and established meaning. The implication seems to be that this isn't mere design, but (Good) Design following the carefully defined Principles thereof. It draws a distinction between everyday design and Good Design, which must apply the Principles given by Rams and hold itself to a higher standard.

Next, consider a question asked as an introduction to a discussion of space tourism:

(30)   Who was the first Space Tourist?

The discussion that follows this occurrence aims to distinguish space tourists from astronauts and others that have gone into space for other reasons. This is a section header, but it is safe to assume that it is not written in title case due to the lowercase "was" and "first." It is a clear example of CFC instilling legitimacy and focusing attention on the contrast between the most salient meaning and others that are possible. The meaning of this question can be restated as "Who was the first *proper* space tourist?" for a definition of "proper" that is understood or, in this case, that is to be expanded upon shortly. Similarly, we have:

(31)   […] its amazing to me to hear people compare themselves with the lives of Stars

As above, this refers not to just any stars but to the most famous and most salient Stars; those that immediately come to mind. These two examples illustrate capitalization serving to specify and restrict possible readings just as Ghomeshi concludes that CFR does.

It is worth noting that most but not all examples of nonstandard capitalization collected fall into one of the two categories discussed in this section. It can also more broadly affect the rhythm or meter in which prose is read, as here:

(32)   Yeah yeah yeah No no no no No no no no

This usage often overlaps with capitalization for emphasis and for assertion that a noun is proper, but this is a clean example of how it is slightly different. This kind of capitalization forces a desired meter or sentence-level stress pattern and dictates how the sentence is to be "sung." It can be considered an element of many of the previous examples, but this particular usage highlights it as a distinct phenomenon. There were also several other capitalizations discovered that did not fit into one of the two

categories, but they were largely isolated and/or likely mistakes and thus do not warrant individual consideration.

5. Discussion

This section examines the findings presented above and makes an attempt to draw conclusions about informal usages of capitalization online. This venture set out to define and explore the concept of Contrastive Focus Capitalization, consider the ways it and other nonstandard capitalizations are used in online discourse in English and apply those findings to a broader discussion of the evolution of the relationship between spoken and written language in the Digital Age. Even with its imperfect and incomplete data collection methods, the venture to find examples of CFC in real online text posts was successful. The grammar and semantics of CFC are reasonably consistent and I argue that the phenomenon is meaningfully distinct from the more common practice of taking liberties with what can be considered a proper noun. The scope of CFC and creation of proper nouns were different and different meanings were implied. CFC and novel proper nouns handle compounding differently, occur in different (but not mutually exclusive) contexts and appear to encode different kinds of information. The comparison of CFC to Ghomeshi's CFR proved informative: as hypothesized, a striking number of similarities were found between the apparent meanings of the two structures. However, the scope of the two phenomena did not align as neatly as was hypothesized.

The current analysis faces some significant challenges when attempting to describe the scope of CFC. Some observations are certainly possible: CFC is most common on nouns, likely due to interaction with the other capitalization phenomenon discussed. It cannot target proper nouns. Both these observations contrast with the scope of CFR, which is used robustly on a wide variety of syntactic categories including proper and common nouns, verbs, adjectives and even prepositions. Furthermore, the curious behavior of CFR in idioms does not appear to be replicated for CFC although more study is needed to confirm this difference. Perhaps the largest open question uncovered by this inquiry is how to distinguish between CFC and capitalization for simple emphasis, if such a distinction can even be drawn usefully. One hypothesis may be that lexical categories like adjectives and adverbs have limited access to CFC because their capitalization will be interpreted emphatically, but other categories like nouns and verbs (which can't take on information about degree in the same way) can be contrastively focused by CFC. Under this hypothesis, *Good* would mean "very good," but *Design* would have to mean "the

established, salient form of design" because "very design" is not a meaningful interpretation. This hypothesis predicts that CFC can be used on verbs and contentful prepositions as well as nouns, which could be investigated empirically in future research. It is also possible that capitalization for emphasis and CFC are not considered to be distinct phenomena by their users, and that CFC simply includes emphasis as one of its many meanings.

While none of the specific examples used in the paper came from the NUS SMS corpus, that data also offered some useful context. Compared to the blog posts and similarly published texts of GloWbE, SMS messages' use of capitalization was much less predictable. Capital letters were often used for emphasis as described by Danet (2014), but these usages were much harder to parse. It was less clear what was a deliberate choice and what was the result of the text-input method used, and Lee's (2016) discussion of assuming intentionality in speakers informed the decision to exclude this data from ultimate consideration. Therefore the hypothesis that CFC occurs in text messages would need to be examined with more sophisticated text-parsing software or in a laboratory setting.

Another observation worth discussing is that the combined usages of CFC and proper noun creation in concert produce an interesting effect in names of demographic groups, especially marginalized ones. It's not always clear (or important) which of the two processes is behind this, but it is fairly common to see communities like Black, Gay, Deaf, etc. rendered with a capital letter when the meaning is the community rather than the characteristic that defines it. Consider again this data from section 4.1:

(19) Look at Black Africa
(20) Queer Hero strategy: successful people from the various academic disciplines who were homosexuals will have their sexuality identified in class/curriculum materials
(21) I know many Black Americans who have gone on a Sojourn to Africa
(22) […] the vision I have to achieve equality for Deaf people
(23) If people think there is even a trace of a chance of this leading to criticism of Gay people, it will die.

This capitalization, possibly via a combination of CFC and proper nounhood, has the effect of legitimizing the group and drawing attention to its status as a marginalized group deserving of respect. The historically-anomalous definitions of proper nouns discussed by Lee (2016) have left English with some inconsistencies in this regard: Hispanic, Asian-American and Jewish are national or religious identities and are therefore

proper nouns, but black, gay and deaf are descriptors and thus are not. This, along with the concept of "discourse focus" discussed by Maia (2018), illustrates the kind of real-world impact these usages of capitalization can have. If capitalization can draw a reader's attention to certain words in a sentence and/or characteristics of those words – and the previous section together with Maia's results appear to support that assertion – then treating these words like proper nouns and/or applying CFC to them could have measurable impact on perceptions of those groups as legitimate, institutional and of a kind with the more traditionally capitalized (and also more traditionally respected) groups based on religion or national identity. This is an empirical question, and the existence of this effect would need to be confirmed in a controlled experiment. In the meantime however, this specific application of nonstandard capitalization offers compelling motivation to continue exploring the issues presented in the current paper.

6. Scope

Having considered some of the patterns of use of capital letters in Web-based English discourse, it is prudent to examine the scope of this investigation and explore recommendations for future study. This paper is a first attempt at formalizing a pattern of use of capitalization called Contrastive Focus Capitalization which appears to serve many of the same semantic roles as Ghomeshi's Contrastive Focus Reduplication. It proposes and explores some additional patterns discovered in the data using methods of data analysis developed to look at CFC specifically. Those methods are insufficient to make most statistical claims. The selected examples are illustrative of the range of uses and meanings of CFC, but they are not a statistically representative sample and little is known about the distributions or frequency of any of the phenomena discussed. The process necessitated searching through a great many automatically-highlighted capitalizations for those most relevant to the current inquiry. It was possible to filter out sentence-initial capital letters and some other common capitalizations such as the word *I*, but with the available technology is was not possible to filter out names, places, initialisms or any of several other similarly common types of capitalization except by selecting only the most relevant data.

The method proved effective in finding enough data to support the claims of this paper and to draw some broader conclusions about English Web-based capitalization in general. However, there was almost certainly sampling bias introduced by the speed needed to search a reasonable amount of data in the time available. Texts with unusually many proper

nouns would be less likely to be selected from, as the higher number of capitalizations surrounding the relevant ones would make it harder to pick them out visually. For instance, many texts discussing sports were so full of team and player names that most of the nouns were capitalized (and therefore highlighted by the computer). An effort was made to look more closely at these texts to avoid such sampling bias, but this compensation surely fell short of eliminating the effect entirely. A more statistically representative data set could alleviate this concern in future research.

Additionally, the size of even the 1% sample of GloWbE made it unfeasible to review all the available data this way. The data was organized alphabetically by country, so a higher-than-chance percentage of examples used come from Australia and the United States. The former was the first to appear, and the latter proved to offer the heaviest concentration of instances of CFC. Therefore, while the sample data was truly global the selections were largely not. Future analyses could confirm whether the phenomena discussed are generalizable to the rest of the English-speaking world, but again, this paper does not claim to be statistically representative of any particular subsection of the data available. Rather, it argues that these usages exist and that capitalization in online discourse exhibits characteristics not found offline or in spoken English.

The particular data examined was in English, and other languages would no doubt display different patterns. It seems likely that each language used on the Internet has to some extent developed analogous idiosyncrasies that exist separately from any spoken equivalent, but the specifics of those patterns would be based on the histories, conventions and orthographical systems of those languages. This data was also within a certain range of formality and register. Most of the texts examined were blog posts, which tend to share some very broad characteristics in terms of writing style and are likely not perfectly representative of Web-based English more generally. Future researchers might expand data collection into social media, discussion boards and other forms of informal, online communication. Claims made about apparent meaning or likely implications have not been empirically validated, and a controlled experiment would be able to determine if those claims represent the perceptions of most English speakers. As of now they are the perceptions of one speaker in particular: a White, educated, cisgender male native speaker of General American English. A fairly simple experimental design could confirm which groups of English speakers share these perceptions and begin to explore the phenomena discussed here in the broader context of world Englishes.

Another possible direction for future study could be to compare usages of capitalization over time. Precise publication dates are not included in GloWbE and the nature of data on Web-based English excludes texts more than a small number of decades old. However, there are examples of references to capitalization aloud that are older than that, such as is found the song "Ya Got Trouble" from the play and film *The Music Man* (Wilson 1962). That song includes the lyric "Trouble with a capital T," and the live version premiered in 1957. Older examples like this one raise an important question about CFC and its proposed novelty. The current data suggests that CFC is used in online discourse in a reasonably consistent and widespread way, but it does not address whether this usage is truly new. It appears to have features and meanings that set it apart from other usages of capitalization in computer-mediated, spoken and written communication, and this analysis treats it as a feature of English CMC specifically. The possible existence of a larger historical pattern in the development of CFC would need to be considered via a dataset not limited to modern Web-based English, but the historical examples that come to mind are not so similar to CFC in their form and meaning that they appear to be the same phenomenon. Contrast the meanings of "Trouble with a capital T" (Wilson 1962) with "[…] in service of, like, The People with capital letters" (Gross 2019). The former means roughly "a lot of trouble," while the latter points to a specific interpretation of the phrase "The People" as contrasted with other, less salient meanings. More study is needed to confirm the novelty of CFC, but the historical trends in use of capitalization discussed above and by Lee (2016) support the assertion that, while capitalization has carried semantic meaning for some time, usages before and outside of CMC largely reflect changing written standards and lack the contrastive focus that characterizes CFC.

A notable omission from this analysis is the usage of caps lock, or all-caps. In investigating the idea of CFC, many conventions and forms with various degrees of similarity to the definition given above ("the capitalization of the first letter of a word other than a proper noun sentence-internally to convey one of several closely related meanings: legitimacy, salient nature and/or the intent for an established or understood meaning") were encountered. The decision was made that this practice of capitalizing entire words, phrases and sentences was sufficiently distinct not to be included. The phenomenon is also already better-studied than those covered here. See Danet (2014) for an overview of some of the forms deemed too unlike CFC for inclusion here, including all-caps. Given that this initial analysis supports the idea that CFC is being utilized intentionally by English-speaking Internet users at least in the United States, it would be worthwhile to examine its productivity and perception

in a laboratory setting. An experimental design similar to that of Maia (2018) could help separate the emphasis aspect of CFC from the more nuanced semantics that draw upon the work of Ghomeshi. Such a laboratory experiment could gather self-report data about writers' intent, compare the perceptions of CFC with those of CFR and other grammatical structures and document which usages of CFC are judged to be acceptable or unacceptable by its users.

7. Conclusion

Overall, there is preliminary but strong evidence that CFC is emerging as a meaningful grammatical construction of written, Web-based English. It is used in many different ways and contexts, but there is enough cohesion to support its continued study as an organic development of online English writing. The fact that such a thing can exist has powerful implications for the future of computer-mediated communication. It suggests that Internet users feel a degree of ownership over their writing styles and a willingness to experiment that, while not altogether new, is more common and accessible than ever before. As spoken language, written language and computer-mediated language continue to exist side by side, assumptions about their increasingly intricate relationship must be interrogated. Techniques used in the past, such as examining written language as a proxy for what was being spoken at the time of its creation, may become less reliable as such development continues. Spoken and written language will likely always remain closely related, but we must not let intuitive ideas about how they do or do not fully capture each other have undue effect on our inquiries. While the forms and peculiarities of Web-based language are certainly receiving a fair amount of attention in the field of linguistics, it is important to note that new patterns are always forming and their continued study will only offer more insight into the ways in which one of humanity's oldest inventions – language – is augmented, enhanced and adapted by one of its newest – the Internet.

REFERENCES

Chen, Tao, and Min-Yen Kan. "Creating a Live, Public Short Message Service Corpus: the NUS SMS Corpus." *Language Resources and Evaluation*, 2012, doi:10.1007/s10579-012-9197-9.
Crystal, David. *Internet Linguistics: a Student Guide*. Routledge, 2011.
Crystal, David. *Language and the Internet*. Cambridge University Press, 2011.

Danet, Brenda. "Computer-Mediated English." *The Routledge Companion to English Language Studies*, by Constant Leung and Brian V. Street, Routledge, 2014, pp. 146–156.

Davies, Mark. (2013) Corpus of Global Web-Based English: 1.9 billion words from speakers in 20 countries (GloWbE). Available online at https://corpus.byu.edu/glowbe/.

Discourse Constructions of Youth Identities. Benjamins, 2003.

Eisenstein, Jacob. "Systematic Patterning in Phonologically-Motivated Orthographic Variation." *Journal of Sociolinguistics*, vol. 19, no. 2, 2015, pp. 161–188., doi:10.1111/josl.12119.

Evans, Vyvyan. *Emoji Code: the Linguistics behind Smiley Faces and Scaredy Cats*. PICADOR, 2018.

Ghomeshi, Jila, et al. "Contrastive Focus Reduplication in English (The Salad-Salad Paper)." *Natural Language & Linguistic Theory*, vol. 22, no. 2, 2004, pp. 307–357., doi:10.1023/b:nala.0000015789.98638.f9.

Gross, Terry, host. 'Never Look Away' Asks: Why Make Art? Who Is It For? Fresh Air, NPR, 6 Mar. 2019. NPR, www.npr.org/2019/03/06/700663706/never-look-away-asks-why-make-art-who-is-it-for

Gunraj, Danielle N., et al. "Texting Insincerely: The Role of the Period in Text Messaging." *Computers in Human Behavior*, vol. 55, 2016, pp. 1067–1075., doi:10.1016/j.chb.2015.11.003.

Jaffe, Alexandra. "Introduction: Non-Standard Orthography and Non-Standard Speech." *Journal of Sociolinguistics*, vol. 4, no. 4, 2000, pp. 497–513., doi:10.1111/1467-9481.00127.

Kalman, Yoram M., and Darren Gergle. "Letter Repetitions in Computer-Mediated Communication: A Unique Link between Spoken and Online Language." *Computers in Human Behavior*, vol. 34, 2014, pp. 187–193., doi:10.1016/j.chb.2014.01.047.

Lampert, Martina. "Say, Be like, Quote (Unquote), and the Air-Quotes: Interactive Quotatives and Their Multimodal Implications." *English Today*, vol. 29, no. 04, 2013, pp. 45–56., doi:10.1017/s026607841300045x.

Lee, Jerry Won. "The Politics of Intentionality in Englishes: Provincializing Capitalization." *Critical Inquiry in Language Studies*, vol. 13, no. 1, 2016, pp. 46–71., doi:10.1080/15427587.2015.1126524.

Matsumoto, Mallory E. "From Sound to Symbol: Orthographic Semantization in Maya Hieroglyphic Writing." *Writing Systems Research*, vol. 9, no. 2, 2017, pp. 99–122., doi:10.1080/17586801.2017.1335634.

Metcalf, Allan A. *OK: the Improbable Story of America's Greatest Word*. Oxford University Press, 2012.

Na'aman, Noa, et al. "Varying Linguistic Purposes of Emoji in (Twitter) Context." *Proceedings of ACL 2017, Student Research Workshop*, 2017, doi:10.18653/v1/p17-3022.

Newman, John G. "The spread of the s-plural in Middle English (1150-1420): A corpus study." (1999).

Rodríguez-Hidalgo, Carmina, et al. "Expressing Emotions in Blogs: The Role of Textual Paralinguistic Cues in Online Venting and Social Sharing Posts." *Computers in Human Behavior*, vol. 73, 2017, pp. 638–649., doi:10.1016/j.chb.2017.04.007.

Tagliamonte, Sali A. "So Sick or so Cool? The Language of Youth on the Internet." *Language in Society*, vol. 45, no. 01, 2016, pp. 1–32., doi:10.1017/s0047404515000780.

Wilson, Meredith. *The Music Man*. *YouTube*, 1962, www.youtube.com/watch?v=LI_Oe-jtgdI.

"Upside-Down Face Emoji." *Emojipedia*, emojipedia.org/upside-down-face/.