Article

Sophia Burnett*

Offsetting love and hate: The prosodic effects of the non-standard 1sg in tweets to Boris Johnson and Jeremy Corbyn over four days of the UK general election

https://doi.org/10.1515/eujal-2022-0012

Abstract: There is no punctuation in English endowed with attenuating qualities which could function as the contrary of "!", and in a language with no speech levels, such paucity of expression can come at a cost, especially online. This paper on the non-standard 1sg in English — *i* — aims to demonstrate that the use of this novel form is both conscious and meaningful, indeed it is a variation carrying its own linguistic mechanisms. Using linguistics and statistics we will see: a) how the use of the lowercase variant, which in English is an aberration, can have prosodic effects on the utterance, and thus signify a feeling. And, b) how textometry allows us to reveal the use of this non-standard 1sg in a contrastive corpus of tweets addressed directly to (@) Boris Johnson and Jeremy Corbyn, over four days of the UK general elections, by testing its frequency with several collocations, among others: slurs and hate speech, hapax, and conjunctions. The results obtained tend towards the confirmation that the non-standard 1sg is used not only to signal youth, but also as a precautionary implement deployed when weighing in on divisive topics, amounting to publishing a statement with a caveat.

Keywords: 1sg, CMC, Variation

Résumé: Il n'existe pas en anglais de ponctuation dotée de qualités atténuantes qui pourrait fonctionner comme le contraire de "!", et dans une langue sans niveaux de parole, une telle indigence d'expression peut avoir un coût, surtout en ligne. Cet article sur le 1sg non-standard en anglais — *i* — vise à démontrer que l'utilisation de cette nouvelle forme est à la fois consciente et significative, il s'agit en effet d'une variation portant ses propres mécanismes linguistiques. En adoptant une approche linguistique et statistique, nous verrons : a) com-

***Corresponding author: Sophia Burnett,** CY Cergy Paris Université, Cergy-Pontoise, France.
E-Mail: sophia.burnett@cyu.fr

ment l'utilisation de la variante minuscule, qui en anglais est une aberration, peut avoir des effets prosodiques sur l'énoncé, et ainsi signifier un sentiment. Et, b) comment la textométrie nous permet de révéler l'usage de ce 1sg non standard dans un corpus contrastif de tweets adressés directement à (@) Boris Johnson et Jeremy Corbyn, durant quatre jours des élections générales britanniques, en testant sa fréquence avec plusieurs collocations, entre autres : insultes et discours de haine, hapax, et conjonctions. Les résultats obtenus tendent à confirmer que le 1sg non standard est utilisé non seulement pour signaler la jeunesse, mais aussi comme une mesure de précaution déployé lorsque l'on prend position sur des sujets qui divisent, ce qui revient à publier une déclaration sous réserve.

Zusammenfassung: Im Englischen gibt es keine abschwächenden Satzzeichen die als das Gegenteil von "!" fungieren könnten, und in einer Sprache ohne Sprachstufen kann ein solcher Mangel an Ausdrucksmöglichkeiten seinen Preis haben, insbesondere im Internet. In diesem Beitrag über das nicht standardisierte 1sg im Englischen — *i* — soll gezeigt werden, dass der Gebrauch dieser neuen Form sowohl bewusst als auch bedeutungsvoll ist, da es sich um eine Variation mit eigenen linguistischen Mechanismen handelt. Mit Hilfe von Linguistik und Statistik werden wir sehen: a) wie die Verwendung der Kleinschreibung, die im Englischen eine Abweichung darstellt, prosodische Effekte auf die Äußerung haben und somit ein Gefühl ausdrücken kann. Und b) wie die Textometrie es uns ermöglicht, die Verwendung dieses nicht standardisierten 1sg in einem kontrastiven Korpus von Tweets zu zeigen, die direkt an (@) Boris Johnson und Jeremy Corbyn gerichtet sind, über vier Tage der britischen Parlamentswahlen, indem wir die Frequenz mit verschiedenen Kollokationen testen, unter anderem: Verunglimpfungen und Hassreden, Hapax und Konjunktionen. Die Ergebnisse deuten darauf hin, dass das nicht standardisierte 1sg nicht nur verwendet wird, um Jugend zu signalisieren, sondern auch als Vorsichtsmaßnahme, wenn man sich zu kontroversen Themen äußert, was der Veröffentlichung einer Erklärung mit einem Vorbehalt gleichkommt.

1 Born digital

In December 2013, Justine Sacco wrote a tweet to her 170 followers before boarding a flight from London to Cape Town. "Going to Africa. Hope I don't get AIDS. Just kidding. I'm white!" Unbeknownst to her, during the twelve hour flight her tweet went viral, with people the world over voicing their moral condemnation. By the time Sacco stepped off her plane, her tweet and the reactions to it had done irrepar-

able damage to her professional and private life.[1] This case of public shaming may well have served as an event line for the use of the standard (upper-case) subjective first-person singular pronoun in English — I — hereinafter referred to as 1sg. With its correct punctuation, the majuscule for the country, a correctly rendered acronym and an exclamation mark at the end, there was little space for deniability, no linguistic disclaimers allowing Justine Sacco to distance herself from her own words. It was all there and so was she, no less than three times in the short message. Implicitly before *Going (I'm going),* then explicitly *I,* and *I'm*. To quote Heath (2018) "it is unsurprising that an orthographic convention for marking prosody has emerged (Baym 2015, Crystal 2011, Werry 1996, Zappavigna 2012, Ferrara et al 1991)".

Since 2013 the non-standard (lower-case, an aberration in English) 1sg – i – has appeared online in "born-digital data"[2], and one online demographic in particular uses the non-standard form as sociolect and that is youngsters; both as signalling, i.e., "immediate and unhesitating identification" (Jaffe 2000) for reasons such as Sidani (2016) describes "Teenagers in marginalized communities (...) can identify assistance and companionship using of social networks", but also for the prosodic 'slouch' it affords an utterance.

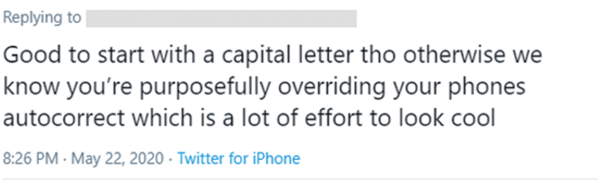

Figure 1: Overriding for a minuscule is common knowledge

The use of the non-standard 1sg alluded to in Figure 1 needed to be distinguished from other pragmatic uses explored in this and upcoming papers, hence the novel nomenclature 'active nonchalance', because upon observation it is clearly a crafted (active) manner of looking cool (nonchalance). The question of which came first, the intentional dismantling of grammatical norms for the purpose of generational differentiation, or the cool sound of the inherent prosody, is a problem of indefinite causal order.

Active nonchalance was an entry point phenomenon that led to this study whose purpose it is to render observable evidence that the use of *i* is not random but meaningful, and conveys a tone of voice comparable to someone who is not

1 https://www.nytimes.com/2015/02/15/magazine/how-one-stupid-tweet-ruined-justine-saccos-life.html

2 A definition of the term "Born Digital" https://www.oclc.org/content/dam/research/activities/hiddencollections/borndigital.pdf

fully assertive; in contrast to the standard form *I*, which would be (in subvocalized speech) clearly enunciated and with a primary tonal emphasis. The hypothesis being that the use of the non-standard 1sg is consciously employed in the semantic organization of the lexicon for its distancing properties, and deployed in online situations of risk. "The systematic nature of most orthographic deviations on social media sites suggests that the observed range of variation is not due to carelessness or typos, but should rather be analyzed as stylistic" (Crystal 2011). Heath (2018) quite rightly supplements this with "(and linguistic)". In order to understand the prosodic significance of *i* and its possible meaning in a phrase, it is necessary to move away from the Saussurian linguistic ideal[3] and venture into what he called "external linguistics", i.e., sociolinguistics and psycholinguistics, as part of the "ever expanding research into the linguistics of verbal exchange within sociolinguistics and related fields (...) that has facilitated a renewed interest in paralanguage (and a continuing one in prosody)." James (2017).

1.1 The sound of text

Gadet (2006) states that oral language is natural whereas written language is learned, and contrasts the attributes of the oral and the written. Since then however, social media and other CMC have proliferated, and the boundary between oral and written has become fluid, especially in the direction of oral to written. It would seem that people publishing online, in particular on public networks are finding ways in which to render their writing more 'oral'.

Social media is read by millions of people on a daily basis engaging their inner voice to read their timeline. Tweets are often crafted precisely for their potential to evoke emotion, either hammering home illocutionary goals, i.e., USING ALL CAPS — a phenomenon that Heath (2018) analyzed in depth drawing on Twitter corpus data to show how users employ single-word capitalization in positions indicative of emphatic stress and semantic focus — or as we will see with the lower-case 1sg, sanding the corners of them down. As Petkov and Belin (2013) stated, "Silent reading involves different levels of processing (orthographic, semantic, syntactic, phonological, emotional) but also seems to involve our little inner voice". This phenomenon has been studied long before the appearance of CMC, beginning with St Augustin (397) and by the way of French psychologists Egger and Ballet (1880). "Since then, several studies in experimental psychology have

3 "Its study must disregard everything that is foreign to its organism, to its system, in a word everything that is designated by the term 'external linguistics'." Saussure (1916)

evaluated this proposal and have shown that silent reading involves our inner speech." (Berlotti, Grandchamp, et al. 2016). Modern written English, which is devoid of the speech levels it once had must now rely on the non-syntactical tools of lexicon, discursive environment, and tone to convey pragmatic subtleties. In a similar manner to which humans render dreams (Aaru et al. 2020) using stored memory, I posit that the modulation inherent in subvocal speech is not a monotonous reader and can, on the contrary relate to the reader's brain a great palette of emphases, intensities, registers, dialectal accents, in short all idiosyncrasies of sound elements — real or recorded — already encountered in the reader's life. This similarity is echoed in Fuchs and Krivokapić's study into prosodic boundaries in writing (2016), that indicated a parallel to speech: "(...) there is evidence of cumulative lengthening, such that segments at major boundaries are longer than segments at minor boundaries." In Leech's (2014) work on "superconstraint", he notes situations wherein "both participants are 'leaning over backwards' to avoid the potential discord resulting from following their own goals." With this in mind, and given the concise nature of much CMC exchange, it is quite logical to see the emergence of variations such as the 1sg minuscule, which is — in terms of lexemes — a highly economical pragmatic.

Table 1: Reproduced from Gadet (2006) Variation sociale en Français [Social variation in French]. This table apposes the attributes of social language, both oral and written, at a point in time of great social media expansion.

Oral	Written
Prosody	Punctuation
Evanescence	Permanence
Contextualization	Autonomy
Implication	Detachment
Redundancy	Concision
Natural	Acquired
Directed towards others	Directed towards self
Transparent	Dense
Unclear	Precise

1.2 Phonological hypothesis and theoretical approach

If we concede that most readers perceive a vocalized representation of *i*, we need to determine the quality of this internalized 'oral' production to see if it follows

the same rules as externalized phonology. The core hypothesis herein is that *i* is enunciated less loudly than *I*, with an onset tonic accent that is strongly diminished. We can already confront this notion with existing knowledge in phonology; if we accept that the phonology of subvocal speech follows the same rules as the phonology of voiced speech.

Chomsky and Halle (1968), established that the word "*blackboard*" had an onset tonic accent, but that the inflection fell at the end of the word. However, when the noun was transformed into the noun phrase (two words) *black board*, there was a tonic accent on each word entry, and the inflection rose at the end.

Figure 2: Reproduction of Chomsky and Halle's "blackboard" (1968) The Sound pattern of English.

So let us consider *I* /ˈaɪ/ in a sentence: "I'm not really sure I like chicken." Here the 1sg holds its standard form for the English language: /ˈaɪ/. We can then study the non-standard form in the same sentence: "*I*'m not really sure *i* like chicken." Finally, we can apply Chomsky and Halle's blackboard to our example (chicken) phrase as a framework.

a. *Black board*: It is the space between *black* and *board* that allows for more primary tonal emphasis on the noun and imposes a neutral if not rising inflection on the whole phrase.
b. *Blackboard*: In the absence of space the main tonic accent of the noun *board* is reduced. A descending inflection is imposed on the whole phrase.
c. "I'm not really sure I like chicken.": Here the utterance of /ˈaɪ/ followed by the space allows for a main tonic accent on the consonant of *like*, as well as an opening of the vowel phoneme /aɪ laɪk/. We find a neutral if not rising inflection on the whole phrase.
d. "I'm not really sure *i* like chicken.": If we accept that the subvocal speech of the variant *i* is represented by a 'smaller', 'lower' voice,
e. This has the effect of diminishing the main tonic accent on the opening consonant of *like*.
f. Which has the effect of 'shortening' the space between *i* and *like*.

g. This increases the processing speed: *i* and *like* become almost glued together, *ilike*, a chunking effect which imposes a downward inflection on the whole phrase.
h. The 1sg thus denatured, the diphthong has become a monophthong and can be represented by: ə as in ago // ʌ as in but // ɑ: as in ask.

The English language like any other is bound to evolve its sound pattern. Most often when there are changes, it is in the direction of economy, an easing of articulation. For example, only the so-called rhotic places in North America or in the West of the United Kingdom continue to pronounce the "r" at the end of words. For other English speakers the *r* has become a schwa. It is therefore not surprising to note the departure of a diphthong *I* which, in 'Received Pronunciation' mobilizes a large part of the vocal apparatus, towards a monophthong *i* which is weak to the point of not requiring any modulation at all. Mounin's (1974) definition of linguistic prosody as "foreign to the double articulation but inseparable from the discourse" is pertinent here because the ultimate goal of this research is that of a better understanding of online discourse, in this case written.

If the prosody of *i* can cast a shadow — as we have just seen, thanks to phonology — we can test its distribution thanks to syntax. In English, the pronoun is always subject to a verb; and thanks to intransitive verbs, even in a short, two-word sentence, e.g., I exist. "According to one interpretation," says Dubois (2002) "the predicate is a verbal phrase, whether it consists of a verb alone, or of the verb and one or more elements subordinate to it." This renders the 1sg syntactically powerful.

Since *I* is a personal pronoun and thus a recurring subject in much writing, it has a large potential for syntactic influence. Let us now explore the idea that the impact of this prosody is also broad; perhaps having a prosodic influence on almost the entirety of a multi-clause sentence, not just the surrounding lexemes. Thus the lowercase 1sg could rely on syntax to extend the effects of its prosody to the whole statement; and this for other reasons that go beyond the purely phonological as we have just seen above thanks to Chomsky and Halle. *I* is always an argument of a predicate. Thus, *I* is found in predicates with 1, 2 and 3 arguments. For example:

– I exist
– I cook for her children
– I gave the dog a bone

Now let's replace the *I* with an *i*, and ponder the prosodic effect.

– i exist
– i cook for her children
– i gave the dog a bone

Again, the main tonic accent is on the opening consonant of the words *cook* and *gave*. The 1sg thus denatured, it is now possible to represent it as ə /ɑː/ or /ʌ/. This has the effect of 'shortening' the space between *i* and *cook* and *i* and *gave*. This increases the reading speed, and as in the example above, *i* cook and *i gave* become almost glued together – icook / igave, with a downward inflection throughout the phrase. Whereas in the example in Figure 3, the given phrase illustrates a use of *i* in the first and second part of the phrase, opening both clauses. The prosodic effect is applied throughout thanks to the repetition of the lowercase 1sg at the beginning of the second clause. This acts as a beacon that repeats the signal, reinforcing the prosody.

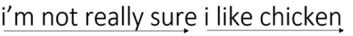

Figure 3: Second clause reinforcement of the prosody over the whole utterance

If we take the sentence "i cook for her children" and lengthen it by making the direct object a subordinate clause: "I cook for her children who are on holiday for two weeks", minimizing the typography of the personal pronoun at the beginning of the sentence: "i cook for her children who are on holiday for two weeks", we can see how the prosody of *i* could have an instance on the entirety of a long statement.

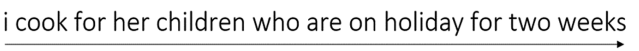

Figure 4: Weak onset prosody with a longer phrase

Unlike the sentence "i'm not really sure i like chicken" this longer sentence does not benefit from a second lowercase 1sg. Can prosody be conveyed and relayed by the arguments of a predicate in an utterance? It may be possible in this way: In this extended sentence, the predicate is 'cook'. Knowing that the predicate is *cook* and the arguments are *i*, *children*, and the subordinate clause, then perhaps the English mother tongue reader intuitively understands this distribution, and upon seeing that the lower case 1sg is part of it, will naturally extend the same prosodic effect to all the other arguments. One of the questions this raises is: Do users of the 1sg minuscule favour simple sentences or, on the contrary, is it deployed precisely because one has more words to communicate and that it is the fact of spending time, of lingering as it were, in a potentially hostile place that is the reason for a cautious stance.

In an attempt to answer this, the corpus will be explored in order to see if a salient syntactic structure co-occurs with the lowercase 1sg. This will be achieved by identifying sentences that include a token of the most frequent

subordinating conjunctions — an advantage for this research being that they are invariable words. After analysing the tweets in textometry[4], it is then possible to establish if the corpus contains more occurrences of *i* in sentences including a subordinating conjunction, or co-occurrences with abusive or hateful words.

1.3 Suprasegmental properties

Suprasegmental analysis is important for addressing the prosody of the lowercase 1sg which follows time old principles. "From the Old English period through to today, all content words must be at least two morae long." (Hogg 1992). Additionally, the fact that the pronoun in question is composed only of a vowel is important because it is the vowels of a morpheme that vocalize mores. "A syllable onset (the first consonant or consonants of the syllable) does not represent any mora. The syllable nucleus represents one mora in the case of a short vowel, and two morae in the case of a long vowel or diphthong." (Howell and Van Borsel 2011). In all languages, as Borrell and Salsignac (2002) point out, "Rhythm facilitates syntagmatic splitting because pauses often appear at the boundaries of syntactic groups."

In the sentence used as an example, "I'm not really sure *i* like chicken", the 1sg is followed by a consonant, and the *I* in *like* does not carry morae. If we accept that the denatured sound of *i* can be represented by a schwa, ɑ: or ʌ, it is very easy at the palate/tongue level to continue an utterance in which the letter following the *i* is a consonant, especially those that are not velar, uvular, or pharyngeal, such as [t / b / d / l / k] .

Followed by a vowel, which in English are necessarily voiced, or a consonant that requires more exhalation such as *h*, the *i* is slightly harder to enunciate. So it is, technically, easier to say *i love you* than *i hate you*. To voice *i love you*, the speaker only needs to preserve the exhalation of *i* and add a slight lingual contact on the alveolar ridge: [ʌlʌvju :] To say *i hate you*, given that the "h" in English is voiced, the speaker has to perform a double exhalation [ʌheɪtju:].

The personal pronoun in English is always followed by a verb or modal. It can be followed by another form only after the addition of a comma, i.e., "I, a weary

4 "Textometry is already well rooted in social science studies and quantitative linguistic research [10][11], mostly developed in France with numerous pioneers, Pierre Guiraud, Charles Muller, Jean-Paul Benzécri, Ludovic Lebart and André Salem." (MacMurray and Leenhardt 2012).

traveller with a grey beard". But this is a literary form of writing; it would be considered archaic used in conversation on Twitter, except knowingly for its idiosyncratic effect. Rodero and Potter (2017) confirm that prosody is commonly used as a didactic tool to amplify certain speech components to a child or adult learning a language; in which case it is customary to exaggerate the existing prosody, because it is thought easier to retain information that has been expressed in this way. Adults too rely on prosody for semantic and syntactic processing (Falk 2014). This is all the more useful as a practice in English, which is a language where the simple change in the distribution of the tonic accent within a word can transform a verb [record] into a noun [record]. Cruttenden (1997: 13) says of tonal stress for highlighting syllables, that of "the three features (pitch, length, and loudness), pitch is the most effective, and loudness the least." Indeed it is the length of the syllable that is manipulated when playing with and transforming standard typography.

Accentuating the existing tonal accent is a means to ensure people retain information. We have all, at one time or another, had to communicate an important piece of information — a name or a number, for example — perhaps over a poor phone connection, or during a loud party, and the most important parts of the phrase, the ones that we would like the listener to remember, are usually over-articulated, longer, and with a louder onset. "Grouping serves an informational function allowing for packaging of notes and tonal events into larger relevant temporal units." (Falk 2014). Just like the unstressed syllable of a Shakespearean iamb, the parts of less interest or no interest are reduced in juxtaposition; they can be forgotten. These are the words that serve to build the sentence but are not recognized as semantically rich. This use of prosody can also be implemented in written communication. The prosody of lowercase 1sg is not an exaggeration of stress, but its opposite. The author is consciously deploying the *i*, whilst simultaneously communicating, "Forget I'm here: do not accord too much importance to this statement, nor to me".

2 Methodology

To verify the hypothesis that the non-standard 1sg is used intentionally, and that it is used pragmatically for its attenuative properties, textometry analyses were used to query:

a. The frequency of occurrences of standard and non-standard 1sg
b. Hapax in lowercase 1sg
c. Additional forms in 1sg (vertical vs. horizontal)
d. Subordinating conjunctions

e. Active forms—contrasting salience
f. Occurrence of positive and negative words
g. Occurrences of direct address 1sg lowercase + you1
h. Co-occurrences of non-standard 1sg
    1. Discursive environment of *i* to determine:
    2. Is there a user profile of *i* in this corpus?
    3. The habitual/unhabitual nature of an author's use of lowercase 1sg; a sign of accommodation

For the study of a – h, automated extraction + statistical analysis by textometry was used. For the study of 1–3, manual analysis was required in addition. This manual analysis, plus other offline experiments conducted during this study are too voluminous to be communicated herein.

2.1 Extraction and textometry

The tweets addressed to Boris Johnson and Jeremy Corbyn were extracted four nights in a row from their live feeds, on December 10th, 11th, 12th, and 13th at 7:30 pm, and for a duration of ten minutes. Having previously looked at the activity towards the two accounts in the run-up to the elections, the estimation was that ten minutes would suffice to glean a corpus large enough for a contrastive analysis. Per the hypothesis, an intensification was expected — an increase in occurrences of the lowercase 1sg as the election approached, with more impassioned language and narrowing of the discourse. That the day before, and the evening of the General Election, the vocabulary used would convey either support or contestation.

In real time it was possible to see that the volume of tweets addressed to Boris Johnson was higher than to Jeremy Corbyn. This meant that the same sampling period (10') generated a difference in corpus size of a few hundred lines (tweets). In order to prepare the corpus for textometry, the @BorisJohnson corpus was truncated from the end to work on the same number of tokens. The primary software used was Iramuteq[5], an open-source tool built on R and Python and developed by Pierre Ratinaud at the LERASS Laboratory at the University of Toulouse. "As the origin of its name indicates, IRaMuTeQ is an Interface to R for Multidimensional Analysis of Texts and Questionnaires." (Pincemin 2018). The extracted raw data is cleaned and saved to an .xl file. Iramuteq espe-

5 http://www.iramuteq.org/

cially requires the removal of any asterisks which could lead to confusion during the segmentation. The introduction of the text must be marked by four asterisks, and each variable by one asterisk; then conversion to .txt allows transfer of the corpus to the software. A small difficulty with using Iramuteq for this corpus was that some functionalities allowing visualizations in particular (grouping by topicality) only worked well when the corpus was lemmatized in lower case; but such queries were useless to this study if majuscule and minuscule 1sg were counted in the same way.

TXM[6], another open-source tool, developed by Serge Heiden at the IHRIM Laboratory at ENS Lyon was used to query co-occurrences in particular. TXM requires that the .csv file and the .txt file be in the same folder at the time of the transfer, and the multiplication of corpus versions requires meticulous labelling in order not to make source errors. An open-source[7] tool called VOYANT[8], which uses many of the same algorithms as Iramuteq, including Pearson and Alceste, was also implemented. The tool, which was developed by Stéfan Sinclair (McGill University) and Geoffrey Rockwell (University of Alberta) offers powerful queries and visualizations, and because it runs externally (web-based), it was a flexible and lightweight addition that proved useful for control analyses.

3 Corpus

In order to find a large volume of opinionated @ tweets and a hostile environment, the impending UK general election was chosen; indeed several thousand tweets could be extracted @BorisJohnson and @JeremyCorbyn, i.e., tweets posted publicly on the profiles of the politicians in question. Of course although the messages were all '@' Boris Johnson and Jeremy Corbyn, no response was expected from the politicians, indeed the pages became ersatz of a topic board where people came to post their grievances and or support about and for the two candidates.

6 https://txm.gitpages.huma-num.fr/textometrie/
7 https://github.com/sgsinclair/Voyant
8 https://voyant-tools.org/

Table 2: Sample of BJ 10/12 full corpus (including standard and non-standard 1sg)

488	Call me old fashioned but I prefer not to respect referendums won by election fraud and Russian meddling.
489	plus I've got learning difficulties and ceroble palsey
490	Think I might start supporting the left. According to Twitter they are all experts on the NHS, trade agreements, tax loopholes, employment law, who tells the biggest fibs, bunnies and sherbet.
491	deep fry the shit out of it then maybe i'll chuckle.
492	I can't find my phone, do you have it?
493	Don't tell me what I should or shouldn't do...
495	Good morning. It's still get brexit started not done. You carry on saying it and I'll carry on refuting it.
496	Wish I could vote, it would be for you. Epatriot from Canada
497	I would never vote for you. Your just a liar, Jeremy gets my vote.
498	If you keep chanting the same words, it doesnt make it more likely.I don't trust you
499	I thought we were having it because Theresa lost the majority
500	I'd rather shit in my hands and clap.

A political corpus was chosen because it easily afforded access to empassioned tones, a chronology with an event, and two very different candidates promising discourse heterogeneity.

Table 3: Sample of JC 10/12 full corpus (including standard and non-standard 1sg)

830	you're so cute i love you!!!!!
831	I've never read more shit from bbc reporters in my life
832	Horrid man...as I keep saying dangerous man
833	I listened to your pledge at climate hustings last week. Are you going to have a word with your boss?
834	I know I am wasting my time but Can you tell us what the EU changed to make you change from Anti EU to now pro EU. It is confusing you were shouting how awful their system was. So please explain. This should have been asked by the like of Andrew Neil.
835	This is fake and you know it, I wouldn't trust you to run a bath let alone this great country!
836	I can understand why Santa's little helpers tried to clean up their shit on your twitter account while you slept.
837	You really dropped the ball on this fuvker, didn't you?
838	Labour setup from start to finish. No wonder the mum now wants the media to back off – in case we find out the full facts. But I doubt if you've checked to see if it was true. This is what Labour bias looks like.
839	Spoil ur ballot paper in protest 2 the utter shite this country has 2 offer #GeneralElection19
840	i was canvassed by the tory party at my door in north ayrshire the canvassar said if i voted for labour

Four extractions were made, each of ten minutes, (~ 2000 tweets) between December 10, 2019 and December 13, 2019. Election day was December 12.

Table 4: Structure of the corpus of tweets comprised of 323,514 words addressed to Jeremy Corbyn and Boris Johnson, candidates in the UK general election, December 2019.

JC10	Tweets @ Jeremy Corbyn 10/12/19	BJ10	Tweets @ Boris Johnson 10/12/19
JC11	Tweets @ Jeremy Corbyn 11/12/19	BJ11	Tweets @ Boris Johnson 11/12/19
JC12	Tweets @ Jeremy Corbyn 12/12/19	BJ12	Tweets @ Boris Johnson 12/12/19
JC13	Tweets @ Jeremy Corbyn 13/12/19	BJ13	Tweets @ Boris Johnson 13/12/19

The two men campaigning to become the new British prime minister were Boris Johnson for the Conservative Party, and Jeremy Corbyn for the Labour Party. Both men had impassioned supporters but also many detractors. Boris Johnson was deemed responsible for Brexit, which was the most divisive event in recent UK history. Johnson was also criticized for being a *toff*, from the British upper class, educated at Eton and Oxford, where he was member of the Bullingdon Club[9] in the 1980s, which represented the height of British elitism. Quoting Homer in Greek during television interviews[10], and visibly comfortable with the media since his days as Mayor of London, Johnson's persona was divisive.

Jeremy Corbyn, having long supported the cause of the Palestinians, was accused of anti-Semitism by people in his own party. Corbyn was a more reserved party leader, and many Labour party members criticized him for not showing more 'backbone' to the opposition, and for being non-committal about Brexit. In the events that occurred during the four days of the extraction of these tweets, and that surface in the lexicon to both candidates, there was of note:

10/12 – BBC debate: heated debate between the two men about Brexit. Also, Johnson appears in a parody of a scene from the romantic comedy *Love Actually*, promising to finalize Brexit.

11/12 – Last pitches: Corbyn urges 'vote for hope', Johnson promises to 'make the NHS (National Health Service) his priority'. Johnson hides in a refrigerated room to avoid questions from a TV reporter.

12/12 – Election day.

13/12 – Boris Johnson is the new Prime Minister

9 https://www.theguardian.com/politics/2019/jul/07/oxford-bullingdon-club-boris-johnson-sexism-violence-bullying-culture

10 In 2013, Johnson was invited to a televised interview at the Melbourne Writers Festival during which he recited an extract of Homer's Iliad in ancient Greek.

4 Data

4.1 Standard 1sg

The use of the standard 1sg shows a similar progression in both corpora, with a gradually declining frequency. The standard 1sg is important here, in that its very presence allows us to draw one simple conclusion in regard to the conscious implementation of the non-standard form. Given that it has been present in popular culture since at least 2014 when American rap artist Kendrick Lamar used the sign for his single artwork[11], and given the rapidity of online memetics (Smith and Copland 2021) the fact that the lower-case 1sg has not completely replaced the majuscule, (cf. Figure 5), tells us it is indeed a variant, not a new standard — enriching, not zero-sum. "Written language norms are pluralised to the extent that different styles of writing can be deemed appropriate in different environments." (Androutsoupoulos, 2011). Note that the personal pronoun in both its forms is most present in the fray of election night.

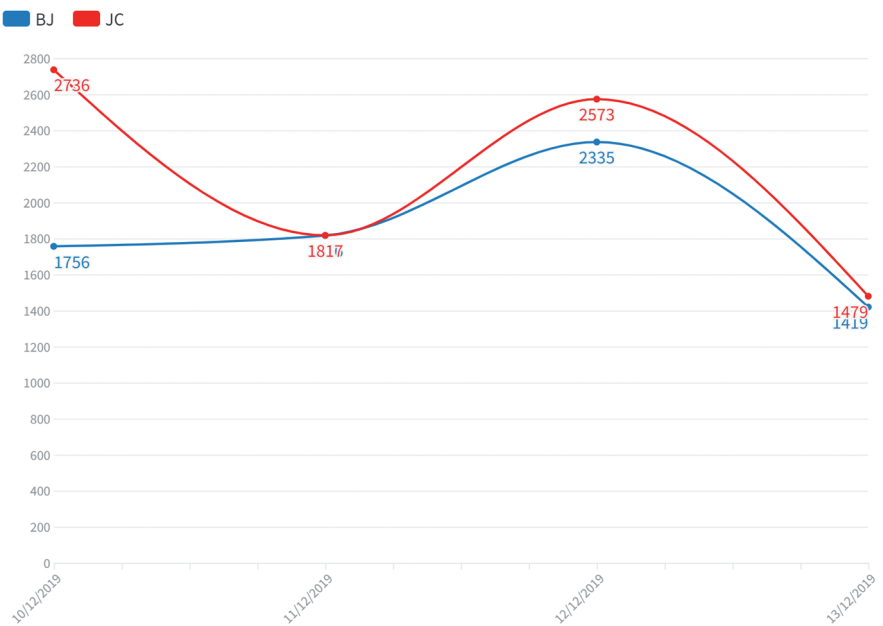

Figure 5: Occurrences of the standard 1sg within the full corpus

11 https://acclaimmag.com/music/kendrick-lamar-releases-cover-art/

4.2 Non-standard 1sg

There is a significant rise of the non-standard form @JeremyCorbyn around the election date and an almost mirror image @BorisJohnson. In both corpora the use of the non-standard form does not stem from accommodation, since the corpus was comprised of @ tweets replying to the politicians' accounts, both of whom communicate with standard punctuation and grammar containing few if no idiosyncrasies, and especially devoid of any lowercase 1sg. Unlike the standard 1sg distribution, here the implementation of the non-standard form encounters a complete reversal between day one and two of extraction. The fact that the distribution of lowercase 1sg (cf. Figure 6) varies so much from that of the standard 1sg (cf. Figure 5) is eloquent of some form of conscious implementation.

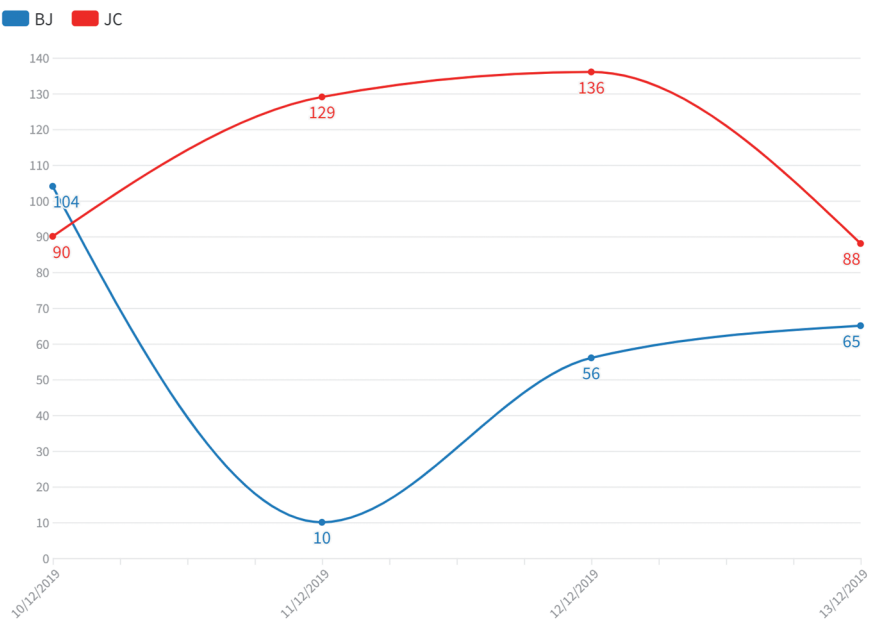

Figure 6: Distribution of lowercase 1sg over four days @ BJ and JC in full corpus

4.3 Saliences and general statistics

Table 5: Synthesis of the general statistics showing the most salient active and supplementary forms. The tables contrast sub-corpora from 10–13 December 2019.

BJ10 contains 51,885 words	JC10 contains 55,391 words
Lexical density: 0.034	Lexical density: 0.030
Average n° words per phrase: 13.9	Average n° words per phrase : 15.6
Most frequent: brexit (333); vote (322); like (201)	Most frequent: vote (347); labour (324); hope (249)

BJ11 contains 30,710 words	JC11 contains 37,213 words
Lexical density: 0.056	Lexical density: 0.051
Average n° words per phrase: 14.3	Average n° words per phrase: 17.1
Most frequent: fridge (265); vote (238); voting (164); like (148); hope (117)	Most frequent: vote (293); labour (249); know (163); want (154); don't (149)

BJ12 contains 38,083 words	JC12 contains 41,542 words
Lexical density: 0.047	Lexical density: 0.046
Average n° words per phrase: 14.9	Average n° words per phrase: 15.1
Most frequent: vote (359); voted (237); brexit (199); people (194); boris (172)	Most frequent : labour (602); vote (592); voted (374); just (293); i'm (197)

BJ13 contains 34,397 words	JC13 contains 34,293 words
Lexical density: 0.051	Lexical density: 0.056
Average n° words per phrase: 17.1	Average n° words per phrase: 16.6
Most frequent: borisjohnson (494); voted (176); hope (169); congratulations (162); brexit (156)	Most frequent: thank (351); labour (227); jeremy (219); hope (205); country (202)

Vote ranks first in saliences towards both accounts, however Iramuteq considers it a noun for Johnson and a verb for Corbyn, who in general has more forms including the imperative, which could imply confidence or at the least, implication. The hashtag #votetactfully does not appear since it is not lemmatized — nevertheless there is the presence of *vote labour* as well as VOTE which exemplifies the prosody of the shouted imperative. It suffices to look at these first levels of salience to understand that the two corpora deal with the same theme.

However, the data soon conveys that public discourse towards Boris Johnson departs from the lexicon of politics to focus on topics that are peripheral and im-

pertinent, and this is a response to the politician's own very fragmented rhetorical stratagem of 'deadcatting'[12]. The word *fridge* figures prominently, due to the fact that Johnson had sought refuge in a refrigerated room on live television. In contrast to the @JeremyCorbyn corpus, the top saliences are devoid of party reference (Tory/Conservative).

In all there are five forms related to the incentive to vote, and three that allude to it. Two forms require explanation: In the Johnson corpus, *FBPE*[13], which is a hashtag acronym for Follow Back Pro Europe; a hashtag seeking to unite remainers.

4.3 Conjunctions

Although there is co-occurrence of subordinating conjunctions + lowercase 1sg the results are not compelling. In both corpora, the use of coordinating conjunctions is higher the day after the vote than two days before. The full corpus 1sg minuscule shows @BorisJohnson receives more subordinating conjunctions than the JC account, even after disambiguation with those that can be easily placed at the opening of a sentence. It would seem however, that in this corpus there is an inverse correlation between the high number of non-standard 1sg towards JC and the low quantity of conjunctions present.

4.4 Positive and negative co-occurrences

The televised advertisement by the Conservatives parodying the film Love Actually, invalidated the word love as a query for the 'positive' word groups from day 1 before cleaning. The prominent co-occurrences for positive words on December 10, 2019 for Boris Johnson are: brexit (22); let's (22); make (22); united (22); idea (21). Two days before the election, the positive messages look ahead to a future of cooperation (united/let's), construction (make/idea), articulated around his election promise, that of finalizing Brexit for those who voted for it in the referendum.

In the prominent co-occurrences for negative words @JeremyCorbyn, *hope* ranks first for both positive (63) and negative (28) words. This is caused by the citizen takeover of political discourse, in this case the end-of-campaign message

12 https://dbpedia.org/page/Dead_cat_strategy

13 https://www.theguardian.com/media/2018/jan/17/fbpe-what-is-pro-eu-hashtag-spreading-across-social-media

'*vote for hope*'. Both corpora bring together so many slurs and insults, from the quaintly archaic to the shockingly post-modern, and often tailored to each participant. Specific to Johnson there is *toff* (16). One of the insults specific to Corbyn, *commie*/*communist* (9) alludes to his left-leaning rather than centrist socialist politics. The data from the full corpus non-standard 1sg shows a variety of personalised insults for both men.

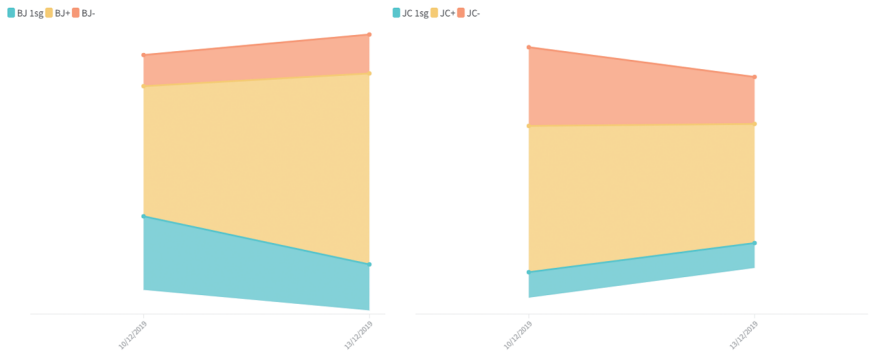

Figure 7: The positive and negative lexicon in juxtaposition to occurrences of non-standard 1sg, before and after Johnson's victory.

Positive lexicon towards BJ after his victory correlates with a decline in use of the non-standard form (Figure 7). On the whole, JC received a much higher volume of emotive words — whether positive or negative — than BJ; and since JC also received an inversely proportionate (to BJ) high number of non-standard 1sg, then there is a clear correlation between emotional language and the deployment of the non-standard form. The drop in use of the non-standard 1sg towards BJ after his election also supports the hypothesis of the miniscule being used as an attenuative device. Once one's champion's victory has been validated by the democratic process of a whole nation, there is less need for the addition of a caveat to a message of support in a hostile environment—one can shout one's alliance with less fear of reprimand.

4.5 Direct address

There are some tweets such as "i love u I HOPE you resign" that exist likely in response to "I LOVE YOU please don't resign" a phrase appearing multiple times in the initial extraction. At first glance, and if one pays no regard to orthographic variation, it is a phrase that makes no sense. However, given all that has been demonstrated thus far in this study, the lower-case 1sg offsets the positive input

of the word *love* to the extent that it carries an invisible negation (*I don't love you*). In this light it serves as example of the semantic-pragmatic functional division of the two 1sg forms, and this is perhaps the reason for the cautious algorithm; *don't* is classified as negative as well as *resign*, which parses as "don't do something bad". don't (-) + resign (-) = (+) Considering *love* as a positive word when it is not co-occurring with the word *actually* is reliable for this sub-corpus. The word *thank* occurs 351 times, and with a left/right concordance, we can see that it is affixed on the right by *you* and a subordinate clause of sincere gratitude — and so it was necessary to check for hyperbolic co-occurrences such as *amazing* (which could flag sarcasm) or the word *nothing*: "thanks for nothing". The top co-occurrences for negative words are: *day* (47); *hope* (33); *jeremycorbyn* (28); *think* (20). "So many of us are determined to keep the hope alive. Best wishes, Jeremy." vs "I hope we can stop you."

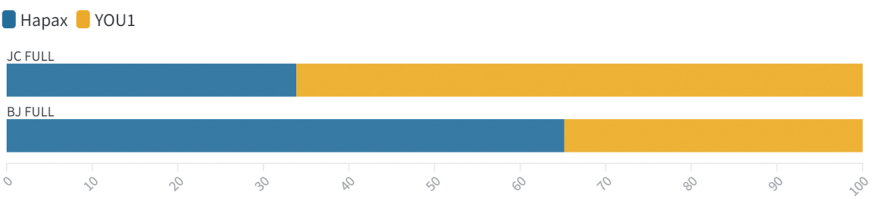

Figure 8: Proportion of hapax vs you1, from the non-standard 1sg full corpus

You in English does not differentiate between singular and plural. Here however since the corpus is composed of messages to a named user account, there is no doubt as to the exclusivity of the leadership. It is the man and not the political party he represents — and indeed when someone refers to it, the party in question is clearly cited, e.g. Conservatives/Tories/Labour.

Table 6: Standard and non-standard direct address (You1) with minuscule 1sg to both accounts.

	BJ	BJ	JC	JC
	You1	you1	You1	you1
10/12/19	11	53	0	71
11/12/19	0	6	0	72
12/12/19	18	34	0	61
13/12/19	40	46	46	148
	69	*129*	*46*	*352*

Exploring the sub-corpus comprising only the lowercase 1sg, and targeting the direct address word *you*, reveals a greater ratio of hapax to 1sg minuscule @BorisJohnson, which does coincide with the fragmentation seen in the similarity analysis since the greater the number of distinct topics (fragmentation of the semantic field), the greater the potential for hapax. The standard personal pronoun *I* was favoured in juxtaposition with direct address @BorisJohnson. This symmetry of self-reference in the standard form (I – You) could be a manifestation of accommodation, in the sense of 'standing up to someone of import'.

4.6 Hapax

Gadet (2006) speaks of verticality and horizontality of reading, and indeed a hapax may act as a vertical, more semantically loaded word, due to its unique nature in a corpus or indeed a simple phrase (Williams et al., 2015). The tweets addressed to @JeremyCorbyn are longer thus per Zipf's law it is unsurprising[14] that this corpus contains the most hapax. However, tweets to @BorisJohnson contain more hapaxes when in conjunction with the non-standard 1sg.

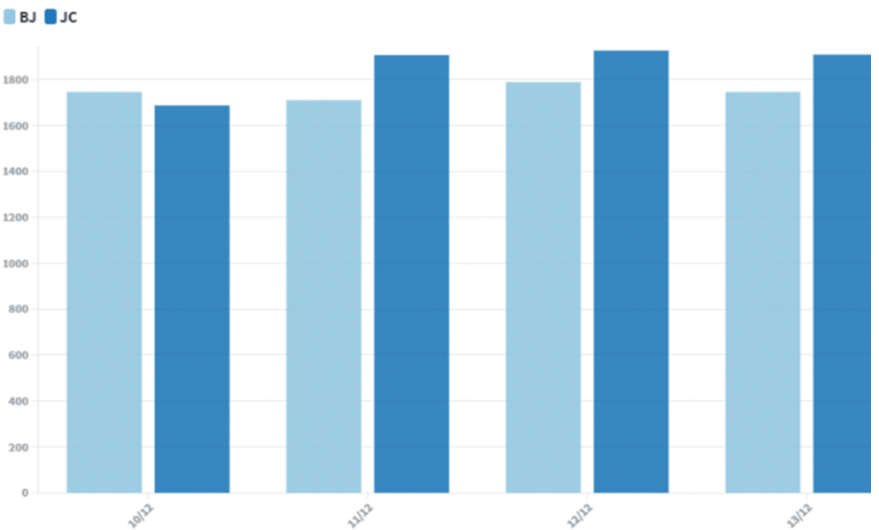

Figure 9: Hapax co-occurrence with the 1sg minuscule to BJ and JC

14 https://xlinux.nist.gov/dads/HTML/zipfslaw.html

4.7 Manual analysis

Context based accommodation can provide evidence of conscious implementation of the lower case 1sg, showing it is not just an unthoughtful usage for speed and concision.

Ten tweets were taken from each Johnson/Corbyn extraction of December 10 and 13, in total forty tweets containing the non-standard 1sg. The criterion for selecting these tweets was co-occurrence with love or hate lexicon, strong opinion, and insults. The authors' profiles were then searched in Twitter's native search engine, and each author's public page analysed to gauge whether their use of the non-standard form was their habitual orthography. It showed that the visibly young authors (assumed after identifying discursive environment: profile picture, biography, emoji use) used the minuscule indiscriminately, on their main public feed as well as in replies. Authors identified as older showed accommodation — using the standard form in their main public feeds. For example this person used the non-standard 1sg to address JC (cf. Table 3, line 830), however their profile page presented standardized orthography.

> I wrote about being a woman (because it's about the only identity card I have to play against the poor minorities I'm pretending to help).

Figure 10: Fully standardized orthography on one author's profile page showing evidence of accommodation.

5 Discussion

In the extractions over the four days there was a mirrored plot of the number of non-standard 1sg to both accounts, with the use of the non-standard form being clearly higher in number to JC, in turn inversely proportionate to BJ. The use of the standard 1sg shows a similar progression in both corpora, with a gradually descending frequency, but with a clear peak of intensified use on election day. In the heat of the moment personal opinion reigns, whether it be in a loud or small voice. In both sub-corpora the use of non-standard 1sg is prevalent on the day of, and just after the vote. This being a moment wherein emotions run high, the result supports the hypothesis that the non-standard form behaves like an emotionally inflected form used as attenuative protection in order to safely state one's mind or share one's opinion. There are similarities in first-order saliences in both sub-corpora due to the event-imposed lexicon, however there is more word count per sentence in JC and more hapax, which is a natural product of the probability of a longer phrase. A global and binary + and − sentiment analysis shows that for the

integral corpus, on the first and last days, JC receives more positive terms than BJ. On the applied sample, BJ receives more slurs, and more variants of *i* + be/have and the modals *will* and *would*. In both corpora, the use of coordinating conjunctions is much more frequent the day after the vote than two days before. The non-standard 1sg co-occurs more with subordinating conjunctions in BJ. JC receives more direct address (you1), and many more non-standard forms of the lowercase incipit "y" after the election. BJ receives more hapax in the non-standard 1sg corpus; and in both, the rate of hapax is inversely proportional to the rate of the non-standard opening "you". The number of words per sentence shows a similar pattern to that of the non-standard 1sg in JC, with a peak over 12 December. Word count increases progressively in BJ, thus JC has more words per sentence than BJ at first, and then the trend reverses. The data from the full corpus *i* shows a variety of personalised insults for both men. However, the full BJ corpus received a vast majority of slurs + 1sg minuscule.

The recorded results tend towards the conscious deployment of the non-standard 1sg in these corpora. In all cases the presence of the non-standard 1sg is strongly correlated (directly or inversely) to the phenomena of: a) Event paroxysm b) Slurs c) Frequency of hapax and d) Phrase length.

Event paroxysm is the most telling for research into the conscious implementation of the non-standard 1sg and its pragmatics, followed closely by slurs. Indeed the hypothesis of *i* being used for its attenuative effects is supported by these intense fluctuations — when the agora becomes excited, the less reckless deploy their typographic 'shields'. The unexpected distribution of the lower case 1sg in Figure 5 is eloquent, and one could say that it was as dangerous to show love and support to Jeremy Corbyn as it was to assail Boris Johnson with hate speech.

On the other hand, at this stage, the data is not yet convincing for the notion put forward in section 1.2, suggesting a syntactic relay of the prosody of *i* thanks to the predicates. Although there is co-occurrence of subordinate conjunctions + lowercase 1sg, and a similar correlation between the high number of non-standard 1sg towards JC and the low quantity of conjunctions present, more detailed analysis will be required. Perhaps the user knows intuitively, or has learned by practice, that the attenuating power of 1sg cannot maintain its effect on a complex sentence. It could also be that a person susceptible of addressing a politician and using non-standard typography to do so, is of a demographic that would get straight to the point and be done with oratorical precaution. The fact that this research was never intended to provide demographic analyses does have its limits. Indeed there are surely meaningful results to be uncovered by affording a whole paper to this single query of conjunctions, and using powerful methods such as geolocalized demographics, as used by researchers such as Jacob Eisenstein (2015). Another limit to this study is that during the analysis of author profile, some of the authors system-

atically favoured the lower-case form for opening characters in other @ replies. It would have been useful to quantify this particular form of accommodation, although such an endeavour was beyond the scope of this particular study. Also, for the moment is not yet possible to rule out that it is simply that user uptake is not homogenous. As Maybaum (2013) states, "the innovation-decision process (Rogers 1995), which recognizes that an individual's decision to adopt an innovation is not instantaneous, but progresses through five stages: 1) knowledge, 2) persuasion, 3) decision, 4) implementation, and 5) confirmation. At any point in the innovation-decision process, the individual may reevaluate his or her previous behavior and choose to adopt or reject the innovation."

6 Conclusion

In this contrastive corpus, textometry supports the hypothesis that the non-standard 1sg is used in places of danger, during critical moments, and to address people not otherwise addressed, perhaps using language that could see the author flagged, banned, or blocked by other users. The prosodic effect is the same as for the young sociolect of 'active nonchalance', but the pragmatics differ completely. This is not used as a pure indexical, rather it is a form that has emerged through evolutionary necessity to rectify the paucity of the English language that is both devoid of speech levels, and has only one 1sg (as opposed to, for example, Japanese containing eight) which — in an excitable medium such as Twitter that puts the Self up-front-and-centre — can be cleverly used to offset the possible reactions to personal statement whether it be love or hate, simple expression of opinion, use of slurs, or need for deniability. Although these results contribute to a better understanding of the pragmatics of self-representation and attenuation, this is only a first step in addressing the various implementations of the non-standard 1sg.

Just as "most prosodic studies also differentiate two different levels of happiness (elation vs. happiness) and sadness (despair vs. sadness) (Bänziger & Scherer 2005; Burkhardt & Sendlmeier 2000)", Heath (2018), there would seem to be differing utilizations of the non-standard 1sg.

One of these differing implementations I have called the 'petit-Moi' [small Ego], and is a novel variant of what James (2017) referred to as "an expression of the social semiotics of participant interlocution". In the case of the "Petit Moi", the concept of self (signified) as the signifier "i" possesses the referent "i". It is a representation of a 'lessened-Self' self that implicates the real-life physical self beyond parole. The corpora for these further studies are comprised of (anonymized) posts on social media platforms and on mental health support sites.

Finally it must be said that the applications of research in this domain are not trivial. Advances, for example, in defining the prosodies of these subtle phenomena serve to improve text-to-speech software used by people who do not always feel represented by the rather reductive artificial phonetics.

References

Androutsopoulos, Jannis. 2011. Language change and digital media: a review of conceptions and evidence. In Kristiansen, Tore & Coupland, Nikolas (eds.). *Standard languages and language standards in a changing Europe* (1) 145–159. Oslo: Novus Press. https://tinyurl.com/2w53hfu8 (accessed 26 August 2022)

Aru, Jaan, Francesca Siclari, William A. Phillips & Johan F. Storm. 2020. Apical drive — A cellular mechanism of dreaming? *Neuroscience & Biobehavioral Reviews* (119) 440–455. 10.1016/j.neubiorev.2020.09.018 (accessed 26 August 2022)

Augustine, of Hippo, Saint., & Warner, Rex. 1981[AD 397]. *The Confessions of St. Augustine*. New York: Penguin Group.

Bänziger, Tanja & Klaus R. Scherer. 2005. The Role of Intonation in Emotional Expressions. *Speech Communication*. 10.1016/j.specom.2005.02.016. (accessed 26 August 2022)

Baym, Nancy K. 2015. Social Media and the Struggle for Society. *Social Media + Society*. 10.1177/2056305115580477 (accessed 26 August 2022)

Borrell, André & Salsignac, Jeanne. 2002. Importance de la prosodie en didactique des langues [Importance of prosody in language didactics]. In Renard, Raymond (ed.) *Apprentissage d'une langue étrangère/seconde (2): La phonétique verbo-tonale*. [*Second/foreign language learning (2)* : *Verbo-tonal phonetics*] 163–182. Louvain-la-Neuve: De Boeck Supérieur. 10.3917/dbu.renar.2002.01.0163. (accessed 26 August 2022)

Burkhardt, Felix & Sendlmeier, Walter F. 2000. Verification of acoustical correlates of emotional speech using formant-synthesis. *SpeechEmotion*, 151–156. http://web4.cs.columbia.edu/~julia/courses/old/cs6998-02/burkhardt00.pdf (accessed 26 August 2022)

Chomsky, Noam & Halle, Morris. 1968. *The Sound Pattern of English*. New York: Harper & Row.

Cruttenden, Alan. 1997. *Intonation*. New York: Cambridge University Press.

Crystal, David. 2011. Internet Linguistics. 1st ed. Oxfordshire, UK: Taylor and Francis. https://www.perlego.com/book/1607289/internet-linguistics-pdf. (accessed 26 August 2022)

Dubois, Jean. 2002. *La linguistique textuelle* [Text linguistics] Paris: Armand Colin

Eisenstein, Jacob. 2015. Systematic patterning in phonologically motivated orthographic variation. *Journal of Sociolinguistics*, 19(2), 161–188. 10.1111/josl.12119 (accessed 26 August 2022)

Falk, Simone. 2014. Communicative functions of rhythm in spoken discourse – the case of radio broadcasting, *Cahiers de praxématique,* 61. http://journals.openedition.org/praxematique/1904 (accessed 26 August 2022)

Ferrara, Kathleen, Hans Brunner & Greg Whittemore. 1991. Interactive written discourse as an emergent register. *Written Communication* 8 (1) 8–34. doi:10.1177/0741088391008001002. (accessed 26 August 2022)

Fuchs, Susanne & Jelena Krivokapić. 2016. Prosodic boundaries in writing: Evidence from a keystroke analysis. *Frontiers in Psychology* (7)1678. doi:10.3389/fpsyg.2016.01678. (accessed 26 August 2022)

Gadet, Françoise. 2006. *La Variation sociale en français* [Social variation in French] Paris: Ophrys.
Heath, Maria. 2018. Orthography in Social Media: Pragmatic and Prosodic Interpretations of Caps Lock. Proceedings of the Linguistic Society of America. Linguistic Society of America. doi:10.3765/plsa.v3i1.4350. (accessed 26 August 2022)
Hogg, Richard. 1992. Phonology and Morphology. In Hogg, Richard (ed.). *The Cambridge History of the English Language*. Vol. 1. Cambridge, UK: Cambridge University Press.
Howell, Peter & Van Borsel, John. 2011. *Multilingual Aspects of Fluency Disorders*. Bristol: Multilingual Matters.
Jaffe, Alexandra. 2000. Introduction: Non-standard orthography and non-standard speech. *Journal of Sociolinguistics*, 4: 497–513. https://doi.org/10.1111/1467–9481.00127 (accessed 26 August 2022)
James, Allan. 2017. Prosody and paralanguage in speech and the social media: The vocal and graphic realisation of affective meaning. Linguistica, 57(1), 137–149. doi:10.4312/linguistica.57.1.137–149 (accessed 26 August 2022)
Leech, Geoffrey. 2014. *The pragmatics of politeness,* New York: Oxford University Press.
Loevenbruck, Hélène, Grandchamp, Romain, Rapin, Lucile, Nalborczyk, Ladislas & Dohen, Marion 2018. A cognitive neuroscience view of inner language: to predict and to hear, see, feel. In Langland-Hassan, Peter & Vicente, Agustín (eds.). *Inner Speech: New Voices*. Oxford: Oxford University Press.131–167.
MacMurray, Erin & Leenhardt, Marguerite. 2012. Textometry and Information Discovery: A New Approach to Mining Textual Data on the Web. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.217.9320&rep=rep1&type=pdf (accessed 26 August 2022)
Maybaum, Rebecca. 2013. Language Change as a Social Process: Diffusion Patterns of Lexical Innovations in Twitter. Annual Meeting of the Berkeley Linguistics Society. Linguistic Society of America. 10.3765/bls.v39i1.3877.
Mounin, George. 1974. *Dictionnaire de la linguistique* [*Dictionary of linguistics*] Paris: PUF.
Perrone-Bertolotti, Marcela, Grandchamp, Romain, Rapin, Lucile, Baciu, Romain & Lachaux, Jean-Philippe. 2016. Langage Intérieur [Inner speech]. In Pinto, Serge & Sato, Marc. (Eds.) *Traité neurolinguistique. Du cerveau au langage* [On neurolinguistics. From the brain to language]. 109–125. Louvain-La-Neuve: De Boeck Supérieur. https://hal.archives-ouvertes.fr/hal-01467107/file/Perrone_Bertolotti_etal_2016_LangageInterieur.pdf (accessed 26 August 2022)
Pincemin, Bénédicte. 2018. Sept logiciels de textometrie [Seven textometry software tools]. https://halshs.archives-ouvertes.fr/halshs-01843695/file/sept_logiciels_de_textometrie_180718a.pdf (accessed 26 August 2022)
Rodero, Emma & Potter, Robert F. 2017. Melodic variations to stimulate your attention*, Human Communication Research,* 43 (3) 397–413.
Saussure, Ferdinand de. 1985[1916]. *Cours de linguistique générale*, Paris : Payot.
Sidani, Jaime. 2016. The Association between Social Media Use and Eating Concerns among US Young Adults. *Journal of the Academy of Nutrition and Dietetics*, 116 (9) 1465–1472. 10.1016/j.jand.2016.03.021 (accessed 26 August 2022)
Smith, Naomi and Copland, Simon. 2021. Memetic Moments: The Speed of Twitter Memes. *Journal of Digital Social Research,* 4 (1) 23–48. 10.33621/jdsr.v4i1.95. (accessed 26 August 2022)
Werry, Christopher C. 1996. Linguistic and Interactional Features of Internet Relay Chat. *Pragmatics & Beyond New Series*. Amsterdam: John Benjamins Publishing Company. 10.1075/pbns.39.06wer. (accessed 26 August 2022)

Williams, Jake Ryland, Paul R. Lessard, Suma Desu, Eric Clark, James P. Bagrow, Christopher M. Danforth, & Peter Sheridan Dodds. 2014. Zipf's Law Holds for Phrases, Not Words. arXiv. 10.48550/ARXIV.1406.5181. (accessed 26 August 2022)

Wilson, Deirdre & Sperber, Dan. 2004. Relevance Theory. In *Handbuch Pragmatik*. Liedtke, Frank & Tuchen, Astrid (eds). Stuttgart: J.B. Metzler.10.1007/978-3-476-04624-6_8 (accessed 26 August 2022)

Zappavigna, Michele. 2012. *Discourse of Twitter and social media.* London: Bloomsbury.