Structural Onomatology for Username Generation: A Partial Account

Francico S. Marcondes   José João Almeida   Paulo Novais

ALGORITMI Centre — Department of Informatics, University of Minho, Portugal
francisco.marcondes@algoritmi.uminho.pt, jj@di.uminho.pt, pjon@di.uminho.pt

Abstract

The username hints for most of the on-line social networks are mostly unpleasant for human beings since they are mostly a simple name variation followed by numbers. This paper shows that it is possible to generate human likable usernames through heuristics guided by structural onomastics. The objective then is to conceive heuristics as such and check its availability in Twitter in order to verify if is it possible to generate a sufficiently big and available username data-set that is able to justify the transitions from unpleasant to a pleasant username suggestion. This paper finds that it is possible to generate 8281 handles on average through the proposed heuristics and their permutations, therefore, the number of various possibilities is comfortable. This is a partial account since not all possibilities were explored and some improvements are required, but suits for a proof of concept and to indicate paths.

1 Introduction

Username suggestions provided by account managers are not often captivating for the users, these, however, are selected by the users for avoiding wasting time searching for an appealing one. In short, username selection is seen more as a bureaucratic step for participating with a community than an identity element [5]. Nevertheless, when choosing between a not appealing username and an engaging one, the users will probably elect the second.

This paper research problem is to generate human likeable usernames in order to improve the quality, from the human perspective, of suggestions given by social media account managers.

It is not clear the reasons that further developments on username generators were not carried out. Even the account managers of huge enterprises like Google, Twitter and Facebook provide suggestions with very simple heuristics such as joining the first and last names with a number stream. On creating a new account, Twitter and Facebook does not even ask for their users, assigning automatically a name+numbers usernames to be eventually changed *a posteriori* at expense of losing social interaction links (see figure 1 for a reference).

One possibility is that this was never an actual issue for these companies nor a major problem for their users, then it did not worth to handle this out. Another possibility is to avoid the problem of unexpectedly generating an offensive suggestion. Nevertheless, usernames relates with the profile's popularity [8].

This *paper objective* is to present a set of usernames (*handle*) generation heuristics and verify their availability concentration on Twitter for generating an "appealing" ranking for providing username suggestions.

In: Rudolph S., Marreiros G. (eds.): Proceedings of STAIRS, Santiago de Compostela, Spain, august 2020, published at http://ceur-ws.org

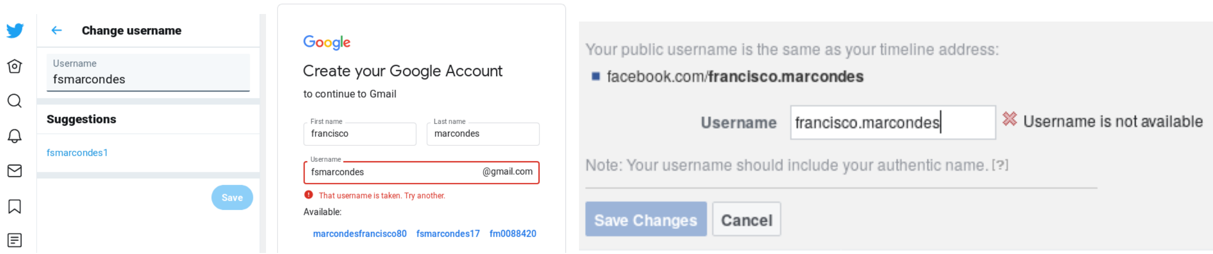


Figure 1: Instances of account managers username suggestions provided by Twitter, Goggle and Facebook.

A search on `ACM DL` did not returned any relevant literature. The query `(username AND generat*) (NOT (username OR generat*))` retrieved 7,049 results in January, 2020. By reading the titles and abstracts of the 100 most relevant papers (as classified by the `ACM DL`), only [4] draws near to this paper research problem. Nevertheless, it aims to generate pronounceable random words and not human appealing usernames. The forward snowballing over [4] also did not also retrieved any relevant literature. The literature profile correlates with account managers state of the art (as in figure 1) suggesting this subject as quite unexplored. Most of retrieved papers, such as [7], aims to extract information such as gender and language from nicknames or to meet a same person on several social media [11, 16].

The need for username generation is due to the fact that the most common usernames were already taken. The difficulty then is to generating human likable usernames that were not yet used.

Username in Twitter is called *handle* being also a Twitter `URI`. This means that it is possible to verify a Twitter handle by fetching the `URL https://twitter.com/`*`handle`* (returning the `404` error the username is considered available). This checking readiness is the main reason for focusing this paper on Twitter, yet, presumably, the same results can be generalized for other the social-media.

In addition, for this paper only American names were considered, therefore it cannot be expected to reach similar results outside of this scope. Whereas there could be universal heuristics, presumably most of them are related to a language and a culture, especially from a structural perspective. For instance, the use of diminutive for nickname formation is common for several cultures including Latin (*Chico* in Portuguese), Saxon (Franky in English) and Oriental (*Furan-kun* in Japanese), however, the different structure requires particular heuristics.

This paper does not also consider the structural difference between male and female nicknames. It is possible to guess that Franky stands for Frank and Frankie for Francine but for properly assess such structures a data-set like LDC2012T11 *cf.* [3] should be further explored considering the weighted relation between gender and nickname structure. Additionally, even the used data-sets providing popularity weight information, they are not considered in this paper. Finally, name order for composed nicknames or the prevalence of name/surname derived nicknames is also not considered.

2 Pseudonym, Nicknames and Usernames

Nicknames and usernames are pseudonym types distinct to each other [13] shaping its own onomastic category [1]. The username relates to the nickname in the sense that both share etymological motivations and they depart from each other as nicknames result from interaction whereas usernames are demanded to participate within a community. Also, like given names and "proper" pseudonyms, usernames are chosen but different from them it must be unique. In addition, usernames do not necessarily refer to a person but also to an idea or an account content *i.e.*, usernames are not necessarily an anthroponym. By not being necessarily anthroponym, the username research should focus more on structure than in semantics [5]. The structural approach suits usernames because they can be considered as linguistic exceptions [6] as they may never say aloud or be part of a syntactic context; they are not committed to grammar and orthography rules (including gender distinctions) [13].

A Twitter username[1] is case-insensitive alpha-numeric 4 to 15 char length[2] string in the form `username::=[a-z|0-9|_]`. With 300+ million active users [12] a major problem is to find a suitable unique

[1]https://help.twitter.com/en/managing-your-account/twitter-username-rules, fetched in Jan. 2020

[2]In the Twitter rules it is written that the handle is between 1-15, however it only accepts new usernames between 4 and 15. This is, probably, for avoiding *username squatting i.e.*, the act of selling social media accounts with associated earned value that had created a black-market for *rare handles*. For a reference, it was offered around $50,000 for the username `@N` in 2014 [10]; lesser rare handles, up to three letters, were traded by couple hundred dollars [2]. Currently, a Twitter account in the black-market vary from couple cents to dozens of dollars due to followers' number, SMS verified and account age.

username that is still available without resorting to numbers and non-name elements. Highlight that Twitter holds both username and nickname, this paper focus on the username. In addition, there is an important onomastic distinction between Google and Twitter/Facebook usernames as reveled in 1. For creating a Google account, the user must define its own username that is unchangeable afterward. Twitter and Facebook use a different strategy, they set a generic username through a simple heuristic and assign it to the new account; if the user wants, the username can be changed afterward (to the expense of losing all linked references [9]).

It must then to be verified the relation between nicknames and Twitter handles. For verifying this relation, given that Twitter's active users are around 300Mi in 2020 [12] a sample of 30k random nicknames was retrieved from LDC's Nicknames data-set [3] and checked against Twitter's user-base for availability. The results are that 1763 (5.87% of the sample) of plain nicknames are available in Twitter. This result suggests that nicknames are being widely used as handles in Twitter, therefore, using structural nickname formation strategies suits this paper intent on generating name-based usernames and strengthens the idea that nicknames and usernames are structurally related.

3 Structural Heuristics for Nickname Generation

The name formation uses two data sets, one first-names[3] another with last-names[4]. For a reference, by joining one first name with one last name $(first \times last)$ it is possible to generate $1,68923e10$ names. It is also common for occidental names to be formed by composed names and with two or three surnames, increasing the amount.

Certainly, there are "common names" that most of the generated usernames will be already taken. The idea is to verify if there are a name combination and nickname heuristics that are more likely to be available as a handle on Twitter. Therefore all names were considered within the same probability. The name builder has a function signature as `buildName(gender, compound:Boolean, surnames:Integer):Name`. The command `generateName(random.choice(['Male','Female']), random.choice([True,False]), random.randint(1,3))` was then used for building the sample.

Then, given a name, the username generation starts. As presented in table 1a there are several elements that can be used for composing a username. This paper focuses on generating usernames based on *personal names* as elements. When usernames are formed upon personal names, they share the same structure and rules of a nickname [13] getting into a form like <`nickname`><`name`> [5].

Structurally, a nickname is formed by *abbreviation*, *modification* or *name portions*, however, there are also nicknames without any clear formation rule. For the first case scenario, a set of heuristics based on structural onomastics can be used for generating nicknames. For the second case scenario, a data-set based approach must be used. Fortunately, this second scenario matches with contracted nicknames and then implements the *contraction* heuristic. A straightforward and convenient structural onomastic typology for nickname formation was found in [15] and adapted as depicted in table 2 guiding the heuristic development. A graphical description of the way that these heuristics relates to themselves are presented in figure 1b. For this paper it was possible for formulating suitable, human likeable, heuristics for *Separation*, *Portions*, *Initials*, *Contraction*, *Diminutive* and *Fancy* but not for *Swapping*, *Phonetic*, *Dropping* and *Combination*. The focus is given for the first group.

Any word generator may unexpectedly produce bad-words and this is not an exception. For an instance, names such as `Analee` and `Nazifa` may produce bad words by picking the first four letters. For handling this problem a blacklist[5] is used as a strategy, even not being the perfect solution is aids on avoiding, at least, the most scandalous situations. It must be stressed that username suggestions are supposed to be presented privately for each user, therefore, even names in the "gray area" may be presented being up to the human to choose it or not. This will vary according to each user's personality. Therefore, all generations are filtered for bad-words.

Finally, for limitations, the lack of papers that were found for this subject results in a preliminary set of heuristic rules that must be further developed until they can reach some maturity. The proposed heuristic for this paper is suitable yet not fully developed in the sense that every time a new adjustment emerges. For an instance, a rule picking the three first letters of name with the pattern consonant-vowel-vowel (CVV) suits for `GIO[vanni]` but not for `JOA[n]`, then, for this paper, this rule was dropped yet a "smarter" rule may be conceived on future works. Nevertheless, some heuristics overlaps filling some gaps of each other, for instance, the gap left by CVV rule dropped from the heuristic on the *portion* heuristics is mostly filled by the *separation*. Therefore, understanding how the presented heuristics interact with themselves is also important for this paper.

[3] 104,110 names from the US Social Security without duplicated names https://www.kaggle.com/kaggle/us-baby-names/version/2
[4] 162,254 surnames from the US 2010 https://data.world/uscensusbureau/frequently-occurring-surnames-from-the-census-2010
[5] 1704 English bad-words https://www.freewebheaders.com/full-list-of-bad-words-banned-by-google/

a)

Element Type	Characteristics	Instance Elem.	Example
Titles	Tendency to appear first, be followed by a personal name.	dr, ms, just, real	`justKrista`
Personal names	Tendency to appear first, be followed by a surname.	chris, mike	`chirsAdams`
Determiners	Tendency to appear first, be followed by a noun.	the, that, big	`bigJoe`
Person. pronoun	Tendency to appear first, be followed by a verb.	i, your, my	`iDrinkOJ`
Org. suffixes	Tendency to appear last, following an organisation name.	uk, news	`girlAtNY`
Circumflexes	Tendency to appear both first and last, adjacent to a name.	x, xo, _	`xOliviax`
Number	Tendency to appear last, following a personal name.	(all numbers)	`Johnson78`

b)

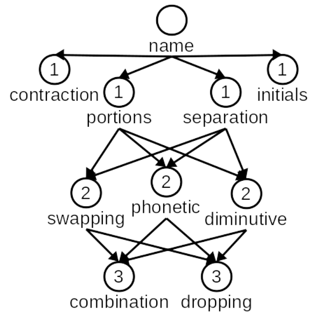


Table 1: a) Typology of common elements present in usernames *cf.* [5]. The highlighted row shows this paper's focus. b) A heuristic formulation for structural onomastic nickname formulation based on rules presented in table 2 (the number within each circle is the 'L.'). A nickname may stop at each level or may proceed to the next for encompassing more features. The *Fancy* rule-set as a special case was not included in this picture.

I.	L.	Heuristic	Example	Description
✓	1	Initials	ZS from Zachary Smith	The first letter of each name.
✓	1	Portions	Liz from Elizabeth	A nickname may come from the front, end or middle of name.
✓	1	Separation	Mary-Ann from Maryann	If a name is a composition of two other names then split.
✓	1	Contraction	Ike from Eisenhower	*Ad hoc* formation, usually due to socio-historical circumstances.
×	2	Swapping	Bill from Will	Swap letters for the first letter of a name portion.
×	2	Phonetic	Bob from Robert	Like swapping but based on the phonetic structure.
✓	2	Diminutive	Charlie from Charles	Include terminations such as -EE or -Y in a name portion.
×	3	Dropping	Fanny from Frances	Dropping such as R or H within consonant compounds.
×	3	Combination	Miz from Mary Elizabeth	A combination of the nicknames of a compound name.
✓	*	Fancy	Markus or from Mark	Creative possibilities, a general heuristic cannot be envisioned.

Table 2: Nickname formation rule-set adapted from [15]. 'I' shows the heuristics discussed in this paper. 'L' is the heuristic's transformation level, for an instance, a *diminutive* (level 2 heuristic) suits better to a name *portion* (level 1) than to a name (level 0), see table 1b for reference. The *Fancy* Rule Name is highly *ad hoc* being able to encompass several heuristic possibilities, therefore it may be placed on any level according to the defined heuristics and may involve some other typologies as presented in 1a.

Name Portions. Name portion-based nicknames can emerge from the front of a name, from its back and from its middle. Most of these formations can be reduced into a letter trinomial composed by vowels (V) and consonants (C), such as, for the name, *CHARLES* the front trinomial is CCV, the middle CVV, and the back CVC (for name generation purposes the letter 'Y' is considered a vowel). Eventually, the fourth and fifth letters must be also considered, such as on CCC formations, as for the front of *CHRISTINE*. For the middle-name portion, good heuristics could not be conceived for this paper. There were then created 32 rules except for those that did not suit after a brief "human likability" inspection (that were dropped). Follows a code snippet:

```
trinomial = name[:3]
if _isCVV(trinomial):
    candidates.append(name[:2])     # JO[an]
#     candidates.append(name[:3])     # JOA[n] <-- DROPPED
if _isVCC(trinomial):
    ...
    if len(name) > 3 and _isVowel(name[3]): candidates.append(name[:4])   # ALLI[sson]
```

Name Separation. It is common to a compound name to become a name such as Mary Ann becoming Maryann that can be easily re-splat when a nickname is emerging. Therefore, structurally, name separation is looking for inner-names within a name in the form `innernames.append([n for n in dataset if n in name])`. Name *separation* and *portions* are qualitatively distinct yet structurally similar (a separation is a portion of a name). On average, *separations* produces 2 ± 2 names that were not within *portions* against 7 ± 3 that are. This suggests that these heuristics have a high structural relation yet important qualitative differences. Nevertheless, in practice, these heuristic names can be used interchangeably.

Name Contraction. There are two possibilities for name contractions, one is a result from a socio-historical process such as *Greta* from Margaret, another one is a result of *portion, swapping and dropping* letters combination such as *Mike* from Michael. Therefore, for this paper, name *contraction* is used for denoting nicknames whose formation rule cannot be properly determined from a structural perspective. Therefore, for handling *contraction* a bag-of-words strategy will be used. Among a set of evaluated nickname data-sets (only five were found) the American English Nickname Collection (LDC2012T11) [3] presents the higher rate of non-obvious nicknames, being then chosen. Highlight that data-set is copyrighted and cannot be freely distributed. Also, it presents some compound nicknames such as `Johnny Boy` from *John* (being the space replaced with an underscore) and it may present quite uncommon nicknames such as `James` from *Monroe*.

For understanding the benefit of using that approach, the other heuristics proposed in this paper were used for generating nicknames and compared to the nicknames proposed by the data-set, `[i for i in h if i in c]` where `c` ∈ `Contraction` and `h` ∈ `Heuristic`). In short, the other heuristics do not over-set this one, meaning that it provides a set of non-obvious nicknames adding value when joined with the other heuristics for username formation. As a result, whereas *portion* and *separations* holds a high structural relation between each other, *contraction* does not.

Name Diminutive. The strategy is to include the terminals -Y, -IE, -EY, -EE, -IN and -KIN at the end of names, also, by doubling consonants that are not H or R and replacing C and Q with K (becoming -KY, -KIE, *etc.*). A caution to be taken is with names ending on I or Y as it would become something like -IY or -YEY (some of these terminations would suits as fancy name yet not for this one). For this heuristics, it was defined 40 rules. Being a level two ruleset, it suits better for name portions than to full-names. For instance, *Burnam* as **full-name** results in `Burnamy` and `Burnamey` whereas as **portions** becomes `Burny` and `Burney`.

```
if name[-1] == 'i' or name[-1] == 'y':
    candidates.append(name[:len(name)-1]+'kin')
    if name[-2] != 'y' and name[-2] != 'i':
        candidates.append(name[:len(name)-1]+'ee')
    ...
if _isConsonant(name[-2]) and name[-2] != name[-3] and name[-2] != 'r' and name[-2] != 'h':
        candidates.append(name[:len(name)-1]+name[-2]+name[-1])
```

Fancy Names. There is not a "correct" set of fancy variations for nickname formation, in short, creativity is limited by human likability. Therefore a strict research on fancy username generation involves getting people feedback (out of scope for this paper). For a proof of concept, three rule types is being used. The first is fancy characters, such as the use of "_" for creating nicknames like `_MARY` and `M_A_R_Y` and letter replacing such as '3' for 'E' and '4' for 'FOR', *e.g.* `R3DFORD` and `RED4D`. The second is "foreignnessization" by including letters such as 'H' and 'G' after the last name vowel for strength and 'US' and 'UM' for "latinization". The third is based on the repetition of name portions such as `JAJA` from *Janet* and `CICI` from *C[hr]ISTINE*. These rules are extremely *ad hoc*, for this paper, there were defined 26 of them, for a snippet:

```
candidates.append(''.join([c+'_' for c in name])[:-1])
if 'for' in name: candidates.append(name.replace('for', '4'))
...
if name[-1] in vowel:
    if name[-2] != 'u': candidates.append(name[:len(name)-1]+'us')
...
if _isVCV(trinomial): candidates.append((name[1:3]+name[1:3]).capitalize())
```

3.1 Nickname Evaluation

For internal evaluation, it was built a name sample with 168922 (0.00001% of the population). The idea is to verify how many nicknames, on average, each generator creates. The results are presented in table 3.

Considering a total of 91 nicknames and an availability rate on Twitter of 5.87% (see section 2), the estimation is that, for each name, it is possible to find around 5 nicknames. This is a quite narrow margin to work with. Then the topic must be further explored. The probability for a not available username varies according to the range of people that they encompass. For instance, a handle such as `F` encompasses, at least, all names starting with 'F' whereas `FSMarcondes` is quite less embracing. In short, as general a nickname is, less probable to be available due to its naming scope. For numbers, table 4 shows the availability rate in Twitter for the proposed heuristics. Ba aware that table data may be a little biased due to Twitter's policies changes over the time. For instance, it is not currently possible to set a handle less than 4 characters[2] yet these existing accounts may still be suspended and removed[1], therefore *initials* availability may raised.

a)

Nickname Generator	Sample				Avail.
	AVG	±	Max	Min	
Name	3	1	5	2	0.47%
Initials	3	1	5	2	3.85%
Portions	18	5	31	2	4.23%
Separation	9	3	26	0	1.25%
Contraction	4	5	40	0	5.87%
Diminutive	24	10	63	0	26.89%
Fancy	39	13	81	6	59.27%
SUB-TOTAL 1 (raw)	101	29	197	24	-
SUB-TOTAL 2 (no rep.)	92	26	177	23	-
TOTAL (no bad words)	91	26	177	23	-

b)

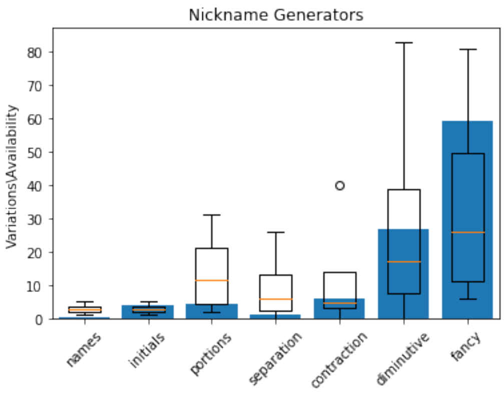


Table 3: a) Descriptive statistics for the proposed nickname generators. The *initials* depicts the shape of built names. For a proper benchmark, all strategies received the same name data-set (even higher level heuristics such as *diminutive*). SUB-TOTAL 1 is the summation of all generated nicknames and SUB-TOTAL 2 excludes repetitions form the summation and TOTAL filters for bad-words. It is important to highlight that bad-word filtering reduced one nickname on *average* and then is significant. For the availability, for each heuristic, a non-repeating random sample of 30k nicknames was gathered and checked against Twitter's for availability. b) The nickname generators data plotted.

Another availability variable is the *appeal*, *i.e.*, how a likable a nickname is. Table 4 reveals that the most appealing username is the *proper name* itself, followed by *separations* that is also mostly *proper names*. These is followed by *initials*, *portions* and *contractions*. Curiously, *diminutive* is not a high appealing username yet is a highly appealing nickname, supposedly, it is not expected to a person to call himself with a diminutive; *fancy* result on the other hand was expected, presumably, due to its *ad-hoc*ness. Therefore, on the other hand, as more specific is a nickname more likely to be available.

4 Username Composition and Suggestion Ranking

A common feature on usernames is to compound a nickname with a name (*e.g.* `BillGates`) or with another nickname (*e.g.* `JLo`). This brings products into handle generation, therefore, it is possible, on average, to produce 8281 (91×91 from table 3) username suggestions for each name; ranging from 529 to 31329 variations. For assessing the availability for compositions as such, the products were generated for each pair of heuristic resulting into sets composed by elements formed as $< h_1 >< h_2 >$ and $< h_2 >< h_1 >$, where h_1 and h_2 are two heuristics. From each set, a subset with 30k (around (0.00001% of Twitters active users *cf.* [12]) of non-repeating nicknames within a length of $4 \leq x \leq 15$ were randomly selected and checked over Twitter in March 2020; the results are presented in table 4a.

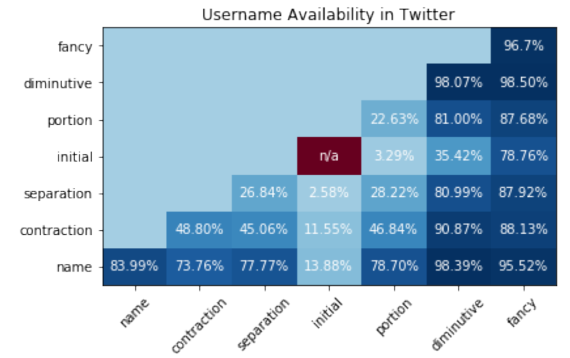
Table 4a shows that usernames with *initials* tends to be more appealing, supporting the idea that smaller handles are preferred to longer ones, except, as shown in table 4c, that the longer handle is a personal name. This means that `EDijk` is a more appealing username than `EdsgerDijkstra`, yet it is as appealing as `Edsger` or `Dijkstra`. The *diminutive* and *fancy* heuristics followed the pattern raised in table 3.

As for appealing, the difference between *name* and *name-name* availability, may be explained by considering that repetition for the former requires a two-name homonym. Also, it must be considered the difference between "common" and "rare" names and sociological issues in creating names such as marriage, *e.g.* it is more likely to exist a German-German name than to a German-Japanese, therefore names such `BernonKoyama` tends to be less common whereas splitting them, they tend to be equally common. In short, since this paper sample is artificial, it may be "sociologically biased" and further studies should be performed with an actual name sample for refining the *name-name* availability.

For a suggestion ranking, the results gathered in tables 3 and 4a can be normalized and used as an *appeal index* (depicted in 4c). According to the proposed index, as higher is the appeal of a generator, higher is the human likability potential, yet, also, that the "best" usernames are probably already be taken. Therefore a trade-off must be considered for suggesting. For this paper, the selected compositions are *Por-Por* (#8, 22.63%), *Sep-Sep* (#10, 28.84%), *Sep-Por* (#11, 28.22%) and Dim-Ini (#13, 35.42%); see suggestion instances in table 4b.

For some highlights, the suggestions is a starting point for attaining a valuable handle *e.g.* the `DeGeGene`

a)



b)

Name	Oprah Gail Winfrey
Suggestions (109)	`OprPrah, -, WinOpra, WOprey`
Name	Ellen Lee DeGeneres
Suggestions (91)	`ElleDeGe, ElleGener, DeGeGene, DeGGyE`
Name	Willard Carroll Smith
Suggestions (224)	`MithWill, CarrolWilla, OllWill, SmiteyW`
Name	Angelina Jolie Voight
Suggestions (58)	`IeJoli, JoliAngeli, AnVo, VJollee`
Name	Francisco Supino Marcondes
Suggestions (272)	`SupPino, MarconSup, CoFrancisc, FSuppey`

#01 Sep-Ini	0.00706888	#02 Name	0.0071633	#03 Ini-Por	0.00901419	#04 Separation	0.01905134
#05 Con-Ini	0.03164555	#06 Name-Ini	0.03802946	#07 Initials	0.05867811	#08 Por-Por	0.06200336
#09 Portion	0.06446972	#10 Sep-Sep	0.07353824	#11 Sep-Por	0.07731926	#12 Contraction	0.08946507
#13 Ini-Dim	0.09704636	#14 Con-Sep	0.12345875	#15 Con-Por	0.12833573	#16 Con-Con	0.13370589
#17 Name-Con	0.20209316	#18 Name-Sep	0.21308006	#19 Name-Por	0.21562814	#20 Ini-Fancy	0.21579253
#21 Sep-Dim	0.22190245	#22 Por-Dim	0.22192985	#23 Name-Name	0.23012208	#24 Por-Fancy	0.24023221
#25 Sep-Fancy	0.24088978	#26 Con-Fancy	0.24146516	#27 Con-Dim	0.24897241	#28 Name-Fancy	0.26171283
#29 Fancy-Fancy	0.26494588	#30 Dim-Dim	0.26869951	#31 Name-Dim	0.26957627	#32 Dim-Fancy	0.26987766
#33 Diminutive	0.40983232	#34 Fancy	0.90333811	#35 Ini-Ini	n/a		

Table 4: a) Twitter availability for generated nicknames (sample of 30k for each composition). b) Selected instances of username suggestions from *Por-Por*, *Sep-Sep*, *Sep-Por* and Dim-Ini composition (one for each). The values within parentheses are the total number of suggestions retrieved by the selected heuristics. c) *Appeal Index* ($\bar{\mathrm{A}}$) — ordered availability data presented in tables 3 and 4a normalized.

may become the `DeGGene` (also available in Twitter) that perhaps pleases that person. Eventually, appears "finished" likable handles such as `MithWill` and `WOprey`. Among the suggestions there are both "fun" and "serios" suggestions, respectively, `IeJoli` and `AnVo`. It was also noticed that bigger names compositions are more like to present bad suggestions due to the 15 character length restriction. Finally, for a personal account, all the presented suggestions for the last instance name are quite acceptable, they are not dream usernames but better than those presented in figure 1. This suggests by one side that this paper has succeeded in presenting a proof-of-concept (TRL-3 [14]) that there is room for explore before recurring to number streams. By another side that the proposed heuristics can be improved in future works.

5 Conclusion

This paper has shown that it is possible to generate a comfortable diversity of human likable nicknames to be explored before recurring to number streams as it is being done so far. Also, structural onomastics suits for guiding the generation heuristics for username creation. For this paper, there was proposed a handle suggestion system that (1) generates nicknames based on the products of six onomastic categories, and (2) ranked them through the *appeal index* ($\bar{\mathrm{A}}$) (after checking its availability in Twitter).

For future works, it is suggested to expand, improve and refine the proposed heuristics for generating more and better human likable usernames suggestions. That may include the middle portion of a name, phonetic coincidences and name combinations and may consider other properties such as gender, nationality, *etc.* Also, there can be proposed ontology relating name, nicknames and usernames through several parameters. For instance, `Franky` is a male English diminutive for `Frank` that relates to `Chico` in Portuguese which in turn is a diminutive for Francisco. Finally, submit the usernames for human inspection, without neglecting cultural differences, for improving the heuristics and the appeal index.

Some questions that remains open are the appealing difference between username patters, e.g. is <first-name><last-name> more appealing than <compound-name><middle-name> or <last-name><first-name>? Is there an important bias comparing automatic generated names, as used in this paper, and people actual names? Is it possible to generate structurally odd nicknames by feeding a neural network?

Acknowledgements

This work has been supported by FCT Fundação para a Ciência e Tecnologia within the RD Units Project Scope: UIDB/00319/2020

References

[1] Katarzyna Aleksiejuk, 'Internet names as an anthroponomastic category', *Annex Secció*, **3**, 243–255, (2014).

[2] Mattha Busby, 'You can buy anything on the black market: including twitter handles', *The Guardian*, (Apr. 2018).

[3] Vitor R Carvalho, Yigit Kiran, and Andrew Borthwick, 'The intelius nickname collection: quantitative analyses from billions of public records', in *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 607–610. Association for Computational Linguistics, (2012).

[4] Heather Crawford and John Aycock, 'Kwyjibo: Automatic domain name generation', *Softw. Pract. Exper.*, **38**(14), 15611567, (November 2008).

[5] Eleanor Crocker, *Whats in a Username?: An exploration of wordlist-based methodologies for the structural analysis of online usernames*, Master's thesis, University of Edinburgh, Scotland, UK, 2018.

[6] Vivian De Klerk and Barbara Bosch, 'The sound patterns of english nicknames', *Language Sciences*, **19**(4), 289–301, (1997).

[7] Aaron Jaech and Mari Ostendorf, 'What your username says about you', in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 2032–2037, Lisbon, Portugal, (September 2015). Association for Computational Linguistics.

[8] Paridhi Jain and Ponnurangam Kumaraguru, '@ i to@ me: An anatomy of username changing behavior on twitter', *arXiv preprint arXiv:1405.6539*, (2014).

[9] Paridhi Jain and Ponnurangam Kumaraguru, 'On the dynamics of username changing behavior on twitter', in *Proceedings of the 3rd IKDD Conference on Data Science, 2016*, CODS 16, New York, NY, USA, (2016). Association for Computing Machinery.

[10] Selena Larson, 'The not-so-secret black market of twitter handles', *ReadWrite*, (Jan 2014).

[11] Yongjun Li, You Peng, Zhen Zhang, Hongzhi Yin, and Quanqing Xu, 'Matching user accounts across social networks based on username and display name', *World Wide Web*, **22**(3), 1075–1097, (2019).

[12] Ying Lin, '10 twitter statistics every marketer should know in 2020', *Oberlo*, (Nov 2019).

[13] Jako Olivier, 'Twitter usernames: Exploring the nature of online south african nicknames', *Nomina Africana*, **28**(2), 51–74, (2014).

[14] ESA TEC-SHS. Technology readiness levels handbook for space applications, 2008.

[15] Wikipedia contributors. Nickname — Wikipedia, the free encyclopedia, 2020. [Online; accessed 12-February-2020].

[16] Reza Zafarani and Huan Liu, 'Connecting corresponding identities across communities', in *Third International AAAI Conference on Weblogs and Social Media*, (2009).