An expanded demographically informed analysis of Toontown Rewritten chat message data

E. Ciereszynski

January 18, 2024

Abstract

This paper presents a study which expands upon previous demographically informed linguistic research performed on chat message data gathered from the online game Toontown Rewritten. Prior research in this area has not been informative and it has been hypothesized that this is primarily due to a lack of data. The study's research questions ask if findings will become more extensive and informative if corpus size is significantly increased and a broader set of linguistic metrics is examined. In order to answer these questions, a new, larger demographic message corpus was compiled, merged with the original demographic corpus, and subsequently analyzed through the calculation of population and linguistic metrics and statistical testing. Outcomes were extremely similar to those yielded by prior research with only a small handful of novel significant results predominantly related to differences in message length and sentiment between genders. It is hypothesized once again that a lack of data remains a chief limitation, but future work is not currently planned in this domain due to multiple obstacles to future data collection and the marked uniformity of results across multiple studies.

Keywords: Toontown, Toontown Rewritten, Disney, NLP, natural language processing, demographics, sociolinguistics

Contents

List of Tables

List of Figures

1 Introduction

1.1 Background

This study constitutes the next step in demographic and linguistic research related to the online game Toontown Rewritten [Toontown Rewritten team, 2014]. A fair amount of demographic and linguistic Toontown research has previously been conducted, the most relevant studies being a major demographic exploration published in January 2022 [Ciereszynski, 2022a], a solely NLP-related investigation published in April 2022 [Ciereszynski, 2022b], and a demographic NLP study published in July 2023 [Ciereszynski, 2023]. These studies utilized statistical and correlative methods and reported on demographic and linguistic metrics. Few notable or significant results have been yielded by previous work, and the present study is thus important and relevant due to these major gaps. One of the primary conclusions of the demographic NLP study was that the demographic corpus was far too small and that this dearth of data brought about the study's lack of interesting results, and it was hypothesized that compiling a largely expanded corpus would be necessary for further exploration to be productive.

It has also been hypothesized that the treatment of too many demographic characteristics and not enough linguistic metrics has contributed to the existence of this gap in findings. The present research and its methodology have been designed with two key goals in mind: first, to mitigate these factors and therefore make meaningful contributions to the sphere of Toontown-related research, providing some sort of baseline findings as none currently exist, as well as possibly to the larger sphere of natural language processing and its intersection with demographic research, and second, to attenuate the practical issue of insufficient data. The study's research questions ask if findings will change or become more informative with a largely expanded corpus and the examination of a more extensive set of linguistic metrics, and this will be addressed through the calculation of these metrics and the performance of statistical analysis.

1.2 What is Toontown Rewritten?

Toontown Rewritten is an MMORPG launched in 2014 [Toontown Rewritten team, 2014] as a fanmade reincarnation of Disney's Toontown Online after the original game closed its digital doors in 2013 [Disney, 2003]. In Toontown, the player creates a character known as a Toon and completes tasks in order to advance through the game. Toons are highly customizable and the player is able to choose from many options for species, colours, clothing items, and weapons, known in the game as Gags. It is possible to communicate in the game through text chat, which relies on a whitelist of words, numerals, and symbols. I have been involved with Toontown across its various forms since 2006.

1.3 Objectives

The primary objective of this iteration of research is to discover relationships or phenomena which could not previously be identified in demographically informed linguistic analyses due to limited data. Research in the Toontown sphere remains largely exploratory due to the paucity of comprehensive or consistent significant results gathered from previous work. The size of the demographic corpus which is utilized in this study has been greatly expanded from that of previous corpora in hopes of accomplishing this objective.

1.4 Research questions and hypotheses

In order to systematically investigate the new demographic corpus, various research questions and hypotheses were formulated, although they are broad due to the aforementioned nature of the Toontown research landscape. I ask if there will be significant changes to population demographics and linguistic characteristics when a corpus markedly larger than previous corpora is analyzed, if expanding the set of linguistic metrics to be investigated will produce consequential or notable results which were not previously able to be uncovered, and if subgroups of the population who exhibit divergent or atypical linguistic behaviours will become apparent during analysis. In the large demographic study published in January 2022, various bundles of co-occurring demographic characteristics which seemed to be perpetuating various countercultural trends were identified [Ciereszynski, 2022a], and it is hypothesized that Toons possessing some of those characteristics may display linguistic behaviour which differs from that of the overall population, despite the fact that some of the demographic variables utilized in that study are no longer being considered in the current research.[1] Examining these research questions in detail is crucial for the advancement of Toontown-centric research and may also hold implications for analysis at the intersection of demographic research and corpus linguistics, particularly in online contexts.

2 Literature review

It is paramount to acknowledge and keep at the forefront of the mind the fact that the individuals whose demographic and linguistic characteristics are being examined in the present study are virtual characters in an online world and that their characteristics frequently do not map or cannot be mapped onto the human individual who created them and pilots them in-game. A body of research related to this very particular context does not appear to currently exist. This, however, permits an interesting avenue of inquiry to arise: will the linguistic behaviours of virtual characters exhibiting a range of demographic characteristics, some of which can be directly mapped onto humans, display some sort of correspondence with findings from preexisting sociolinguistic research on human subjects in an online context? Of the four demographic variables treated in this study, only gender can undergo this mapping, as the remaining three are unique to Toontown's universe and have no close human analogue. Fortunately, gender is one of the most frequently examined demographic variables in sociolinguistic research and extensive work has been conducted in this area, including in online spaces.

Previous research findings on the relationships between gender and lexical variation in the social media sphere have been somewhat varied. Some previous research has found that women are significantly more inclined than men to use abbreviations and contracted forms online and have a broader lexical register [Schler et al., 2006, Bamman et al., 2012, Ling, 2005a], but other research has discovered the opposite in terms of contracted forms [Baron, 2004], while other investigations have uncovered only minute differences in the kinds of language that men and women use on social media [Palmer, 2012]. In self-reported survey data collected for research conducted during my undergraduate studies, nonstandard linguistic forms were preferred nearly equally across genders [Ciereszynski, 2018]. Within certain sociolinguistic contexts, women are frequently identified as the primary users of nonstandard forms [Labov, 2001, Trudgill, 1972]. Labov and Trudgill both observe men's inclination toward traditionally standard forms in situations where lexical norms are unclear, while women tend to favour nonstandard variations when facing lexical ambiguity and when nonstandard forms carry a sense of covert prestige. In his landmark 1972 study, Trudgill highlights that covert prestige reflects "the different sub-cultures [sic] within [this] society" [Trudgill, 1972, p. 194]. Women tend to prefer nonstandard variations when they have covert prestige despite being "considered incorrect" [Leith, 1997] and when norms are not overtly prescribed, and to prefer standard forms when prestige is overt [Labov, 2001, Trudgill, 1972]. Considering the social media space and, more broadly, digital and virtual universes as subcultures within modern society, the concept of covert prestige becomes highly relevant to the linguistic dynamics within these particular circumstances. Given that overt norms and prestige for lexical variants on social media are often not explicitly prescribed, especially due to the constant flux in the dynamics of many social media spaces [Labov, 2001], it can be inferred that women may exhibit a greater tendency than men toward nonstandard forms, including acronyms and abbreviations, which may or may not carry covert prestige.

Notable gender-related results have also emerged in other linguistic areas. Women tend to send longer messages than men both in terms of word count [Rosenfeld et al., 2016, Ling, 2005b] and character count [Ling, 2005a]. Women may also send more messages than men [Rosenfeld et al., 2016], send messages which are more lexically dense [Rafi, 2008], and show significantly more positive emotion across diverse contexts and topics [Sun et al., 2020]. It will be investigated in the present study if any of these findings are relevant or true for the demographic corpus despite its unique characteristics and population.

Previous Toontown-related demographic and linguistic research has unfortunately yielded few interesting or useful results. A comprehensive demographic research paper published in January 2022 did reveal large bundles of demographic characteristics which were statistically more likely to co-occur and that certain groups, such as male Toons and Toons under 110 Laff, may be perpetuating various trends [Ciereszynski, 2022a], but a sizable portion of those findings are not applicable to the present research design because

[1] Refer to Section 3.2 for a detailed explanation underlying changes over time in Toontown demographic research design.

many of the demographic variables analyzed in that study are no longer being treated. An explanation of these changes and the rationale underlying them is provided in the methodology section below. A preliminary demographically informed linguistic analysis of Toontown chat message data published in July 2023 did not return noteworthy results [Ciereszynski, 2023], and it was hypothesized that this was due to the small size of the corpus. This motivated the undertaking of the present analysis utilizing a largely expanded corpus. Specific results and data from previous analyses are referred to and compared with new results and data in this paper.

3 Methodology

3.1 Data collection

The corpus utilized to carry out this analysis consists of the original demographic corpus, which includes 4,000 chat messages compiled by hand between August 2022 and January 2023, merged with a new demographic corpus consisting of 6,000 messages compiled between August and October 2023. Demographic corpora contain chat messages and various pieces of demographic information related to the speaker of the message. Two of the most glaring limitations of previous demographic NLP research in the Toontown sphere are a lack of data and corpora which are quite small by the standards of the broader field of corpus analysis. These disadvantages are interconnected. The small sizes stem from the fact that chat messages and the corresponding demographic details of the speaker must be observed in-game and typed out by hand, which is a laborious and unpredictable task. An acute awareness of these factors and their detrimental impact on productive research was the primary motivation underlying the creation of this extended corpus of 10,000 chat messages.

3.2 Demographic variables

Four demographic variables were treated in this study. When a chat message was recorded, the species, gender, missing Gag track, and Laff of the speaker were recorded along with the message. Demographic data was subject to cleaning in Python prior to analysis, which consisted of locating and resolving typos which had been inserted into the corpus during manual data collection. The `pandas` and `numpy` packages were used to manipulate data.

There are 11 animal species from which a player can pick when creating their Toon, and Toons can be male or female.[2] Laff is a quantitative metric by which a Toon's health is measured. Toons enter the game with 15 Laff points and can increase their maximum Laff up to 140 through a combination of storyline tasks and additional side activities. As Toons progress through these tasks, they gain additional Gag tracks, which are the game's weapons. It is possible to possess six of the seven Gag tracks. When a Toon acquires their sixth and final track, they progress into the game's final storyline playground. A Toon's missing track is the seventh track which they did not choose. Toons who had not yet progressed to the final playground and thus did not have a singular missing track were recorded as having no missing track. Figure 1 provides a visual representation of the seven Gag tracks. The Toon below is missing lure.

Figure 1: Toontown Rewritten Gag panel

Some previous demographic Toontown work has included additional demographic characteristics, such as a Toon's name tag and organic Gag track. The primary motivation for eliminating such characteristics as variables is that they are mutable. It is very easy to change these characteristics of a Toon at any point in the game. It is not methodologically sound to conduct demographically-focused research in which

[2] Originally, a Toon's gender could be easily determined through two primary visual characteristics: male Toons do not have visible eyelashes while female Toons have long eyelashes, and there are many clothing items and styles which are gender-locked. However, as of October 14, 2023, both of these characteristics have been done away with. In update 3.10.3, eyelashes became a toggled characteristic and all gender locks on clothing were abolished. The expanded corpus was completed prior to the implementation of these changes. In light of this major update, it will be difficult to utilize gender as a demographic variable in similar studies in the future. I am not yet certain how this will be navigated.

individuals can modify some of their relevant demographic characteristics at will over a longer timescale if specific individuals are not being tracked or monitored over this timescale. This can compromise the study's internal validity. Laff was retained as a demographic variable despite its mutability because it has been robustly demonstrated in previous work to possess strong predictive power.

3.3 Linguistic metrics

Seven linguistic metrics were calculated in this analysis: message length in words, message length in characters, word length, usage of first-person singular pronouns, type-token ratio, sentiment, and subjectivity. Token frequencies, both across the entire corpus and across demographic groups, were also examined. Message length in words and characters, sentiment, and subjectivity have been analyzed in previous research, but word length, first-person pronoun usage, and type-token ratio have never been investigated. VADER, a sentiment analysis tool specifically attuned to text data from online contexts, was utilized to calculate sentiment scores. The `TextBlob` library was utilized to calculate subjectivity.

It had been my intention to employ named entity recognition in this research, but unforeseen major issues in the output of the `spaCy` library's entity recognizer during analysis rendered this impossible. This situation will be elaborated upon in the discussion section.

Chat message data was subject to preprocessing in Python prior to analysis. This consisted of removing capitalization, punctuation, and superfluous whitespace from all messages in the corpus. Each message was subsequently split into individual word tokens. Python's `string` package was utilized to accomplish much of this text preprocessing.

3.4 Statistical methods

Statistical analysis was carried out using one-sample and two-sample t-tests. These tests were effectuated in Python using the `SciPy` library. Tests were not performed for every single permutation of demographic variable levels due to high computational intensity as well as very little variation in results across the large majority of population subgroups. Detailed visualizations were constructed for each pair of demographic variables and areas to investigate further by way of statistical testing were selected based on these plots.

The level of significance α was maintained at 0.005 throughout this study. There are multiple reasons for selecting such a low value. It functions as a makeshift Bonferroni correction, which is judicious because multiple tests are being performed, thereby increasing the chance of making a type I error. In addition, many of the subgroups of the population which underwent testing are extremely small in size, and lowering α helps to reduce the risk of calculating a false positive result when working with very little data.

4 Results

4.1 Demographic results

It is important to be aware that all figures below refer to amounts of chat messages across various groups as opposed to actual sizes of these demographic groups. For example, $n = 3663$ in the table below indicates that 3,663 of the 10,000 chat messages collected were uttered by cats, not that 3,663 individual cats were observed. The population size would be much smaller than 10,000 if this were the case because many individuals had more than one of their chat messages recorded.

4.1.1 Species

Demographic results are nearly identical to previous findings. In the first two demographic studies which have been conducted in Toontown Rewritten, female Toons or their messages were slightly more prevalent, but in the present study, the genders have nearly reached a perfect equilibrium. This may be due to the increased size of the corpus.

Species	n
Bear	458
Cat	3663
Crocodile	474
Deer	842
Dog	1554
Duck	758
Horse	176
Monkey	280
Mouse	993
Pig	175
Rabbit	627

Table 1: Species demographics

4.1.2 Gender

Gender	n
Male	4957
Female	5043

Table 2: Gender demographics

4.1.3 Missing Gag track

Missing track	n
None	2465
Toon-up	731
Trap	3198
Lure	538
Sound	165
Drop	2903

Table 3: Missing track demographics

4.1.4 Laff

Laff distribution, visualized in Figure 2, displays a striking similarity to that observed in previous investigations, with small peaks around 15, 50 to 55, and 60 to 65 Laff, a larger jump between approximately 100 and 110, and a sudden sharp increase at 140, the maximum amount of Laff that a Toon can achieve. The average Laff value is 94, whereas it was 92 in the most recent demographic NLP analysis.

4.2 Linguistic results

4.2.1 Message length

Message length was measured in words and characters. The average length of a message in the corpus was 4.26 words and 19.37 characters. These averages are very similar to previous findings in this area. The first non-demographic corpus analysis yielded a mean message length of 3.70 words and 16.55 characters and the recent demographic NLP research executed with the original, non-expanded version of the corpus yielded a mean message length of 4.28 words and 19.57 characters.

Species	$\bar{x}_w$	$\bar{x}_c$
Bear	4.38	19.95
Cat	4.18	18.98
Crocodile	4.46	20.46
Deer	4.23	19.42
Dog	4.37	20.18
Duck	4.31	19.56
Horse	3.93	17.94
Monkey	4.10	17.94
Mouse	4.31	19.18
Pig	4.19	18.56
Rabbit	4.21	19.29

Table 4: Message length across species

Gender	$\bar{x}_w$	$\bar{x}_c$
Male	4.34	19.81
Female	4.18	18.94

Table 5: Message length across gender

Missing track	$\bar{x}_w$	$\bar{x}_c$
None	4.26	19.86
Toon-up	4.21	19.03
Trap	4.27	19.35
Lure	4.18	18.88
Sound	4.46	20.13
Drop	4.25	19.11

Table 6: Message length across missing track

A stabilization effect appears as the number of messages increases. There is less variability and the mean message length remains closer to that of the corpus overall as the density of observations, messages in the context of this study, increases. Figures 3 and 4 closely mimic each other in this sense. Values of Laff for which there are larger amounts of messages tend to fall close to the population mean in terms of message length.

A handful of significant statistical results were achieved in this area. Below 100 laff, male Toons ($\bar{x}_c$ = 19.81) have a higher average message length in characters than female Toons ($\bar{x}_c$ = 18.94) ($t(5803) = 3.01$, $p = 0.003$), and male Toons without Toon-up have a higher average message length in both words and characters ($\bar{x}_w$ = 4.46, $\bar{x}_c$ = 20.12) than female Toons without Toon-up ($\bar{x}_w$ = 3.73, $\bar{x}_c$ = 16.92) ($t(729) = 3.59$, $p = 0.0004$, $t(729) = 3.12$, $p = 0.001$).

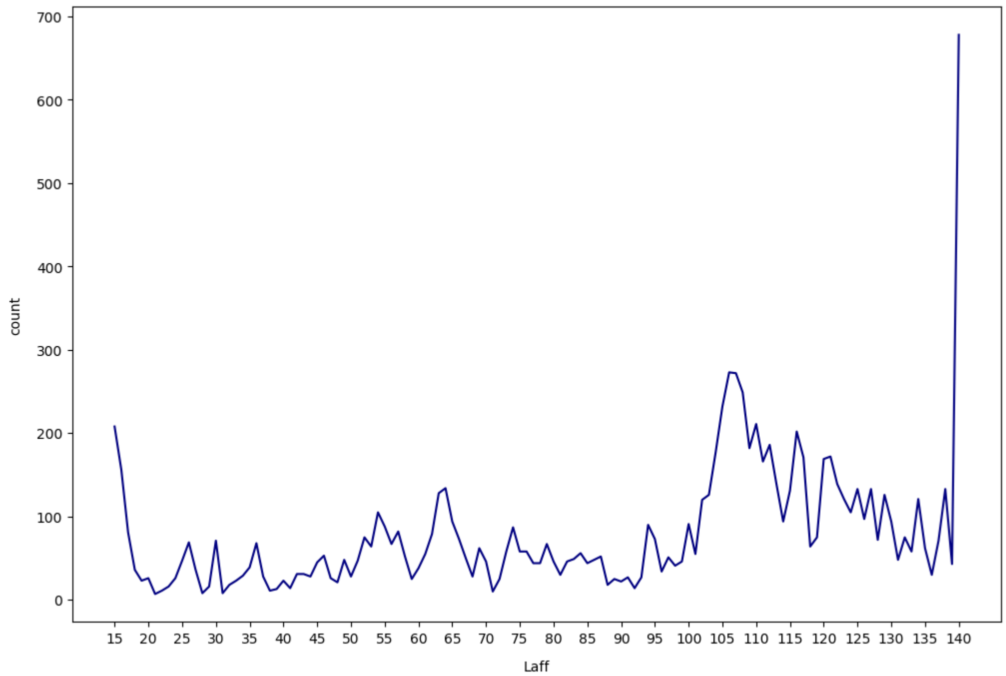

Figure 2: Laff distribution

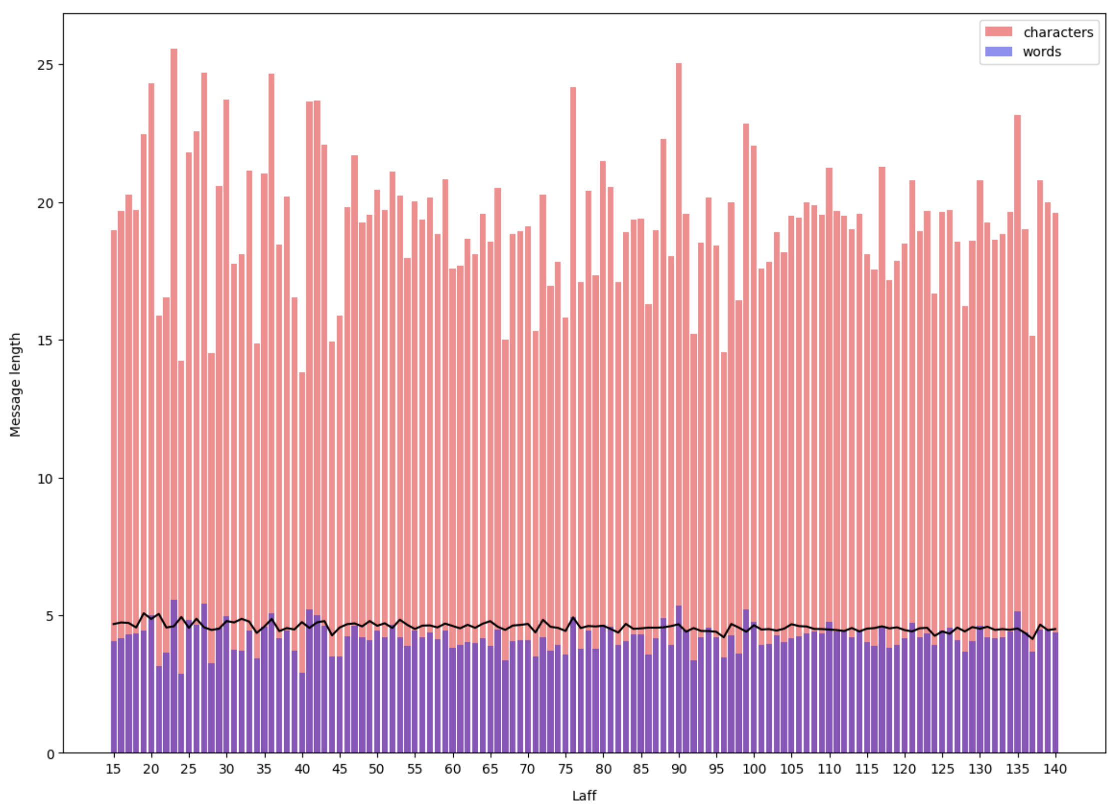

Figure 3: Message length across maximum Laff points

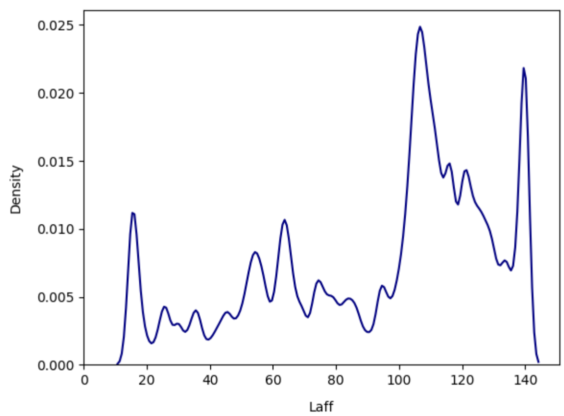

Figure 4: Density of messages across Laff

4.2.2 Word length

Average word length was calculated by dividing the length of a message in characters by its length in words. Word length or duration can be indicative of the complexity of a message or its subject matter. The average word length for the corpus overall is 4.46 characters. For comparison, the average length of a word in the English language in written texts is usually calculated to be approximately 4.7 characters [Norvig, 2013, Wylie, 2021]. The slightly shorter length observed here seems to make logical sense as this is an online chat context in which many abbreviations and acronyms are being regularly utilized. No subgroup of the population displayed an average word length which differed significantly from the population mean.

Species	l
Bear	4.41
Cat	4.43
Crocodile	4.53
Deer	4.48
Dog	4.53
Duck	4.50
Horse	4.44
Monkey	4.48
Mouse	4.37
Pig	4.24
Rabbit	4.51

Table 7: Word length across species

Gender	l
Male	4.42
Female	4.49

Table 8: Word length across gender

Missing track	l
None	4.58
Toon-up	4.45
Trap	4.44
Lure	4.43
Sound	4.43
Drop	4.39

Table 9: Word length across missing track

A stabilization effect can be observed in Figure 5 once again as the density of observations tends to increase along with Laff, which results in the word length remaining closer to the population mean.

4.2.3 Token frequencies

Rank	Token	Frequency
1	i	1679
2	the	1008
3	you	859
4	to	825
5	a	735
6	beans	732
7	for	612
8	is	539
9	it	467
10	my	461
11	ty	424
12	u	392
13	me	383
14	im	382
15	lol	370
16	and	368
17	do	361
18	in	328
19	so	315
20	on	312

Table 10: Top 20 most frequent tokens

The 20 most frequently occurring tokens closely mirror those observed in previous investigations in this area. When compared to other lists of the most frequent tokens in large English-language corpora [Oxford Dictionaries, 2011], there is some

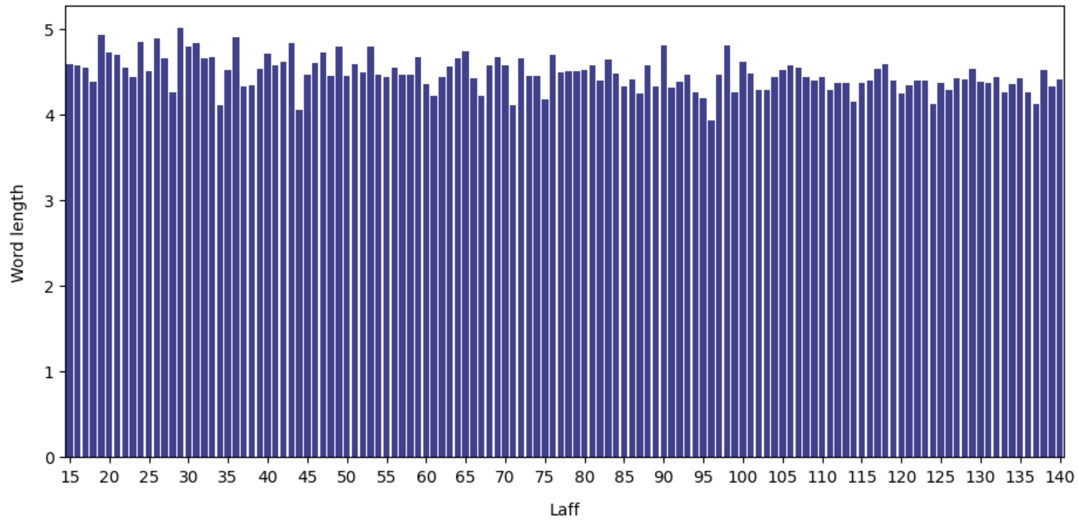

Figure 5: Average word length across Laff

divergence, likely due to the online chat origin of the corpus in tandem with the presence of digital slang such as “ty” and “lol”. The top tokens are also nearly identical across subgroups of the population. Tables of the most frequently appearing tokens for each demographic group, with the exception of individual Laff subgroups, have been included as an appendix to this paper. It is judicious to provide an explanation for the exceptionally high frequency of the term “beans” among the top tokens. There exists an area of the game where Toons are nearly always congregated in large numbers due to the fact that it has become a trend to randomly provide jellybeans, which function as Toontown’s currency, as a gift to those idling in that location. A significant amount of data collection was done in this zone as there tends to be a steady stream of in-game chat occurring there at most times of day. Interestingly, as can be observed in Tables 42-45 located in the appendix, the raw frequency and relative proportion of the token “beans” tend to become smaller as Laff increases, and it does not appear at all among the top 20 tokens for Toons with the maximum Laff of 140.

4.2.4 First-person pronoun usage

Usage of first-person pronouns was quantified by dividing the total number of appearances of first-person singular pronouns by the total number of tokens to obtain a proportion. For the present study, the set of first-person singular pronouns was defined as follows:

$$\text{fpp} = [\text{i, id, ive, im, my, mine, me, myself}]$$

Punctuation was removed from the members of the set as it was removed from the messages in the corpus. The first-person pronoun usage statistic for the corpus was calculated to be 0.07. This calculation was subsequently performed for demographic subgroups. No individual group showed divergence from the population in this regard.

Species	$\bar{f}$
Bear	0.07
Cat	0.07
Crocodile	0.07
Deer	0.07
Dog	0.06
Duck	0.07
Horse	0.07
Monkey	0.05
Mouse	0.08
Pig	0.08
Rabbit	0.07

Table 11: First-person pronoun usage across species

Gender	$\bar{f}$
Male	0.07
Female	0.07

Table 12: First-person pronoun usage across gender

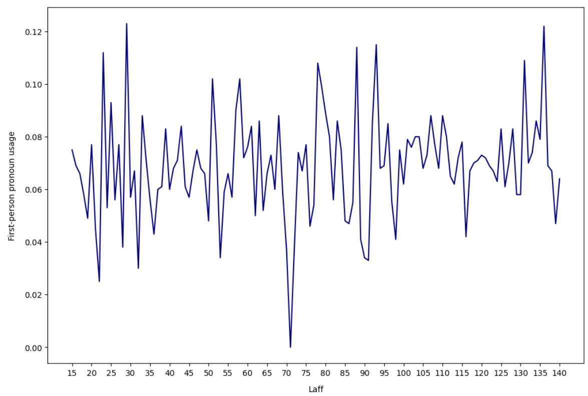

Figure 6: First-person pronoun usage across Laff

Missing track	$\bar{f}$
None	0.07
Toon-up	0.08
Trap	0.07
Lure	0.07
Sound	0.09
Drop	0.07

Table 13: First-person pronoun usage across missing tracks

The sudden plunge in Figure 6 occurs due to the fact that there are only ten messages in the corpus from Toons who possess 71 Laff points. No first-person singular pronouns are uttered in those ten messages.

4.2.5 Type-token ratio

Type-token ratio, henceforth referred to as TTR, is a measure of the number of unique words in a text corpus. It is a ratio of the total types, or unique tokens, to the total number of tokens in the corpus. A TTR of 1.0 would indicate that every single word present in the corpus appears only one time, and as the TTR approaches zero, more and more repetition of tokens is occurring. The TTR of the expanded message corpus in this study is 0.11. This is a low lexical ratio, indicating somewhat low lexical variety. Among the demographic groups being examined, TTR displayed a consistent tendency to decrease and tend toward the overall corpus TTR as message count increased. For example, as displayed in Table 14, horses and pigs display a noticeably high TTR, but messages from these two groupd comprise only 1.76% and 1.75%, respectively, of the corpus, whereas cats achieve a TTR very close to that of the corpus, but message from cats comprise more than a third of the total messages recorded. Figure 7 once again displays a stabilization of TTR and a movement towards the corpus TTR as the density of messages tends to increase along with Laff.

Species	TTR
Bear	0.37
Cat	0.17
Crocodile	0.36
Deer	0.30
Dog	0.25
Duck	0.31
Horse	0.52
Monkey	0.44
Mouse	0.27
Pig	0.47
Rabbit	0.33

Table 14: Type-token ratio across species

Gender	TTR
Male	0.16
Female	0.14

Table 15: Type-token ratio across gender

Missing track	TTR
None	0.20
Toon-up	0.36
Trap	0.18
Lure	0.36
Sound	0.51
Drop	0.19

Table 16: Type-token ratio across missing tracks

4.2.6 Sentiment

Average sentiment for the expanded corpus is 0.10, which is slightly positive. This value is very similar to the average sentiment of the first demographic corpus ($\bar{x} = 0.11$) and the non-demographic corpus ($\bar{x} = 0.08$). Of the 10,000 chat messages, 34.1% were deemed by the VADER analyzer to be positive, 12.3% to be negative, and 53.6% to be neutral. These percentages are also very similar to those calculated for the previous two corpora.

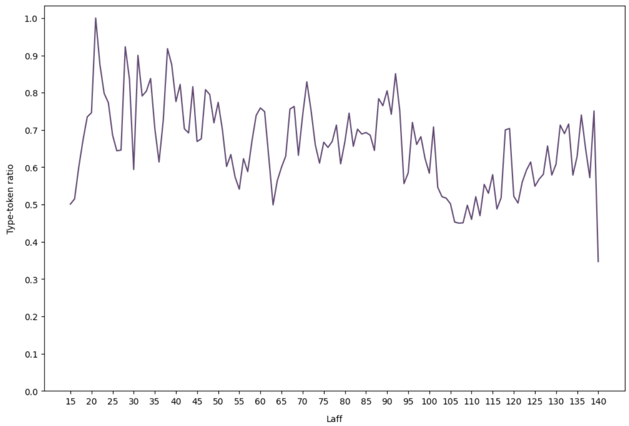

Figure 7: Type-token ratio across Laff

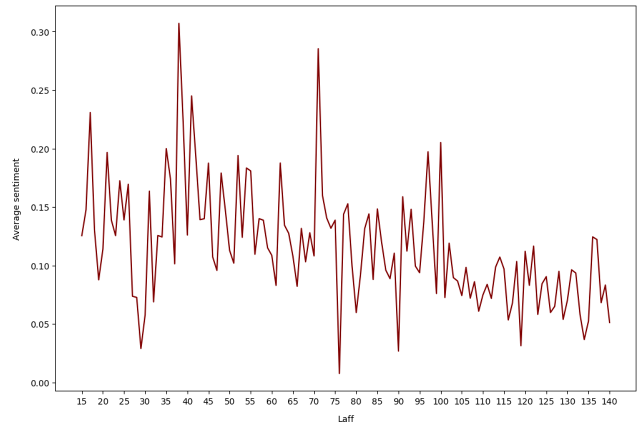

Figure 8: Average sentiment across Laff

Species	$\bar{x}$
Bear	0.13
Cat	0.09
Crocodile	0.14
Deer	0.11
Dog	0.09
Duck	0.10
Horse	0.12
Monkey	0.12
Mouse	0.10
Pig	0.10
Rabbit	0.14

Table 17: Sentiment across species

Gender	$\bar{x}$
Male	0.09
Female	0.11

Table 18: Sentiment across gender

Missing track	$\bar{x}$
None	0.14
Toon-up	0.05
Trap	0.10
Lure	0.08
Sound	0.06
Drop	0.09

Table 19: Sentiment across missing tracks

Various significant results were achieved in terms of sentiment. One-sample t-tests determined that Toons missing Toon-up ($\bar{x} = 0.05$) and Toons without a missing Gag track ($\bar{x} = 0.14$) had average sentiment scores which differ significantly from that of the overall corpus ($t(2464) = 7.16$, $p = 1.03 \times 10^{-12}$ and $t(730) =$ -4.63, $p = 4.24 \times 10^{-6}$), more negative in the case of Toons without Toon-up and more positive in the case of Toons without a missing track, as well as that messages from female bears have a higher sentiment score ($\bar{x} = 0.18$) than the overall corpus ($t(151) = 3.17$, $p = 0.002$). A two-sample t-test determined that, among Toons without a missing sixth Gag track, female Toons ($\bar{x} = 0.17$) have a higher sentiment score than male Toons ($\bar{x} = 0.12$) ($t(2643) = 4.59$, $p = 0.002$).

4.2.7 Subjectivity

The mean subjectivity of the expanded corpus is 0.19. This is nearly identical to those of the first demographic corpus ($\bar{x} = 0.19$) and the non-demographic corpus ($\bar{x} = 0.20$). No population subgroup differed significantly from the overall corpus mean. The previously described stabilization effect in terms of NLP metric values and Laff points is present once again as is visible in Figure 9.

Species	$\bar{x}$
Bear	0.18
Cat	0.19
Crocodile	0.22
Deer	0.20
Dog	0.19
Duck	0.19
Horse	0.20
Monkey	0.17
Mouse	0.20
Pig	0.18
Rabbit	0.17

Table 20: Subjectivity across species

Gender	$\bar{x}$
Male	0.19
Female	0.20

Table 21: Subjectivity across gender

Missing track	$\bar{x}$
None	0.19
Toon-up	0.19
Trap	0.19
Lure	0.18
Sound	0.18
Drop	0.20

Table 22: Subjectivity across missing tracks

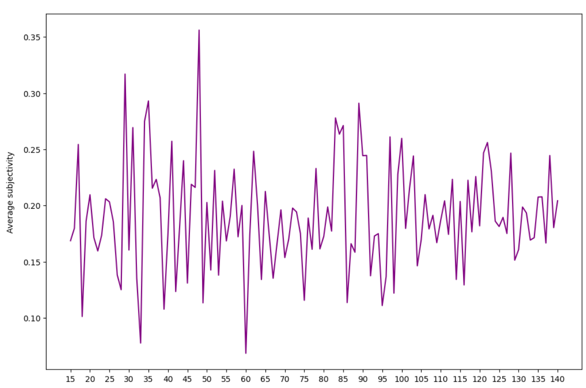

Figure 9: Average subjectivity across Laff

5 Discussion

Unfortunately, few notable or interesting results were yielded by this analysis despite the significant expansion of the demographic corpus. Only a small handful of statistically significant results were achieved and there was little divergence from the results of previous research in this area. The data display a pronounced stabilization effect between Laff points and quantitative metrics, as was also observed in the preliminary demographic NLP study published in July 2023 [Ciereszynski, 2023], wherein values of linguistic metrics remain closer to the overall corpus mean as Laff increases due to the tendency of message frequency to increase in tandem with Laff. Figure 4 displays the density of messages across the range of Laff values.

In terms of statistically significant results which were achieved, male Toons without Toon-up display longer message length in both words and characters than female Toons without Toon-up, and among Toons below 100 Laff, male Toons display longer message length in characters than female Toons. Additionally, sentiment scores for Toons without Toon-up and Toons without a missing Gag track differed significantly from the overall corpus score, with the sentiment scores for these subgroups being lower and higher than the corpus score, respectively. In the large 2022 Toontown demographic exploration, being male and missing Toon-up were found to be some of the characteristics defining a group of Toons who appeared to be on the vanguard of a set of countercultural trends [Ciereszynski, 2022a], so it is likely that they may also exhibit various sorts of divergent linguistic behaviour.

Where direct or analogical comparison are possible, many of the results observed in the present study do not align with previous findings in sociolinguistic research. The distribution of most frequent tokens is nearly identical across all groups when compared to each other and the corpus itself[3] with neither male nor female Toons appearing to prefer contracted or abbreviated forms. Message lengths and counts were also nearly identical between male and female Toons, as was word length. As mentioned previously, below 100 Laff, male Toons sent longer messages than female Toons. Sentiment score did not differ significantly between male and female Toons as a whole, although the average sentiment of female bear Toons was significantly higher than that of the corpus and female Toons without a missing Gag track had a higher average sentiment score than male Toons without one. However, in terms of message distribution with regard to sentiment

[3] Refer to the appendix for tables displaying the most frequently occurring tokens across demographic groups.

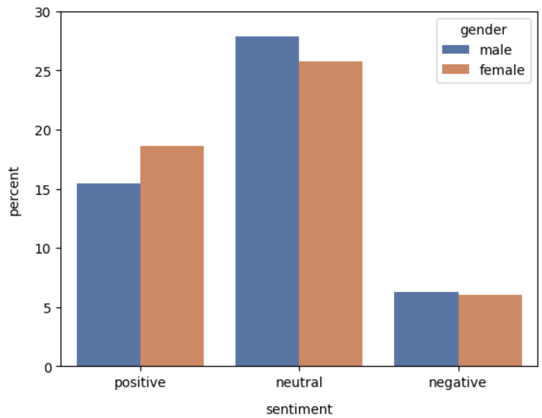

Figure 10: Distribution of sentiment judgment by percentage across gender

judgment (positive, neutral, and negative), female Toons do display a significantly larger frequency of positive messages, or messages with a sentiment score greater than zero, than male Toons ($\chi^2(1, N = 10{,}000) = 35.60$, $p < 0.001$), mirroring previous sociolinguistic findings that women may tend to show significantly more positive emotion across variegated topics. Figure 10 displays the distribution of sentiment judgments across gender.

Named entity recognition was an intended area of linguistic investigation in the present research, and it was hypothesized that examining the different kinds of named entities about which demographic groups spoke and the frequencies at which they spoke about them could be revealing and informative. However, unforeseen major issues in the output of the entity recognizer from the `spaCy` library obstructed this analysis. The entity recognizer was unable to correctly interpret contracted forms, consistently parsing online abbreviations such as "tysm", "idk", "lol", and "ily" as names of people and organizations. When applied to each message, much of what the entity recognizer returned was puzzling and unusable. To provide a handful of examples, words such as "yummy" and "kinda" were judged to be names of people, "skeleton across the road" was recognized as an organization, "hyacinth" as a date, "second language" as referring to a length of time, and "creepy" and "lesbian" as nationalities, religions, or political groups. It is likely that a sizable portion of these glaring issues stem from the online chat origin of the corpus, but it does not seem possible to explain many of them so simply. Deeper investigation into pertinent algorithms and significant fine-tuning will be necessary if named entity recognition is to be applied to Toontown data in future research.

In light of the outcomes of the present analysis as well as previous research, it appears necessary to place this area of investigation on the back burner for the time being. The primary limitation of the previous demographic NLP analysis was the size of the first demographic corpus, yet expanding its breadth to 10,000 messages from the original 4,000, as well as examining multiple new linguistic metrics, was able to produce only a small handful of reportable results. In theory, the logical strategy for future research would simply be to collect even more data, as this corpus arguably still remains somewhat minimal in length by linguistic research standards, and ideally I would be keen on continuing research in this area. However, given the laboriousness of collecting data by hand, the unpredictability of player behaviour on a day-to-day basis, and the significant changes to the game's gender system as explained in the methodology section, beginning data collection once again for further analyses does not seem practical or productive. It is possible that the population of Toontown is simply quite linguistically uniform at its core and that demographic distributions will continue to remain similar over time, at least in terms of the characteristics and variables which have been investigated, or that more significant differences do exist but the resources and protocols for collecting the proper type or amount of data to reveal those differences are not presently accessible.

The application of these data to other contexts, primarily machine learning, was previously planned, but given the uniformity of the results and lack of any sort of breakthrough provided by the expanded corpus, this may not be sensible. This being said, this is not the end of the road for the demographic corpus. Machine learning applications are not out of the question, although much strategizing about an ideal and fruitful research design will be necessary, and I would also like to create something dynamic and visual with the expanded corpus, such as an interactive dashboard.

6 Conclusion

The principal objective of this analysis was to dig deeper into previous demographic linguistic research conducted on Toontown Rewritten chat message data by way of significantly expanding the demographic corpus and examining a larger set of linguistic metrics. Previous research published in July 2023 conducted on a smaller demographic corpus yielded very few reportable results and it was thus hypothesized that conducting similar analyses on a much larger dataset could be more successful. To test this hypothesis, a second demographic corpus consisting of 6,000 in-game chat messages and various demographic characteristics of the speaker of each message was compiled and subsequently merged with the original demographic corpus of the same structure collected earlier in 2023. A handful of significant results were achieved, but the vast majority of results were extremely similar to those obtained from prior Toontown demographic and NLP research. Population demographics were nearly identical and a marked stabilization effect for quantitative linguistic results as Laff increased was observed once again. Some statistically significant results indicate that certain subgroups of Toons, namely Toons without Toon-up and male Toons below 100 Laff, may display divergent linguistic behaviour in certain areas, which aligns with the results of a large demographic study published in January 2022 [Ciereszynski, 2022a]. Female Toons also displayed a significantly larger proportion of positive messages than male Toons, reflecting some existing real-world sociolinguistic findings. It is quite possible that corpus size was once again the primary limitation of this analysis and that the construction of an even larger demographic corpus would yield a more extensive or variegated set of conclusive results, but due to multiple factors, such as the difficulty of collecting data by hand and major changes to the game's gender system, and in light of the fact that, at this point, two analyses have produced nearly identical results across the majority of their areas of investigation, further research in this specific domain is not currently planned. Extensive strategization will be necessary to ascertain how best to conduct future Toontown research at the intersection of demography and linguistics.

7 References

References

[Bamman et al., 2012] Bamman, D., Eisenstein, J., and Schnoebelen, T. (2012). Gender identity and lexical variation in social media. *arXiv: Computation and Language*.

[Baron, 2004] Baron, N. (2004). See you online: Gender issues in college student use of instant messaging. *Journal of Language and Social Psychology*, 23:397–423.

[Ciereszynski, 2018] Ciereszynski, E. (2018). Lexical variation and gender in informal social media environments. Department of Linguistics, McGill University.

[Ciereszynski, 2022a] Ciereszynski, E. (2022a). An exploration of Toontown Rewritten demographics. `https://github.com/c-z-c-z/data-analytics/tree/main/Toontown-Rewritten-demographics`.

[Ciereszynski, 2022b] Ciereszynski, E. (2022b). Toontown Rewritten corpus analysis. `https://github.com/c-z-c-z/natural-language-processing/tree/main/Toontown%20Rewritten%20corpus%20analysis`.

[Ciereszynski, 2023] Ciereszynski, E. (2023). A preliminary demographically informed analysis of TTR chat message data. `https://github.com/c-z-c-z/natural-language-processing/tree/main/TTR-demographic-NLP-analysis`.

[Disney, 2003] Disney (2003). Toontown Online. Computer software.

[Harris et al., 2020] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362.

[Honnibal et al., 2020] Honnibal, M., Montani, I., Landeghem, S. V., and Boyd, A. (2020). spaCy: Industrial-strength natural language processing in Python.

[Hutto and Gilbert., 2014] Hutto, C. and Gilbert., E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*.

[Labov, 2001] Labov, W. (2001). *Principles of Linguistic Change, Volume 2: Social Factors*. Wiley-Blackwell.

[Leith, 1997] Leith, D. (1997). *A Social History of English*. Routledge.

[Ling, 2005a] Ling, R. (2005a). *The length of text messages and use of predictive texting: Who uses it and how much do they have to say?*

[Ling, 2005b] Ling, R. (2005b). *The Sociolinguistics of SMS: An Analysis of SMS Use by a Random Sample of Norwegians*, pages 335–349.

[Loria, 2018] Loria, S. (2018). textblob Documentation. *Release 0.15*, 2.

[McKinney, 2010] McKinney, W. (2010). Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61.

[Norvig, 2013] Norvig, P. (2013). English Letter Frequency Counts: Mayzner Revisited or ETAOIN SRHLDCU. `http://www.norvig.com/mayzner.html`.

[Oxford Dictionaries, 2011] Oxford Dictionaries (2011). The OEC: Facts about the language. Retrieved from `https://web.archive.org/web/20111226085859/http://oxforddictionaries.com/words/the-oec-facts-about-the-language`.

[Palmer, 2012] Palmer, J. (2012). The role of gender on social network websites. University of Central Florida.

[Rafi, 2008] Rafi, M. (2008). SMS Text Analysis: Language, Gender and Current Practices. *SSRN Electronic Journal*.

[Rosenfeld et al., 2016] Rosenfeld, A., Sina, S., Sarne, D., Avidov, O., and Kraus, S. (2016). WhatsApp usage patterns and prediction models. *ICWSM/IUSSP Workshop on Social Media and Demographic Research*.

[Schler et al., 2006] Schler, J., Koppel, M., Argamon, S., and Pennebaker, J. (2006). Effects of age and gender on blogging. pages 199–205.

[Sun et al., 2020] Sun, B., Mao, H., and Yin, C. (2020). Male and female users' differences in online technology community based on text mining. *Frontiers in Psychology*, 11.

[The pandas development team, 2020] The pandas development team (2020). pandas-dev/pandas: Pandas.

[Toontown Rewritten team, 2014] Toontown Rewritten team (2014). Toontown Rewritten. Computer software. `https://www.toontownrewritten.com`.

[Trudgill, 1972] Trudgill, P. (1972). Sex, Covert Prestige and Linguistic Change in the Urban British English of Norwich. *Language in Society*, 1:179–195.

[Virtanen et al., 2020] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and Contributors, S. . (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17:261–272.

[Wylie, 2021] Wylie, A. (2021). Benchmark readability against the BBC. `https://www.wyliecomm.com/2021/01/benchmark-readability-against-the-bbc/`.

A Most frequent tokens by demographic group

A.1 Species

Table 23: **Cats**

Rank	Token	Frequency
1	i	632
2	the	323
3	to	313
4	you	291
5	a	240
6	beans	193
7	for	190
8	u	181
9	is	179
10	my	175
11	it	175
12	im	144
13	me	139
14	lol	135
15	and	132
16	ty	131
17	do	123
18	on	120
19	so	114
20	in	112

Table 24: **Dogs**

Rank	Token	Frequency
1	i	234
2	the	165
3	you	132
4	to	128
5	a	126
6	beans	114
7	is	101
8	for	89
9	my	76
10	and	69
11	u	66
12	it	62
13	do	60
14	that	59
15	me	55
16	are	54
17	in	54
18	on	54
19	this	52
20	so	48

Table 25: **Mice**

Rank	Token	Frequency
1	i	197
2	the	93
3	a	79
4	to	78
5	you	75
6	beans	68
7	for	63
8	is	60
9	it	54
10	ty	52
11	me	47
12	so	42
13	lol	41
14	im	40
15	have	39
16	u	38
17	my	36
18	in	36
19	on	35
20	need	35

Table 26: **Deer**

Rank	Token	Frequency
1	i	145
2	the	98
3	beans	95
4	you	72
5	to	70
6	for	68
7	a	64
8	ty	49
9	im	48
10	lol	38
11	and	35
12	is	34
13	do	33
14	my	32
15	me	29
16	it	27
17	in	27
18	bean	26
19	we	26
20	need	25

Table 27: **Ducks**

Rank	Token	Frequency
1	i	118
2	the	98
3	you	87
4	to	54
5	a	53
6	beans	50
7	for	49
8	it	44
9	is	37
10	do	36
11	my	36
12	ty	33
13	me	30
14	im	30
15	on	29
16	have	28
17	this	27
18	in	26
19	we	25
20	of	24

Table 28: **Rabbits**

Rank	Token	Frequency
1	i	99
2	you	67
3	the	65
4	to	65
5	for	54
6	beans	49
7	ty	37
8	a	36
9	is	33
10	lol	30
11	my	29
12	it	25
13	im	25
14	me	24
15	need	23
16	what	21
17	so	21
18	u	20
19	thank	20
20	and	18

Table 29: **Bears**

Rank	Token	Frequency
1	i	82
2	the	53
3	a	40
4	beans	37
5	you	36
6	for	33
7	it	28
8	to	27
9	is	24
10	ty	21
11	me	20
12	lol	20
13	and	18
14	u	18
15	that	18
16	im	17
17	in	17
18	just	16
19	get	16
20	be	15

Table 30: **Crocodiles**

Rank	Token	Frequency
1	i	76
2	beans	53
3	to	47
4	the	43
5	a	42
6	you	41
7	is	30
8	my	28
9	for	26
10	it	24
11	and	24
12	ty	22
13	lol	21
14	do	21
15	me	20
16	in	20
17	get	17
18	so	16
19	just	16
20	this	16

Table 31: **Monkeys**

Rank	Token	Frequency
1	beans	46
2	i	33
3	the	32
4	a	25
5	you	22
6	for	22
7	is	18
8	to	16
9	ty	13
10	on	13
11	it	13
12	need	112
13	my	12
14	and	11
15	what	11
16	we	10
17	have	10
18	do	10
19	thank	10
20	get	10

Table 32: **Horses**

Rank	Token	Frequency
1	i	25
2	you	18
3	a	16
4	the	15
5	to	15
6	my	12
7	are	10
8	beans	10
9	ty	8
10	is	8
11	me	8
12	it	8
13	and	7
14	all	7
15	good	6
16	that	6
17	toon	6
18	do	6
19	can	6
20	not	6

Table 33: **Pigs**

Rank	Token	Frequency
1	i	32
2	the	23
3	you	18
4	beans	17
5	is	15
6	a	14
7	for	14
8	to	12
9	my	12
10	ty	11
11	how	9
12	do	9
13	need	8
14	that	7
15	hi	7
16	too	7
17	it	7
18	this	7
19	we	7
20	in	7

A.2 Gender

Table 34: **Male**

Rank	Token	Frequency
1	i	823
2	the	513
3	to	403
4	a	403
5	you	393
6	beans	330
7	for	298
8	is	282
9	my	227
10	it	220
11	and	203
12	im	197
13	u	196
14	me	195
15	in	180
16	do	163
17	on	162
18	ty	155
19	so	151
20	this	150

Table 35: **Female**

Rank	Token	Frequency
1	i	850
2	the	495
3	you	466
4	to	422
5	beans	402
6	a	332
7	for	314
8	ty	269
9	is	257
10	it	247
11	my	234
12	lol	227
13	do	198
14	u	196
15	me	188
16	im	185
17	and	164
18	so	163
19	what	157
20	thank	150

A.3 Missing track

Table 36: **No missing track**

Rank	Token	Frequency
1	i	376
2	beans	296
3	the	283
4	you	238
5	to	209
6	for	191
7	a	172
8	ty	138
9	my	133
10	is	128
11	it	97
12	thank	96
13	im	94
14	and	89
15	me	88
16	so	86
17	do	83
18	in	76
19	lol	73
20	here	72

Table 37: **Toon-up**

Rank	Token	Frequency
1	i	128
2	the	59
3	to	55
4	a	48
5	is	42
6	for	41
7	my	41
8	u	39
9	it	38
10	im	37
11	you	36
12	on	35
13	so	30
14	me	30
15	and	28
16	that	25
17	in	24
18	go	24
19	beans	23
20	do	23

Table 38: **Trap**

Rank	Token	Frequency
1	i	544
2	the	293
3	you	260
4	to	257
5	a	255
6	beans	208
7	for	187
8	is	180
9	u	156
10	ty	145
11	it	141
12	my	141
13	me	130
14	lol	121
15	and	117
16	im	113
17	do	109
18	on	105
19	that	104
20	get	101

Table 39: **Lure**

Rank	Token	Frequency
1	i	96
2	the	49
3	you	44
4	to	43
5	beans	37
6	a	35
7	is	32
8	it	27
9	need	26
10	do	25
11	for	25
12	me	24
13	my	24
14	lol	23
15	u	23
16	how	20
17	im	20
18	all	18
19	get	18
20	and	17

Table 40: **Sound**

Rank	Token	Frequency
1	i	39
2	the	20
3	you	20
4	is	11
5	my	11
6	a	11
7	me	11
8	in	10
9	here	9
10	for	9
11	it	9
12	beans	8
13	to	8
14	this	7
15	and	6
16	are	6
17	do	6
18	u	6
19	what	6
20	was	5

Table 41: **Drop**

Rank	Token	Frequency
1	i	490
2	the	304
3	you	261
4	to	253
5	a	214
6	beans	160
7	for	159
8	it	155
9	is	146
10	lol	129
11	u	122
12	do	115
13	ty	115
14	im	114
15	my	111
16	and	110
17	in	107
18	so	101
19	me	100
20	what	89

A.4 Laff groupings

Table 42: **Low** (15-64) ($n = 2461$)

Rank	Token	Frequency
1	i	381
2	beans	288
3	the	284
4	you	240
5	to	201
6	for	189
7	a	169
8	ty	133
9	my	130
10	is	130
11	thank	95
12	it	90
13	and	90
14	im	90
15	me	90
16	so	87
17	do	79
18	lol	74
19	in	74
20	bean	71

Table 43: **Medium** (65-99) ($n = 1643$)

Rank	Token	Frequency
1	i	260
2	beans	211
3	the	190
4	you	174
5	to	131
6	for	128
7	a	115
8	ty	100
9	lol	80
10	is	79
11	it	74
12	me	69
13	do	66
14	my	66
15	and	62
16	thank	59
17	so	56
18	im	53
19	need	51
20	are	50

Table 44: **High** (100-139) ($n = 5218$)

Rank	Token	Frequency
1	i	918
2	the	484
3	to	433
4	a	395
5	you	390
6	is	295
7	for	270
8	it	259
9	u	247
10	my	245
11	beans	230
12	im	212
13	me	202
14	lol	201
15	and	192
16	in	189
17	do	187
18	ty	186
19	on	180
20	get	164

Table 45: **Maxed** (140) ($n = 678$)

Rank	Token	Frequency
1	i	114
2	to	60
3	a	56
4	you	55
5	u	51
6	the	50
7	it	44
8	is	35
9	on	34
10	do	29
11	im	27
12	for	25
13	and	23
14	in	23
15	like	23
16	no	22
17	have	22
18	me	22
19	what	21
20	my	20