The Comparisons between the Language Used by Male and Female Peers in Twitter

Arissa Handini Soedjono

Abstract

Internet linguistics is currently making its way to a clearer path to discover categorizations, features, and particularities to its subject. This article attempts to give contribution to the ongoing linguistic research on gender and online communication. The platform we choose to analyze in this analysis is the microblog, Twitter. We use corpus linguistics as the approach to our analysis. Corpus linguistics involves natural texts, which are texts that are generated from real life data. The corpus of this study consists of 2,627 female tweets and 908 male tweets. The female tweets consist of 30,590 words while the male tweets consist of 10,405 words. For the main theories, we use language and gender theories in face-to-face communication as proposed by Lakoff (1975) and Coates (2004) as well as theories of online communication proposed by Herring (1994) and Crystal (2011).The result shows differences in the use of pronouns, similarities in the use of abbreviations, and differences and similarities in the use of vulgar words.

Keywords: Effort; Social structure; Patriarchy; Capitalist Society, Oppression

1. Introduction

Language in the Internet is continually developing. This is supported heavily by new waves of online social networks that allow people to express opinions and thoughts. One of the social networks is Twitter. The most apparent attribute to observe the language in Twitter is by comparing the features of male and female tweets, emphasizing on the use of vocabularies. These vocabularies can be observed from a corpus. A corpus according to Sinclair (2004: 19) is "a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research". Based on this definition, the corpus of Twitter is the main source for analysis and discussion used to determine the comparisons between male and female tweets.

Language and gender is the main theory that is used for the interpretation. As we know, language in the Internet is a combination of spoken elements and written elements. Therefore, traditional gender theories proposed by Lakoff (1975) and Coates (2004) as well as gender online communication theories proposed by Crystal (2011) and Herring (1994) will be incorporated in this research. The idea is to integrate both characteristics to find the features in male and female tweets. According to the well-known theory by Lakoff (1975), the typical female language is marked by the use of certain linguistic features such as hedging devices, tag questions, intensifiers and qualifiers, so-called trivial lexis, empty adjectives and rising intonation on declaratives. On the other hand, Coates (2004: 110 & 133) stated that the typical male language is identical to the use of swear words and competitive style. These features, however, are found in the face-to-face context. Meanwhile, according to Herring (1994: 3 & 4), the male style is characterized by "adversariality: put-downs, strong often contentious assertions, lengthy and/or frequent postings, self-promotion, and sarcasm while the female-gendered style, in contrast, has two aspects which typically co- occur: supportiveness and attenuation".

There are two articles that relate to this article. The first study was conducted by Colley and Maltby (2007) from the University of Leicester. The second study was conducted by Ekasari (2008) from Airlangga University. Colley and Maltby (2007) focus on how the Internet has changed men's and women's lives. They use participants that posted comments in the BBC News website in the

discussion topic "Has the Internet changed your life?". Ekasari (2008) examines the study of the role of adjectives semantically found in the articles of women's magazine *InStyle* and Men's magazine *Details*. She uses Biber's theory of adjectives which are categorized as descriptor and classifier. The two articles mentioned previously focus on different kinds of platforms and linguistic intentions. In this article, we focus on the vocabularies to examine the comparisons used by male and female peers in Twitter.

2. Methodology

In this research, we use qualitative method. According to Creswell (2009: 2), a qualitative research involves individual interpreting which means that the researcher is required to make interpretations of the meaning of the data. We used text corpus of Twitter because it consists the written conversation needed for this research. We use the peers in our university because most of Twitter users are young adults. They all range from 21-22 years old.

Since information in Twitter rapidly changes, we chose a particular month which is April. In order to ease my research, we used an online website that assists us on retrieving older tweets called Snap Bird. From this website we chose to search someone's timeline and typed their Twitter username. After that, we browse the tweets and then copy and paste the ones we needed to Microsoft Word. The tweets that we collected were the ones that do not involve replying and/or retweeting. In other words, we did not collect conversational tweets. We only collected personal tweets that conveyed the users' thoughts, ideas, opinion, and emotions. Then we compile them according to the dates before finally converting it to .txt. After compiling tweets individually, we compile the tweets to all female and all male categories and also convert it to .txt. It is important to note that there are greater female tweets collected than male tweets. There were 2,627 female tweets and 908 male tweets that were collected. The female tweets consist of 30,590 words while the male tweets consist of 10,405 words.

3. The top 10 words used in male tweets vs. female tweets

The top ten words are the most frequent words used by both male and female users. In determining the top ten words, we use a software called AntConc. To use this software, we have to insert the male and female data that are in the form of .txt. After that, we click the Word List tab. Before beginning the search, we have to treat all data as lowercase to avoid repetition of the same word. Finally, we click the Start button. Here are the results of the top ten words listed by AntConc:

No.	Male	Female
1	april	i
2	ini	aku
3	di	t
4	d	yg
5	yg	di
6	t	ga
7	ga	ada
8	itu	co
9	co	http
10	http	the

Table 1: Top Ten Words

The most frequent word in the male category is April. However, we do not use this word in our research because we only inserted it as a determiner of the month in which the tweets are being retrieved. Therefore, we ignore the word April in both male and female categories. According to the table, the most frequent word in the female category is *I*. *I* is the English first-person pronoun that indicates individuality. This also applies to the word *aku*, the Indonesian word of *I*. We can see that neither *I* nor *aku* is in the top ten of the male category. This shows that females use the word *I* and *aku* often because they are considered more self-centered than males.

The second most frequent word in the male category is *ini*, which means *this*. This word can go alongside with the opposite word *itu*, which means *that*. Both of the words appear in the male category but not at all in the female category. This shows that males use demonstrative pronouns that reflect indirectness more often than females.

The pronoun *I* and the article *the* in the female category indicate the high use of English language rather than Indonesian language by female users. This is opposite to the male category where it does not show any English language in the top ten. This occurs because there are more female students in the English Department of Airlangga University than male students. In addition, female is more adaptive in learning new languages. As a result, female peers in our research use English in their tweets more than male peers.

4. The most distinctive features in the male and female tweets

From these top ten words, we can determine the features that are most distinct. These features are divided into three parts: personal pronouns, abbreviations, and vulgar words. As mentioned in the methodology, there are more female tweets collected than male tweets. This is because females post tweets more than males do. To compensate the differences, we normalize the number of tweets by counting words per minute (WPM) of each word. The formula is:

$$\frac{Frequency}{Total\ Number\ of\ Word\ Tokens} \times 1000000 = WPM$$

Frequency: The frequency of words found in Word List of AntConc

Total Number of Word Tokens: The total number of word tokens of both male and female tweets

We use this formula to determine the *words per minute* (WPM) which is used as the final determiner of the differences of male and female tweets on the use of personal pronouns, abbreviations, and vulgar words. The total number of word tokens in the female category is 30,590 while the total number of word tokens in the male category is 10,405. These numbers apply to all of the following sub-chapters. We will start by calculating the pronouns.

4.1 Personal pronouns

Pronouns can be used to determine the degree to which people focus on or relate to others. In this case, we can determine how male and female users of Twitter use personal pronouns as an individual representation or how male and female users of Twitter use personal pronouns to relate to others in their tweets.

Word	Male		Female	
	Freq.	WPM	Freq.	WPM
I	49	4709.27	391	12781.95
me	11	1057.18	96	3138.28
aku	58	5574.24	327	10689.76
saya	41	3940.41	41	1340.30
gue	0	0	9	294.21
gw	0	0	39	1274.92

Table 2: The list of male and female tweets using first-person pronouns

The data above shows that female use first-person pronouns such as *I*, *me*, and *aku* more often than male. It shows that females are more representative than males. The word *gue* is the informal word of *I* in Indonesian language. This word is often used in the regions of West Java. Most people who use that word, especially teenagers and young adults, abbreviate it as *gw* in writing. We can see that both of the words are not found in the male category. This is because the users of this research are not from West Java, but from East Java, where Javanese language is often used. However, this does not apply to the female category. Although the female users are from East Java, they sometimes use *gue* and *gw* as a pronoun. This shows how females are more adaptive to other languages than males.

The only difference is the use of the word *saya*. From the table it is clear that male use the word *saya* more often than female. *Saya* also means *I* in Indonesian. But in contrast to the word *gue*, *saya* is a formal word. This indicates that male like to be conveyed as having manners. Overall, female use more first-person pronouns more than male do.

Word	Male		Female	
	Freq.	WPM	Freq.	WPM
you	22	2114.36	182	5949.65
kamu	8	768.86	34	1111.47
km	10	961.07	39	1274.92

Table 3: The list of male and female tweets using second person pronouns

According to the table above, females use more second person pronouns such as *you*, *kamu*, and *km* more than males. This shows that females are direct in expressing something towards other people. On the other hand, males tend to be indirect in expression. One possible reason is because female use Twitter as a place for communicating their thoughts of other people without confronting the person in face-to-face communication. In other words, the word "you" can be referred to any reader of the tweets.

Before selecting the third person pronouns *he* and *she*, we must look at the concordances because the word *he* can also mean referring to someone. Here, we find there are 27 words but only 20 meant the actual pronoun in English.

Word	Male		Female	
	Freq.	WPM	Freq.	WPM
dia	6	576.64	17	555.73
he	2	192.21	20	653.80
she	1	96.10	16	523.04

Table 4: The list of male and female tweets using third person pronouns

The data above shows that males use the word *dia* more often than females. *Dia* is a third person pronoun in Indonesian. However, females use the English third person pronouns *he* and *she* more often than males do. This also confirms the previous explanation that female use English more than male do. Third person pronouns are used in Twitter to discuss about other people.

In conclusion, female users use more pronouns that encode the relationship between the writer and the reader (especially first person and second person pronouns) while male tend not to refer it. Meanwhile, male and female use the same amount of third person pronouns that only differ in the language choice which is Indonesian or English. Moreover, the use of pronouns is related more to female style of writing than to male's where female "present things in a relational way" (Argamon, Koppel, Fine, & Shimoni). Biber (1995) has explained that women's writings are more "involved" while men's writings are more "informational". The use of first-person pronouns and second person pronouns proves that females use the "involved" style while males prefer to use generic pronouns to refer someone. This conclusion also accords with my previous findings of the typical male and female tweets where female use Twitter for more personal purposes than male.

4.2 Abbreviations

According to Magnan (2008: 24), abbreviations are one of the many characters of CMC. As we know, an abbreviation is used for shortening a certain word. Its use became more popular due to the growth of SMS (Short Message Service). This happens because SMS only provides few characters for a person to use. The same occurrence happens in Twitter. Since Twitter only allows 140 characters, abbreviations are commonly found. Also, according to Danet & Herring (2007: 173), tendency towards using abbreviations is greater in synchronous mode rather than asynchronous mode. This is because people in synchronous mode type faster. Since Twitter comprises both aspects, it is likely that users will also type fast, therefore using abbreviations. Here, I find abbreviations are used equally both by male and female. These abbreviations even have different styles, depending on the writer.

Word	Male		Female	
	Freq.	**WPM**	**Freq.**	**WPM**
yg	88	8457.47	275	8989.86
km	10	961.07	39	1274.92
jd	21	2018.26	27	882.64
bs	1	96.10	16	523.04

Table 5: The list of male and female abbreviations

In addition, the abbreviations above are heavily influenced by Indonesian language. People in Indonesia, especially young people, are used to use abbreviation their spelling. The notion of spelling certain words correctly (without being abbreviated) is commonly perceived as a formal language. Moreover, certain words that are not abbreviated can be perceived as an angry tone. As a result, young people tend to avoid the usage.

4.3 Vulgar words

According to Coates (2004: 86), the use of vulgar words or swear words is commonly linked to males than females. The vulgar words we find in Twitter, however, are almost the same between male and female. The only thing that differs both is the spelling variety.

Word	Male		Female	
	Freq.	**WPM**	**Freq.**	**WPM**
fucking	2	192.21	3	98.07
fuck	2	192.21	2	65.38
fuckin	-	-	1	32.69
fuaaaakkkk	-	-	1	32.69
fuaaakkk	-	-	1	32.69
fuaaakkkk	-	-	1	32.69
fuk	-	-	1	32.69
fuucck	-	-	1	32.69
wakakakakafuak	-	-	1	32.69
cok	4	384.43	-	-
jancokan	2	192.21	-	-
jancok	-	-	1	32.69
hancok	-	-	1	32.69
hancik	-	-	1	32.69
cuk	-	-	1	32.69
cukk	-	-	2	65.38
jancik	-	-	1	32.69
jancukan	-	-	1	32.69
wancuk	-	-	1	32.69

Table 6: The list of male and female vulgar words

Huffaker & Calvert (2005) stated that males will use language that is more resolute than females. As we can see on the table above, the resoluteness can be seen through the spelling: males are more homogeneous in writing the word *fuck* and *jancok* (a vulgar word in Javanese) while females tend to change their writing style by adding and/or changing the letters. This is done in order to accentuate emotion i.e. *fuaaaakkkk*, *fuk*, *fuucck*, *wancuk*.

We can notice that the words *fuaaaakkkk* and *fuucck* are written in reduplicating letters. Similar to the previous sub-chapter, this kind of occurrence is called by Danet & Herring (2007: 169) as "eccentric spelling". The goal of the user is to replicate spoken pronunciation in written language. Eccentric spelling also applies for the word *jancuk* or *jancok*. In writing these two words, we can see that male and female users have preferences. Male uses *jancok* while female uses *jancuk* or *jancik*. The *o* vowel gives more stress than the *u* vowel, making it look harsher in writing. Therefore male use it more than female. Meanwhile, *jancik* with the vowel *i* is preferred by female to refine the rudeness.

The use of vulgar words in Twitter contradicts Lakoff's (1975) theory about female language being passive. In fact, female are more expressive in showing their emotions by using vulgar words. The reason to this is because online communication has given these peers "freedom and flexibility" (Huffaker & Calvert, 2005) in their writing.

5. Conclusion

This article has shown the comparisons between male and female tweets. The first is in the use of pronouns. In this case, male and female are similar and different in terms of the degree to which people focus on or relate to others. For the first-person pronouns and the second person pronouns, females tend to be more self-centered and direct in addressing people while males are the opposite. Meanwhile, both male and female are equal in quantity in the use of third person pronouns. The second is the use of abbreviations. The result shows both male and female use abbreviations equally in their tweets. This is necessary due to the limitation of characters that Twitter provides. The abbreviated words are mostly Indonesian words. The third is the use of vulgar words. For this matter, both male and female use vulgar words in their tweets but in the female category, the vulgar words undergo spelling diversions called eccentric spellings.

6. References

Argamon, S., Koppel, M., Fine, J., & Shimoni, A. R. (n.d.). Gender, Genre, and Writing Style in Formal Written Texts. Retrieved June 3, 2012, from http://u.cs.biu.ac.il/~koppel/papers/male-female-text-final.pdf

Biber, D. (1995). Dimensions of Register Variation: A Cross-linguistic Comparison. Cambridge: Cambridge University Press.

Coates, J. (2004). Women, Men and Language Third Edition. Harlow: Pearson Education.

Colley, A., & Maltby, J. (2007). Impact of the Internet on our lives: Male and female personal perspectives. Computers in Human Behavior , 2005-2013.

Creswell, J. (2009). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches Third Edition. California: SAGE Publications.

Danet, B., & Herring, S. C. (2007). The Multilingual Internet: Language, Culture, and Communication Online. New York: Oxford University Press.

Ekasari, A. A. (2008). A Study of the Role of Adjectives Semantically Found in the Articles of Women's Magazine InStyle and Men's Magazine Details. Surabaya: Airlangga University.

Herring, S. (1994). Gender differences in computer-mediated communication: Bringing familiar baggage to the new frontier. Keynote talk at American Library Association Annual Convention. Miami, Florida.

Huffaker, D. A., & Calvert, S. L. (2005, January). JCMC Home. Retrieved March 27, 2012, from Journal of Computer-Mediated Communication: http://jcmc.indiana.edu/vol10/issue2/huffaker.html

Lakoff, R. (1975). Language and Woman's Place. New York: Harper & Row.

Magnan, S. S. (2008). Mediating Discourse Online. Amsterdam: John Benjamins Publishing Company.

Sinclair, J. (2004). Corpus and Text-Basic Principles. In M. Wynne, Developing Linguistic Corpora: a Guide to Good Practice (pp. 1-20). University of Oxford: AHDS Literature, Languages and Linguistics.