

An Investigation of Different Language Choice through Personal Pronouns in the Twitter

Nur Hafiz Abdurahman

LPDP Awardee, Universitas Tanjungpura, Indonesia

e-mail: arahmanhafiz@gmail.com

Abstract:

This study aimed to find evidence regarding the use of personal pronouns in the discourses produced by males and females. Personal pronouns were chosen as the object of analysis, as several studies has suggested them as one of the features that may distinguish the gender of the authors. This study analysed publically available corpus, Rovereto Twitter N-Gram Corpus (RTC), utilized by Herdagdelen (2013). It is gender-of-the-author tagged, which makes the author's gender analysis easier. The corpus was analysed using AntConc (Anthony, 2014). From AntConc's concordance analysis, it was found that women utilised more personal pronouns, especially the ones that can create closer bond. On the other hand, men have greater tendency to distant themselves using generic pronouns than women. In conclusion, men and women in this study may use personal pronouns differently.

Keywords: *Personal Pronoun, Twitter, AntConc*

1. INTRODUCTION

In academic writing, different language choices are almost invisible. On the other hand, informal conversation can usually depict the different involvedness level of gender because of its 'interactional nature' (Argamon et al., 2003, p. 322). In English articles written by Persian speakers, female writers used the same pattern as their native counterparts, while male ones were affected mostly by their native language (Seyyedrezaie and Vahedi, 2017). In addition, Hosseini and Tammimy (2016) and Jasmani, et al. (2011) found that verbs and pronouns may also provide distinctive gender-oriented information. For these reasons, I intend to investigate the different utilisation of language by men and women through Twitter which is increasingly popular and provides more natural interaction unlike other written discourses such as essays. The strength of Twitter, therefore, is it may provide a natural condition for people to either be 'involved' or 'informative' (Argamon et al., 2003) in their produced-tweets without being afraid to be interrupted.

The study of how gender distinguish the language utilisation has been in the spotlight since Lakoff's (1975). From then, many researchers have tried to find evidence that gender plays an important role in the produced discourse (Tannen, 1994; Cameron, 2003; Talbot, 2003; Cunha et al., 2014). These studies found that men tend to be more straightforward than women. Nonetheless, Cameron (2003, p. 465) noted that the language and gender stereotypes 'has often begun from folklinguistic stereotypes'. Thus, a research is often set by an agenda, and the results cannot avoid to recirculate these stereotypes. This article, as a result, admits that this study also fell into the similar paradox. Nonetheless, it is important to note that this study does not deny other possible factors that may affect the language utilised in the data.

This paper has two key aims. Firstly, to find evidence whether involved features, particularly personal pronouns, are more frequent in women-authored discourses as found in many studies (Argamon et al., 2003; Argamon et al., 2007; Newman et al., 2008; Bamman et al., 2014). Secondly, I will address how each gender utilises language through frequent collocation of the most common personal pronouns. One thing to note, however, is that this paper only aims to present phenomena as additional evidence of gender differences in Twitter. Thus, it does not necessarily mean that gender is the distinguishing factor in discourse.

1.1 Research Questions

There has been much literature providing evidence that men and women use different linguistic features in their discourses, especially in social media. It is also evident that personal pronouns are more favourable by females (Hosseini and Tammimy, 2016). Nevertheless, it is rare to find how these devices and the semantic preferences accompanying them differ between genders. Therefore, the following research questions will guide my study:

- Do females from Herdagdelen's (2003) Rovereto Twitter N-Gram Corpus (RTC) use personal pronouns in their tweets more than males?

- If yes, how do their language choices through the utilisation of personal pronouns differ?

2. LITERATURE REVIEW

2.1 Approaches to Language and Gender

Language and gender have been studied by many researchers using various approaches. The first is deficit approach, labelled to Lakoff's (1975) *Language and Woman's Place*. Lakoff (1975) mainly discussed women's language in everyday conversation. This approach sees women as a weaker group and is often criticised because it indirectly suggests women to 'speak like men if they want to be taken seriously' (Coates, 2004, p. 6). The second approach is the dominance approach which still sees women as the oppressed group (Coates, 2004).

Unlike the previous approaches, the third approach sees that men and women are two different values; thus, they speak differently (Coates, 2004, p. 6). In this approach, each gender was seen equal, and any differences that occur are due to the gender's 'culture' differences, not their social position. Although the difference approach is greatly criticised when it is applied to the talk among people of different genders (Coates, 2004, p. 6), Cameron (1992, p. 61) argued that gender is a social construction itself. In other words, just like how people of different cultures may speak and use similar language differently, men's and women's using language differently should be seen as a distinguishable social phenomenon. For this reason, dynamic approach – social constructionism – then emerged.

It is suggested from the fourth approach that language should not be dichotomized based on masculine/feminine point of view; nevertheless, I will adopt different approach because I would like to see the differences in the tweets produced by each gender. However, this preference does not mean that I agree that men and women in general speak differently. It is merely due to the limitation of the data, postings in Twitter, which only has one social variable, gender.

2.2 Discourse and Corpus

Before moving further into the use tweets as a written discourse, the definition of discourse and corpus will first be discussed. Discourse is the produced language used in communication. Cameron and Panovic (2014, 3-4) summarised three possible definitions of discourse: (1) 'language above the sentence'; (2) 'language in use'; and (3) a social practice with language as a meroneither spoken or written although recently computer-mediated discourse –produced in computer-mediated environment, such as email, social media and online chatting– becomes popular too.

The record of these divisions is often called as a text. Partington *et al.* (2013, 3) summed a text as a 'by-product', 'the record' or 'the trace of discourse action'. In other word, although discourse, especially the spoken one, occurs in real time, it can be reproduced for analysis in form of the text. The analysis of this text is commonly called discourse analysis. It studies how 'the language is used to influence the beliefs and behaviour of other people' (Partington *et al.* 2013, 3).

To sum up, discourse is the language itself. While sentence can be produced without any certain relation to certain context, discourse can often be related to a particular context. Discourse can be divided into two dimensions, spoken and written, although recently the emergence of internet and computer also sheds light to a new dimension, computer-mediated discourse, for any discourses produced in the computer/internet related environment. Examples of computer-mediated discourse is email, online chatting, or status in social media.

Corpus is a collection of texts, and in this project, the texts are students' papers. Corpus linguistics usually focuses on the analysis of the authentic texts. It often analyses the texts quantitatively. For this reason, Hyland (2009, 110) claimed corpus linguistics and discourse analysis, which usually uses qualitative approach, as 'perfect bedfellows'. While corpus commonly consists of a huge amount of texts, discourse analysis often focuses on a small number of cases. Applying a corpus-based research for a huge amount of texts, and from the findings selecting some special cases and analysing these may be time-saving and more meaningful. In addition, Partington *et al.* (2013, 5-6) also argued that corpus linguistics may provide useful evidences for discourse analysis.

2.3 Twitter as a Computer-mediated Discourse Research Tool

The growth of computer-mediated discourse has been in line with the growth of Internet; they are increasing. Unlike the traditional discourse, computer-mediated discourse (CMD) such as the one in Twitter is written discourse which works like the spoken one. Crystal (2006) introduces the term 'netspeak' for any computer-mediated discourse. The inevitable growth of this 'netspeak' is a promising area for linguistic research.

Moreover, unlike traditional spoken discourses in which the power has been an issue when comparing two genders' linguistic choices, Web 2.0 as a 'user-generated platform' (O'Reilly, 2005), gives more freedom for users to write what they want. This is one strength of using Twitter as the corpus. In addition, a large amount of dataset from social media allows researchers 'to analyse the frequency of individual words' (Bamman *et al.*, 2014, p. 136). This possibility allows not only to find different features such as pronouns, but also to track any lexical words accompanying them. For these reasons, I would like to present individual pronouns used by each gender to see any different semantic utilisation by each gender.

2.4 Use of Pronoun as a Gender Representative

During the past 40 years, much information regarding gender differences in language use has become evident. In the academic writing, however, the use of pronouns does not necessarily represent the gender of the author. It seems to be a common agreement that the use of pronouns in academic writing greatly depends upon the discipline it was written. In many hard science articles, such as physics, there is a sense to distance for the author's findings in the article. Hence, first-person pronouns, i.e. *I* and *we*, are used less often. The third person pronoun, *it*, is more prevalent. Geertz (1983, cited in Harwood, 2005, p. 1208) used the term 'author-evacuated' for this phenomenon. These authors often choose this method regarding

the findings to avoid being falsified by other fellow academics of the discipline. Moreover, Hyland (2001) also argued that the eradication of self-mention in research article aims to gain acceptance by the readers. Nevertheless, he emphasised the self-mention as in the personal pronoun. Nevertheless, the use of *I* and *we* in academic writing tends to be emphasised as these can be used as author's promotional devices, not the gender representative (Harwood, 2005, p.1226).

Apart from academic writing, several recent studies suggest that although there may be some gender stereotypes, it is evident that men and women in general use language differently. In the spoken discourse, females are more likely to be 'affective' than males (Holmes 1993). In written discourse, such as blog, Newman et al. (2008) study found that men and women use function words such as pronouns, intensifier and articles differently. In general, a woman has greater tendency to use more pronouns and intensifiers while a man uses articles more often (Newman et al., 2008, p. 231).

Similarly, Argamon et al. (2003) introduced the term 'involvement' for the use of pronouns and intensifier and 'informational' for function words such as determiner and articles. These terms were used to distinguish different characteristics between two genders' writing. According to Argamon et al. (2003), pronouns, especially the devices which refer to relationship between the writer and the reader, are favoured more in women's writing. These differences may not be absolute, but in comparison, women generally use them more (Baker 2008). In addition, although Bamman et al. (2014) argued that it is hard to make a clear distinguishable notions using the 'involved' and 'informational' terms for gender-specific discourses distinction, their findings in the use of pronouns in Twitter also supported the notion of women being involved by using devices such as personal pronouns.

Argamon et al. (2007) went further by conducting an automated blog analysis. In their research, Argamon et al. (2007) comprised an English corpus of approximately 140 million words from blogs. Similar to the previous research (Argamon et al., 2003), the algorithm can quite accurately classify the gender of the author at around 80.5% accuracy by tracking some specific features of each gender's writing (Argamon et al., 2007, p. 6). In addition, Burger et al. (2011, cited in Bamman et al., 2014, p. 137) also came into conclusion that the automated prediction is more accurate than human's when it comes to classifying gender of the author. This literature has emphasised the importance of such features in the gender-based discourses.

In short, it is evident that men and women generally have different characteristics in utilising language, in this case English, and these characteristics can even be used to trace the gender of the author. This article intends to see how these linguistic features differ through the use of personal pronouns in men's and women's produced tweets.

3. RESEARCH METHODOLOGY

3.1 The Data

I applied a corpus-driven analysis (Baker, 2010) by analysing a naturally occurring data to see different language utilisation by Twitter users. By using corpus, it is easier to see and compare different utilisation by both genders. In addition, there are certain patterns that can be drawn by analysing words in ‘naturally occurring language’ (Baker, 2008, p. 76-77). Baker also added that corpus may give unexpected yet insightful cases of how language is utilised (Baker, 2008, p. 81).

The analysed data is a publically available corpus, Rovereto Twitter N-Gram Corpus (RTC), utilized by Herdagdelen (2013). RTC is gender-of-the-author tagged which makes gender distinction process easier. In addition, using existing corpus is useful to ‘scholars who may have neither the resources to assemble large teams of researchers nor the time and computing know-how to develop and use new tools’ (Mautner, 2007, p. 52).

Although the RTC consists of approximately 75 million Twitter posts, I randomly chose 1000 tweets for each gender using excel RAND formula from 1-gram tweets of RTC. From these 2000 tweets, there are approximately 12781 word tokens of the males’ tweets and 13592 word tokens of the females’ ones. These tweets were separately analysed, so there are two corpora: the female-produced and the male-produced tweets. For ethical purpose, any users mentioned in the corpora will be replaced by *username*. The analysis was conducted in 2016.

3.2 Approach and Analysis

A mixed method research was employed to answer the research questions. Unlike Argamon’s (Argamon et al., 2003), this study analysed only one variable of the produced discourses – personal pronouns – and provided concordance analysis to present the phenomena. In addition, Argamon et al. (2003) explored gender differences in writing of various genre while this article explored the differences in Twitter. For the first question, a quantitative method was utilised in order to compare the differences; the findings were normalised by presenting per 1000 words number. This quantitative result would be used to answer my first research questions.

One of the most well-known tools for assessing a corpus is AntConc (Anthony, 2014). It is famous because it is free and user-friendly. I utilised AntConc to explore the use of personal pronouns. Firstly, personal pronouns used in each gender’s corpus were tracked and counted; the non-standard variations such as *’em* and *u* were also tracked. After that, The results of the frequent words collocating the personal pronouns such as *I*, *you* and *it*, which are dominant in both corpora will be presented.

Finally, to answer the second question, I employ qualitative method by discussing semantic preferences accompanying the three most frequent personal pronouns (*I*, *you* and *it*). This discussion will be accompanied with *AntConc*’s concordance of *I love* in both corpora.

Through the corpus-driven approach, I intend to explore how each gender utilises language by looking to the personal pronouns and the verb accompanying them. One thing to note, however, I did not intend to make a baseless speculation regarding the findings nor do I intend to generalize that every female uses personal pronouns more than men. Instead, I would like to present them as linguistic phenomena.

4. FINDINGS

4.1 The Use of Personal Pronouns by both Gender

The overall use of personal pronouns is displayed in table 1 below. Because male and female corpora has different word tokens, 12781 and 13592 respectively, the normalized numbers are shown in table 1 as well.

Table 1. The use of personal pronouns in both genders' tweets

Personal Pronoun	Male		Female	
	Total occurrences	Occurrences /1000 words	Total Occurrences	Occurrences /1000 words
I	564	44	1100	81
me	135	11	144	11
we	99	8	21	2
us	36	3	0	0
you	231	18	234	17
he	16	1	33	2
him	9	1	15	1
she	13	1	33	2
her	14	1	24	2
they	20	2	21	2
them	19	1	12	1
it	184	14	213	16
Total	1340	105	1850	136

In general, each gender uses almost all personal pronouns. *I* was used the most in each corpus although women use it nearly twice more often than men. On the second rank, *you* occurs in each corpus quite equally. Non-human single third person pronoun, *it*, ranks the third and occurs a bit higher in the female-authored corpus. *Me* ranks fourth and occurs quite equally in both corpora.

Some other points from table 1 above merit attention. Firstly, the occurrences of third-person pronouns (with *it* being the exception) are the least of all; *he*, *him*, *she*, *her*, *they* and *them* occur once in a thousand words of male's tweets and almost twice in the female-posted ones (table 1). Although the females use these devices twice as many as men, I cannot say that it is generally true because of small differences between two corpora.

Secondly, first-person singular devices (*I* and *me*) are used more by females (92 occurrences per 1000 words). This phenomenon may reflect that women try to be more 'involved' than men by creating closer relationship using these devices.

Argamon et al. (2003, p. 330) argued that these pronouns are used to ‘encode the relationship between the writer and the reader’.

It can also be seen that women use more personal pronouns in their tweets. A big gap of occurrences (136 of women’s as opposed to 105 of men’s tweets) in both corpora means that women utilise more personal pronouns than men; however, men used first-person plural (we and us) more than five times than women. This is interesting to note because Argamon et al. (2003, p. 329) speculated that:

‘The greater use of plural pronouns reflects the tendency of male authors to encode classes rather than individualized entities and may also serve as a depersonalization mechanism that reduces the specificity of reference to gender, number, and personhood’.

In other words, men might use first-person plural pronouns to distant themselves from being only an individual. This is different from female users who utilised greater number of *I* and *me* than the plural counter-parts (*we* and *us*).

Taken together, these findings suggest that women utilise more personal pronouns, especially the ones that can create closer bond. On the other hand, men have greater tendency to distant themselves using generic pronouns than women. Nonetheless, it must be taken with extreme caution as there is a possibility that such utilisations were made due to other social factors, such as individual’s personality (Hosseini & Tammimy, 2016).

4.2 Semantic Preferences Accompanying Personal Pronouns

Firstly, I will show frequent words collocating *I*, *you* and *it* from both corpora. These pronouns were chosen due to having the three highest frequencies. The words are the most frequent two words to the right (2R) of the pronouns. After that, I will discuss some selected concordances based on these findings.

Table 2. The frequent words (2R) accompanying *I*, *you* and *it*

Rank	Male			Female		
	Personal Pronouns			Personal Pronouns		
	I	you	it	I	you	it
	collocating words (2R)			collocating words (2R)		
1	<i>have</i>	<i>like</i>	was	don't	were	was
2	don't	<i>know</i>	<i>stinks</i>	<i>miss</i>	I	is
3	can't	we	on	<i>love</i>	<i>username</i>	up
4	was	not	<i>made</i>	was	<i>guys</i>	the
5	<i>really</i>	I	is	just	<i>feel</i>	never
6	<i>need</i>	will	a	<i>have</i>	don't	<i>looks</i>
7	<i>love</i>	which	would	<i>wish</i>	can	I'm
8	am	where	<i>works</i>	<i>know</i>	are	<i>hurts</i>
9	<i>wanna</i>	were	will	<i>feel</i>	you	you
10	<i>think</i>	<i>username</i>	when	<i>want</i>	<i>xx</i>	won't

Note: *username* is the replacement for any username mentioned using '@' function in Twitter; the *italicized* words are content words.

From table 2 above, while using *I*, there is only one word, *love*, that expresses feeling in male-authored corpus. On the other hand, *miss*, *love*, *wish* and *feel* frequently collocate *I* in the female's postings. I will show the concordance of *love* used in both corpora later in this section. In addition, *feel* and *hurts*, other verbs that may express desire or feeling, were also frequent to accompany *you* and *it* in the female-authored tweets.

Interestingly, there is a use of *xx* (sometimes also *xoxo*) which accompanies *you* in the female corpus. This word is usually used to represent kisses (X) and hugs (O) which can also be said as words that express positive feelings. I cannot find any example use of *xx* or *xoxo* in the male corpus, so it is not possible to discuss about the use of *xoxo* in the male corpus. Nevertheless, it is evident that women use more content words to express their feeling, such as *miss* and *love*, than men. While *I* and *you* are 'linguistic devices that solidify relationships' (Argamon et al., 2003, p. 322), the semantic preferences of *miss*, *love*, *wish*, *feel* and *xx* possibly indicate that women use more affective words when talking about relationships than men.

To illustrate the use of the affective words in more detail, I now examine cases of how each gender utilises *I* and *love* from table 3 and table 4 below.

Table 3. Concordance of *I love* in male corpus (appendix 1)

1	home mummy my brothers bullyin me... niiiight	<i>I love</i>	everyone in the world
2	98,@username wooww!!! thank you!!	<i>I love</i>	you na!!!
3	sleep 101,"Grr Im cravin a cuddle off jake	<i>I love</i>	that dog , don't want him to grow this game! Heheh. Rain has stopped now
4	ername it works !! haha.. 140,Would you rather... 247,"@username Oh Snap!! You like playing	<i>I love</i>	now
5	Xbox?	<i>I love</i>	playing Xbox , but we got no online

Table 4. Concordance of *I love* in female corpus (appendix 2)

1	is graduating tomorrow! 29,"@username thx	<i>I love</i>	yours, too <3 was up??"
2	you '@username You guys did it mayynne!!!!	<i>I love</i>	the album
3	39,crying like a baby as p.s	<i>I love</i>	you
4	78,@username haii	<i>I love</i>	u so much how cum ur not
5	115,crying like a baby at p.s	<i>I love</i>	you
6	123,@username Don't dye your hair dark!	<i>I love</i>	it the colour it is
7	you and i didn't want to wake up	<i>I love</i>	you girl
8	275,i gotta cry... bby	<i>I love</i>	u... please talk to me...
9	310,"@username OHHHHHHH (L) omggg, ahh.	<i>I love</i>	himmm and i love youu. awk, it
10	OHHHHHHH (L) omggg, ahh. i love himmm and	<i>I love</i>	youu. awk, it makes me sad and
11	her! that description is totally perfect though	<i>I love</i>	it!"

Overall, table 3 reveals that men utilise the word *love* towards non-human objects such as *dog*, *game* and *Xbox*. This means that although *love* conveys their feelings, this feeling might not be used to solidify relationship. In addition, line 1 illustrates that they also use *love* as a statement to unspecified humans, i.e. people in general.

On the other hand, table 4 indicates that women tend to use love to depict affection towards human beings because of the human-referring pronouns that follow. Although a number of concordances are not equal, these tables present how semantic utilisations between men and women are different.

In short, it is evident that men and women might interact differently in the high interaction linguistic context, in this case Twitter, where the whole *netizens* who has access to their tweets can see. Nevertheless, the difference in power in Twitter needs to be exercised cautiously as the other background of the users who tweeted were unclear.

5. DISCUSSION AND CONCLUSION

This study set out to better understand the differences between men and women in utilising language. The data presented here is still limited in term of the corpus as well as the representativeness of each gender. The reality may differ when a bigger number of corpora were utilised. The variable is another limitation because the linguistic features characterising gender cannot be generalised from only the use of personal pronouns; there are a lot of features needed to take into account (Hosseini & Tammimy, 2016; Bamman, 2014; Newman et al., 2008; Argamon et al., 2007; Argamon et al., 2003).

In addition, when compared individually per gender, it is also possible for each gender to utilise personal pronouns differently. Apart from gender, other values – ideology, culture and personality – of each individual may affect their word choices too (Hosseini & Tammimy, 2016).

This study only aimed to find a supporting evidence that ‘informational’ and ‘involvement’ devices are utilised differently by each gender. Moreover, this study does not generally mean that each gender utilises language differently, but that gender is one of the social variables to affect individual’s language utilisation.

Returning to the questions posed at the beginning of this study, it is now possible to state that quantitatively it is female who uses more personal pronouns. Furthermore, using different approach model, men and women utilised language differently (though not absolutely); men might utilise the content words only to express feeling, whereas women are more likely to both express their feelings or desires and create closer ties using specific words such as *love*, *miss* and *feel*. One thing to note, however, the finding of this study does not mean that gender play the most important role in writing.

6. REFERENCES

- Anthony, L. (2014). *AntConc* (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from: <http://www.laurenceanthony.net/> [Downloaded 10 October 2015].
- Argamon, S., Koppel, M., Fine, J. and Shimoni, A.R. (2003). Gender, genre, and writing style in formal written texts. *Text-*, 23(3), pp. 321-346. Available from <http://www.degruyter.com/view/j/text.1.2003.23.issue-3/text.2003.014/text.2003.014.xml> [Accessed 2 November 2015].
- Argamon, S., Koppel, M., Pennebaker, J.W. and Schler, J. (2007). Mining the blogosphere: Age, Gender, and the Varieties of Self-expression. *First Monday*, 12(9).
- Baker, P. (2008). 'Eligible' Bachelors and 'Frustrated' Spinsters: Corpus Linguistics, Gender and Language. In Harrington et al., pp. 73-84.
- Baker, P. (2010). Corpus Method in Linguistics. In Litosseliti, pp. 93-116.
- Bamman, B., Eisenstein, J. and Schnoebelen, T. (2014). Gender Identity and Lexical Variation in Social Media. *Journal of Sociolinguistics* 18(2): pp. 135–160.
- Burger, J.D., Henderson, J., Kim, G. and Zarrella, G. (2011). Discriminating gender on Twitter. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 1301–1309. In Bamman, B., Eisenstein, J. and Schnoebelen, T. (2014). Gender Identity and Lexical Variation in Social Media. *Journal of Sociolinguistics* 18(2): pp. 135–160.
- Cameron, D. (1992). *Feminism and Linguistic Theory*. Basingstoke: Macmillan.
- Cameron, D. (2003). Gender and Language Ideologies. In: Holmes, J., & Meyerhoff, M. (eds.). (2003). *The Handbook of Language and Gender*. Oxford: Blackwell, pp. 447-467.
- Cameron, D., & Panovic, I. (2014). *Working with Written Discourse*. Los Angeles: SAGE.
- Charles, M., Pecorari, D. and Hunston, S. (eds.). (2009). *Academic Writing: At the Interface of Corpus and Discourse*. London: Continuum.
- Coates, J. (2004). *Women, Men and Language*, 3rd ed. London: Pearson Education.
- Crystal, D. (2006). *Language and the Internet*, 2nd ed. Cambridge: Cambridge University Press.
- Cunha, E., Magno, G., Goncalves, M.A., Cambraia, C. & Almeida, V. (2014). He Votes or She Votes? Female and Male Discursive Strategies in Twitter Political Hashtags. *PLOS One*, 9(1), pp. 1-8.
- Eckert, P. and McConnell-Ginet, S. (1995). Constructing meaning, constructing selves: Snapshots of language, gender, and class from Belten High. In Hall et al., pp. 469–507.
- Eleta, I. and Golbeck, J. (2014). Multilingual use of Twitter: Social networks at the language frontier. *Computers in Human Behavior*, 41, pp. 424-432.
- Hall, K. and Bucholtz, M., eds. (1995). *Gender Articulated: Arrangements of Language and the Socially Constructed Self*. New York: Routledge.
- Harrington, K. (2008). Perpetuating Difference? Corpus Linguistics and the Gendering of Reported Dialogue. In Harrington et al., pp. 85-102.

- Harrington, K., Litosseliti, L., Sauntson, H. and Sunderland, J., eds. (2008). *Gender and Language Research Methodologies*. Hampshire: Palgrave Macmillan.
- Harwood, N. (2005). 'Nowhere has anyone attempted.. In this article I aim to do just that': A corpus-based study of self-promotional *I* and *we* in academic writing across four disciplines. *Journal of Pragmatics*, 37, pp. 1207-1231.
- Herdagdelen, A. (2013). *Rovereto Twitter N-Gram Corpus*. Available from http://clic.cimec.unitn.it/amac/twitter_ngram/en.1grams.gz [Accessed 2 December 2015].
- Herdagdelen, A. (2013). Twitter N-Gram Corpus with Demographic Metadata. *Language Resources & Evaluation*, 47, pp. 1127-1147.
- Holmes, J. (1993). Women's talk: The Question of Sociolinguistic Universals. *Australian Journal of Communication*, 20(3), pp. 125-148.
- Holmes, J., & Meyerhoff, M. (eds.). (2003). *The Handbook of Language and Gender*. Oxford: Blackwell.
- Hosseini, M. and Tammimy, Z. (2016). Recognizing Users Gender in Social Media using Linguistic Features. *Computers in Human Behavior*, 56, pp. 192-197.
- Hyland, K. (2009). Corpus Informed Discourse Analysis: The Case of Academic Engagement. In Charles, M., Pecorari, D. and Hunston, S., eds., *Academic Writing: At the Interface of Corpus and Discourse*. London: Continuum. Pp.110-128.
- Jasmani, et al. (2011). Verbs and Gender: The Hidden Agenda of a Multicultural Society. 3L: Language, Linguistics, Literature, Vol 17 (2011)
- Lakoff, R. (1975). *Language and Woman's Place*. New York: Harper & Row.
- Litosseliti, L., ed. (2010). *Research Method in Linguistics*. London: Continuum.
- Mautner, G. (2007). Mining Large Corpora for Social Information: The Case of Elderly. *Language in Society*, 36(1), pp. 51-72.
- Newman, M. L., Groom, C. J., Handelman, L. D. and Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes*, 45(3), pp. 211-236.
- O'Reilly, T. (2005). *What is Web 2.0*. Available from <http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html?page=1> [Accessed 2 January 2016].
- Partington, A., Duguid, A. and Taylor, C. (2013). *Patterns and Meanings in Discourse: Theory and Practice in Corpus-assisted Discourse Studies (CADS)*. Amsterdam: John Benjamins.
- Schmid, H.-J. and Fauth, J. (2003). Women's and Men's Style: Fact or Fiction? New Grammatical Evidence. Paper delivered at *Corpus Linguistic Conference*. Lancaster University, UK.
- Seyyedrezaie, Z.S. and Vahedi, V.S. (2017). Projecting Gender Identity Through Metadiscourse Marking: Investigating Writers' Stance Taking In Written Discourse. *Indonesian Journal of Applied Linguistics*, Vol 6, No 2 (2017)
- Sun, Y., Wang, N., Shen, X., & Zhang, J. (2015). Location information disclosure in location-based social network services: privacy calculus, benefit structure, and gender differences. *Computers in Human Behavior*, 52, pp. 278-292.

- Talbot, M. (2003). Gender Stereotypes: Reproduction and Challenge. *In*: Holmes, J., & Meyerhoff, M. (eds.). (2003). *The Handbook of Language and Gender*. Oxford: Blackwell, pp.468-486.
- Tannen, D. (1994). *Gender & Discourse*. Oxford: Oxford University Press.