**IN-GROUP IDENTITY IN ONLINE MASS MEDIA**

**A CASE STUDY OF ‘CITY OF PIGS’ BLOG POSTING**

Prihantoro

Universitas Diponegoro

prihantoro2001@yahoo.com

**Abstract**

‘City of pigs’ is a blog posting that once was a trending topic in Indonesia as it addressed the dirtiness of Bandung. This topic was not new since many newspaper has published news and articles about the some places in Bandung that is consistently dirty. However, what made this posting a trending topic was the author’s background (she was a foreigner), the text itself, and readers responses. Pros and cons did take place. As for the cons, many believe she addressed to specific communities, and established herself as separate group. In this paper I seek to describe how such assumption was derived. By using corpus approach, the text was processed and pronouns (with the contexts) were retrieved and analyzed. Pronoun reference analysis shows that some of the references are made exophorically, causing the readers have to use their metalinguistic knowledge, in this case the superiority of Caucasian race. At this point, multi interpretation is more likely to occur. The references for some pronouns are also generalized such as Muslims, People of Bandung, Asians and Indonesians, though some are implied. Quantitatively, the generalization is not much, but for some readers, this tends to be a crucial importance that cause them to respond the posting negatively.

Keywords: blog, discourse, group identity, pronoun, text processing, keyword in contexts

1. **INTRODUCTION**

In this paper, the linguistic data is obtained from a blog posting, ‘city of pigs’ addressed specifically to dirtiness of Bandung, one of the big cities in Indonesia. The posting for few days was a trending topic in Indonesian mass media. Many called this posting controversial (that was why it became a trending topic) and somewhat offensive. Consider the following excerpt from the posting:

“Bandung – the city where people think pig meat is too dirty to eat, yet they

live in more filth than the animal itself. It could be a lovely city. Yet it looks

(and smells) like a disaster area”

“You see, in here they can’t have a normal trash can like us”

“Let me remind you of the many refugee camps in Somalia. They are not dirty.

Because people don’t have what to throw. Not that’s poverty. You are just a pig

and money are not an excuse.

As the nature of blog allows responses, it turned out to get both positive and negative responses. From the negative responses, many readers believe that this posting is harsh, offensive and racist, even though to some extent, some people acknowledge that the there is a portion of truth value in it. Some accepted the critic, but criticized the way it was written:

“The writer sings praises for her efforts and w as just a few sentences short of

canonizing herself in her ‘Quixotic’ efforts w hile she lambastes the people

of Bandung”

“This is a very fine piece I hope every one would read. Your tone comes out harsher

than I’d like, but I do agree that w e need the treatm ent”

“Well, y eah, somehow people in this country relate Caucasians to superiority”

I will focus on the latest part of the readers belief (the posting is racist). Some of the comments mentioned about the Caucasian, European and specific countries, which refer to countries in western hemispheres. While in the posting, the author also mentioned Bandung, Asian and Indonesian. Finding and analysis section of this paper will describe how the use of pronouns in this posting can derive such identity assumption. Here, corpus approach is employed as it has been proven useful in discourse analysis as described by Conrad & Biber (2008), and also Sinclair (2006). Some of the literatures are reviewed in section two.

1. **LITERATURE REVIEW**

Electronic and online media have become center of attention as it shows pattern of interaction in a community. The online media itself is one of the objects of discourse studies (Mathesson, 2005). Internet is a facility where people can get connected, reducing time and space factors. Netizens (internet citizens, a.k.a Internet users) can get connected to other netizens on different time zone and continent without having to travel and meet the individual in person. The information can be shared via a website. However, not all people are capable of designing a website plus a web hosting service requires money. A blog (weblog) is a more user friendly spot where people can share information without having to be able to design a web. It is already equipped with web publishing tool; therefore no coding or scripting knowledge is required. Some of the web functionalities are developed to support interactive communication. The social function of a blog is better known than its informative function, as people may also respond to the posting.

**About Pronoun**

Language is often related to identity (Edwards, 2009), by means of community. However, the use of language can be different person to person. Therefore, the identity itself must be related to the discourse created by an individual (in this case, blogging discourse media). One of the aspects that builds discourse coherence is pronoun.

Besides its grammatical function to build cohesion and coherence, as described Halliday & Hasan (1976), the relation of pronoun and the background of speakers have already come into attention of Brown & Gilman (1960), which is more popular as T-V (Tous and Vous) relation. The use of pronoun in specific discourse has also received serious attention: such as the use of exclusive and inclusive pronoun in academic discourse as described by Harwood (2005). Recently, in discourse analysis, the use of pronoun is becoming more essential as it represents how the author puts him/herself on one side over another, as it is well described by Coleman & Ross (2010). As for this, pronoun reference is a crucial importance.

As pronoun is used to replace the reference noun, then the reference is ideally recoverable. The reference can be made anaphorically (before the pronoun), cataphorically (after the pronoun). In computational linguistics, recovering anaphoric and cataphoric pronoun are still quite a challenge, as described in (Wolf, Gibson, & Desmet, 2004). One that is interesting, is that the reference can also be made exophorically (beyond the text).

As for the latest (exophoric), it is quite difficult to recover the pronoun independently (solely depending on the text), without understanding the background of the writer. Even by a qualitative approach, multi interpretation is still most likely to take place. Recovering this kind of pronoun reference to social reality (Schriffin, 2008) is not as easy as it seems. How the readers understand the text (and how they decide the exophoric reference) is largely affected by this factor.

1. **METHODOLOGY**

**Research Data and Digital Text Processing**

Linguistic data can be composed of speech or written texts. These days, the data are digitalized. In terms of data preservation and maintenance, digital format has some advantageous over the manual format. Archiving linguistics data in physical form definitely requires enormous time, space and effort, not to mention the maintenance cost yet. In terms of processing, the digital format provides significant supports for researchers such as; data retrieval, classification and extraction.

The digital data processing had received a considerable amount of criticism in the beginning (Mc Enery & Hardie, 2012), addressed particularly to the Brown corpus by Chomsky around 1960’s. At that period, the criticism did sound, as the technological supports were not significant. Most of the processing stages at that time still had to be done by hands (manually). These days, with the advancement of technology, the analysis of linguistic data can be performed, to some extent, automatically by text processing software. In the next sub section, I describe the nature of the data (as digital text), and the corpus processing software that I used to process the data.

**Research Procedure**

The data in this paper is obtained from a blog posting ‘city of pigs’ (the nature of blog as on-line social media allows internet users to freely read the blog and to leave comments as well). This data is processed by AntConc (Anthony, 2006), freely available corpus processing software. The posting on HTML format is converted to .txt format, a format the AntConc can process. The digital text is processed as raw corpus, which means no annotation is performed. The first processing is obtaining a general description of the corpus. In this processing, the description is quantitative. The second processing is meant to understand the specific distribution of pronouns, the linguistic devices that became one of the sources of controversy. Still, the analysis is quantitative. The third stage of the processing is identifying the pronoun references. The analysis here is qualitative, which is by using concordance function to recognize the keyword (pronoun) in context. Concordance is a very basic function in corpus linguistics that presents keywords in context so that users can analyze the context (Adolphs, 2006). These three processing are the building blocks to support my argument that the sense of in-group identity is strong in this posting, presented in section 4 of this paper.

1. **FINDINGS AND DISCUSSION**

As the corpus is uploaded, AntConc can perform several language processing tasks. One that I describe here to begin is ‘word list’ function. It tokenizes the strings and determines both number of tokens and types in the text file. As for this, the file is shown on the upper left with .txt extension. At this point, it is necessary to explain why I use the term ‘token’ and ‘type’. Token is actually a more proper terminology, as ‘word’ in this sense, means all word forms. Therefore, words that are derived or inflected from one lexeme is distinctive tokens (e.g: ‘go’, ‘went’, and ‘gone’ are three distinct tokens though English speakers are quite aware that these tokens are the word forms of a lemma <go>). The lemma <go> itself is called ‘type’. Figure 1 described the type, token, and its frequency of occurrence:

**Figure 1. Word List in [city of pigs.txt**



 A slight overview of the word list in figure 1 indicates the followings: The text is composed of almost 2000 tokens (1929), and the number of types is almost a third (713). Among the tokens, the most frequently occurring token is ‘the’ (occurrence=58) and ‘a’(occurrence=55). It reconfirms what is already common in other corpus works related to English. Definite and indefinite articles usually outnumber the other tokens in terms of frequency. What is also common in English text is the high frequency of function words. Up to the 10th rank, the figure shows the domination of function words to content words. Pronoun ‘I’ (occurrence=52) is on the 4th rank. This shows that this posting was written in highly personal style. In the next sub section, I will describe the quantitative distribution of ‘I’ and the other pronouns.

**Quantitative Distribution of Pronouns**

 Pronouns in English can take different forms depending on their syntactic function. Their distribution as subject, object or possessive pronoun is shown by table 1:

**Table 1. Quantitative distribution of Pronouns**

|  |
| --- |
| **Pronoun** |
| **S** | **O** | **Poss Adj** | **Poss** | **Reflexive** |
| I (52) | Me (15) | My (12) | Mine (0) | Myself (0) |
| We (7) | Us (3) | Our (3) | Ours (0) | Ourselves |
| You (20) | Your (5) | Yours (0) | Yourself(1) |
| They (27) | Them (15) | Their (16) | Theirs (0) | Themselves Mine (0) |
| He (7) | Him (1) | His (1) | Himself Mine (0) |
| She (0) | Her (0) | Hers (0) | Herself Mine (0) |
| It (25) | Its (0) | - | Itself (1) |

 The distributional property of pronouns is shown by table 1. The occurrence of possessive and reflexive pronouns is extremely low, close to zero occurrences. There are some pronouns (subject-object and possessive adjectives) where the occurrence is also zero, such as ‘its’, ‘her’ and ‘she’. Some pronouns, though not zero occurrence, are low in occurrence, especially ‘we’, ‘us’ and ‘our’. Pronouns that are shared are presented by the same column (e.g, ‘you’ as subject and object pronoun). Therefore, to understand their syntactic distribution, contextual analysis is required. Consider concordance extract for ‘you’ as presented by figure 2.

**Figure 2. Concordance for ‘you’**

****

Concordance function shows the keyword(s) in context(s). By using the contexts, it is possible to deduce the concordance in figure 2, that ‘you’ is used more frequently as subject. Token ‘you’ are used as objects only in line (4) and (19). The pronouns that are frequently used to show in-group identity (by exclusion from the group) are 2nd and 3rd plural forms such as ‘you’, ‘they’, ‘them’ and their variants. This of course requires further analysis, but the recapitulation of pronouns showed that the occurrence of such pronouns (as presented by table 1) is not extremely high or too low. However, gaps are noted. For instance, the occurrence of ‘you’ is 27, but the occurrence of ‘we’ is only seven (7). As I have commented previously, the occurrence of pronouns must be supported by contextual analysis of references that the pronouns refer. In the next sub section, I show the references of some pronouns that are significant in terms of in-group identity markers.

**Pronouns as In-Group Identity Markers: Recovering Reference**

 I will begin by discussing the references for 1st person plural pronouns such as ‘we’,’us’ and, ‘our’. One difficulty about this in-group identity pronoun is that the same word forms are used to indicate inclusiveness and exclusiveness. In different language, the word forms are distinctive; such as *kami* (exclusive) and *kita* (inclusive).

In the posting, the references of ‘we’ were made really clear: her family. She wrote the posting with her experience going around parks in Bandung with her husband and her baby as well. The reference is made cataphorically. Consider some fragments of the posting as presented by table 2:

**Table 2. The context of ‘we’.**

|  |  |
| --- | --- |
| a | Since **we** couldn’t take a walk in any park, **I and my husband** decided to take the baby for a walk in a government office near his parent’s home |
| b | So for privacy sake **we** will call it Grumpy Scientists place. |
| c | Due to bad weather, **we** didn’t take a walk in Monday. |

 Pronoun ‘we’ is used in the first clause with no reference. However, the subject of the second clause has made the references clear. This made the reference for the rest of ‘we’ to get recovered anaphorically. Now, I will move on to the contexts of the other pronouns, ‘us’ and ‘our’, which are (3 occurrences each). Please see table 3:

**Table 3. The Context of ‘us’ and ‘our’**

|  |  |
| --- | --- |
| a | You see, in here they can’t have a normal trash can like **us** |
| b | Since the **city council** already knew their **people** they made sure that the trash cans are put in solid concrete. After being unable to remove them and sell them, **our** **active citizens** just broke them to pieces out of pure spite. |
|  c | **My husband** was this time in charge of pushing the baby stroller **and I** was on to do some cleaning…. Here Id like to mention I could NOT possibly gather every garbage I see on **our** way. Complete cleaning of an area of 10x10m would take an hour or more. Two hundred meters into **our** journey and I already had filled one big 1m wide and 1,50m trash bag |

At this point, some referring expressions help to recover the reference of the pronoun. The reference of pronoun ‘us’ in (a) might be the source of controversy. The reference is not clear, and I strongly believe that it is made exophorically. The term exophoric means that the reference is not stated explicitly in the text; it is implied. No anaphoric reference was detected. It is also cataphorically difficult to recover the pronoun as the distance is noted. As for this, multi-interpretation is more likely to occur. The reference may be; the author and the reader, or the author and her family (but the distance between the pronoun and the reference suspect is too far), or people from western hemisphere (as compared to the eastern hemisphere).

Many disagreement in response of her writing believe that the word ‘us’ refers to the latest, as she is from one of the countries in Europe. This derives from readers looking at her national background. She is a Bulgarian. This national background is proved to affect the responses. In addition, she frequently made comparison of everything to western style, for instance, bathing. For some responses, that what might justify the assumption that she was a ‘racist’ blogger.

This pronoun resolution is quite different from ‘our’ in (b) that refers to the referring expression- ‘active’ citizens of Bandung (this term is aimed at people in Bandung who intended to steal the trash bin, but since it is composed of concrete, they destroyed it) - and in (c) that refers to the author and her husband. For these the references were made explicit and specific.

 What is quite significant is the reference she made for 2nd person singular. As it has been commented they are not too dominant. Zero occurrences are made for female, and only nine (9) are made for male. Among the nine, seven (7) is made and among seven, six are made into one person, and another one is for animal. Consider table 4:

**Table 4. The context of ‘he’**

|  |  |
| --- | --- |
| a | **A rat,** longer than 20 cm, running free in the house, eating whatever **he** can get (sometimes the cat) is a daily routine |
| b | Now there is a new mayor, **Ridwan Kamil,** who got a foreign (USA) education. **He** is trying to change Bandung into a place, where humans can live \*. **He** is trying to do everything in the same time  get the filthy food carts off the streets, plant some gardens and find a solution for the trash. |
| c | I’m not much into politics, but I kind of like **the new mayor**. You can see that idealistic belief in him, that **he** can make a change. I think Bandung and Indonesia in general need more people like that. Yet again, Im not interested in politics at all. I dont know if **he** is that good or **he** just got me hooked up on another lie. Politics tend to do that. |

Table 4 (a) shows that the reference of ‘he’ is made for a rat. She was trying to describe how filthy the lifestyle so that rat can roam freely. The use of ‘he’ (to refer to a rat) is actually not completely correct since it is an animal (though animals can be male or female). The reason for using male pronoun (to a negative polarity animal) is still unclear. The reason for this might be the fact that she is a female or another reason.

 The reference for (b) and (c) are the same. It is made to Ridwan Kamil, the new major in Bandung. The comments are positive here (as shown by the context). The foregrounding of Ridwan is made clear by the relative clause ‘who got foreign (USA) education’. This can be just an act of giving information, but it can also be an implication to say that it takes someone from western hemisphere or someone who got foreign (again western) hemisphere that at least attempts to do positive things.

The foregrounding of USA Education addresses positively to Ridwan made some readers generalized that it takes someone from western hemisphere (like the author of the blog), or someone who already got foreign education (like Ridwan Kamil) to clean Bandung.

It is interesting to see the distributional plot of the context of ‘he’ that refers to Ridwan Kamil. It is located on the beginning, and almost they very end. In the beginning the attitude is completely positive, but at the very end, she stated her doubt as she said ‘I dont know if **he** is that good or **he** just got me hooked up on another lie. Politics tend to do that’. Third person plural pronouns occurrences is quite high; they (27), them (15), their (16).

**Table 5. The Context of ‘They’, ‘their’ and ‘them’**

|  |  |
| --- | --- |
| a | **Bandung the city where people** think pig meat is too dirty to eat, yet **they** live in more filth than the animal itself. It could be a lovely city. Yet it looks (and smells) like a disaster area. You see, in here **they** can’t have a normal trash can like us.  |
| b | **They** just kept throwing their garbage AROUND the trash cans, on the road or just next to their house, turning it into a decaying, stinky, slimy hill, where **their** children play. After that the three groups united in raising a voice of complain that its dirty, because of the government. A smaller group blamed God, other blamed communism.  |
| c | Give **them** awareness that their trash shouldnt be on the street. I am one person, I dont have networking or influence on people. **They** just see me as some crazy white woman, which makes me feel sad. But the mother in me knows that change should happen in the mind of the individual first and until **Indonesian people** are ready for change no one can force **them** into it. So I’ll pack my cases and move my family away from this toxic, hazardous place and just pray that people in here learn a simple rule: |

In table 5, you can see that ‘they’ refer to people in Bandung. It indicates that they are associated with, filth, dirt and ignorance when it comes about the environment. A generalization is made, and this what made some people responded negatively. The phrase ‘think pig meat (pork) is too dirty to eat’ implied multiple references, and Muslim and Jewish are two of them as for them eating pork is prohibited. However, since the population of Muslim in Bandung is extremely high as compared to Jewish, the point of generalization goes to the first. Therefore, those who live more filth than the animal, refers to the same reference. The identity that she created for a particular identity is shown by these expressions.

The generalization is enlarged as she also mentioned ‘Asian’. Consider the reference of ‘they’ as shown in (a). In (b), the result is not much different. Alhough ‘they’, ‘their’ or ‘them’ (see c and d) specifically narrow down to people who behave negatively, still some negative comments root from generalization she made first.

 Interestingly, she narrowed the reference down from Asian to Indonesian in (c). The pattern can be described as follow: the people of Bandung >> people who are careless and ignorant about the environment >> Asians >> people who are careless and ignorant about the environment >> Indonesians. How people responded and believed that this posting is racist might derive from the generalization she had made (people of bandung, Asians and Indonesians).

 Marking in-group identity can be done in different ways: exclusion, inclusion, or both. When the author did exclusion, she can use pronouns like ‘they’, ‘them’ and etc. Inclusion can be done by using pronouns like, ‘we’, ‘our’ and etc. In this posting none of the inclusion literally refers to Europeans or people from western hemispheres; it is implied. There are two possibilities about why her posting is considered racist. One, it may derive from some pronouns where the reference is not clear. Readers took the decision to connect this to his nationality background. Second, it may derive from the reference pattern she made where generalization occur in the beginning and the end (some in the middle). Even though most of the references are made specific but people see the generalization (Muslim, people of bandung, Indonesians, and Asians), and it strongly affects the judgment. The fact that the number of generalization is not as many as the specific reference does not stop some people to think that this posting is racist.

1. **Conclusion and Recommendation for Further Research**

The exclusivity of one group over another is often shown by different means: but one thing for sure, language is the media of its delivery. Quantitatively the use of in-group identity pronoun (and the opposite) is not many (as shown by dispersion plot); however, it greatly affects how people reacted as it can be seen from the comments. However, the group that she refers to can linguistically be multi interpretable as the reference is sometimes made exophorically. Decision to respond writing is not only made under the influence of the content, but also from the way it is written. And often, the controversy does not particularly derive from the content, but the style of how it is written. Blog is an open-access facility where everyone can write and respond. Internet users can always respond to a blog posting in a variety of ways; but unfortunately, it is not always about the content. The style of a writing must be well considered before it is posted online: one of the linguistic aspects that require careful consideration is the preference of pronouns. Another aspect that is quite important to research is the referring expressions as It has wider coverage than pronoun. However the interpretation of pronoun or referring expression still rely heavily on metalinguistics aspects; therefore a more detailed qualitative analysis is required, not only to the text, but also to the author and the readers as well as background is a crucial importance.

# Bibliography

Adolphs, S. (2006). *Introducing Electronic Text Analysis.* London: Routledge.

Anthony, L. (2006). Concordancing with AntConc: An introduction to tools and techniques in corpus linguistics. *JACET Newsletter* , 155-185.

Brown, R., & Gilman, A. (1960). The Pronouns of Power and Solidarity. Dalam T.-A. Sebeok, *Style in Language* (hal. 253-276). Cambridge: MIT Press.

Coleman, S., & Ross, K. (2010). *The Media and Public: Them and Us in Media Discourse.* Sussex: Wiley-Blackwell.

Conrad, S., & Biber, D. (2008). Register Variation: A Corpus Approach. In D. Schriffin, D. Tannen, & H. Hamilton, *The Handbook of Discourse Analysis* (pp. 175-197). Oxford: Blackwell.

De fina, A., Schriffin, D., & Bamberg, M. (2008). *Discourse and Identity.* Cambridge: Cambridge University Press.

Edwards, J. (2009). *Language and Identity.* Cambridge: Cambridge University Press.

Halliday, J., & Hasan, R. (1976). *Cohesion in English.* London: Blackwell.

Harwood, N. (2005). Inclusive and Exclusive Pronoun in Academic Writing. *Applied Linguistics 26 (3)* , 343–375.

Mathesson, D. (2005). *Media Discourses.* Berkshire: McGraw-Hill.

Mc Enery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice.* Cambridge: Cambridge University Press.

Schriffin, D. (2008). From Linguistic Reference to Social Reality. In A.-D. Finna, D. Schriffin, & M. Bamberg, *Discourse and Identity* (pp. 103-134). Cambridge: Cambridge University Press.

Sinclair, J. (2006). *Trust The Text: Language, Corpus, and Discourse.* London & New York: Blackwell.

Wolf, F., Gibson, E., & Desmet, T. (2004). Discourse coherence and pronoun resolution. *Language and Cognitive Processes, 19(6)* , 665-675.