
A Corpus-based Study of the Stereotypical Construction of Sundanese People

Susi Yuliawati

Department of Linguistics, Faculty of Cultural Sciences, Universitas Padjadjaran, Bandung, Indonesia

Email address:

susi.yuliawati@unpad.ac.id

To cite this article:

Susi Yuliawati. A Corpus-based Study of the Stereotypical Construction of Sundanese People. *Communication and Linguistics Studies*.

Vol. 6, No. 3, 2020, pp. 40-46. doi: 10.11648/j.cls.20200603.11

Received: July 7, 2020; **Accepted:** July 27, 2020; **Published:** August 13, 2020

Abstract: This article uses corpus linguistics methods and theories to study how the Sundanese depicted as people with courteous characters in a 2.9 million-word corpus of *Manglé*, a Sundanese magazine, published between 1958 and 2013. The study examines the usage patterns of Sundanese words denoting ‘courtesy’ and ‘discourtesy’ in the corpus by employing a mixed-method research design. Using the corpus software WordSmith Tools, the analysis of word frequency found that the courtesy category is lexically more diverse, i.e., containing more lexical units, than the discourtesy category. Besides, the courtesy lexemes are more frequently used than the discourtesy lexemes. Based on collocation analysis, the top three most frequent words signifying courtesy, i.e., SOMÉAH ‘nice and welcome’, MARAHMAY ‘cheerful’, and DARÉHDÉH ‘pleasant and friendly’, have the semantic preference of friendliness; social actions, states, and processes; and people. On the other hand, the semantic preference of the top three most frequent words signifying discourtesy, i.e., BAEUD ‘sullen’, JAMEDUD ‘surly’, and KURAWAUD ‘surly’, is predominantly unfriendly traits. The analyses demonstrate that Sundanese people in the corpus of *Manglé* are constructed as a friendly community portrayed to have some personality traits such as favorable, friendly, and welcoming, particularly to visitors and strangers. The result seemingly constructs the stereotype of the Sundanese ethnic group that is commonly known among the other ethnic groups in Indonesia as respectful and friendly people.

Keywords: Collocation, Corpus Linguistics, Lexemes, Stereotype, Sundanese

1. Introduction

With the population of 36,701,670, Sundanese is the second-largest ethnic group in Indonesia, predominantly living in the western part of Java Island. Indonesia itself is the third populous country in Asia after China and India and also home to 1,331 ethnic groups, according to the 2010 Population Census. It makes the country among the world's most diverse. Among the other ethnic groups, Sundanese people are stereotypically recognized to be friendly, warm, and polite. As stated by a prominent Sundanese critic, Ajip Rosidi, Sundanese people are well known to be amiable, cheerful, kind-hearted, and having a good sense of humor [1]. In line with that, Sampeliling also claimed that the Sundanese people's stereotype is gentle, polite, respectful, and brave [2]. Thus, many Sundanese elites became famous diplomats of Indonesia, and the most popular ones are Mochtar Kusumaatmadja and Marty Natalegawa. Based on beliefs arising from a culture that constructs the ethnic stereotypes,

the present paper studies the identity of the Sundanese community from a corpus linguistics perspective, by investigating words used in media that depict the Sundanese people concerning courteous and discourteous characters.

According to Wierzbicka, there is a close connection between the lexicon of a language spoken by society and the life of it [3]. The notion indicates that vocabulary is a substantial index of the culture of a society. In other words, vocabulary may provide precious clues to the apprehension of culture. A notable example may be recognized when we find difficulty in finding an equivalent word in other languages. For example, English does not have a word corresponding to the Sundanese noun *seseherahan*. The word refers to a traditional ceremony when the bridegroom family brings gifts such as clothes, footwear, food, religious articles, and make-up set to the bride and her family before the wedding ceremony, signifying his responsibility as the head of the prospective family. English also does not have a word corresponding to the Japanese word *miai*, which refers to a

formal occasion when the prospective bride and her family meet the bridegroom and family for the first time [3]. The example of these cultural-specific words reflects ways of living and ways of thinking of a particular society. Therefore, language is regarded as a remarkable device for the construction and maintenance of culture. Through language, people create and share beliefs, values, attitudes, identities, and categories. An investigation into how society uses language thus can expose the critical aspects of culture.

Bednarek and Bublitz stated that one of the well-established approaches to comprehend the culture of a society is to investigate the vocabulary of its language, which serves a dual function, i.e., to reflect as well as to define the cultural concepts of a society [4]. They also argue that the function is indispensable in establishing and reinforcing the system of ideological beliefs and values that build the cultural identity of society. For example, they studied a fun-related ideology in US-American and British cultures by examining the usage of cultural keyword *enjoy* in corpora of US-American and British English by using a corpus linguistic approach. The analyses show that the usage pattern of the word *enjoy* supported the notion that grammar promotes the ideology of fun. All the samples of the word use in the contexts constructed the message: 'having fun is good' and had led to an establishment of the cultural pattern of *having fun* as a natural and elemental socio-cultural asset in the cultures of the UK and US.

The other research studying culture through the investigation of words was conducted by Schönefeld [5]. It is a cross-linguistic study on the cultural concept of HOT by investigating three keywords in English (E), Russian (R), and German (G), i.e., E *hot*, R *gorjač/žark*, and G *heiß*, using a corpus analytical technique, namely collocation, a technical term referring to habitual co-occurrences of words reflecting repetitive experiences of the speakers/writers as well as their culturally shared knowledge. Based on the data taken from British National Corpus (BNC), Russian Corpus of Tu bingen University, and the IDS Mannheim: COSMAS II, they found significant differences in what speakers of English, Russian, and German associate with the respective forms of HOT. The differences were undeniable evidence that the languages at issue did not all apply the same metaphorical mapping indicating that they have different cultural and folk models as a basis for specific conceptualizations.

Unlike Bednarek and Bublitz and Schönefeld, Millezia studied cultural differences based on words in use from a specific type of discourse, i.e., political discourse, in American English, British English, and Italian [4-6]. She analyzed the usage pattern of the words *terror/terrore* and *terrorism/terrorismo* in a spoken corpus built from the speeches of George W. Bush, Tony Blair, and Silvio Berlusconi during 2005. They revealed that the co-occurrence patterns of the words under investigation in political discourse varied across cultures. The American and the British cultures used two different phrases to denote the same concept, i.e., *war on terror* and *fight against terrorism* respectively, and the Italian culture used *lotta al terrorismo* that was similar to the

British way. The study also revealed that Bush used the word *terror* more frequently than *terrorism*, while Blair and Berlusconi more frequently used the word *terrorism* and *terrorismo*. It indicates that the words and phrases chosen by people demonstrate their beliefs, expectations, evaluations, and the universe of discourse.

The other related study using the Sundanese language to study the relationship between language and culture is from Yuliawati and Hidayat [7]. Their research studied the construction of women in the corpus of Sundanese magazine *Manglé* by investigating the usage of five Sundanese nouns denoting women, i.e., *geureuha*, *mojang*, *pamajikan*, *wanita*, and *wanoja*, spanning from 1958–2013. They found that among the five words denoting women, *wanoja* was the only word with steadily increasing frequency in the magazine. Based on the analyses of collocation and semiotics, the study also found that *Manglé* increasingly depicted *women* as independent, i.e., their presence was closely related to their existence in the public sphere. The finding supported the notion proposed by Cameron that gender construction is mediated through language and discourse [8]. In other words, language is a powerful device to construct meaning socially.

Based on the previous research, the present study aims to study the stereotype of Sundanese people using the corpus-based approach by examining the lexemes denoting 'courtesy' and 'discourtesy' in the available Sundanese corpus, i.e., the corpus of the Sundanese magazine *Manglé*, published between 1958 and 2013. The present writer is aware of the tendency that scholars dominantly regard stereotypes, including ethnic stereotypes, as unfavorable because they are potential to lead to conflicts [9-15]. However, according to the opposite opinion, stereotypes are conceptualized as positive. They serve as the first step to the contact between cultures; they prepare people from one culture to potential clashes with other cultures. Concerning that matter, I expect the present research contributes to the study of ethnic stereotypes examined through a linguistic perspective, particularly corpus-based study. Furthermore, the result may give an in-depth understanding of the cultural characteristics of the Sundanese ethnic group.

2. Methods

Corpus-based research generally involves a qualitative analytical technique and a quantitative one to analyze real patterns of use in natural texts. Thus, the present study employs the research approach that is widely known as a mixed-method research design. The research design integrates quantitative and qualitative approaches to provide a deeper understanding of a research problem than either approach alone. As stated by Greene et al., combining the two paradigms in research is valuable for building a comprehensive description and providing answers to a broader range of research questions [in 16]. About the design model, the present writer at first performed quantitative research. Then, I analyzed the result and built an explanation of the result with the qualitative research. The quantitative

analysis was utilized in two stages: word frequency analysis to identify word occurrences in the corpus and the significance test of MI score ≥ 5 within a 4–4 window span and with a minimum frequency of 3 to determine significant collocates. To help carry out the statistical measurement, I used a corpus tool, namely WordSmith Tools 6.0. The significant collocates of words denoting ‘courtesy’ and ‘discourtesy’ were then analyzed using semantic preference theory to create words’ semantic profiles, used as the basis to discuss the construction of the stereotype of Sundanese people in the texts.

For the present research, I used the available Sundanese corpus constructed by Yuliawati [17]. The corpus was built from texts in a Sundanese magazine, *Manglé*, published between 1958 and 2013. By employing a calculator size sampling and the technique of proportional cluster random sampling, 92 editions of the magazine were compiled to construct the corpus of *Manglé*. The size of the corpus is 2,940,537 words and 131,570 types (distinct words). The magazine is regarded monumental in the development of Sundanese media mainly because of its objective to preserve Sundanese culture. Besides, in Indonesia, *Manglé* is the longest-running media written in a local language that has been in a continuous publication from 1957–present, i.e., the magazine has run in more than half of the century. In the beginning, *Manglé* was a monthly magazine. Then, in 1965 it changed into a bi-monthly magazine; in 1969, it published three times a month, and since 1971 it has published every week. The magazine contains several rubrics such as entertainment and human interest, history and culture, religion and education, and news reports.

As stated by Biber and Reppen, the corpus linguistics approach is associated with four major characteristics. First, the research is empirical to describe the actual patterns of language in use. Second, the research investigates a big and principled collection of natural text, known as corpus. The corpus designed and constructed represents a target domain of language use. Third, the research involves a far-reaching use of computer analysis employing either automatic or interactive techniques. Fourth, the research commonly combines quantitative and qualitative analyses [18]. Due to these characteristics, research employing corpus linguistics as a methodology is an empirical study of language in use which findings have greater generalizability and validity than would otherwise be achievable. Corpus linguistics approach is also applied to study language from many different perspectives such as phonology, morphology syntax, semantics, pragmatics, and sociolinguistics.

Corpus linguistics has several distinctive analytical techniques, such as word frequency, collocation, and semantic preference. The approach regards that word meanings are often created by the associations that the words participate in, alongside other words they frequently co-occur, rather than by the words in isolation [19, 20]. In this case, words tend to appear with certain words accompanying them in particular contexts, indicating the patterns of co-selected words that speakers and/or writers conform to [19]. Thus, the approach considers meaning as a social construction [17]. A corpus

analysis to identify meaning based on this principle is known as the analysis of collocation. The term refers to a lexical relation between two or more words co-occurring within a few words of each other in running text. For example, the word PROVIDE frequently co-occurs with words referring to precious things that people need, such as *help* and *assistance*, *money*, *food* and *shelter*, and *information* [21]. In this case, the word PROVIDE is known as the node word, the word being investigated, while the words *help*, *assistance*, *money*, *food*, *shelter*, and *information* are called the collocates, the co-occurring words in the corpus.

Therefore, the word meaning can be described by using the collocational analysis. According to Stubbs, the collocational meaning resulted from the shared semantic features in a set of collocates, and its node refers to semantic preference [19]. Therefore, he stated that semantic preference is related to a lemma or word form and a set of semantically related words. For instance, in the *British National Corpus* (BNC), the semantic preference of the word RAISING is work and money, determined from its collocates such as *income*, *prices*, *wages*, *earning*, and *unemployment* [22]. Concerning the semantic preference analysis, this study used semantic categories from UCREL Semantic Analysis System (USAS) to create the semantic profiles of words referring to ‘courtesy’ and ‘discourtesy’ in the *Manglé* corpus.

3. Results and Discussion

The stereotype of Sundanese people as friendly people may result from the social values and cultural teachings, derived from the indigenous knowledge of the Sundanese community. Some Sundanese proverbs, for example, say *luhur budi handap asor* ‘Sundanese people have to be virtuous and modest’ and *soméah hade ka semah* ‘Sundanese people have to be respectful to guests’. They respectively teach the Sundanese people to be virtuous and humble, and also to be considerate and amiable to others from various backgrounds. Such proverbs, which are a part of the indigenous knowledge system, are the underlying foundation for the people of Sunda behaving in a certain way. This reason, the study investigates the stereotypical construction of courteous Sundanese people from the usage patterns of words denoting ‘courtesy’ and ‘discourtesy’ in the *Manglé* corpus. The research is expected to provide linguistic evidence that constructs the Sundanese stereotype as a friendly community.

The present study is corpus-based research rather than a corpus-driven one, which means that the words selected for examination were determined by the researcher, rather than chosen from, for instance, a top-ten list of frequent words. In this case, the words were selected based on the combination of the researcher’s intuition, interviews with some Sundanese natives, and dictionary. Grounded in words collected from the above information, I found nine words denoting courtesy and seven words denoting discourtesy in the *Manglé* corpus, as shown in Table 1.

The courtesy lexemes are *soméah* ‘nice and welcome’, *marahmay* ‘cheerful’, *daréhdeh* ‘pleasant and friendly’, *akuan*

‘welcome’, *amis budi* ‘amiable’, *galéhghéh* ‘nice and welcome’, *géréccék* ‘talkative and genial’, *suranyéh* ‘friendly and considerate’, and *saréséh* ‘nice and welcome’. On the other hand, the discourtesy lexemes are *baeud* ‘sullen’, *jamedud* ‘surlly’, *kuraweud* ‘surlly’, *camberut* ‘sullen’, *jamotrot* ‘sulky’, *baketut* ‘dour’, and *ngagadeud* ‘unfriendly expression’.

Table 1. The frequency of words denoting courtesy and discourtesy.

No	Courtesy	Freq.	No.	Discourtesy	Freq.
1	<i>soméah</i>	125	1	<i>baeud</i>	68
2	<i>marahmay</i>	108	2	<i>jamedud</i>	31
3	<i>daréhdéh</i>	46	3	<i>kuraweud</i>	28
4	<i>akuan</i>	12	4	<i>camberut</i>	17
5	<i>amis budi</i>	10	5	<i>jamotrot</i>	17
6	<i>galéhghéh</i>	7	6	<i>baketut</i>	14
7	<i>géréccék</i>	2	7	<i>ngagadeud</i>	12
8	<i>suranyéh</i>	1	8	-	-
9	<i>saréséh</i>	1	9	-	-

From the analysis of word frequency, it does not only show that the courtesy category is lexically more abundant than the discourtesy category, but it also demonstrates that in total, the courtesy lexemes occur more frequently than the discourtesy lexemes. If the total frequency of words in both categories is calculated in percentage, the percentage of the courtesy lexemes is 63%, while the discourtesy lexemes are 37% in the Manglé corpus (shown in Figure 1). In other words, the texts in the Manglé corpus discuss courteous characters more frequently than discourteous characters.

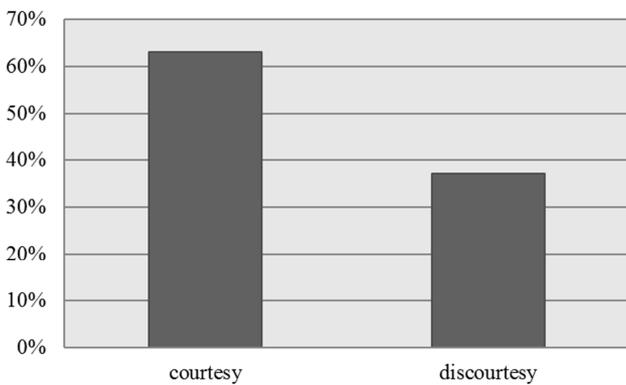


Figure 1. The total frequencies of the courtesy and discourtesy lexemes in percentage.

Considering the contents of the magazine that dominantly discuss Sundanese culture and also the people, it suggests that the concept of courtesy is fundamental for the Sundanese that conditions the way they think and act, as stated by Sapir that the particular language that the people speak much influence their conceptualization of the world [in 23]. On the other hand, language is viewed as a symbolic guide to culture and has a significant role in constructing reality [24, 25]. Thus, the result of frequency analysis may indicate that the courtesy lexemes

used in the *Manglé* corpus have constructed the Sundanese people's stereotype as friendly people.

As presented in Table 1, the top three most frequent words in courtesy category are *soméah* (125), *marahmay* (108), and *daréhdéh* (46), while the words in discourtesy category are *baeud* (68), *jamedud* (31), and *kuraweud* (28). To obtain a deeper understanding of the conceptualization of courtesy and discourtesy in Sundanese, collocation analysis is used to create semantic profiles of the top three most frequent words in both categories. As already explained in the methods, to investigate word meanings based on the pattern of collocation, a lexical relation between a word and its co-occurring words, the study uses a significance test of MI within a 4–4 window span. A co-occurring word is categorized as a significant collocate if the MI score is less than or equal to 5, and the frequency is 3 at a minimum.

Table 2. The collocates of the top three most frequent words in the courtesy category.

<i>soméah</i>	MI	<i>marahmay</i>	MI	<i>daréhdéh</i>	MI
<i>pikaconggaheun</i>	13.31	<i>béar</i>	13.63	<i>akuan</i>	14.39
<i>akuan</i>	12.95	<i>pasemon</i>	13.08	<i>soméah</i>	12.01
<i>ramah</i>	12.42	<i>paromanna</i>	12.97	<i>amis</i>	10.74
<i>daréhdéh</i>	12.01	<i>pasemonna</i>	12.26	<i>budi</i>	9.33
<i>béar</i>	11.42	<i>hégar</i>	11.58	<i>semah</i>	8.84
<i>semah</i>	10.57	<i>budi</i>	9.83	-	-
<i>marahmay</i>	9.78	<i>soméah</i>	9.78	-	-
<i>sopan</i>	9.48	<i>amis</i>	9.51	-	-
<i>ngabageakeun</i>	9.06	<i>beungeutna</i>	8.99	-	-
<i>hormat</i>	8.94	<i>témbong</i>	8.72	-	-
<i>hadé</i>	8.70	<i>némbongkeun</i>	7.84	-	-
<i>budi</i>	8.30	<i>beunget</i>	7.56	-	-
<i>bageur</i>	7.88	<i>katémbong</i>	7.56	-	-
<i>pribumi</i>	7.39	<i>imut</i>	6.69	-	-
<i>geulis</i>	7.28	<i>hade</i>	5.98	-	-
<i>handap asor</i>	6.88	-	-	-	-
<i>imut</i>	6.48	-	-	-	-

Table 2 lists the significant collocates of the courteous lexemes: *soméah*, *marahmay*, and *daréhdéh*, generated by WordSmith Tools within the threshold mentioned above. Among those words, *marahmay* is the node word with the highest number of significant collocates, while *daréhdéh* has the least number of significant collocates. The result justifies the research conducted by Shin and Nation who discovered that the more frequent the node word, the higher the number of collocates [in 26]. In this case, the node word *soméah* occurs more frequently than *marahmay* in the corpus, and *soméah* also has a higher number of collocates than *marahmay*. In brief, it suggests that the word *marahmay* occurs with more different words than *soméah*.

Using the USAS semantic analysis, the collocates of the node word *soméah*, *marahmay*, and *daréhdéh* can be grouped into several semantic categories.

Table 3. The semantic categories of the lexemes SOMÉAH, MARAHMAY, and DARÉHDÉH

A. SOMÉAH	
friendliness	<i>pikaconggaheun</i> ‘familiar’, <i>akuan</i> ‘nice and welcome’, <i>ramah</i> ‘friendly’, <i>daréhdéh</i> ‘pleasant and friendly’, <i>béar</i>

respect	'cheerful', <i>amis</i> 'agreeable', <i>marahmay</i> 'cheerful', <i>bageur</i> 'warm-hearted'.
people	<i>hormat</i> 'respectful', <i>handap asor</i> 'respectful'
happy	<i>sémah</i> 'guest', <i>pribumi</i> 'host'
politeness	<i>ngabageakeun</i> 'to delight', <i>imut</i> 'smile'
evaluation/good	<i>sopan</i> 'polite', <i>hadé</i> 'good'
social actions, states & process	<i>budi</i> 'manners'
judgment of appearance	<i>geulis</i> 'beautiful'
B. MARAHMAY	
body	<i>pasemon</i> 'face', <i>paromanna</i> 'his/her face', <i>pasemonna</i> 'his/her face', <i>beungeutna</i> 'his/her face', <i>beunget</i> 'face'
friendliness	<i>béar</i> 'cheerful', <i>soméah</i> 'friendly and welcome', <i>amis</i> 'agreeable'
seem/appear	<i>tembong</i> 'to appear', <i>nembongkeun</i> 'appearing', <i>katembong</i> 'seem to be'
happy	<i>hégar</i> 'happy', <i>imut</i> 'smile'
social actions, states & process	<i>budi</i> 'manners'
evaluation/good	<i>hadé</i> 'good'
B. DARÉHDÉH	
friendliness	<i>akuan</i> 'nice and welcome', <i>soméah</i> 'friendly and welcome', <i>amis</i> 'agreeable'
social actions, states & process	<i>budi</i> 'manners'
people	<i>sémah</i> 'guest'

The semantic categories demonstrate that the word *soméah* has the semantic preference of friendliness, respect, people, politeness, happy, evaluation/good, a judgment of appearance, and social actions, states, and processes. The most compelling evidence demonstrates that *soméah* is strongly associated with words that convey friendliness (e.g., *pikaconggaheun* 'familiar', *akuan* 'nice and welcome', *ramah* 'friendly', *daréhdéh* 'pleasant and friendly', *béar* 'cheerful', *amis* 'agreeable', *marahmay* 'cheerful', and *bageur* 'warm-hearted'), and thus the prevalent semantic preference of *soméah* is friendliness. Another critical point is that *soméah* is closely associated with people in a particular relation, i.e., a host-guest relationship (e.g., *sémah* 'guest', *pribumi* 'host'). It indicates that a host-guest relationship is the marker of the word usage of *soméah*.

Unlike *soméah*, the word *marahmay* is more strongly associated with words relating to human body predominantly the face (*pasemon* 'face', *paromanna* 'his/her face', *pasemonna* 'his/her face', *beungeutna* 'his/her face', *beunget* 'face') than with words depicting friendliness (*béar* 'cheerful', *soméah* 'friendly and welcome', *amis* 'agreeable'). Besides, *marahmay* also has the semantic preference of seem/appear (*tembong* 'to appear', *nembongkeun* 'appearing', *katembong* 'seem to be') denoting that *marahmay* is a friendly characteristic that is noticeable especially from the face. These are some samples of the word usage taken from the concordance demonstrating this concept.

Meanwhile, the usage of *daréhdéh* in the corpus is more similar to *soméah*. It has the semantic preference of friendliness (*akuan* 'nice and welcome', *soméah* 'friendly and welcome', *amis* 'agreeable'), social actions, states & process (*budi* 'manners'), and people (*sémah* 'guest'). It suggests that the word *soméah* is closely associated with a friendly characteristic, particularly in a host-guest relationship. However, *daréhdéh* has less semantic preferences than *soméah*, which means that the word *daréhdéh* has less semantic categories than *soméah*. Generally speaking, the words *soméah*, *marahmay*, and *daréhdéh* are mainly used in the corpus to denote friendly traits expressed particularly through manners (*budi*) in the context of a host-guest relationship. However, the word *semah* does not only refer

literally to guests, but also to others from different backgrounds.

Concerning the lexemes in the discourtesy category, the collocation analysis focuses on the words *baeud*, *jamedud*, and *kuraweud*. The result shows that *baeud* is the word with the highest number of collocates, while the lowest number of collocates is found in the word *jamedud*, as presented in Table 4. Because the occurrences of the discourtesy lexemes in the corpus are much less than the courtesy lexemes, the number of collocates of the discourtesy lexemes is also small.

Table 4. The collocates of the top three most frequent words in the discourtesy category.

<i>baeud</i>	MI	<i>jamedud</i>	MI	<i>kuraweud</i>	MI
<i>jamedud</i>	12.05	<i>baued</i>	12.05	<i>haseum</i>	13.19
<i>Emin</i>	10.84	-	-	<i>Pa</i>	6.28
<i>ngadengekeun</i>	10.63	-	-	-	-
<i>imut</i>	7.78	-	-	-	-
<i>semu</i>	7.60	-	-	-	-
<i>seuri</i>	7.23	-	-	-	-

Based on the USAS semantic analysis, these are the semantic categories for the node words *baeud*, *jamedud*, and *kuraweud*.

Table 5. The semantic categories of the lexemes BAEUD, JAMEDUD, and KURAWAUD.

A. BAEUD	
politeness	<i>jamedud</i> 'surly'
happy	<i>imut</i> 'smile', <i>seuri</i> 'laugh'
body	<i>semu</i> 'face'
sensory	<i>ngadangukeun</i> 'listening'
B. JAMEDUD	
politeness	<i>baeud</i> 'sullen'
C. KURAWAUD	
politeness	<i>haseum</i> 'surly'
people	<i>Pa</i> 'Sir/Mr'

The word *baeud* has the semantic preference of politeness, happiness, body, and sensory. On the other hand, the semantic preference of *jamedud* is only politeness, and *kuraweud* is politeness and people. All node words are strongly associated with words depicting the level of politeness, which in this case is an unfriendly feature (*jamedud* 'surly', *baeud* 'sullen',

haseum 'surly'). The significant difference is that the word *baeud* in the *Manglé* corpus is strongly associated with a word depicting unfriendly features and with words that are in opposition, i.e., *imut* 'smile' *seuri* 'laugh'.

Different from lexemes in courtesy category, the discourtesy lexemes are rather difficult to interpret further because they do not have enough significant collocates to examine resulted from their low frequencies in the *Manglé* corpus. However, this is the linguistic evidence of courtesy/discourtesy concept revealed from the real samples of language used in Sundanese, which were recurrently co-selected by the speakers/writers. From the *Manglé* corpus, the study demonstrates that courteous characteristics of the Sundanese are more intensely discussed than the discourteous characteristics. It can be seen not only in terms of the word occurrences in the corpus, but also the lexical diversity and meanings. All things considered, the final analysis suggests that these findings provide compelling linguistic evidence that the words in courtesy category construct the identity of the Sundanese more strongly than the words in discourtesy category, as stated by Cameron language is a powerful device to socially construct meaning [8]. Additionally, the findings seem to support the stereotype of the Sundanese people that are widely known as courteous people among other ethnic groups in Indonesia.

4. Conclusion

The present study analyzed the *Manglé* corpus to examine Sundanese words' usage patterns denoting 'courtesy' and 'discourtesy'. The study focuses on the analyses of word frequency collocation and semantic preference. The results demonstrate that words in the courtesy category are lexically more diverse than the discourtesy categories, i.e., there are nine words denoting courtesy and seven words denoting discourtesy. Besides, the total courtesy lexemes occur 63% while the discourtesy lexemes are 37%, indicating that the words in courtesy category are more frequently used than the words in the discourtesy category. With the further focus on the top-three most frequent words in both categories, the courtesy words of SOMÉAH, MARAHMAY, and DARÉHDÉH have the semantic preference of friendliness; social actions, states, and processes; and people, while the discourtesy words of BAEUD, JAMEDUD, and KURAWÉUD predominantly have the semantic preference of unfriendly characteristics. The analyses demonstrate that Sundanese people in the corpus of *Manglé* are constructed as a friendly community portrayed to have some personality traits such as favorable, friendly, and welcoming, particularly to guests and strangers. The finding seemingly constructs the stereotype of the Sundanese ethnic group that is commonly known among the other ethnic groups in Indonesia as respectful and friendly people. As a final point, the present study argues that language is a mechanism that plays a significant role in constructing meanings.

References

- [1] Rosidi, A. (2009). *Manusia Sunda*. Indonesia, Bandung: Kiblat.
- [2] Sampeliling, A. R. (2017, November 13). *Berbagi stereotip kelompok etnis di Indonesia: Membangun atau merugikan*. Retrieved from <http://connect.upnyk.ac.id>.
- [3] Wierzbicka, A. (1997). *Understanding Culture through Their Key Words: English, Russian, Polish, German, and Japanese*. New York: Oxford University Press.
- [4] Bednarek, M. & Bublitz, W. (2007). *Enjoy! The (phraseological) culture of having fun*. In P. Skandera (Eds.), *Phraseology and culture in English*. Berlin & New York: Mouton de Gruyter.
- [5] Schönefeld, D. (2007). *Hot, heiß, and gorjachij: A case study of collocations in English, German, and Russian*. In P. Skandera (Eds.), *Phraseology and Culture in English*. Berlin & New York: Mouton de Gruyter.
- [6] Millezia, D. (2006). *Classifying Phraseology in a spoken corpus of political discourse*. *ESP Across Culture*, 3, 41-65.
- [7] Yuliawati, S. & Hidayat, R. S. (2018). *Construction of woman in the Sundanese magazine Manglé (1958-2013): A corpus-based study of metalinguistic signs*. In M. Budianta, M. Budiman, A. Kusno, & M. Moriyama (Eds.), *Cultural Dynamics in a Globalized World*. London: Routledge.
- [8] Cameron, D. (1992) *Feminism and linguistic theory*. Basingstoke: Macmillan.
- [9] Brislin, R. (1993). *Understanding culture's influence on behavior*. Fort Worth: Harcourt Brace College Publishers.
- [10] Crystal, D. (1997). *Cambridge encyclopedia of the English language*. Cambridge: Cambridge University Press.
- [11] Gu, J. (2009). *Theorizing about intercultural communication: Dynamic semiotic and memetic approach to intercultural communication (a commentary)*. In eds G. M. Chen, R. Heisey. *Intercultural Communication Research* (pp. 109-115). Beijing: Higher Education Press.
- [12] Gudykunst, W. B., & Kim, Y. Y. (1995). *Communicating with Strangers: An approach to intercultural communication*. In ed. Stewart J. New York, McGraw Hill.
- [13] Kohls, L. R. (1984). *Survival kit for overseas living*. Yarmouth, M. E. Intercultural press.
- [14] Rosch, E. (1978). *Principles of categorization*. In eds E. Rosch, Lloyd, B. Hillsdale. *Cognition and categorization* (pp. 27– 48). New Jersey: Lawrence Erlbaum Associates Publishers.
- [15] Valdes, J. M. (1993) *Culture bound*. Cambridge, MA: Cambridge University Press.
- [16] Litosseliti, L. (2010). *Research method in linguistics*. London & New York: Continuum International Publishing Group.
- [17] Yuliawati, S. (2018). *Kajian Linguistik dan Semiotik: Perempuan Sunda dalam Kata*. Bandung, West Java, Indonesia: Refika Aditama.
- [18] Biber, D. & Reppen, R. (2015). *The Cambridge handbook of English corpus linguistics*. UK: Cambridge University Press.

- [19] Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford, New York, & Toronto: Oxford University Press.
- [20] Stubbs, M. (2002). *Words and phrases: Corpus studies of lexical semantics*. Blackwell Publishing.
- [21] Cheng, W. (2012). *Exploring corpus linguistics. language in action*. London & New York: Routledge.
- [22] Baker, P., Hardie, A. & McEnery, T. (2006). *A glossary of corpus linguistics*. Edinburg: Edinburg University Press.
- [23] Saeed, John. L. (2016). *Semantics*. USA & UK: Blackwell Publishers, Ltd.
- [24] Clair, R. N. (1986). *Language and the social construction of reality*. *Language Sciences*, 2 (2).
- [25] Bo, C. (2015). Social constructivism of language and meaning. *Croatian Journal of Philosophy*, 15 (43), 87–113.
- [26] Webb, S. & Kagimoto, E. (2010). *Learning collocations: Do the number of collocates, position of the node word, and synonymy effect learning?*. *Applied Linguistics*, 32 (3), 259-276.