

Examine persuasion strategies in Chinese on social media

Yu-Yun Chang
Graduate Institute of Linguistics
National Chengchi University
yuyun@nccu.edu.tw

**Po-Ya Angela Wang, Han-Tang Hung,
Ka-Sîng Khóo, and Shu-Kai Hsieh**
Graduate Institute of Linguistics
National Taiwan University
{d08142001, r07142007,
r07142006, shukaihsieh}@ntu.edu.tw

Abstract

Studies based on linguistic theories have reported that internet celebrities employ various persuasion strategies in their speeches. Since persuasion is closely related to pragmatics which involves people's cultural background or world knowledge and makes it harder for researchers to capture a specific pattern, various features are proposed by different researchers from a speaker's perspective. In this paper, we tackle whether these theory-based features can be implemented quantitatively with statistic measurements from both speakers' and readers' perspectives in large datasets. We have examined these features based on cohesion and coherence which can help investigate the language use of the internet celebrities syntactically, semantically and even pragmatically. The results present that the extracted features in this study have an effect in identifying the high-influence speakers, and can be further implemented to predict influencers on social media automatically.

1 Introduction

With the growth of the internet, more and more people become internet celebrities or opinion leaders of their social networks. It is found that these people have high influence on the public's ways of thinking or behaviors by delivering a powerful speech with different persuasion strategies, which express their own confidence to the points they made (Dillard and Marshall, 2003)

and attempt to change the opinions or behaviors of the audience (Halmari and Virtanen, 2005).

Although the manipulation of persuasion strategies by a speaker will have impacts on readers, it is not guaranteed that readers will behave as expected. Readers may perceive different susceptibilities based on their age, gender, culture, personality, and other cognitive factors (Orji et al., 2015; Ciocarlan et al., 2019; Oyibo et al., 2017). That is, each reader may present different degrees of effectiveness even to the same persuasion strategy. This kind of inconsistency further indicates that speakers' and readers' commitments to the proposed persuasion strategy may differ from each other. While making a discourse commitment, speakers would expect that readers follow and accept the expressed proposition or belief (Gunlogson, 2008; Hamblin, 1970; Stalnaker, 1978). On the other hand, after receiving the message delivered by the speaker, readers would decide whether their actions will behave as expected by the speaker (Katriel and Dascal, 1989). Thus, the process of being persuaded is regarded as a reciprocal behavior (Cialdini, 1987).

Various linguistic features for delivering persuasion strategies have been proposed. However, most of them focus on investigating from speakers' point of view instead of readers'. To assess whether the performed persuasion strategies have successfully evoked the expected behaviors from readers, the analysis need not to only consider the perspectives from speakers, but also from readers. Readers' responses can

be treated as a tool to assess the stance made by the speaker (Cornillie, 2018). If the readers agree with the speaker’s points of view, it is natural that readers will reward back with positive responses; and vice versa.

This study targets on examining persuasion from the perspective of which theory-based linguistic features implemented into the speech delivered by speakers can be mostly accepted by readers to attain the goal of persuasion. We explore persuasive features grammatically and semantically based on cohesion and coherence in discourse, and extract features automatically from large-scale datasets with the least human efforts. In this study, we inspect the features of persuasion strategies used in social media, and conduct an exploratory analysis with quantitative approaches for assessment. The assessed high-influence strategic features in this study can be further applied to detect and identify the online potential influencers in affecting readers’ behaviors massively.

2 Previous Works

The interactions between speakers and readers usually involve the process of persuasion, which is one of the topics that are concerned by researchers in the field of pragmatics. The role of pragmatics focuses on how utterances are applied in interaction (Schmidt and Richards, 1980), and how speakers follow the written maxims of conversation in order to cooperate with readers and to be socially accepted (Cutting, 2002).

Searle (1969) addressed that persuasion is a directive speech act that speakers impose the intention on the readers to let them perform certain actions. Additionally, Nweke (2001) considered persuasion as a systematic approach which aims at changing or emphasizing opinions, attitudes, beliefs or values for a positive response. Researchers (Lakoff, 1981; Searle, 1969; Banerjee and Carrell, 1988; Brown and Levinson, 1987) further stated that persuasion is a face threatening act, which would threaten readers’ negative face if they do not take the action as desired by the speaker. Therefore, since a reader’s

negative face may be threatened by the speaker’s persuasion act, the reader would decide whether he/she accepts the speaker’s viewpoints or not (Taillard, 2002).

In order to let readers agree with speaker’s suggestions to the largest scale, the speaker needs to be strategic while expressing his/her intentions on the statements. Hovland et al. (1953) were the first group of researchers that initiated the study of persuasion and attitude change at Yale University. They studied on the components of what led to a successful persuasion communication by asking “Who says what to whom and with what effect?”. The components of a persuasion process include presentation, attention, comprehension, acceptance, and retention. An effective persuasion would induce readers to make the favorable responses to the speaker (Hegtvedt and Johnson, 2017). However, steps that lead to a successful persuasion are difficult to be fulfilled and consistent with changed attitude (McGuire, 1985; Severn et al., 1990). Severn et al. (1990) also pointed out there are different hurdles to a persuasive communication, such as language, education, belief consistency, and lack of trust.

We target on the notions of cohesion and coherence in this paper. Cohesion helps capture persuasive features grammatically and semantically (Halliday and Hasan, 1976); while coherence grabs not only the semantic information (van Dijk, 1977), but also the pragmatic concept of a discourse (Widdowson, 1978). In addition, Brown and Yule (1983) point that coherence is an interaction between readers and the discourse. Thus, by applying both cohesion and coherence, we will be able to examine persuasive features with different levels of discourse analysis.

In the concept of cohesion, there are four categories, including reference, ellipsis/substitution, conjunction, and lexical cohesion (Halliday, 1994; Halliday and Hasan, 1976; Halliday and Matthiessen, 2004; Martin, 2008). Among them, ellipsis/substitution will not be covered in this study, since the quality of retrieved items heavily depends on the performance of the parser and may influence the assessment result.

In the category of lexical cohesion, since it can be further presented in various forms (Strauss and Feiz, 2014), we take two of them into consideration: repetition and simile.

As for coherence, it has been addressed that it signals deeper pragmatic information such as world knowledge (Redeker, 1990; Taboada and Mann, 2006), and can be applied for topic interpretation which is usually presented via collocations (Newman et al., 2010; Lau et al., 2013; Nokel and Loukachevitch, 2015). There are various association measures mentioned in previous studies to calculate collocations statistically, such as Mutual Information (MI) (Church and Hanks, 1990), t-score (Church et al., 1991), log-likelihood (Dunning, 1993), Dice (Dice, 1945), and logDice (Rychlý, 2008). Each of these measures has its own advantages, and can be implemented to retrieve a list of words which highly collocate with a target word. To sum up, by applying quantitative methods in this study on cohesion, we will be able to inspect which grammatical and semantic strategies or techniques characterize high-influence speakers with statistical evidence; and on coherence, we will be able to observe the topics that may attract increasing popularity in discourse level.

3 Methodology

In this study, to examine the linguistic features of high-influence speakers quantitatively from speaker-reader interactions, we extract posts from PTT corpus which collects various discussion threads from an online bulletin board system, PTT. PTT is one of the prevalent social media frequently used in Taiwan which contains numerous discussion boards. We target on posts published on four discussion boards, which are Boy-Girl, Marriage, BabyMother, and WomenTalk. These boards are highly visited daily by PTT users, and have the characteristics that people would post their real-life dilemmas and seek for suggestions or comments from others. People who are interested on the issues mentioned within a post can reply the post directly and express their own opinions or thoughts. The replied post will be automatically assigned with

a “Re:” label to the post title. Readers are allowed to comment on posts via “push”, “boo”, and “comment” to show their preferences or judgments to the post content, which refer to positive, negative, and neutral responses respectively.¹

We focus on analyzing all the “Re:” posts from the four boards for the reason that we are able to scrutinize and track down the discrepancies between high-influence and low-influence speakers from their replies to the discussion starters. We target speakers who have at least two posts and whose number of replied posts lies above the third quartile of all the replied post in the board, to make sure each speaker have enough posts to retrieve persuasive features for comparison. Among these filtered speakers, only the top three speakers who have the highest Positive Feedback Degree (PFD) scores are regarded as the high-influence speakers; and the rest as the low-influence. A PFD is defined as the following by calculating a speaker’s:

$$\frac{\text{number of “push”} - \text{number of “boo”}}{\text{total number of replied posts}} \quad (1)$$

A total of six features are taken into consideration in this study, which are reference, conjunction, simile, repetition, collocation, and the average length of a post. It is known that reference can be presented via pronouns, and previous studies indicate that high-influence speakers tend to use the pronoun “you” more often; thus, we focus on exploring the first, second, and third personal pronouns. A list of keywords for conjunctions (e.g., 所以 ‘so’, 其實 ‘actually’, and 最後 ‘eventually’) and similes (e.g., 如同 ‘as’, 譬如 ‘such as’, and 似乎 ‘seem’) are provided. Since repetition is the most simple and direct instance of lexical cohesion (Halliday, 1994) that can be used to express the important topics widely discussed by the speakers, we further take the repeated words as target words to find their collocated words for observing other relevant and frequently mentioned topics. Additionally, we consider that high-influence speakers would use

¹The data and scripts employed in this research will be released to the public.

more persuasion strategies and have a longer length of post in expressing their own opinions than low-influence speakers; therefore, the length of a post is added as a feature. These features listed below illustrate how they are collected and calculated in this study.

- reference, conjunction, and simile:
The occurrence of each feature in a post is extracted.
- length of a post:
The number of word tokens within a post is calculated.
- repetition:
To find the representative word repetitions of high-influence and low-influence speakers, we collected a list of word types observed from the posts published by both groups of speakers. The raw frequencies of the word types occurred in the two groups are calculated respectively. Next, we subtract the raw frequencies of low-influence speakers from high-influence speakers to get the weight of each word type, and sort the subtraction result from the highest to the lowest. If the result is positive, it means the word type favors high-influence speakers more; and vice versa. Thus, we take the top 30 words with the highest scores as the representative repetition words for the high-influence speakers; and the last 30 words with the lowest scores for the low-influence speakers.
- collocation:
The repetition words observed from the two groups of speakers are taken as target words to explore their collocates. We apply the python NLTK package to measure and extract collocation with pointwise mutual information algorithm, and the window size is set to two. If a target word has more collocates than the other, it refers that the target word may occur in an amount of various contexts and may evoke more diverse topics. In order to know the number of collocation pairs of each target word, we

set up a collocation threshold by taking the pairs that occur at least twice into consideration. For pairs that occur only once are not taken into account, since it may only happen by chance to be collocated with the target word.

We perform a statistic test and an association measurement for test evaluation, which are independent-samples t-test and biserial correlation (Sheskin, 2011) respectively, on five features (excluding repetition) between the high-influence and low-influence speakers. Since repetition words are filtered based on the subtracted weights, the values are no longer parametric which is the element required while performing the two tests. The analysis of repetition words will be discussed in the next section.

Firstly, in t-test, it examines whether the two groups are from the same population by comparing the differences between the means of the groups. The higher the t-value produced from the test, the bigger the differences are between the two provided groups. Therefore, we would like to examine whether there are differences between the high-influence and low-influence speakers. As for biserial correlation (r_b), it measures the relationship between two groups, and further presents which feature favors the high-influence or low-influence speaker. It is an association measurement that takes a continuous variable (Y) and an artificially dichotomized variable (e.g. the high-influence (X_1) and low-influence speakers (X_0)), where the dichotomies have an underlying continuum. It is derived from point-biserial correlation (r_{pb}) which its dichotomized variable is natural categories (e.g. dead and alive). It ranges from -1 to +1, where -1 denotes a perfect negative association, +1 presents a perfect positive relationship, and 0 indicates there is no association between the two groups. The algorithm is listed in (2).

$$r_b = \left[\frac{\bar{Y}_1 - \bar{Y}_0}{S_Y} \right] \left[\frac{p_0 p_1}{h} \right] = \frac{r_{pb} \sqrt{p_0 p_1}}{h} \quad (2)$$

where \bar{Y}_1 and \bar{Y}_0 are the average data pairs grouped as X_1 and X_0 respectively; in addition,

p_0 and p_1 refers to the proportion of data pairs categorized into X_0 and X_1 respectively; and S_Y describes the standard deviation of the continuous variable Y , and h is the height of the standardized normal distribution at point z which divides the two proportions p_0 and p_1 .

4 Results and Analysis

Before taking a look at the results produced by the statistic tests, a summary of the extracted posts, number of tokens and speakers of the two groups of speakers from the four discussion boards, is provided below in Table 1.² It is observed from the table that although the number of posts, tokens, and speakers are higher in the low-influence speakers, each of them often only have two posts. On the contrary, the authors in the high-influence group have higher average posts in each board.

The results of t-test and biserial correlation are elaborated in the below two sections, 4.1 and 4.2. The discussion of repetition words and collocation is listed in section 4.3.

4.1 T-test: same population or not?

The statistic results that show the differences between the high-influence and low-influence speakers over five features on the dataset are presented in Table 2. It is observed that all the features have significant effects while comparing the two groups of people, except collocation. This implies that features such as reference, conjunction, simile, and length of a post can help identify the two groups of speakers, but not for the collocation feature.

Since collocation does not tell a significant difference between the high-influence and low-influence speakers in the entire dataset, we further explore into detail, and apply t-test on the four discussion boards independently to see whether the two groups of speakers still have the same behavior. The t-test results of collocation on the four boards are listed in Table 3. It is observed that except Marriage board, the other

boards all have a significant effect ($p < 0.05$). This addresses that the collocation feature can help categorize two different groups of speakers, but it is less effective while performing on Marriage board. This may due to the reason that the number of retrieved collocations is quite sensitive to t-test and the collocations extracted for both groups of speakers on Marriage board are quite similar. However, this does not infer that the topics intrigued by the high-influence and low-influence speakers are the same.

Lastly, we can further take a deeper look at the reference feature which includes three types of personal pronouns. The results of three personal pronouns are listed in Table 4. From the table, it shows that all the three personal pronouns are frequently used and addressed by the high-influence speakers than low-influence speakers with a significant effect. Although previous researchers indicated that the high-influence speakers frequently addressed the second-person pronoun in posts (Liu, 2017; Strauss and Feiz, 2014), it is discovered in this experiment that they also use the first-person and third-person pronouns quite often to increase their persuasiveness. This might be affected by the length of a post. Since the length of a post of high-influence speakers is significantly longer than that of low-influence speakers as shown in Table 2, it is possible that with longer contexts, the opportunities of employing coreference with different types of personal pronouns increases.

4.2 Biserial Correlation: the relationship of the two groups

After performing t-test on the features to examine whether there is a difference between the population of the two groups of speakers, a biserial correlation measurement is further applied to these features to investigate which feature favors the high-influence or low-influence speakers more. The coefficients (r_b) calculated by biserial correlation are displayed in Table 5. It shows that all the features have relatively stronger positive correlations with the high-influence speakers, except collocation which has a less positive correlation. The distributions of these features

²For the tables presented in this paper, the “high” word in the Groups column represents the high-influence speakers, and the “low” is used for indicating the low-influence speakers.

Board	Group	Post	Token	Speaker
Boy-Girl	high	248	260946	3
	low	1155	524580	902
Marriage	high	21	24999	3
	low	132	88948	105
BabyMother	high	28	32314	3
	low	32	11684	26
WomenTalk	high	48	46225	3
	low	331	139528	252
Total	high	345	364484	12
	low	1650	764740	1285

Table 1: A summary of the extracted posts from the four discussion boards.

Features	Groups	Mean	Standard Deviation	t-value	Significance
reference	high	59.01	59.52	10.045	$p < 0.01$
	low	25.89	32.16		
conjunction	high	29.63	20.57	12.82	$p < 0.01$
	low	14.56	16.24		
simile	high	2.68	2.26	14.54	$p < 0.01$
	low	0.85	1.43		
length of a post	high	1056.48	707.73	18.74	$p < 0.01$
	low	463.48	490.29		
collocation	high	35.75	59.12	1.82	$p = 0.07$
	low	24.68	30.24		

Table 2: The statistic results of t-test over five features for the high-influence and low-influence speakers.

Boards	Groups	Mean	Standard Deviation	t-value	Significance
Boy-Girl	high	94.40	95.24	2.50	$p = 0.015$
	low	46.70	38.02		
Marriage	high	17.37	11.42	-0.90	$p = 0.37$
	low	22.03	25.65		
BabyMother	high	18.23	10.34	6.47	$p < 0.01$
	low	4.83	4.10		
WomenTalk	high	13.00	8.31	-2.43	$p = 0.018$
	low	25.13	25.52		

Table 3: The statistic results of t-test on collocation over each board for the high-influence and low-influence speakers.

in the two groups of speakers are displayed in Figure 1 with boxplots. As observed from the figure, all the figures show a positive slope from the low-influence to the high-influence speakers;

and among these, the slope of the collocation only leans slightly on the high-influence speakers.

Additionally, we also compute biserial corre-

Personal Pronouns	Groups	Mean	Standard Deviation	t-value	Significance
First-person	high	22.42	30.69	7.13	$p < 0.01$
	low	10.26	17.36		
Second-person	high	22.80	20.08	12.21	$p < 0.01$
	low	9.11	12.31		
Third-person	high	13.79	17.32	7.55	$p < 0.01$
	low	6.51	10.20		

Table 4: The statistic results of t-test over the three personal pronouns under reference feature for the high-influence and low-influence speakers.

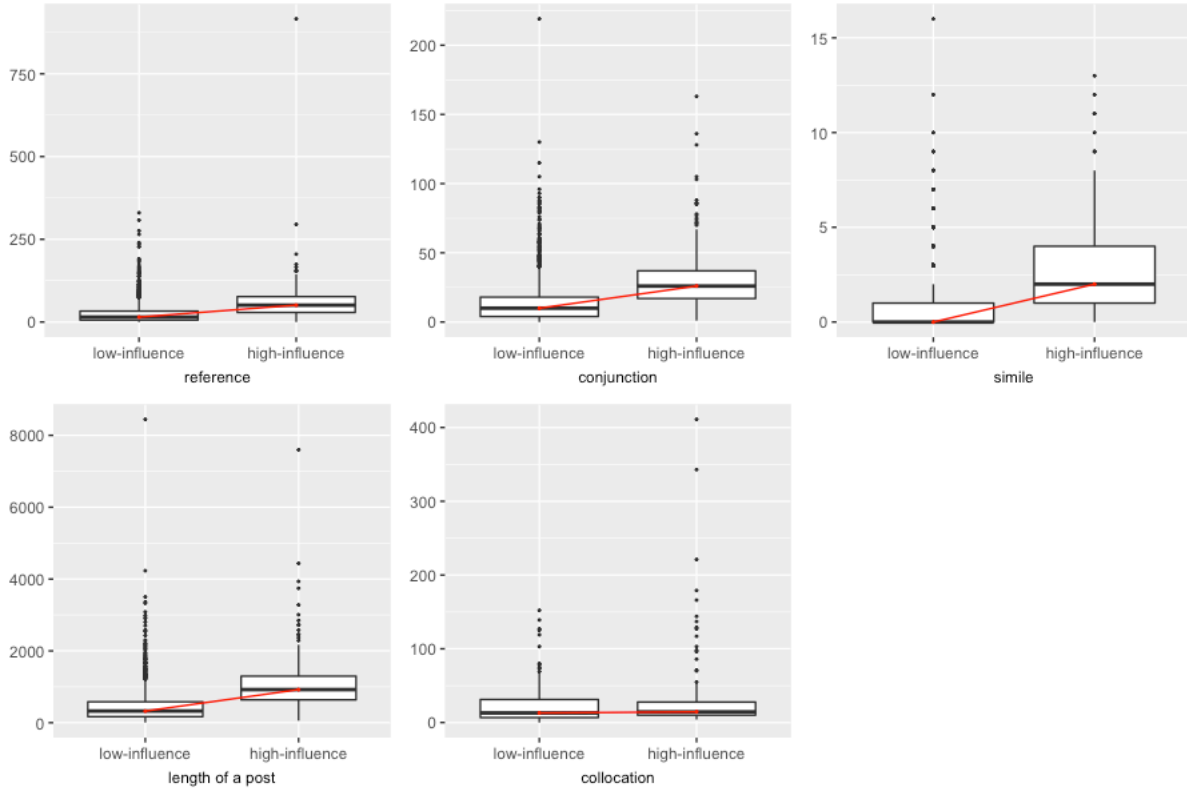


Figure 1: The distributions of five features in the high-influence and low-influence speakers.

Features	r_b
reference	0.46
conjunction	0.47
simile	0.59
length of a post	0.47
collocation	0.10

Table 5: The statistic results of biserial correlation over five features for the high-influence and low-influence speakers.

lation on the three personal pronouns for a more detailed observation on the reference feature. The coefficients are as listed in Table 6, and a visualization of the slopes between the two groups of speakers are displayed in Figure 2. The coefficients presented in the table describe that all the three personal pronouns have a positive correlation to the high-influence group as well. This indicates that the three personal pronouns are frequently used by the high-influence speakers instead of the low-influence speakers. Further-

more, among the three personal pronouns, the second-person pronoun has the highest correlation of all. This also coincides with the statements illustrated by previous studies, which described the high-influence speakers tend to use the second-person pronouns more often than the low-influence speakers.

Features	r_b
First-person	0.33
Second-person	0.51
Third-person	0.34

Table 6: The statistic results of biserial correlation over three personal pronouns for the high-influence and low-influence speakers.

4.3 The analysis of repetition and collocation

In addition to the above statistic results, we can further take a look at the interplay of repetition words and their collocates, and see how they elicit various topics.³ Among the four boards, there are some discrepancies among the topics that are frequently talked about by the high-influence speakers and the low-influence speakers. For example, in Boy-Girl, the high-influence speakers focus more on describing the mental situations and struggles from both sides of a relationship that people will encounter, such as 坦白說 ‘to speak frankly’, 心裡 ‘in the heart’, 感情裡 ‘in the relationship’, 想追 ‘want to court (someone)’, 明明知道 ‘(someone) apparently knew that’, 退回朋友 ‘resolve upon a friendship’, and 異性朋友 ‘friends of the opposite sex’; however, the low-influence speakers emphasize on illustrating the situations on one side of the relationship only, such as 男方認為 ‘boyfriend/husband considers that’, 醜化男方 ‘vilify the boyfriend/husband’, 女方騙錢 ‘scammed by the girlfriend/wife’, and 女方父母 ‘the girlfriend/wife’s parents’.

In Marriage board, the topics mainly stated by the high-influence speakers are wife and children, like 老婆說 ‘wife says’, 抱老婆 ‘hug

the wife’, 老婆好好 ‘(my) wife is good’, 照顧小孩 ‘take care of the children’, and 教小孩 ‘teach the children’; whereas the low-influence speakers target on topics related to husband and the wife’s mother-in-law (e.g., 幫老公 ‘help the husband’, 先生同意 ‘the husband agrees’, 婆婆反悔 ‘the wife’s mother-in-law goes back on something’, and 嗆婆婆 ‘talk tough to the wife’s mother-in-law’). This also reveals the contradiction of Taiwan’s culture, where females nowadays focus more on themselves rather than contributing their life to the husband or mother-in-law.

For BabyMother, topics surrounded on the healthiness of a baby during pregnancy are what the high-influence speakers are interested in (e.g., 染色體檢查 ‘chromosome examination’, 染色體異常 ‘chromosome abnormality’, 唐氏症超音波 ‘Down Syndrome anatomy ultrasound’, and NIPT 正常 ‘Non-Invasive Prenatal Test (NIPT) normal’). By contrast, the low-influence speakers like to share experiences on how to take care of a baby (e.g., 餵寶寶 ‘feed the baby’, 寶寶哭 ‘the baby cries’, 喝奶 ‘drink milk’, and 安撫奶嘴 ‘calm (the baby) with a pacifier’).

As for WomenTalk, the high-influence speakers address topics related to culture and women’s self benefits (e.g., 女性文化 ‘feminine culture’, 厭女文化 ‘misogyny culture’, 個人利益 ‘personal benefits’, and 個人年薪 ‘personal annual salary’; on the contrary, the low-influence speakers talk about negative words put on female (e.g., 台女EZ ‘Taiwanese girl is easy to get laid’, 罵母豬 ‘scold the female as a sow’, and 醜台女 ‘ugly Taiwanese girl’).

5 Conclusions

In this study, we have extracted features of persuasive strategies based on linguistic theories in discourse level, and examined these features with quantitative methods to investigate whether they reveal an significant effect in identifying the high-influence speakers in social media. From the presented results, these features are able to differentiate the two groups of speakers, which indicate that there is a difference in

³The repetition words are underlined in all the provided examples.

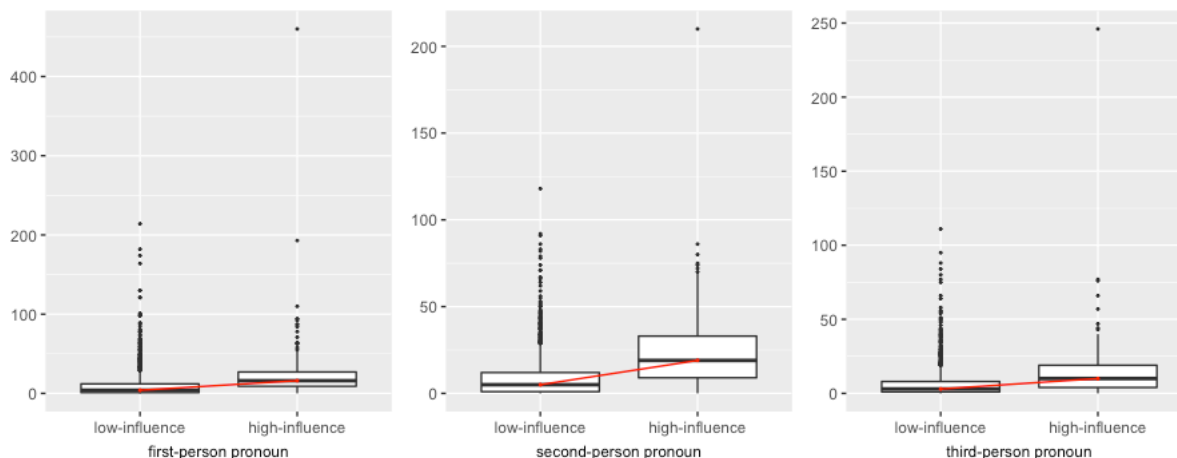


Figure 2: The distributions of three personal pronouns in the high-influence and low-influence speakers.

the usage of persuasion strategies between the two groups of speakers. In addition, the high-influence speakers use these strategies more often in their posts than the low-influence speakers. Moreover, the information provided by the repetition words and their collocates show that the topics brought up by two groups of speakers are distinct from each other.

The findings of this study provide statistical evidence and analysis to examine and support that the theory-based linguistic features can be applied to characterize the high-influence speakers quantitatively in large datasets. We have investigated the extracted features can be further implemented to predict who has the potential to become popular in social media. For the future study, in addition to the inspected features presented in this research, other factors such as gender and social identify will be explored, to examine whether biases are shown in persuasion strategies as well.

Acknowledgments

We are grateful that reviewers have provided and addressed their insights on this research (e.g., the use of pronouns and other factors in affecting one’s persuasion strategy). The comments and suggestions are appreciated and taken into consideration.

References

- Janet Banerjee and Patricia L. Carrell. 1988. Tuck in your shirt, you squid: Suggestions in ESL. *Language Learning*, 38(3):313–364.
- Penelope Brown and Stephen C. Levinson. 1987. *Politeness: Some Universal in Language Usage*. Cambridge University Press.
- Gillian Brown and George Yule. 1983. *Discourse Analysis*. Cambridge University Press.
- Kenneth W. Church and Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1):22–29.
- Kenneth W. Church, William A. Gale, Patrick Hanks, and Donald Hindle, 1991. *Lexical Acquisition: Exploiting On-line Resources to Build a Lexicon*, chapter Using statistics in lexical analysis, pages 115–164. Lawrence Erlbaum.
- Robert B. Cialdini, 1987. *Social Influence: The Ontario Symposium*, chapter Compliance principles of compliance professionals: Psychologists of necessity, pages 165–184. Psychology Press, 5th edition.
- Ana Ciocarlan, Judith Masthoff, and Nir Oren, 2019. *Persuasive Technology: Development of Persuasive and Behavior Change Support Systems. PER-SUASIVE 2019. Lecture Notes in Computer Science*, volume 11433, chapter Actual Persuasiveness: Impact of Personality, Age and Gender on Message Type Susceptibility, pages 283–294. Springer.
- Bert Cornillie. 2018. On speaker commitment and speaker involvement. evidence from evidentials in Spanish talk-in-interaction. *Journal of Pragmatics*, 128:161–170.

- Joan Cutting. 2002. *Pragmatics and Discourse: A Resource Book for Students*. Routledge.
- Lee R. Dice. 1945. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302.
- James Price Dillard and Linda J. Marshall, 2003. *Handbook of Communication and Social Interaction Skills*, chapter Persuasion as a social skill, pages 479–513. Lawrence Erlbaum Associates.
- Ted E. Dunning. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1):61–74.
- Christine Gunlogson. 2008. A question of commitment. *Belgian Journal of Linguistics*, 22(1):101–136.
- M.A.K. Halliday and Ruqaiya Hasan. 1976. *Cohesion in English*. Longman.
- M.A.K. Halliday and Christian M.I.M. Matthiessen. 2004. *An Introduction to Functional Grammar*. Oxford University Press, 3rd edition.
- M.A.K. Halliday. 1994. *An Introduction to Functional Grammar*. Edward Arnold, 2nd edition.
- Helena Halmari and Tuija Virtanen, 2005. *Persuasion Across Genres: A linguistic approach*, chapter Towards understanding modern persuasion, pages 229–244. John Benjamins.
- Charles Leonard Hamblin. 1970. *Fallacies*. Methuen.
- Karen A. Hegtvedt and Cathryn Johnson. 2017. *Social Psychology: Individuals, Interaction, and Inequality*. SAGE.
- Carl Iver Hovland, Irving L. Janis, and Harold H. Kelley. 1953. *Communication and Persuasion: Psychological Studies of Opinion Change*. Yale University Press.
- Tamar Katriel and Marcelo Dascal, 1989. *From Sign to Text: A Semiotic View of Communication*, chapter Speaker’s Commitment and Involvement in Discourse, pages 275–295. John Benjamins.
- Robin Tolmach Lakoff, 1981. *Analyzing Discourse: Text and Talk*, chapter Persuasive Discourse and Ordinary Conversation, with Examples from Advertising, pages 25–42. Georgetown University Press.
- Jey Han Lau, Timothy Baldwin, and David Newman. 2013. On collocations and topic models. *ACM Transactions on Speech and Language Processing*, 10(3):1–14.
- Donghong Liu. 2017. Study of persuasive rhetorical strategies driven by cultural values. *Journal of Central China Normal University (Humanities and Social Sciences)*, 56(4):104–109.
- James Robert Martin, 2008. *The Handbook of Discourse Analysis*, chapter Cohesion and Texture, pages 61–81. Blackwell Publishing.
- William J. McGuire, 1985. *Handbook of Social Psychology*, volume 2, chapter Attitudes and Attitude Change, pages 233–346. Random House, 3rd edition.
- David Newman, Jey Han Lau, Karl Grieser, and Timothy Baldwin. 2010. Automatic evaluation of topic coherence. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2010)*, pages 100–108.
- Michael Nokel and Natalia Loukachevitch. 2015. Topic models: Accounting component structure of bigrams. In *Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA 2015)*, pages 145–152.
- Okechukwu F. F. Nweke. 2001. *Public Relations Practice: Concepts and strategies*. Hamson.
- Rita Orji, Regan L. Mandryk, and Julita Vassileva, 2015. *Persuasive Technology. PERSUASIVE 2015. Lecture Notes in Computer Science*, volume 9072, chapter Gender, Age, and Responsiveness to Cialdini’s Persuasion Strategies, pages 147–159. Springer.
- Kiemute Oyibo, Rita Orji, and Julita Vassileva. 2017. The influence of culture in the effect of age and gender on social influence in persuasive technology. In *Proceedings of the Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization (UMAP 2017)*, pages 47–52.
- Gisela Redeker. 1990. Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics*, 14(3):367–381.
- Pavel Rychlý. 2008. A lexicographer-friendly association score. In *Proceedings of Recent Advances in Slavonic Natural Language Processing (RASLAN 2008)*, pages 6–9.
- Richard W. Schmidt and Jack C. Richards. 1980. Speech acts and second language learning. *Applied Linguistics*, 1(2):129–157.
- John Rogers Searle. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.
- Jessica Severn, George Edward Belch, and Michael Belch. 1990. The effects of sexual and non-sexual advertising appeals and information level on cognitive processing and communication effectiveness. *Journal of Advertising*, 19(1):14–22.

- David J. Sheskin. 2011. *Handbook of Parametric and Nonparametric Statistical Procedures*. CRC Press, 5th edition.
- Robert Stalnaker. 1978. Assertion. *Syntax and Semantics*, 9:315–332.
- Susan Strauss and Parastou Feiz. 2014. *Discourse Analysis: A multi-perspective and multi-lingual Approach*. Routledge.
- Maite Taboada and William C. Mann. 2006. Rhetorical structure theory: looking back and moving ahead. *Discourse Studies*, 8(3):4423–459.
- Marie-Odile Taillard. 2002. Beyond communicative intention. *UCL Working Papers in Linguistics*, 14:189–206.
- Teun A. van Dijk. 1977. *Text and Context: Explorations in the semantics and pragmatics of discourse*. Longman.
- Henry G. Widdowson. 1978. *Teaching Language as Communication*. Oxford University Press.