

The use of Halliday's Theme-Rheme on Google's Privacy Policy Text: A Forensic Linguistics Research

Sabtra Lesmana^{1✉}, Sawirman², and Fajri Usman³

¹²³ Jurusan Ilmu Linguistik, Fakultas Ilmu Budaya, Universitas Andalas
✉ Limau Manis, Padang, Sumatra Barat-25164, Indonesia
✉ saptralesmana@gmail.com

Article Info

Article History

Received:

Mar 2022

Accepted:

Apr 2022

Published:

Jul 2022

Keywords:

*clause, forensic-linguistics,
misleading-statement,
privacy, violation.*

ABSTRACT

This research aimed to reveal and explain the violations conducted by Google to its user privacy which is implied through its Privacy Policy Agreement statement by analyzing it with Theme and Rheme proposed by Halliday. This qualitative research was done by applying the documentation recording method in collecting data and Halliday's Systemic Functional Linguistics as the basic theory. The result of the data was presented in the table analysis and the form of descriptive interpretation.

The result showed that Google build its clauses with the 4 types of theme construction. From the analyzed clauses, the dominantly built clause is the single constituent with a marked theme where "We" is the most occurred subject. Yet, Google applied unmarked themes mostly in their clauses. With these dominant occurrences of "We", Google construes its existence as the authority holders of the whole agreement. The researchers found that similar finite predicators appeared repeatedly such as "collect" and "use" in building its clauses. These constitute Google's strategy to force the user's permission to let Google freely access their privacy. In addition, the operation of the conjunction "and", extension, expansion, and exemplification constituents represent the plentiful request of Google related to its user's information. Those summarized the violation conducted by Google of its user's privacy.

© 2022 Politeknik Negeri Bali

INTRODUCTION

"Many people accept privacy policies without reading them; study reveals" (Reidenberg, 2015; CET News, 2020). Not all internet consumers comprehend and evolve in their use of current conveniences. These detrimental impacts have far-reaching implications that impact not only people but also society as a whole. Since the internet shortens the distance between people, it allows people to reach out to others quicker via a variety of contact channels. Arbitrarily using language becomes the underlying foundation for anybody, whether knowingly or unintentionally, to create effects of language that cannot be handled and are often irresponsible.

As stated by Auer and Schmidt (2010) that language is formed through a relative form, cooperative, and mostly does not reflect many actions. Relative uniformity is a prerequisite for language to be able to serve its basic purpose (communication); the existing variability serves several other functions. Language is not only a neutral medium for generating subject knowledge, but is a form of social practice that acts to reflect social reality (Silverman, 2000). Language is used by humans to interact with other humans and remain linked. Humans use words to establish their activities and identities. If a crisis arises, humans may use language to track any event, even those that exist in between. Forensic linguistics is one of the fields of linguistics that can be used to uncover linguistic difficulties that have occurred or may occur within people.

Forensic linguistics is a multidisciplinary discipline that is used to answer legal questions about language (Udina, 2017; Leonard, 2017). It enhances legal analysis by strictly applying scientifically accepted principles of language analysis to legal evidence such as e-mails, text messages, contracts, letters, confessions, and recorded speeches (Shuy, 2006). Forensic linguistics itself is one of the fields of science in linguistic studies that has recently become popular to be developed since its presence is capable of being an expert solution in solving deadlocks or being a guide in determining facts.

Any created text is a reflection of the text maker's propensity, and it is, of course, a way for him to manipulate the interlocutor or the object of the text (Asya, 2013). Forensic linguistics is concerned with analyzing a document that is used in a civil or criminal context. On the same pad, Olson (2008) states forensic linguistics is "the application of linguistic expertise to a specific social environment, especially the legal field". Depending on the legal sense, any spoken or written document, such as a suicide note, law, deposition, handwriting, video, or email scam, can be identified as forensic text.

Under the use of Systemic Functional Linguistics, Wawan Gunawan and Aziza (2017) wrote the first publication in the Indonesian Journal of Applied Linguistics. They studied Theme and Rheme progression of undergraduate thesis: investigating meaning-making in academic writing. Whilst, Dian, and Ernie (2017) published a paper titled "The Realization of Interpersonal Meaning in Course Newsletters: a Systemic Functional Linguistic Perspective". Their analysis was inspired by the idea that language structure is made by the institution of a linguistic unit with a systematic construction. Dian and Ernie explored how to extend Halliday's theory of Structural Functional Linguistics to text, precisely allowing a realization of the development of meaning from the clause system. The application of the theory then expanded to the analysis of Google Privacy Policy Text.

From the data material analysis, Andow and Mahmud (2019) released a publication related to the analysis of contradictory facts in the Privacy Policy text. Supported by the enhancement of computerized tools to help detect deceptive information issued by the application developer in Google Play, they then expand the analysis to more than 11000 most downloaded android applications' privacy policy text, yet they only focus on the contradictory statements. Concerning this research, the analysis was focused on the clause of Google Privacy Policy Text.

To define and expand the field of forensic linguistics in a legal setting, this research employs linguistics and forensics in the examination of contract texts. From a linguistics aspect, the

researchers apply Systemic Functional Linguistics specifically the Theme-Rheme theory by Halliday. This theory provides a systematic way to uncover the language mechanism through its sequential position in a clause. As most people do reading the text like focusing on the verb to assume the clause's information, this theme-rheme concept offers the ultimate way to revealing the text's meaning, especially from the text maker or speaker's intention. We recognize that this is essential to explore the creation of lingual units in the language of the contract text of Google Privacy Policy Text and to interpret the messages found in it more thoroughly.

METHODS

This qualitative research includes data collection via the documentation records method (Neuman, 2013). According to the staging of the analysis sequences, it previously concentrated on giving an explanation the identifying the process of developing the construction of Theme and Rheme on the clauses of Google's Privacy Policy Text. Then a qualitative description is made to reveal and explain the violations committed by Google to its users' privacy as suggested in its text agreement by examining Theme and Rheme's perspective. By applying the Theme-Rheme theory in Systemic Functional Linguistics, the analysis of clauses containing violations can be traced through the linguistic elements in Google's User Privacy Policy statement. Since the type of study was a qualitative research design, all aspects of procedures, from methodology to analysis outcome presentation are presented in the form of descriptive analysis. Both analyzed data are transformed into a concise description focusing on the phenomena and their meaning in in-depth detail (Zaim, 2014).

RESULTS AND DISCUSSION

To identify how the user's privacy is being violated as indicated in the text, the clauses must contain the violation indicator constituent that occurred in the system of the text's clause through Systemic Function Linguistics by Halliday. A text can be said to have violated constituents if they fulfil the characteristics of violation, which according to Green (2001) language is deemed violating if it contains representation such as lying, misleading, and falsely denying.

The Operation of Obligatory and Non-Obligatory Constituent

There is a study in linguistics that focuses on clause building units, and a clause is considered intact if it includes complete building components. The building components in question are linguistic units in the form of words that occupy positions in syntax-based categories.

Data C.1	When you use our services, you are trusting us with your information						
When	you	use	our	you are	Trusting	us	with your
			services			information	
WH/Adj	Subj	Finite	Comple	Subject	Predicator	Complement	
	ect		ment	+Finite			
T1	T2	Rheme					
Theme							

Table 1: Data C.1

Data C.2		We understand this is a big responsibility and work hard to protect your information and put you in control							
We	underst and	This	is	a	big	to	protect	And	put you in control.
				responsibility and work hard	information			(*ellipsis of we)	
Subject	Finite	Det.	Predic: Exist	Det: Exist	Nominal phrase:	Complement : Circ.	Conjunction: additive	Compleme nt: Circ.	
Theme	Rheme								

Table 2: Data C.2

According to the SFL, the acquired meaning takes the form of tagging-meaning, which is linked to the language unit depending on its location in the clause, for example, a verb if in a stand-alone linguistic unit the meaning is about stating or doing something or an action. A verb "use" and "understand" which are quoted from data 1 and 2 above have their meanings translated in the dictionary where "use" means 'to do something with a machine, a method, an object, etc. for a particular purpose and "understand" means 'to know or realize the meaning of words, a language, what somebody says, etc'. If a verb is operated into a clause, say by occupying one of the clause structure's obligatory slots, one of which works as a finite, then the meaning of "use" and "understand" will grow in tandem with the existence of additional constituents that take positions before and after the verb.

The Theme-Rheme theorem had been applied to analyze the meaning of the structure of the first clause; whereas the data 1 element consists of the Theme "When you," which is the grammatical category that bears the place of the subject. The fact that the subject of clause data 1 was begun with a WH/Adj indicates that the current topic suggests specific circumstances (particular time), which are underlined by the word "When," implying that the verbs in the sentence would behave by the explanation provided by the Theme. As a result of data 1's clause, the subject "you" will be able to perform something connected to "our services," namely product services provided by Google. However, in addition to the usage of services by "you," Google states in the text contained in the embedded clause a reciprocal action, realized in "you are trusting us with your information". "Complex sentences often involve embedded clauses that separate the important parts of a sentence and require mental reconstruction by the listener" (Blackwell, 2016, p. 62), in the clause of data 1 structure, this embedded clause is in a complement position. The embedded clause at this complement place verifies Google's request to monitor user information represented by "you." Because the word "When" occurred in the Theme clause structure, the form of monitoring of that information begins to take effect "when", since or when "you" agrees and starts using all services owned and related to Google.

Furthermore, the clause of data 2 has a Theme that is realized by the word "We," which in this case is a textual representation of Google. "We" is a meaningful Theme that is the major emphasis of the intent that is centred on this clause. The other builder constituents, both obligatory and non-obligatory, will then describe the entire meaning of what "We" aim to achieve in greater detail. The constituents that follow the subject are the obligatory finite "understand" and the obligatory compliment, which is completely expressed in the clause level by "this is a big responsibility and work hard to protect your information and put you in control." It can be interpreted that data 2 implies that Google "we" wants to emphasize to its users that it understands his duties and

responsibilities as an "undergoer" of the actions referred to in the complement of data 1 "trusting", and states that he wants to give confidence to users that Google remains as a second party who does not have full control over the collected and managed information. This is implied by the non-obligatory constituent in the clause labelled with Circ. "to protect your information," Conjunction: additive "and," and complement: Circ. "put you in control." Google claims in this statement that it ensures the security of its users' information and provides them with the authority to control it.

The two clauses in data 1-2 constitute the first section of the agreement that will tie the user to Google. At first glance, it appears that Google affirms the security of a coalition with it, but the scope and details connected to binding matters are concealed in the subsequent statements, which are becoming increasingly diverse. This variance will be visible in the next analyses. The following data are taken selectively based on the implication in the clause that has a complete meaning which indicates a violation committed by Google. As previously explained in the analysis of data 1-2, Google implies two things in its statement about Google's intentions. The first is Google states that it has the right to obtain user information, and the second is that Google wishes to provide the best service possible by managing the information provided by the user to Google.

From an independent survey conducted, it is known that statistical data contains various responses related to user activity in reading Google's privacy policy. One of the results obtained from the survey stated that of the 88 respondents surveyed there were about 12 people or about 13% of the respondents who stated that they only read the title or a small part of the agreement. One of the questions in the survey also looked at the parameters of users who only read the title. This means the possibility of the user will only touch 1-2 statements contained in the agreement. This corroborates the findings that users will assume what Google means is comprehensively summed up by the title of the second clause.

While the fact is that in the text of the agreement there are further statement items that are not related to Google's goal stated in the opening clause, as well as things that say what scope will be taken and managed by Google, which is just managing and maintaining at the beginning. The opening becomes contradictory with the stated points in the next section of the text. The issue referred to (contradictory facts and pieces of information) is further detailed in numerous clauses chosen as an analysis of clause data indicating a violation.

Data 3	We want you to understand the types of information we collect as you use our services		
We	Want	you	to understand the types of information we collect as you use our services
Subject Theme	Finite Rheme	Complement	Complement: Circ

Table 3: Data 3

In the perspective of the law, a violation is an act that enters the context of a crime. This viewpoint is consistent with (Sawirman et al., 2014) expressed in their book "Linguistik Forensik" p. 86, which states that "aspects of losses incurred are main variables that can be developed and directed

to various forms of action, plus with the indicator of awareness, then this has met the requirements of the crime regardless of the medium." The action stated in the clause of data 3 "We want you to understand the types of information we collect as you use our services" bears an element of intended violation, assumed from the quotation. The statement in data 3 sends a signal to users to guide them to Google's "goodness," to introduce its wishes, which are indicated by the obligatory finite "want" after the obligatory subject "We," to help the user to "understand" what Google is collecting. This conclusion can be construed in the context of circumstance data 3's non-obligatory constituent, which is realized in "the types of information we collect as you use our services". We can see that there is an element of ambiguity in the information provided by Google in this constituent. The word "types" is used with an indefinite plural, implying that the types Google refers to have no bounds (as for the extent to which only Google can decide the reason, as Google underlines with the word "as you use our services"). If there are variables in the agreement states that are unclear or whose scope is undefined, this can be classified as a violation since this inevitably must be accepted by the user.

Data 4 When you're signed in, we also collect information that we store with your Google Account, which we treat as personal information.					
When you are signed in	We	Also	Collect	information that we store with your Google account,	which we treat as personal information
WH/Adj.	Subject	Expansion	Predicator	Complement: Circ.	Complement: Expository
T1 Theme	T2	Rheme			

Table 4: Data 4

The other clause represents the violation conducted by Google of its user's privacy reflected on data 4. It is a Marked Theme with a multiple-Themed clause. If a clause includes a multiple Theme of Marked Theme, it signifies that the clause's major emphasis of meaning begins with the clause's earliest constituents before the subject. The Theme is in charge of the flow of information in the clause. Furthermore, Rheme is concerned with the growth of the breadth of action information suggested by the Theme. The constituents created the two portions of the clause's Theme, including the obligatory subject, which was realized by "When you are signed in, we." These constituents indicate that Google, as represented by "We", will conduct actions that are realized by the obligatory finite verb "collect," but that expansion will occupy the position immediately realized by "also" before finite operates. The presence of the word "also" denotes a broadening of Google's intended activity. When the "you" user is "signed in," it can be deduced that Google will "collect" even more pieces of information. The non-obligatory constituent at the end of the clause "that we store with your Google Account, which we treat as personal information" gives additional meaning to the information referred to by "also". The more private information has become easily accessible and infinitely shareable and transferable, the more monitoring may extend to private spaces, activities, and time (Gritzalis et al., 2014, p. 18). The information gathered by the Google system will subsequently be saved on the user's account's media. Google will then have ownership over this information as well. This includes an invasion

of confidentiality in which Google will record and collect all of the activity that users conduct on their accounts.

Data 5	When you create a Google Account, you provide us with personal information that includes your name and a password.			
When	you	Create	a Google Account	you provide us with personal information that includes your name and a password
WH/Adj.	Subject	Finite	Complement	Complement: Circ.
T1 Theme	T2	Rheme		

Table 5: Data 5

The information given in the non-obligatory constituents of data 4 is then described in further depth by Google in the next clause, which is represented by data 5. Google outlines the criteria associated with "signed in"; that "signing in" may only be done once the user has registered and created an account to use the services provided by Google. This statement indicates that when a user registers a Google account, the user will supply Google with personal information including a name and password. The information is realized utilizing finite's obligatory constituents, which are realized by the "provide." The Oxford Learner's Dictionary defines provide as 'to give something to somebody or make it available for them to use. Google explains what information may be gathered from users once more. Google has underlined the "collect" and "use" actions carried out numerous times in their text. This signals that Google has many requests from users in terms of how they use existing services. If Google stands on its promise at the outset of the agreement that it would not compromise the confidentiality of user data, then this agreement will end at data 3.

Data 6	We also collect the content you create, upload, or receive from others when using our services.			
We	also	collect	the content you create, upload, or receive	from others when using our services.
Subject Theme	Expansion	Predicator	Complement	Circumstance

Table 6: Data 6

Data 7	This includes things like email you write and receive, photos and videos you save, docs and spreadsheets you create, and comments you make on YouTube videos		
This	includes	things	like email you write and receive, photos and videos you save, docs and spreadsheets you create, and comments you make on YouTube videos.
Deictic: Det. Subject	Finite	Complement	Complement: exemplificatory
Theme	Rheme		

Table 7: Data 7

"We" is the most often appearing element in the Theme part of Google's privacy agreement's single-Themed clause. It is reasonable to assume that Google is the top priority in this agreement. This indicates that Google will become the most powerful player in carrying out the actions outlined in the text. Data 6-7 also have a Theme with obligatory constituents realized by "We." In this data, "We" is joined by many other constituents, including the non-obligatory constituent in the form of expansion "also," the predicator "collect," and the obligatory constituent "the content you create We may assume from this obligatory structure that Google is once again attempting to signal that they want to "collect" content created, uploaded, and accepted by users. Google states in the complement position that it will collect not only content created by users, but also content uploaded by users and content received by users from other users. By placing many verbs in a non-obligatory position, Google's wide scope of power over user data is highlighted.

Data 8	We collect information about the apps, browsers, and devices you use to access Google services, which helps us provide features like automatic product updates and dimming your screen if your battery runs low			
We	collect	information	about the apps, which helps us provide browsers, and features like automatic devices you use to product updates and access Google dimming your screen if services, your battery runs low.	
Subject	Finite	Complement	Complement: Expository Circ.	
Theme	Rheme			

Table 8: Data 8

Google is increasing the emphasis in the text of the agreement on the widening of access rights to the security of users' personal data. In the preceding analysis, Google has explicitly stated in this agreement that they will collect information owned by users when activating this Google service; the information gathered also includes personal identification information such as phone numbers, usernames, passwords, and information to payment transactions used by the user. Continuing to data 8, the scope of the access permissions sought by Google covers the user's actions, including the applications used by the user.

Furthermore, Google acknowledges in its statement that the information gathered at this time differs from that indicated in the preceding sections. The intentional distinction is abided by the non-obligatory constituents that follow complement. This constituent is realized by circumstance in the form “about the apps, browsers, and devices you use to access Google services,” and the additional expository constituent, which “helps us provide features like automatic product updates and dimming your screen if your battery runs low.” Google specifies what extra information it plans to gather in the first non-obligatory constituent. Meanwhile, Google provides the reasons for collecting such information in the following non-obligatory constituents. Google gathers data related to the apps, browsers, and devices that users use while using Google services so that Google may synchronize service updates and assist users in using services efficiently by synchronizing information about the devices used, such as minimizing data collection and reducing battery power consumption.

Some of the data used as a guide for analyzing violations committed by Google against its users via statements contained in its Privacy Policy text can be concluded that Google clearly states that every user who will use Google services will comply with the instructions, orders, and requirements set by Google. As with most agreements, at least two parties will be involved. In terms of legal administration, this Google privacy policy agreement is a bit distinctive. When both parties cannot be confronted at the moment of signing the agreement, this is well termed the one-sided approval agreement. One of the disadvantages of this sort of agreement is that it places one of the parties in the most vulnerable position. One of the weaknesses that one of the parties may encounter is that if the components of the agreement are changed, the party who agreed is obligated to accept the change without having the ability to argue or disagree.

CONCLUSION

Privacy policies are commonly used to inform users about the data collection and use practices of websites, mobile apps, and other products and services. However, the average Internet user struggles to understand the contents of these documents and generally does not read them (Liu et al., 2016). According to the results of the clause analysis of the Google Privacy Policy text in the framework of Halliday's Theme-Rheme SFL theory and its application with forensic linguistic investigations, it possibly concluded that Google's strategy in developing the privacy policy text contains elements of coercion that lead to violations of the majority of users' privacy domains. This finding was also supported by Perez's research (2016) which stated that the IoT did not match the statement of the privacy policy according to the empirical practices. As in Google, the modifications in the clause construction related to this are (1) Google applies the majority of unmarked Theme forms with a single constituent as the clause builder constituent with "We" as the realization (2) In Rheme construction, Google uses the verbs "collect" and "use" as the majority of the clause's finite/predicator. When a user agrees to the agreement, these two verbs offer explicit instructions on the activities Google will take regarding the privacy of its users. (3) In the non-obligatory construction section, the dominant constituents that occur in the agreement's text are exemplification and circumstance. This forensic linguistic research is conducted by referring to Halliday's theory of the Theme-Rheme system and is only limited to Google Privacy Policy text. There are some suggestions that the researcher proposed to other future researchers who are interested to conduct forensic linguistic research as well: 1. Forensic linguistic research can be conducted by using Systemic Functional Linguistic theory, especially the clause's Theme-

Rheme analysis. 2. Other researchers are suggested to decide on the other Systemic Functional Linguistic theory that is suitable to analyze other policy or agreement text other than as already applied in this research.

REFERENCES

- Andow, B., & Mahmud, S. Y. (2019). PolicyLint: Investigating Internal Privacy Policy Contradictions on Google Play. *Proceedings of the 28th USENIX Security Symposium*, 585-602.
- Asya, A. (2013). Linguistic Manipulation: Definition and Types. *International Journal of Cognitive Research in Science, Engineering and Education*, 78-82.
- Auer, P., & Schmidt, J. E. (2010). *Language and Space. An International Handbook of Linguistic Variation. Theories and Method*. Berlin: Walter de Gruyter GmbH & Co. KG.
- Blackwell, W. (2016). *Communication in Investigative and Legal Context Integrated Approaches from Forensic Psychology, Linguistics and Law Enforcement*. Chichester, West Sussex: John Wiley & Sons, Ltd.
- Green, S. P. (2001). Lying, Misleading and Falsely Denying. *Hasting Law Journal*, 158-210.
- Gritzalis, D., Kandias, M., Stavrou, V., & Mitrou, L. (2014). History of Information: The case of Privacy and Security in Social Media. *INFOHIST Legal Publications*, 1-25.
- Gunawan, W., & Aziza, F. (2017). Theme and Thematic Progression of Undergraduate Thesis: Investigating Meaning Making in Academic Writing. *Indonesian Journal of Applied Linguistics*, 413-424.
- Halliday, M., & Matthiessen, C. (2004). *An Introduction to Functional Grammar*. London: Hodder Headline Group.
- Leonard, R. A. (2017). Forensic Linguistics Applying The Scientific Principles of Language Analysis to Issues of The Law. *Hostra Law Review*, 881-897.
- Liu, F., Shomir Wilson, Schaub, F., & Sadeh, N. (2016). Analyzing Vocabulary Intersections of Expert Annotations and Topic Models for Data Practices in Privacy Policies. *The 2016 AAAI Fall Symposium Series: Privacy and Language Technologies Technical Report FS-16-04*, 264-269.
- Perez, A. J., Zeadally, S., & Cochran, J. (2018). A review and an empirical analysis of privacy policy and notices for consumer Internet of things. *Wiley*, 1-15.
- Neuman, W. L. (2013). *Social Research Methods: Qualitative and Quantitative Approaches*. Harlow: Pearson Education Limited.
- Reidenberg, J. R.; Breaux, T.; Cranor, L. F.; French, B.; Grannis, A.; Graves, J. T.; Liu, F.; McDonald, A.; Norton, T. B.; and Ramanath, R. (2015). Disagreeable privacy policies: Mismatches between meaning and users understanding. *Berkeley Tech. LJ*, 30-39.
- Sawirman, Hadi, N., & Yusdi, M. (2014). *Linguistik Forensik Vol 1*. Padang: Pusat Studi Ketahanan Nasional Universitas Andalas.
- Shuy, R. W. (2016). *The Language of Fraud Cases*. New York: Oxford University Press.
- Silverman, D. (2000). Analyzing talk and text', in N. K. Denzin and Y. S. Lincoln (eds), *Handbook of Qualitative Research* (2nd ed). London: Sage.
- Udina, N. (2017). Forensic linguistics implications for legal education: creating the e-textbook on language and law. *Procedia - Social and Behavioral Sciences*, 1337-1340.
- Zaim, M. (2014). Metode Penelitian Bahasa: Pendekatan Struktural. Padang: Sukabina Press.8). Contexts and pragmatics learning: Problems and opportunities of the study abroad research. *Language Teaching*, 51(1), 124-137. doi: 10.1017/S0261444815000440