Google-trickers, Yaminjeongeum, and Leetspeak: An Empirical Taxonomy for Intentionally Noisy User-Generated Text

Won Ik Cho
Seoul National University
tsatsuki@snu.ac.kr

Soomin Kim
Seoul National University
soominkim@snu.ac.kr

Abstract

WARNING: This article contains contents that may offend the readers.

Strategies that insert intentional noise into text when posting it are commonly observed in the online space, and sometimes they aim to let only certain community users understand the genuine semantics. In this paper, we explore the purpose of such actions by categorizing them into tricks, memes, fillers, and codes, and organize the linguistic strategies that are used for each purpose. Through this, we identify that such strategies can be conducted by authors for multiple purposes, regarding the presence of stakeholders such as 'Peers' and 'Others'. We finally analyze how these strategies appear differently in each circumstance, along with the unified taxonomy accompanying examples.

1 Introduction

Noisy text usually originates in unintentional errors in the conversion process of speech or signal. They are also observed in the digitized text production along with spelling errors or grammar mistakes (Subramaniam et al., 2009). However, sometimes noises are intentionally inserted by authors. For instance, Airbnb users write reviews in a code-mixed manner to convey negative information about the accommodations to other users who share the language without being offensive to the foreign host. In another viewpoint, users who share the social identity of a specific community such as Reddit or 2ch create and use memes for fun, which play a role as jargon in those communities (Merritt, 2012). Also, the users of broadcasting platforms such as Twitch, or online games, use profanity terms in an undetectable way (Blashki and Nichol, 2005), to express their emotions and at the same time avoid the detection of automatic censoring systems (Märtens et al., 2015).

In this study, we investigate and formulate the stakeholders of such strategies, namely the author, other users in the community who *should* understand the text, and the ones who should *not* understand the text. Though writing and posting text is done by the author, the other two influence how the avoidance strategies are represented. In other words, the intention of generating noisy text varies depending on the existence of each stakeholder, and this variation may also affect the strategy used. This typology has been domain specifically studied in online censorship and resistance (Elahi and Goldberg, 2012; Beato et al., 2014; Mokwena and Banda, 2019), but we focus more on the generalized agents of this 'text gaming' (Haapoja et al., 2020).

We conduct an in-depth analysis to make up a transferable taxonomy. Our target domain is the Korean online space, where web texts are actively generated and rapidly spread across various communities. We intend to provide a unified taxonomy of posting typology and avoidance strategy, based on the observation of online space expressions. Our contribution is as follows:

- Typology for intentionally noisy online text posting with stakeholder attributes
- A unified taxonomy for text gaming typology and avoidance strategies with domain and language-specific examples

2 Related Work

Diverse strategies have been suggested for text hiding, and they can be acknowledged as linguistic and non-linguistic ones. Non-linguistic strategies are studied in coding theory, and the encoded contents may not be comprehended by humans Agarwal (2013); Taleby Ahvanooey et al. (2019). In Agarwal (2013), various strategies for steganography are reviewed and compared, and the non-linguistic approaches adopt codewords (Rahman et al., 2017), unicode (Aman et al., 2017), word replacement (Ahvanooey et al., 2018), etc., that mainly aim to transmit the message without being exposed.

Proceedings of the 2021 EMNLP Workshop W-NUT: The Seventh Workshop on Noisy User-generated Text, pages 56–61
November 11, 2021.

In contrast, linguistic strategies are often visible to humans. Those include sentence-level order replacement (Topkara et al., 2006), pragmatic transformation (Lu et al., 2009), or transliteration (Khairullah, 2019). Taleby Ahvanooey et al. (2019) classified those approaches into semantic, syntactic, and statistic ones, while they were studied in view of steganography rather than intentional text noise. Subramaniam et al. (2009) has handled the issue in a manner closer to our approach, listing up strategies such as deletion of characters, phonetic substitution, abbreviation, using dialects, etc. However, it is less concerned with why the users adopt such techniques, considering possible stakeholders or pragmatic context.

In a wider view, the topics above are closely related to avoidance strategies and code-switching, which are mainly studied in sociolinguistics in terms of language varieties and community jargon. [1] Famous examples are *'mother-in-law'* codes used among Dyirbal language of northern Queensland, where everyday speech style *Guwal* is replaced by an 'avoidance' style *Jalnguy* when certain opposite-sex relatives (especially in-law) are present, maintaining the syntax but alternating the lexical items (Bell, 2013). Similar happens among Basque-Spanish bilinguals, where linguistic code-switching takes place to fill lexical gaps, convey certain attitude, or smooth negative connotations (Barredo, 1997). These cases refer to when the utterance is modified to fit the proper style of the speech community.

Besides, some code-switching may intend to hinder the speaker's genuine message to specific listeners. In previous studies on online censorship and avoidance strategies (Elahi and Goldberg, 2012; Mokwena and Banda, 2019), the authority that hacks the original text is usually assumed as government. Moreover, such authority is not the only factor that users generate noisy text in the online space; *Meme* is a representative case (Merritt, 2012) that games text with less consideration on being censored. We would like to present a typology that can encompass those phenomena.

We focused on that these strategies are observed more frequently beyond the Latin alphabet writing system due to the usefulness of code-mixing and transliteration. Accordingly, we qualitatively analyze the Korean online space, where the usage of the featural writing system Hangul and English code-mixing are both active (Cho et al., 2020). In specific, the featural writing system is advantageous in diverse phonological representations and the agglutinative nature of Korean allows flexible substitution of code-mixed words.

[1] Here, the community does not only refer to online communities but includes a variety of subgroups of language users.

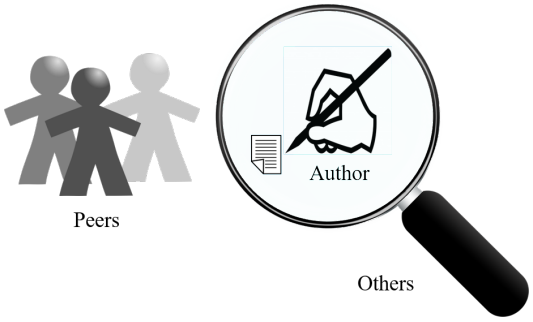

Figure 1: Stakeholders around noisy text generation.

3 Observation

We demonstrate how our observation of Korean online space had led us to set stakeholders of intentionally noisy texts. We build a typology based on them and link it with specific linguistic strategies.

3.1 Stakeholders

The stakeholders of using an avoidance strategy can be categorized into three types (Figure 1).

The first type of stakeholder is the *author*, who writes the text using an avoidance strategy. The author's intention of strategies is to share information with "peers" and prevent "others" from understanding the information. It does not matter whether the text exists temporarily or long-lasting, nor is it private or public. The author is determined at the moment of text posting, while the text type and intention are influenced by other stakeholders.

The other two stakeholders are i) '*who should understand the text*' and ii) '*who should not understand the text*'. Here we indicate the former as '*Peers*' and the latter as '*Others*'. *Peers* are the group to which the semantics of the text should be conveyed, regardless of the properties of the message. *Others* are an individual, a group, or a system that can see the text, but should not recognize its genuine semantics. *Others* could be filtering or censoring algorithms run by the platforms or human users who use decoding (e.g., machine translation) algorithms to translate the text. The conveyed message is content that may challenge or deceive *Others*, or harm the author's (or *Peers*') relationship

	Peers	*Others*	Examples
Tricks	*Yes*	*Yes*	**Google-trickers** in Airbnb Spam message with symbols Messages that avoid censorship
Memes	*Yes*	*No*	**Yaminjeongeum** Community jargons Puns with unknown sources
Fillers	*No*	*Yes*	**Leetspeak** Swear words with number (in game or broadcast chats)
Codes	*No*	*No*	Encryptions Steganography

Table 1: A typology of text posting concerning stakeholders. Bolded are typical ones of each category.

with *Others*.

3.2 Typology

With stakeholders introduced above, we categorize the posting of intentionally noisy texts (Table 1). In all cases, it might be vague if each stakeholder ('*Peers*' or '*Others*') is present/absent, identified/unidentified, or targeted/untargeted. Since we interpret this typology in view of author intention, *Yes/No* is the closest to the presence of each stakeholder, regardless of its identity.

The first case is when both *Peers* and *Others* exist. We call this case **'*Tricks*'**. In this case, people use the avoidance strategy in order to convey a message to *Peers* while not disclosing a genuine intention to *Others*. In Airbnb, *Peers* are viewers of the review who is familiar with the language, and *Others* are eventually the hosts who want to discern the sentiment or toxicity of the review using an automatic machine translation system. Similar happens in spam advertising or online swear words that head other users, though the purpose differs. Furthermore, in a social act such as #*RiceBunny*, Chinese web users use a combination of *Rice* (米, mi) and *Bunny* (兎, tu) to avoid the authority's censorship towards #*MeToo* movement. *Peers* are the ones who are on the same side or the target audience, and *Others* are authorities that seek to block the circulation of such information using various types of censorship.

The second case is when *Peers* exist, but *Others* does not exist. We call this **'*Memes*'**. Here, people use avoidance strategies mainly for fun, while the modified terms become expressions commonly used within a specific community. This meme is generally not intended to deceive or offend the authority or system. [2] In Korean online space, those memes appear in communities such as DC Inside, which corresponds with Reddit or 2ch, in the name of *Yaminjeongeum* (Wikipedia, the free encyclopedia). *Yaminjeongeum* is a composite of *Ya* that comes from *Yagoo gallary* (baseball subreddit) and *minjeongeum* that comes from *Hunminjeongeum* (Official name of Hangul), indicating the writing style within the specific online community. Community users generate and spread puns with textual transformation (addition, deletion, substitution, etc.), intentionally making the original Hangul text noisy. This 'Memes' case also includes other community jargons and puns with unknown sources (Merritt, 2012), along with all the expressions acknowledged by a specific community as a 'code' but not in other communities.

The third case is when *Peers* does not exist but *Others* exist. We call this **'*Fillers*'**. In this case, people use noisy text that can be censored by the authority, albeit they do not intend to let other users grasp its genuine semantics. For instance, some game chats that do not target anyone often contain profanity terms and are banned by the detection system. This can be similarly observed in broadcasting systems or community boards. This led people to create a set of idiomatic terms that are composition of linguistic and numerical symbols, namely 'leetspeak (13375p34k)' (Blashki and Nichol, 2005), to reveal the expression and avoid the censorship. The distinction of 'Tricks' and 'Fillers' would be that, the former has a certain meaning or a target, while the latter is more close to an exclamation that incorporates probably-censored terms. This distinction is more close to that between 'statement' and 'exclamation' of Allwood (2000). Such expressions seem to be small, but are quite frequently found over the online space.

The final case is when *Peers* neither *Others* does not exist. We call this **'*Codes*'**. Private diaries and self-posts all belong to this category. It is encryption only the author can identify, and not for others. Steganography and text watermarking are main candidates of this category. However, we are not investigating this type in the unified taxonomy, since linguistic strategies are seldom utilized here and the resulting texts are usually unreadable.

[2] However, apart from such intentions, societal bias or hate toward a specific group may be inherent in the meme.

	Morphological	Morpho-phonological	Optical	Semantic	Etc.
Tricks	간1편한 만v남 (≈ g1rl 4 u)	ㅈ가튼num (≈ motherfux0r)	ㄱH쓰근 ㅐㄱi (≈ 5h17)	진짜 the love 네요 (← 진짜 더럽네요) [It's really dirty]	가 족같은 장소 (← 족같은 장소) [fxxking place]
Memes	노인코래방 (← 코인노래방) [Coin karaoke]	민숙희 (← 민스키) [Hyman Minsky]	숲틎훈 (← 金長훈) [Jang-Hoon Kim]	과연자학 (← 자연과학) [Natural science]	-메- (← 메이플스토리) [Maple Story]
Fillers	존웃 (≈ lmfao)	이런 ㅁ ㅊ (≈ wtfk)	ㅆ1발 (≈ fuc1<)	^^ ㅣ발	Tlqkf

Table 2: A unified taxonomy with example expressions. ≈ denotes a liberal translation concerning the avoidance strategy, and [ ] denotes a direct translation of the original text indicated with ←. For **Etc.** examples, [fxxking place] contrasts with the original text (favourable place), and [Maple Story] is a simple pun for a Korean online game, Maple Story, used to indicate the game maniacs. We did not find a suitable liberal translation for '^^ ㅣ발' and 'Tlqkf', but their semantics equals 'ㅆ1발'.

3.3 Strategies

There are a total of four types of strategies that are widely observed in the Korean online space. [3] The first is *morphological*, which is the most simple way and includes the sub-strategies such as jumbling characters or words (Perea and Lupker, 2004), sometimes in a code-switched manner. The next is *morpho-phonological*, which distorts the written characters using phonological similarity (Hiruncharoenvate et al., 2015) or glottalization, letting the reader recognize the original text based on the pronunciation. The third one is *optical*, which adopts character substitution or uses redundant consonants in the Korean writing system (Sang-Cheol and Egorova, 2021). The final one is *semantic*, which includes metaphor and sarcasm, sometimes using contrasting homophones.

It is not guaranteed that these strategies comprise the whole approach towards text hiding. However, we deemed this strategy set is sufficient for unifying the expressions with the proposed typology.

4 Unified Taxonomy

Combining the materials in the previous sections, we made up a table for the unified taxonomy of intentionally noisy user-generated texts (Table 2).

Trick words, as mentioned earlier, have a purpose of conveying information to *Peers*, but not to *Others*. Thus, most of these texts contain messages that are adversarial to *Others* (i.e., censoring or translation algorithms), put them on the spot, or disobey the rules.

- '간1편한 만v남' is a teasing term that appears in spam messages, which has a direct meaning of 'girl, easy to meet'. Advertisers insert numbers or Latin alphabets to avoid the spam detectors, so that the information is delivered to the users.
- 'ㅈ가튼num' means 'what a motherfxxxer', where 'ㅈ', '가튼', and 'num' all corresponds with 'male genitals', 'equals', and 'dude', respectively. This variation frequently appears in web text, mainly used to avoid censorship in chatting, comments, or reviews.
- 'ㄱH쓰근 ㅐㄱi' is an optical modification of '개쓰레기' which means a *shxx*. It roughly matches with leetspeak *5h17* (*shxx*).
- '진짜 the love 네요' is pronounced as '진짜 더럽네요' in Korean and it means 'The room is really dirty'. However, Airbnb users use 'the love' to trick the translators and let the host understand it as a favourable message.

Meme words share the purpose of *Tricks* in conveying information to *Peers*, but *Others* to whom their exposure should be avoided is absent. Here, *Peers* can be viewers watching the same broadcast, members of a community, or people in a group. The meme is often to share pleasure with them.

- '노인코래방' is a spoonerism of '코인노래방' (coin karaoke), and it is a pun that uses character-level jumbling in Korean text. Here, '노인' means the elderly, which gives a temporary confusion, but the readers accept the term without any harm of the readability.

[3] We adopt strategies used for analyzing the text of Korean users in Airbnb review posts and apply it to our taxonomy. These strategies are roughly explained here to help the readers understand, but are to be published as a separate article (Kim et al., 2021).

- '민숙희' is a modified term of '민스키', which is a transliteration of economist Hyman Minsky. This sarcastically indicates some chart analysts who shout out market fall. It does not include any profanity terms; thus there is no reason to avoid censorship.
- '숲틎훈' is a variation of '金長훈', which is originally '김장훈', a Korean singer. This usually intends a pun, not harm.
- '과연자학' means *Indeed, self-harm* with direct translation, but it is a spoonerism of '자연과학' which means 'Natural science'. This is a meme widely used by students studying natural science in Korea.

Filler words are not intended to convey information to *Peers*, but are used to avoid censorship. Again, this can be viewed as a kind of exclamation, and it is different from 'Tricks' in that 'Fillers' does not necessarily consider the communication with *Peers*. '존웃' (≈ lmfao) and '이런 ㅁ ㅊ' (≈ wtfk) are usual terms in chats or comments, but they do not necessarily contain semantics that should be protected or conveyed. 'ㅆ1발', '^^ ㅣ발', and 'Tlqkf' all indicates '씨발' (*fxxk*), which are written with optical modification, semantic trick (with positive emoji ^^), or typing Korean in English QWERTY mode. These examples do not head other users, but are under censorship of the system.

Unfortunately, ***Codes*** were difficult to be unified within the adopted strategies. Currently, such texts are more adequately dealt with within coding theory. However, given that recent studies focus on linguistic approaches of text steganography (Taleby Ahvanooey et al., 2019), incorporating these in our taxonomy is arranged as future work.

Discussion Though we categorized the expressions according to the taxonomy, we have not yet investigated the property of each category in detail.

One phenomenon displayed is the temperature of expressions in each posting type. *Tricks* and *Fillers* should avoid the censorship of the authority, that the expressions tend to include social taboos such as profanity terms, swear words, or unethical contents.

Among them, *Tricks* tend to be more informative than *Fillers* since they have a particular purpose of conveying information to *Peers*, in the form of advertisement, secret message, or insulting. It is one of the reasons that more creative avoiding strategies are observable in *Tricks* rather than in *Fillers*, and thus easier to find diverse examples. Such creativity is also actively exhibited in *Memes*, while the aspect is more tilted to pun and that they do not have to take into account *Others*' monitoring.

In total, *Tricks* show hostility towards or harms *Others* or *Peers*, *Fillers* may alert *Others* since they contain profanity terms, but has no particular target. *Memes* are usually punned terms not showing explicit hate, but may contain a cultural or societal bias towards a certain group of people or a person.

Limitation Though we are empirically aware that *Tricks, Memes*, and *Fillers* are prevalent in the online world, used for various purposes, we have yet constructed an annotation scheme or a corpus that lists up detailed typology and corner cases. It is an essential but highly consuming process, which depends on languages and communities.

However, we believe that the characteristics of each text posting type discussed above would help us build a thorough annotation scheme concerning a variety of language phenomena. We expect that constructing such schemes would benefit a lot from our categorization, strategies, and examples, at least in selecting the online space or communities to collect the raw corpus from.

5 Conclusion

In this paper, we scrutinized the posting types and avoidance strategies of noisy web text in the Korean online space. These approaches differ from coding theoretic methodologies such as steganography and more concentrates on decoding using human text understanding. The taxonomy was defined upon the presence of wanted and unwanted readers of the noisy text, and this typology is expected to be transferable to various languages and cultures.

A limitation of our study is a lack of quantitative analysis. As a first step, we aim to annotate the texts collected from various online spaces, e.g., hate speech or biased text, using the built taxonomy. Such application will show how and why toxic comment authors trick the monitoring algorithms, at the same time allowing us to see the correlation between purposes and strategies. We hope our taxonomy can be consolidated as a useful tool to analyze intentionally noisy user-generated texts in online communities.

Acknowledgements

The authors appreciate valuable comments from anonymous reviewers.

References

Monika Agarwal. 2013. Text steganographic approaches: a comparison. *arXiv preprint arXiv:1302.2718*.

Milad Taleby Ahvanooey, Qianmu Li, Jun Hou, Hassan Dana Mazraeh, and Jing Zhang. 2018. AITSteg: An innovative text steganography technique for hidden transmission of text message via social media. *IEEE Access*, 6:65981–65995.

Jens Allwood. 2000. An activity-based approach to pragmatics.

Muhammad Aman, Aihab Khan, Basheer Ahmad, and Saeeda Kouser. 2017. A hybrid text steganography approach utilizing unicode space characters and zero-width character. *International Journal on Information Technologies and Security*, 9(1):85–100.

Inma Muñoa Barredo. 1997. Pragmatic functions of code-switching among Basque-Spanish bilinguals. *Retrieved on October*, 26:2011.

Filipe Beato, Emiliano De Cristofaro, and Kasper B Rasmussen. 2014. Undetectable communication: The online social networks case. In *2014 Twelfth Annual International Conference on Privacy, Security and Trust*, pages 19–26. IEEE.

Allan Bell. 2013. *The guidebook to sociolinguistics*. John Wiley & Sons.

Katherine Blashki and Sophie Nichol. 2005. Game geek's goss: linguistic creativity in young males within an online university forum (94/\/\3 933k'5 9055oneone). *Australian journal of emerging technologies and society*, 3(2):71–80.

Won Ik Cho, Seok Min Kim, and Nam Soo Kim. 2020. Towards an efficient code-mixed grapheme-to-phoneme conversion in an agglutinative language: A case study on to-Korean transliteration. In *Proceedings of the The 4th Workshop on Computational Approaches to Code Switching*, pages 65–70, Marseille, France. European Language Resources Association.

Tariq Elahi and Ian Goldberg. 2012. CORDON–a taxonomy of internet censorship resistance strategies. *University of Waterloo CACR*, 33.

Jesse Haapoja, Salla-Maaria Laaksonen, and Airi Lampinen. 2020. Gaming algorithmic hate-speech detection: Stakes, parties, and moves. *Social Media+ Society*, 6(2):2056305120924778.

Chaya Hiruncharoenvate, Zhiyuan Lin, and Eric Gilbert. 2015. Algorithmically bypassing censorship on sina weibo with nondeterministic homophone substitutions. In *Ninth International AAAI Conference on Web and Social Media*.

Md Khairullah. 2019. A novel steganography method using transliteration of bengali text. *Journal of King Saud University-Computer and Information Sciences*, 31(3):348–366.

Soomin Kim, Changhoon Oh, Won Ik Cho, Donghoon Shin, Bongwon Suh, and Joonhwan Lee. 2021. Trkic G00gle: Why and how users game translation algorithms. In *Proc. ACM Hum.-Comput. Interact. 5*, CSCW2, Article 344 (October 2021).

He Lu, Ma GuangPing, Fang DingYi, and Gui XiaoLin. 2009. Resilient natural language watermarking based on pragmatics. In *2009 IEEE Youth Conference on Information, Computing and Telecommunication*, pages 216–219. IEEE.

Marcus Märtens, Siqi Shen, Alexandru Iosup, and Fernando Kuipers. 2015. Toxicity detection in multiplayer online games. In *2015 International Workshop on Network and Systems Support for Games (NetGames)*, pages 1–6. IEEE.

Emily Merritt. 2012. *An analysis of the discourse of Internet trolling: A case study of Reddit. com*. Ph.D. thesis.

Lorato Mokwena and Felix Banda. 2019. Birds and bees, the 'r' word and zuma's p*nis: Censorship avoidance strategies in a south african online newspaper's comments section. *Sexuality & Culture*, 23(4):1089–1109.

Manuel Perea and Stephen J Lupker. 2004. Can caniso activate casino? transposed-letter similarity effects with nonadjacent letter positions. *Journal of memory and language*, 51(2):231–246.

Mohammad Saidur Rahman, Ibrahim Khalil, Xun Yi, and Hai Dong. 2017. Highly imperceptible and reversible text steganography using invisible character based codeword. In *PACIS*, page 230.

AHN Sang-Cheol and Kyunney Egorova. 2021. Morpho-phonological patterns of recent Korean neologisms. In *Conference on Current Problems of our Time: the Relationship of Man and Society (CPT 2020)*, pages 62–72. Atlantis Press.

L Venkata Subramaniam, Shourya Roy, Tanveer A Faruquie, and Sumit Negi. 2009. A survey of types of text noise and techniques to handle noisy text. In *Proceedings of The Third Workshop on Analytics for Noisy Unstructured Text Data*, pages 115–122.

Milad Taleby Ahvanooey, Qianmu Li, Jun Hou, Ahmed Raza Rajput, and Yini Chen. 2019. Modern text hiding, text steganalysis, and applications: a comparative analysis. *Entropy*, 21(4):355.

Mercan Topkara, Umut Topkara, and Mikhail J Atallah. 2006. Words are not enough: sentence level natural language watermarking. In *Proceedings of the 4th ACM International Workshop on Contents Protection and Security*, pages 37–46.

Wikipedia, the free encyclopedia. Yaminjeongeum. https://en.wikipedia.org/wiki/Yaminjeongeum. Accessed: 2021-08-23.