

**VIRTUALLY SAILING INTO DIGITAL PIRACY: A SOCIOLINGUISTICS STUDY
ON ENGLISH PIRATE IN R/PIRACY COMMUNITY**

*Berlayar secara Virtual ke dalam Pembajakan Digital: Kajian Sociolinguistik terhadap
Pembajakan Bahasa Inggris dalam Komunitas R/Piracy*

Received: 17 October 2023; Revised: 20 May 2024; Accepted: 24 May 2024

Daffa Ramadhan
Universitas Airlangga
daffa.ramadhan-2022@fib.unair.ac.id

How to cite (in APA style):

Ramadhan, D. (2024). Virtually Sailing into Digital Piracy: A Sociolinguistics Study on English Pirate in R/Piracy Community. *Etnolingual*, 8(1), 69—96 <https://doi/10.20473/etno.v8i1.50696>

Abstrak: Tujuan dari penelitian ini adalah untuk memeriksa hubungan antara bahasa dan dinamika komunitas r/piracy. Oleh karena itu, pertanyaan-pertanyaan dalam studi ini adalah: (1) Apa istilah dan ekspresi yang umum terkait laut yang digunakan dalam komunitas r/piracy? (2) Dengan menggunakan istilah-istilah yang berkaitan dengan laut, jenis keyakinan dan nilai-nilai apa yang tercermin oleh anggota komunitas r/piracy? Studi ini menggunakan pendekatan sociolinguistik yang melihat bahasa terhubung dengan norma-norma sosial. Oleh karena itu, CommuNalytics digunakan untuk mengumpulkan bahasa yang digunakan oleh pengguna dan moderator dari komunitas r/piracy dalam unggahan untuk membentuk korpus. Kemudian, *node* dalam korpus secara manual ditandai dengan Sketch Engine. Selain itu, analisis jaringan *built-in* digunakan untuk memeriksa pola komunikasi untuk menentukan aktivitas sosial. Akhirnya, penulis menggunakan teori variasi bahasa Wardhaugh dan Janet. Korpus terdiri dari 3.265 unggahan yang berisi 103.117 kata. Kata-kata yang paling sering ditemukan dalam komunitas ini adalah *torrent* (142), *pirate* (110), *port* (29), *seas* (17), *ship* (15), *mate* (13), *sail* (13), *treasure* (4), *plunder* (3), *mate* (3), *adventure* (2), *map* (2), dan *loot* (1). Studi ini mengungkapkan pentingnya diskusi yang terkait dengan berbagi *file* dan *torrenting*, seperti yang dibuktikan oleh frekuensi tinggi istilah "torrent" di korpus. Penelitian ini menyoroti hubungan antara bahasa dan dinamika komunitas di ruang maya, serta menekankan peran bahasa dalam membentuk dan mengekspresikan identitas dan minat sosial. Studi ini berkontribusi pada pemahaman kita tentang bagaimana bahasa dan masyarakat bertabrakan dalam komunitas daring tertentu, memberikan wawasan tentang aspek sociolinguistik dari Subreddit r/piracy.

Kata kunci: Masyarakat, pembajakan digital, r/piracy, istilah kelautan



Abstract: *The objective of the study is to examine the connection between language and the dynamics of the r/piracy community. Therefore, the questions in this study are: (1) What are the common nautical-related terms and expressions used within the r/piracy community? (2) By using nautical-related terms, what kind of beliefs and values are reflected by the members of the r/Piracy community? This study employs a sociolinguistics approach that views language as intertwined with societal norms. Therefore, Communalytics was used to gather the language used by both users and moderators of the r/piracy community in posts to form a corpus. Then, the nodes in the corpus are manually tagged with Sketch Engine. Additionally, a built-in network analyzer was used to examine the communication pattern to determine social activity. Finally, the author used Wardhaugh and Janet's theory of language variations. The corpus consists of 3,265 posts containing 103,117 words. The most frequent words found in this community are torrent (142), pirate (110), port (29), seas (17), ship (15), mate (13), sail (13), treasure (4), plunder (3), mate (3), adventure (2), map (2), and loot (1). This study reveals the importance of discussions related to file-sharing and torrenting, as evidenced by the high frequency of the term "torrent" in the corpus. This research sheds light on the relationship between language and community dynamics in online spaces, emphasizing the role of language in shaping and expressing social identities and interests. The study contributes to our understanding of how language and society intertwine within specific online communities, providing insights into the sociolinguistic aspect of the r/piracy subreddit.*

Keywords: *Society, digital piracy, r/piracy, nautical-terms*

INTRODUCTION

Several scholars have studied digital piracy (i.e., Kim et al., 2018; Dey et al., 2019; Hui and Png, 2003; Herz and Kiljanski, 2018; Peng et al., 2023), and according to their study, it can be concluded that piracy is illegal in many countries for violating international copyright laws by 'stealing' the work of the content developers. Therefore, often, digital piracy is considered to do more harm than good. For example, digital piracy will reduce the price and sale of its counterpart legal content (Kim et al., 2018; Dey et al., 2019; Hui and Peng, 2003; Herz and Kiljanski, 2018). However, it turns out that there is a good side to piracy. According to Peng et al. (2023), the presence of piracy can sometimes create a win-win solution under the reselling contract for the content developer and the platform. Although it does not necessarily mean that Peng et al. are siding with the act of piracy, the authors give their opinion regarding antipiracy strategies such as strengthening the protection of DRM (Digital Rights Management).

Another study conducted by Wicaksono and Dekar (2017) found that there are several reasons why people engage in digital piracy in Yogyakarta. First, the users found no difference between the original content and its pirated counterparts. Interestingly, the social aspect around its users also plays a major role in consuming digitally pirated goods by creating a "pirating is safe and brings no consequences" discourse, therefore making a



justification for their actions. The extensive cost of original digital goods also justifies why people pirate, especially if they come from economically low-income backgrounds.

While it is interesting that several scholars found the good, the bad, the legal, and the illegal side of piracy activity (and how to take a countermeasure step against it!), it is worth noting that there are only a handful of studies that focus on the people of digital piracy, especially in the linguistics aspect (e.g., Guidi, 2012; Kautsarina, 2017; Lewis, 2015). However, the previous studies didn't explore the usage of language or analyze the social dynamics within a community. In other words, the current research will extend its scope to the community of piracy. This approach provides a fresh perspective on the piracy community, as previous studies primarily focused on the potential impact of piracy on content developers' financial gains. The author would like to add that this research will not discuss the morals and legality of piracy activities, but rather the language and norms shared within the digital piracy community. The digital piracy community in this research will be the r/Piracy subreddit, a community on Reddit dedicated to discussing and sharing information related to piracy, particularly in digital media. An interesting linguistic-related phenomenon the author found in this community is the usage of nautical-related terms that refer to piracy activity within the community. Therefore, this study implies that, theoretically, it contributes to providing a new insight into sociolinguistics by examining the language used in a community, especially the community of illegal activities, to reveal how language adapts to new technologies and social dynamics.

While there are other piracy communities outside of Reddit, however, Reddit was chosen as a social media because, according to M. Arazzi et al.'s (2023) finding, there is an interesting result in which they found that the users on Reddit tend to use formal language whenever possible, indicating that the discussion on Reddit is more civilized than other social media such as Twitter. From a linguistic perspective, Reddit's formal language usage reflects the community's standards and ideals. It is possible that people in the community appreciate civility and professionalism. This could play a significant role in the dynamics of r/Piracy community, especially since this subreddit is about pirating digital media and not a work environment or even a legal discussion. However, being what is assumed to be 'a more



civilized social media' does not necessarily imply that the assumed 'uncivilized' social media is not worth observing. Rather, it highlights the unique characteristics of the Reddit community in terms of language use and the level of civility maintained in discussions. The study does not necessarily imply that other platforms are less civilized or less worth studying, but rather that Reddit stands out in this regard.

Therefore, the questions that arose in the current study are as follows:

1. What are the common nautical-related terms and expressions used within the r/Piracy community?
2. By using nautical-related terms, what kind of beliefs and values are reflected by the members of the community?

RESEARCH METHODS

The author began this research by joining the r/Piracy subreddit and observing the way community members interact with each other. The data in this research is gathered by observing the community, in which the author is on the participating observer spectrum, where the author places himself somewhere out of the way but occasionally interacts with the observed. The author took note of the language used in posts and comments and will try to provide a documentary (screenshot) for this research. The observer method was chosen in this research to allow a more natural approach to collecting data, in a way that the author can observe behavior and interactions in their natural setting rather than relying on self-reported or retrospective accounts. The author visited the community at least once daily by reading through posts and comments while engaging with the community by commenting occasionally.

The author proposed an interview with several community moderators besides relying only on observer participant methods to gain insight into the organization and management of the community, as well as their perspectives on issues related to piracy. Regrettably, the request to carry out an interview has been rejected as per the moderator's statement:

“Search for some older posts containing 'research', 'survey', 'questionnaire' to see what has been deemed acceptable for a post on the sub.



*In short, asking for comments and publicly responding is fine, an online survey that doesn't appear to be phishing is fine, a Reddit poll is fine, but asking for PMs, emails, sharing your own email, overly invasive online questionnaires, or requesting payment or donations will result in a quick ban and your post removed.
I'll leave other mods to reply if they wish, but personally I won't be participating."*

Google Forms might be able to hide the author's email, however, the piracy community tends to avoid Google due to its infamous internet activity tracking nature. However, the good news is that the posts and comments in this community can still be collected through Communalytic, an API tool for collecting posts, comments, and replies from a specific digital community on various social media. To collect the data, the author inserts the designated subreddit name, which, in this study, is r/Piracy, and then the API tools will automatically collect all recent posts containing text up to the last 14 days. In other words, random sampling collection is used to collect the sampling of the whole data. However, it is worth noting that Communalytic automatically excludes reduplication or reposts to reduce redundancy.

Further, using the built-in network analyzer, the author will visualize the communication that happened during the last 14 days, revealing possible social network relationships. Their linguistic behavior was strongly influenced by analyzing social network relationships. Unlike social class categories, network structures are unique and vary between different groups and locations (Milroy, 1987). Therefore, this study falls into the trend study category since the author examines different members and moderators of r/Piracy at different times. The collected data will be cleared of unnecessary text, such as the usernames of the posters or commentators. Once the data has been sorted and cleared, it will be consolidated into a unified corpus. Then, with the help of Sketch Engine, the author will look manually for nautical-related terms and mark them as nodes.

Thus, this research is a sociolinguistics study because it focuses on the language use and communication patterns within online communities, specifically on Reddit. The author wishes to highlight that, as stated by the moderator of r/Piracy, conducting research is allowed as long as the author refrains from requesting or disclosing personal information of any individual, including their own, in any manner. Therefore, even if Reddit encourages its users



to euphemize their usernames, to stand with the ethical code of conducting a sociolinguistics study, this research will protect each user's privacy concern by blurring out every personal or sensitive information such as (but not limited to) username and flairs.

THEORETICAL REVIEW

Sociolinguistics and Digital Community

Sociolinguistics is the study of how language works in casual conversations and media, as well as the societal norms, policies, and laws in language (Wardhaugh and Janet, 2021: 1). Further, it has sociolinguistics emphasized that it is not a study of facts, but rather the study of ideas about how societal norms are intertwined in a daily language use (Ibid). In other words, sociolinguistics views language has strong relations with society. Contrary to Chomskyan linguists' point of view, Labov (in Wardhaugh and Janet, 2021: 4) argues that rather than knowing a speaker's linguistic competence and performance, it is better to understand an individual's linguistic behaviour through the societal knowledge they belong to. In other words, sociolinguistics concerned with real language use, and not the language of some ideal speakers (Ibid).

Sociolinguistics, then, has been introduced into two major perspectives: micro-sociolinguistics, and macro-sociolinguistics. The former investigates how social structure influences the way people talk and how language varieties and patterns of use correlate with social attributes such as class, sex, and age. The latter investigates the attitudes and attachments for the functional distribution of speech forms in society, language shift, maintenance, and replacement, the delimitation and interaction of speech communities. Hence, macro-sociolinguistics are often called the sociology of language.

Since language and society are interrelated, it is worth noting that in different societies, people use different variants of language, that is in sociolinguistics, refers to linguistic changes and differs based on social factors. However, it is worth noting that variation and change are two separate matters, because not all variation is a sign of, or leads to, change (Wardhaugh and Janet, 2021: 199). Linguistics variation refers to stable linguistic differences that can persist over time without leading to change, influenced by factors like socio-



economic class, age, and gender. On the other hand, language change is characterized by a progressive and linear evolution. Identifying and understanding language changes involves studying their initiation, spread, and maintenance (Ibid). The current study focuses on the former.

The society in which the author is interested in the current study is digital society, in which, Sá et al. (2021) explained that it erected from the foundation of information and communication technologies (ICTs), it represents a fundamentally new situation for people and social institutions. The new digital society, then, creates a digital environment that enforce anonymity and assume any identity we choose, without the social constraints of face-to-face interactions. This anonymity allows for more uninhibited expression without fear of social consequences. Additionally, digital communication grants us the ability to selectively choose whom we communicate with and on what terms, giving us unprecedented control over accessibility. The nature of communication itself also has transformed, with faster and more distant phatic communication becoming prevalent. Phatic communication refers to small talk or pleasantries that serve as bonding rituals and regulate discourse. Digital platforms like texting, Twitter, and social networking sites facilitate quick and brief exchanges that fulfill the function of phatic communication. For example, texting a brief message after class or sharing updates on social media about mundane aspects of life.

Therefore, social media is a part of the society, and the significance of studying it is these social media contains communities that are characterized by a diverse range of social institutions and interconnected sets of expectations, beliefs, and moral values about the world, which may intersect but are not necessarily identical.

Pirate English and Its Correlation with Digital Piracy

Sociolinguistics is the study of language use within or among groups that consist of at least two members, have similar purposes (though they might change), and exist for the time being (temporary or quasi-permanent), and the group that we are trying to study is called the speech community (Wardhaugh and Janet, 2021: 63). The speech community shares some kind of communal feeling about a certain linguistic behavior and societal norms (Ibid.). It

means that sociolinguistics views language as playing a crucial role in society by serving as a means of communication and social reproduction.

Furthermore, Wardhaugh and Janet (2015: 20) stated that in social language, it is observed that identities, power, and solidarity coexist. Identities refer to the linguistic construction of membership in social group categories, while power refers to the ability to control events (and people) in order to achieve individual goals. On the other hand, solidarity refers to the common motivation that motivates individuals to form a group. Therefore, identities, power, and solidarity in sociolinguistics are also applicable in maritime, as the author would like to give examples of pirate English. Since the current study examines the language variations in the digital piracy community with real-world pirates, which are evidently analyzed in the result and discussion sections, the language variations of English pirates in the 17th to 18th centuries, according to Mares (2017), have some interesting features. Those features are as follows:

1. Assertion: opening the sentence with 'there', 'here', and 'it'. (e.g.: "Here's me wishing")
2. Negation. ("There be no quarter")
3. Exclusion of the had/have. ("Seed you a man wi' one leg?")
4. Replacement. ("this" replaced by "that", "here" by "there")
5. Characterization/description. ("A hand to")
6. The end of the sentence: 'says I' or 'says you'. ("Properly warn ye be, says I")
7. Third personation.
8. Tense shift in narration: Past-Present.
9. Switched pronouns: nominative for objective. ("I for me, who for whom, we for us")
10. Plain wrong verbs. ("Mr Samuel will have guns to sell, if you gets gold to buy them")
11. Substitution of 'be' for 'do'. ("Be you have any doubloons...")
12. Adjectives after nouns. ("We gave them firearms and cartouches sufficient.")
13. Dropped preposition. ("Yesterday at twain in the afternoon, we weigh'd again, with a breeze...")

Other features that have been found through Mares' study also include pronunciations as follows:



1. Raising the /z/ phoneme in words 'twinty' (twenty) and 'inemy' (enemy).
2. The non-pronunciation of the nasal [ŋ]: 'lanth' (length), or 'mornen' (morning).
3. Medial omission of /ɹ/ in Jamaican 'gens' (against).
4. Final omission in Barbadian and Guyanese pronunciations /biIn/ or /bi:n/ (being).
5. /ɹ/ omission as a feature of Pirate English, exemplified in phrases such as "I a mayner o'non speakin'" (in a manner of speaking).
6. Omitted the pronunciation of pre-consonantal /r/: 'hard' 'word', 'organ' and 'bird'.

The digital piracy community adopted this maritime language in the modern world, as stated by Young's (2007) study that piracy has changed in nature and location over the centuries, adapting to changing times and technologies, indicates that the digital piracy community adopts maritime language as a way to connect with the historical roots of piracy and to express a sense of rebellion against mainstream culture (Burkart and Jonas, 2015). Through an examination of the pirate terminology employed in r/Piracy, it is conceivable that we might uncover the 'Piratical Ethos', encompassing the motivations, philosophies, and attitudes towards intellectual property within this online group. (Lewis, 2015: 46).

Reddit and Piracy

A study focused on Reddit as its research object was conducted by Proferes and the corresponding authors in 2021. According to the respected authors, Reddit is a public platform where discussions and content are visible to anyone, regardless of having an account. Users can vote on posts and comments to determine their visibility. Creating a Reddit account only requires a unique username and password, without email verification. Interestingly, Reddit encourages its users to uphold privacy by using pseudonyms; therefore, it is common for users to have multiple accounts, including the so-called throwaway accounts for temporary use (Ibid.). Acquiring demographic data might be challenging, however, it is widely claimed that the predominant demographic of Reddit users consists of males aged between 18 and 34.

Reddit allows its users to create a subcommunity called Subreddits, or r/. These subreddits are devoted to particular subjects and passions, allowing users to subscribe and



engage, resulting in a customized homepage with content that is specifically targeted to their interests. Subreddits are created and organized by their users, called 'moderators'. Moderators are responsible for enforcing the subreddit's rules and guidelines, ensuring that the content posted is relevant and appropriate for the community, and maintaining a positive and engaging environment for users to participate in. Therefore, each individual subreddit has its own rules, cultures, and norms. Some subreddits even have specific rules for researchers, requiring approval for research-related posts and surveys. Reddit's user agreement prohibits unauthorized data collection, although their API allows access to site content.

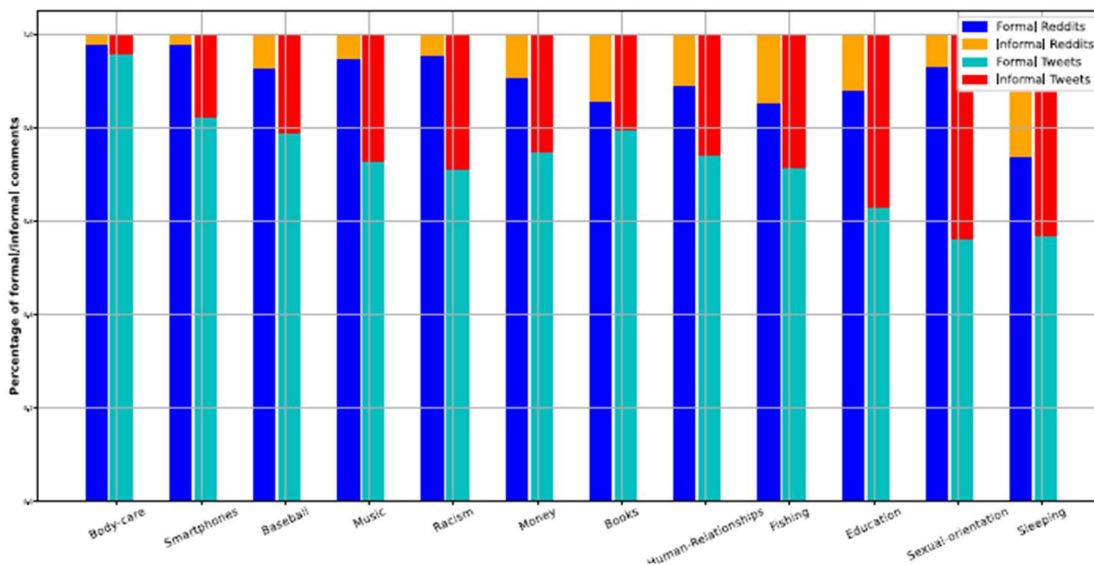
Reddit, however, has almost no limitations for the contents of the subreddits, in a way that it also attracts some activities such as illegal life tips, pornography, and piracy. The Indonesian government has included some content, particularly pornography, in the list of prohibited websites, which is why Reddit is not well-known among Indonesians. However, the one that will be discussed here is the subcommunity in Reddit called the r/Piracy subreddit.

Traditional piracy is commonly perceived as the act of forcefully stealing from the crew or passengers of another ship. (Segrest, 2021: 49). However, it seems that the digital world has adopted Symth's definition of piracy, which is the unauthorized reproduction and distribution of products protected by intellectual property rights, be they physical or digital (Bradley and Kolev, 2022; Fisk, 2009; Sahni and Indranath, 2019). While stricter border control by Customs can effectively prevent piracy in physical goods, it is considerably more challenging to prevent piracy in digital goods. The distribution of digital goods has been demonstrated to be highly efficient and virtually cost-free, as they can be rapidly disseminated across the internet. (Athey and Stern, 2015; Varian, 2003). The copyright laws vary in each country, and in Indonesia, digital goods are considered intellectual property rights and are protected under the Constitutional Law of the Republic of Indonesia Number 28 of 2014 (Astuti and Marpaung, 2021: 1090). Piracy, however, is not the work of an individual but within individuals or communities. The piracy community has long been associated with the illegal distribution of copyrighted materials, and one of the largest communities the author finds (and is accessible by the public) is r/Piracy on Reddit.



M. Arazzi et al. (2023) studied the importance of language and how it can affect the evolution of OSNs (online social networks). M. Arazzi et al. contend that prior to their research, scholars such as Balani and De Choudhury (2015), Schrading, Alm, Ptucha, and Homan (2015), Mohan et al. (2017), and Ferrer, van Nuenen, Such, and Criado (2021) have conducted language studies on OSN. These studies have demonstrated the feasibility and significance of studying language, particularly Natural Language Processing (NLP), because OSN serves not only as a virtual community but also as a platform for shared ideological groups. The OSN M. Arazzi et al. studied two general-purpose social platforms, Reddit and Twitter, in which they studied the popular topics found in both OSNs: smartphones, human relationships, body-care, education, books, money, racism, sexual orientation, baseball, fishing, sleeping, and music. M. Arazzi et al. employ topic modeling, sentiment analysis, and NLP techniques, utilizing tools like DataMiner for social media, to collect the data. As a result, M. Arazzi found on Twitter that more friendly and colloquial language is expected for topics related to personal hobbies such as sports and music, whereas a more sophisticated language is used in politics, the economy, etc. On Reddit, however, it seems like users tend to use formal language whenever possible.

Figure 1 A comparison level of each topic on Twitter and Reddit (M. Arazzi et al., 2023: 12)



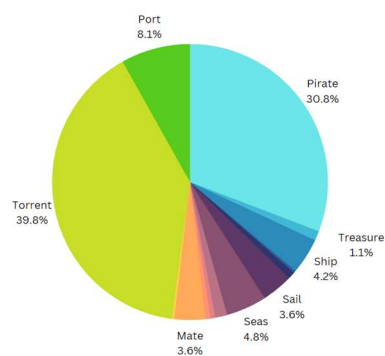
RESULTS AND DISCUSSIONS

Results

Corpus

The corpus in this study was gathered through the use of API kindly provided by Commanalytic to crawl the web. The so-called ‘Reddit Blackout’, unfortunately, made the corpus constrained during which numerous subreddits remained inaccessible and inactive as a form of protest, limiting access to the data collection, including r/Piracy Subreddit. The massive online protest from many subreddit moderators and users was conducted in order to fight the decision of Reddit developers, in which they made Reddit API only accessible via payment methods that cost \$20 million annually. Therefore, the corpus tools can only gather a total of 3,265 posts containing 103,117 words. However, the author can still draw on several nautical-related terms that appeared within the community.

Figure 2 The overall percentage of nautical-related terms (word) in r/Piracy



We can infer the following from the chart:

1. The term "Pirate" appears most frequently in the corpus, with a frequency of 110. This indicates that the concept of piracy is a central theme within the discussions on r/Piracy.
2. Other nautical-related terms such as "Ship" (15), "Sail" (13), "Seas" (17), and "Port" (29) also have notable frequencies, suggesting that discussions on *r/Piracy* encompass various aspects of seafaring and maritime activities.
3. Terms like "Treasure" (4), "Plunder" (3), and "Mate" (13) are associated with the adventurous and treasure-seeking nature often associated with piracy.
4. The word "Torrent" has the highest frequency in the corpus, with 142 occurrences. This

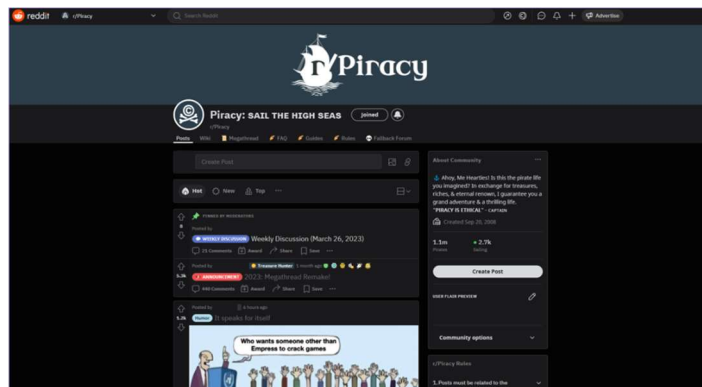
highlights the significance of file-sharing and torrenting discussions within the subreddit, which aligns with the digital piracy aspect.

5. While some terms like "Loot" (1), "Adventure" (2), and "Map" (2) appear with relatively low frequencies, they still contribute to the overall nautical and pirate-themed vocabulary within the corpus.

The author notes that the above corpus only included words commonly used in the community. Therefore, the author must manually examine some nautical terms that are not on the list by applying a case study.

R/Piracy

Figure 3 r/Piracy Homepage



Created on 20th September 2008, the subreddit has a large and active community of more than 1 million users who share a common interest in accessing copyrighted content without paying for it. The author found this to be surprising, especially since this community is open to the public, despite its activity being illegal in many countries. In this subreddit, users can find links to pirated content, discussions on methods for accessing and sharing copyrighted material, tips, and tricks for bypassing digital rights management (DRM) technologies implemented by the software developer or publisher, thanks to astonishing technical expertise by other users, or -in this subreddit, also called as- pirates. This community accepts neither payment nor donation in any kind of form regarding sharing information on pirated digital media, so the whole activity including leading and managing



this community is the work of a volunteer.

The community is led and moderated by a team of moderators who enforce the subreddit's rules, which are quite strict as follows:

1. Posts must be related to the discussion of digital piracy.
 - The focus on file-sharing, articles, and discussions about moral dilemmas with unauthorized distribution, legal developments, challenges, and other related topics are all welcome.
 - Comics, Memes & Screenshots are acceptable, but should be original and of good quality. Images that are too blurry or have a lot of unintelligible text will be deleted.
 - You are not allowed to post any visual representation for seeding / leeching ratio.
2. Don't request invites, trade, sell, or self-promote.
 - No pleading for invites to scene sites, private torrent trackers, DDL forums or other sites with closed registrations.
 - No asking for content downloads, activation keys etc.
 - No trading or selling content, especially not for monetary services.
 - No self-promotion of your own work is allowed unless you expressly receive permission from the moderators.
3. Don't request or link to specific pirated titles.
 - Yes, you can ask generalized questions (e.g., Where can I find games, movies, softwares etc).
 - No, you may not ask for a specific pirated title (e.g., "Where can I download {insert title})?)
 - Yes, you can link to the top-level domain of a site (e.g., <https://1337x.to>).
 - No, you may not link to a specific pirated title (e.g., <http://1337x.to/specific-torrent>).
 - Do not encourage rule breaking by asking which specific title another person is looking for when they make a request.
4. Don't be repetitious, spam, harass others, or submit low-quality posts.
 - Do a search on the internet or this subreddit (Megathread & old posts) first, so as to not be repetitive.
 - Quit spamming the same issues, contentious politics, & misinformation.



- Harassment of others, as well as insulting remarks & inappropriate behavior, will result in a permanent ban without warning.
- Removal of any posts that are of poor quality, lazy, or confusing is entirely up to the moderators.

The content in r/Piracy revolves around sailing the seas, but not always about sailing the seas itself. Users can post news, humor, asking questions, or even new guides not listed in the Megathread. Additionally, there are also posts containing discussions of why the users do piracy.

Topic, Sentiment, and Toxicity

In addition to corpus and r/Piracy, the author would like to add another finding of sentiment and toxicity level to capture its language use, social dynamics, and on-going events at that time. The sentiment analysis highlights the emotional tone and attitudes expressed in the language, revealing the level of positivity, negativity, and neutrality in the community. Finally, the toxicity analysis identifies instances of toxic language and behavior, providing insights into the potential impact of these on the community's dynamics and social relationships.

Table 1.
The Sentiments in r/Piracy Corpus

	Number of Posts	Negative Sentiment	Neutral Sentiment	Positive Sentiment
Sentiment	1719	323 (18.79%)	430 (25.01%)	966 (56.20%)

Table 1 was generated by Commanalytic, using both VADER and TextBlob to analyze posts and comments in English. However, due to excessive amount of reposts, the tools exclude 1819 other post out of 2906 posts. The results are the agreement, or the accuracy between VADER and TextBlob is fair according to Kohen's kappa statistic (with the value of 0.360. Vierra and Garrett: 2005). To clarify, the sentiment analysis indicates that the language employed within the digital piracy group is primarily optimistic, emphasizing the advantages of piracy and the significance of distributing digital content. However, the



analysis also reveals that there are instances of negative sentiment, particularly in response to legal action or criticism from the music and film industries. The sentiment analysis highlights the emotional tone of the language, revealing a sense of excitement and enthusiasm for piracy, as well as a sense of frustration and anger in response to legal action.

Table 2.
The Average of Toxicity Levels in r/Piracy

	Average for dataset	Highest value
Toxicity	0.12692	0.99792
Severe Toxicity	0.00366	0.66678
Identity Attack	0.00227	0.93948
Insult	0.05278	0.99205
Profanity	0.07934	0.99603
Threat	0.00267	0.68014

The API tools analyzed 2999 posts, of which posts without text and posts in languages that were excluded. The higher level of toxicity in this table indicated by how close the value to 1. The toxicity analysis reveals that there are instances of toxic language and behavior in the digital piracy community, including insults, threats, and harassment. These occurrences frequently target those who are considered as opposing the values of the community or who are viewed as a threat to its survival. The toxicity study emphasizes the potential consequences of these occurrences on the community's dynamics and social connections, exposing a feeling of strain and disagreement that can harm the community's overall health. In addition, if we look at the average dataset, this finding proves that Reddit has a moderate use of politeness (i.e. polite language that does not use insults, or even positive cursing).



Discussions

The Reason

Figure 4 A Post with “Discussion” tag

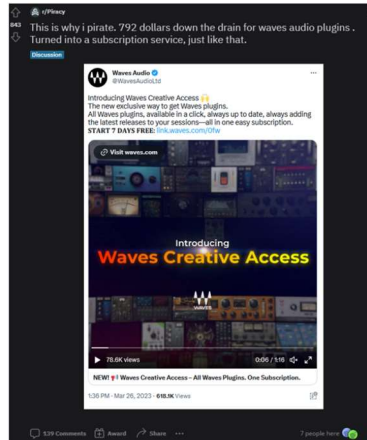


Figure 4 contains a post discussing a user stating that one of the reasons he/she does digital piracy is when a company that provides music editing software decides to move from a one-time purchase to a subscription system. The user argues that this software costs approximately \$800, but suddenly the consumer can no longer update the software anymore.

Figure 5 Time to sail the high seas

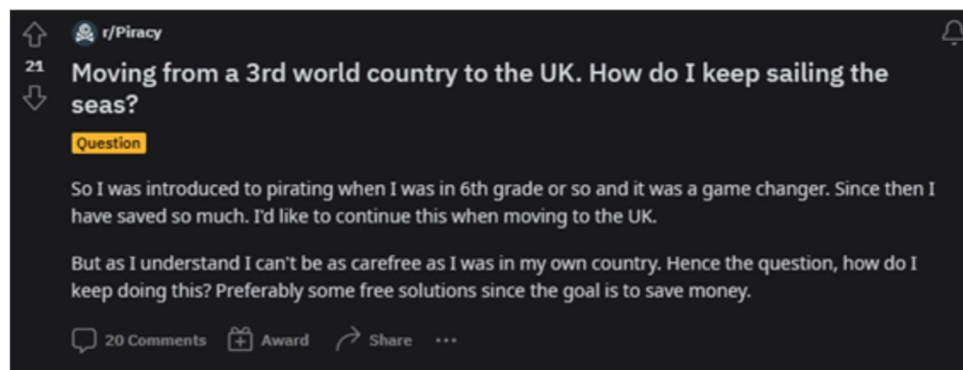


Figure 5 also indicates a similar case where a user finds out that a streaming service provider will be erasing a certain number of contents, leaving the regularly paying users not much of a choice but “sailing the high seas”. Therefore, it can be concluded that “sailing the

high seas” is a phrase used as an idiomatic expression that refers to the movement of a vessel by means of her sails along the surface of the water traveling, particularly the vast and often treacherous oceans. It signifies embarking on a voyage or maritime adventure, typically by ship or boat, that takes place far from the shore or in international waters. Therefore, the "Sailing the high seas" phrase in this community is used to evoke a sense of freedom, exploration, adventure, and rogue for committing international law of digital rights. It is often used metaphorically to symbolize embarking on a new endeavor, facing unknown challenges, or seeking exciting opportunities.

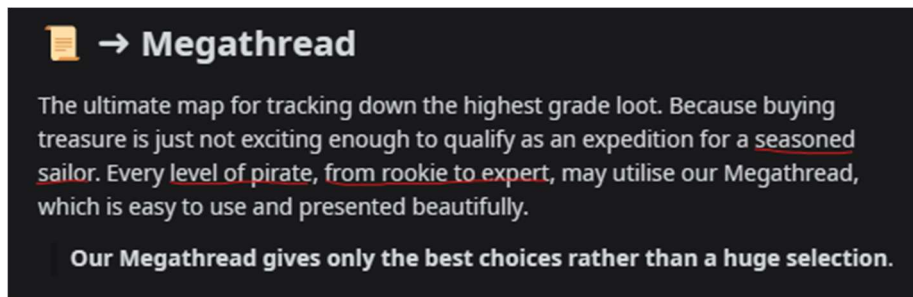
Navigating The Ship and The Expedition Gears

Figure 6 A user concerned with his/her voyage's safety



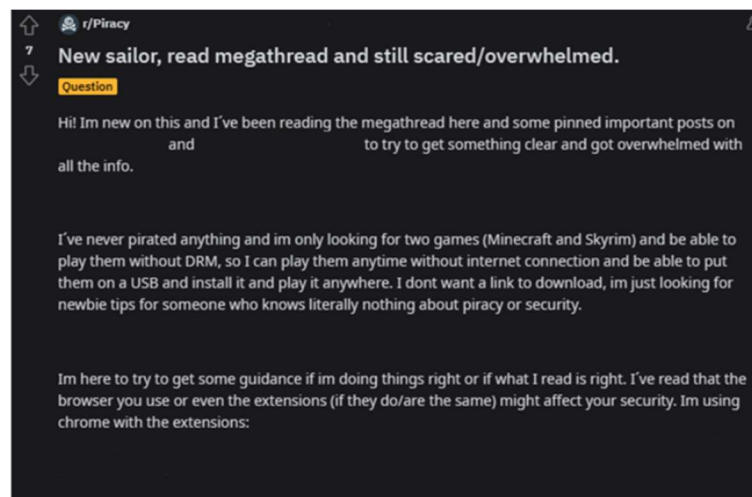
Although the international law on digital rights protections applied to almost every country, it is not surprising that this law is only a writing. However, in some other countries, this law may be much more enforced, resulting in a tight screening of internet activity scanning by IPS (Internet Service Provider), for example, the United Kingdom. For that reason, this community provides the ultimate map to navigate safely, and ultimately to plunder the treasure each user is looking for. The ‘ultimate map’ in this community is called the ‘Megathread’.

Figure 7 Definition of the Megathread



Interestingly enough, “Sailor” or even “Pirates” is used here to describe those who engage in piracy. Pirate (Segrest, 2021: 49) refers to the one who engages in an act of piracy, and piracy itself is an act of robbery or criminal violence at sea by the occupants of one vessel against the occupants of another vessel. Piracy is distinguished from privateering, which is authorized by national authorities and therefore a legitimate form of war-like activity by non-state actors (Ibid). Sea-robbers or an armed ship that roams the seas without any legal commission and seizes or plunders every vessel she meets; their colors are said to be a black field with a skull, a battle-axe, and an hourglass, hence the reason why the r/Piracy’s logo is similar to the flag of pirates. These phrases also highlight the illicit and prohibited aspect of piracy, as pirates are perceived as outlaws or even felons. The term “Sailor”, however, does not have any negative effect on the piracy community. In fact, some Sailors seem to be proud of it as shown in Figure 8.

Figure 8 A sailor out in the fear



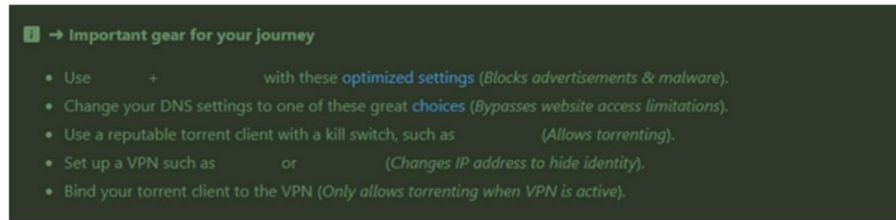
Navigating through the vast sea of terminologies in the realm of piracy can be an overwhelming task for readers. The multitude of unfamiliar terms and concepts can easily lead to confusion. Fortunately, the r/Piracy subreddit comes to the rescue by offering valuable guidance on where to begin and what to focus on. The subreddit provides a helpful banner that serves as a compass, pointing readers in the right direction and shedding light on essential information. By referring to this banner, readers can gain a clearer understanding of the piracy landscape, enabling them to embark on their journey with greater confidence and knowledge.

Figure 9 Clickable Banner containing a figure of what appears to be a pirate and “Matey” in the last word



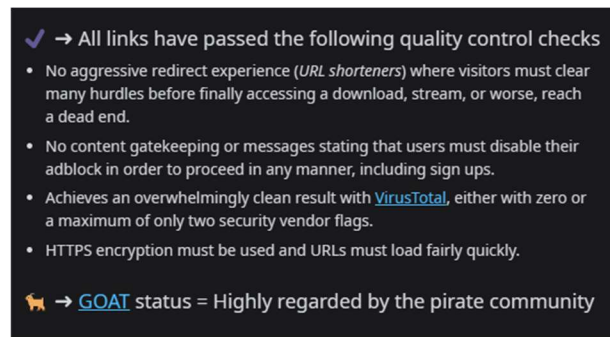
Wiki is a place where one would find useful information regarding piracy, the subreddit, which contains four chapters: the Megathread, FAQ, Guides, and Rules. Among the others, what makes the r/Piracy subreddit unique is Megathread. Megathread is a chapter containing tons of information, including where to find links for accessing the media of one’s desired content; ranging from movies to useful (software) tools, and also tutorials on how to pirate freely without the need to worry about viruses and other potentially dangerous ransomware, while avoiding detection of ISP’s activity tracking. Therefore, in addition to nautical language, the piracy community also uses technical language related to bypassing copyright protection and preventing internet tracking activities. Terms like "VPN" are used to an application for counter-measure internet address tracking, “DNS” is used as a network setting to bypass website access limitations (i.e., blocking). What is interesting, however, is “torrents”. Though it means water that destroys, digital piracy, though still unclear where it stems from, refers to an easy-to-use application from which users can download and upload pirated contents at the same time.

Figure 10 Essential gear before sailing into the seas



In the r/Piracy community, Megathread serves as a comprehensive resource providing the best available and safest choices for accessing digital content. The links provided in the Megathread undergo thorough scrutiny and are endorsed by the community or moderators, guaranteeing a high degree of safety and dependability. Therefore, any links that are not included in the Megathread are typically considered to be unsafe or unauthorized by the community. This emphasizes the significance of adhering to the recommended sources for secure and reliable access to digital content.

Figure 11 Quality Control Statement



A Ship's Hierarchy

The usage of nautical-related language in r/Piracy also signals a particular power structure and social hierarchies' status within the community. As Segrest (2021: 49) mentioned, a pirate is a sea robber, of which the word *pirata* has been formerly taken for a sea captain. Therefore, a ship must have its own crew (Sailor), obviously, a Captain. The picture below circulates around the context where a Sailor posted a question, "What if someday Reddit banned this subreddit?" Then, an individual with a specific title (marked with a red circle) appeared and explained that this community already has its own secondary

sites that can be easily accessed on the right side of the website, should the subreddit go to dust one day while inserting a humorous insult of “Scallywags” in the end. The title in the picture indicates that this individual is the leader of the r/Piracy moderators.

Figure 12 Captain’s on board



Unfortunately, the captain cannot forever rule the ship. The passage below reveals the former captain's hierarchical dynamics. r/Piracy is one of the subreddits that participated in the massive online protest of the Reddit Blackout by making the corresponding subreddit go private, restricting access to the community as a result. The protest, however, cannot go on forever, as the r/Piracy faces the consequences of the removal of the subreddit's captain by Reddit administrators and the subsequent pressure placed on the moderator team to reopen the community.

Figure 13 Captain’s off the board



What is interesting, however, the captain may lose his/her role as a community moderator, but not the ship itself (although, not as a whole). However, the captain also considers the potential impact on the remaining moderators and their loyalty, therefore invites community members to join an alternative platform, referred to as the failback forum, where

they can continue to engage in unrestricted piracy. In addition, the moderator acts as the central communication, and shows the power that he/she holds in the community through Figure 14:

Figure 14 Communication pattern in r/Piracy, generated by Commanalytic

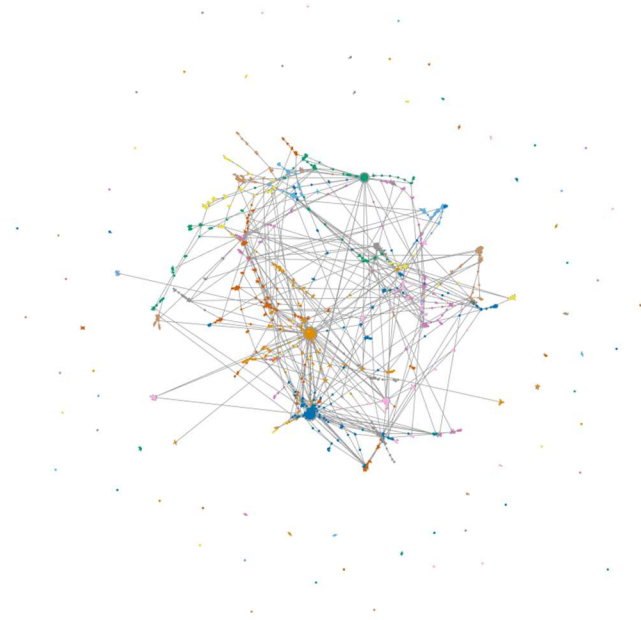


Figure 14 represents a communication pattern in the r/Piracy subreddit. The dots are the users, while the line represents to whom he/she replies or talks. The bigger the dots, the more he/she influences others, which indicates the power they possess in the community. The large orange dot in the middle is the very same subreddit's captain.

Against The Navy

Figure 15 A satirical post against a video game publisher



Users can also post humor to make fun of several circumstances that happened to be a topic in the piracy scene as shown in Figure 15. It is a documentary example of how a humor post is done in r/Piracy. The context situated in the above picture is a company of video game publisher, namely EA is actively trying to threaten people who pirate the game which EA published in 1995. One of the users complained that his/her internet provider is sending him/her a written warning over a letter for committing a violation such as piracy, hence the reason why in several countries, it is necessary to set up a protection of counter measuring the internet provider or even governments' tracking by installing DNS or VPN before pirating.

Figure 16 News about then-uncrackable video games

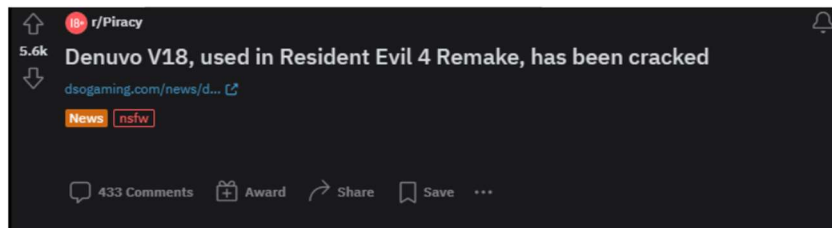


Figure 16 indicates a post containing news of a cracked video game. Cracking refers to bypassing a software DRM (Digital Rights Management) protection in order to make its users able to use the software freely, without having to pay, and the passage narrates a successful act of bypassing a notoriously-hard-to-bypass DRM, namely Denuvo. Game developers and publishers utilize DRM systems such as Denuvo to safeguard their intellectual property and enforce the requirement for users to acquire authorized copies of the game in order to access it. These systems employ various encryption and anti-tamper techniques to make it difficult for hackers and crackers to remove or bypass the protection.

Cracking a DRM system like Denuvo involves skilled individuals or groups who reverse-engineer the software, identify its vulnerabilities, and develop tools or methods to circumvent the DRM's restrictions. Once the encryption of the software is bypassed, it can be distributed and utilized by anybody without restrictions, resulting in rampant piracy.

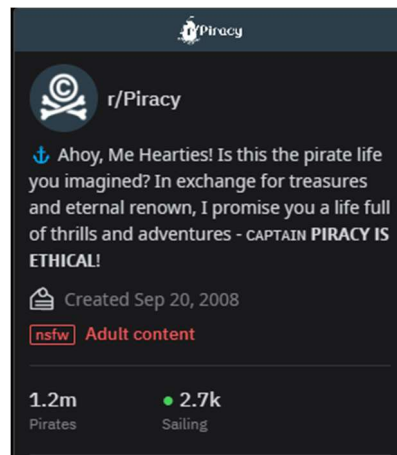
Denuvo is a well-known and challenging DRM system used to protect video games from unauthorized copying and distribution, therefore cracking Denuvo is considered a significant achievement within the digital piracy community, as it allows users to access and use the software without paying for it, which is a violation of the game developers' rights.

CONCLUSIONS

Based on the aforementioned explanation, this investigation yields many discernible conclusions:

1. Language and Culture Connection: Through digital sociolinguistics, it becomes clear that language use on the subreddit reflects the unique culture and identity of the piracy community, with the adoption of specific terminologies and expressions that are exclusive to this online community.
2. Nautical-Related Terms: The presence and significance of nautical-related terms within the language used on *r/Piracy* suggest a strong association between piracy culture and maritime traditions, even if the two have significant differences in terms of the realm where they are doing their activity. The use of such terms reflects the community's desire to maintain a distinct identity and reinforces the romanticized imagery of piracy in popular culture, which can be identified through the *r/Piracy*'s definition in Figure 17:

Figure 17 Description of r/Piracy by The Moderator



Additionally, it is evident that the culture of digital piracy is being passed on to the new generation (or user) through the use of nautical-related terms, especially in the Megathread section that provides tutorials for pirating.

3. Collectiveness: The linguistic patterns observed on *r/Piracy* contribute to the formation of a collective identity among community members. The shared use of specific terms,

phrases, and expressions reinforces a sense of belonging and solidarity within the community.

4. Online Community Dynamics and Language Evolution: The study sheds light on the dynamics of an online community and its impact on language evolution. The language is subject to continuous change and adaptation, including language in *r/Piracy* as new members join, and new piracy-related trends emerge.

These conclusions offer valuable insights into the relationship between language, culture, and online communities within the specific context of *r/Piracy*. They provide a foundation for further exploration of ethnolinguistic dynamics in similar online communities and contribute to our understanding of language use in digital spaces.

However, it is worth noting that this study is limited by its methodology factor that can only be accessed through observant methods, while only relying on frequency appearance in its corpus analysis. Should a fellow researcher place his/her interest in the same topic, the author would like to suggest a more diverse method by implementing interviews, surveys, and Latent Dirichlet Allocation (LDA) analysis for the corpus in order to provide even richer research results.

DISCLAIMER

This study stands neutral in academic fields in the sense that it does not examine the moral, ethical, and legal aspects of the digital goods piracy act, nor does it promote the digital piracy act itself. Therefore, any information related to the piracy user and the piracy tools is redacted.

REFERENCES

- Arazzi, M., Nicolazzo, S., Nocera, A., & Zippo, M. (2023). The importance of the language for the evolution of online communities: An analysis based on Twitter and Reddit. *Expert Syst. Appl.*, 222, 119847.
- Astuti, R., & Devi S. H. M. (2021). Perlindungan Hukum Pemilik Hak Cipta Pembajakan Karya Sinematografi Dalam Grup Chat Pada Aplikasi Telegram. *Jurnal Kertha Semaya* 9(7): 1087-1098.



- Athey, S., & Scott S. (2015). *The Nature and Incidence of Software Piracy: Evidence from Windows*. University of Chicago Press.
- Balani, S., & Choudhury, M.D. (2015). Detecting and Characterizing Mental Health Related Self-Disclosure in Social Media. *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*.
- Bradley, W. A., & Kolev, J. (2023). How Does Digital Piracy Affect Innovation? Evidence from Software Firms. *Research Policy*, 52(3): 1-18.
- Dey, D., Antino K., & Atanu L. (2019). Online piracy and the “longer arm” of enforcement. *Management Science*, 65(3), 1173–1190.
- Ferrer, X., van Nuenen, T., Such, J. M., & Criado, N. (2021). Discovering and Categorizing Language Biases in Reddit. *International AAAI conference on web and social media* (pp. 140–151).
- Herz, B., & Kamil K. (2018). Movie Piracy and Displaced Sales in Europe: Evidence from Six Countries. *Information Economics and Policy* 43, 12–22.
- Hui, Kai Lung, and Ivan P. L. Png. (2003). Piracy and the Legitimate Demand for Recorded Music.” *Contributions in Economic Analysis & Policy* 2(1).
- Hymes, D. (1963). Objectives and Concepts of Linguistic Anthropology. In D. G. Mandelbaum, G. W. Lasker and E. M. Albert (eds.), *The Teaching of Anthropology* (pp. 275–302).
- Kautsarina, K. (2018). Perkembangan Riset Etnografi Di Era Siber : Tinjauan Metode Etnografi Pada Dark Web. *Masyarakat Telematika Dan Informasi*, 8(2), 145. <https://doi.org/10.17933/mti.v8i2.109>
- Kim, A., Atanu L., & Debabrata D. (2018). The ‘Invisible Hand’ of Piracy:: An Economic Analysis of the Information-Goods Supply Chain. *Management Information Systems Quarterly* 42(4), 1117–41.
- Lewis, J. (2015). The piratical ethos in streams of language. *Popular Communication*, 13(1), 45–61. <https://doi.org/10.1080/15405702.2014.977997>
- Mares, M. (2018). Pirate English in The Age of Sail: A Linguistic Analysis of Pirate English and its Importance Within the Multicultural Space of 17th and 18th Centuries Caribbean. *Ucl*. https://www.academia.edu/32601647/Pirate_English_in_The_Age_of_Sail_A_Linguistic_Analysis_of_Pirate_English_and_its_Importance_Within_the_Multicultural_Space_of_17th_and_18th_Centuries_Caribbean



- Mohan, S., Apala G., Michael E. H., Fred P., Ashley S., & Chris P. (2017). The Impact of Toxic Language on the Health of Reddit Communities. *Lecture Notes in Computer Science*, 51–56.
- Peng, S., Bo L., & Shuang W. (2023). *Presence of Piracy and Legal Protection: decisions in the Digital Goods Market Under Different Contracts*. *European Journal of Operational Research*: 1-19.
- Proferes, N., Naiyan J., Sarah C. G., Casey F., & Michael Z. (2021). Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, and Ethics. *Social Media + Society* 7(2).
- Sá, M. J., Santos, A. I., Serpa, S., & Ferreira, C. M. (2021). Digital Literacy in Digital Society 5.0: Some Challenges. *Academic Journal of Interdisciplinary Studies*, 10(2), 1. <https://doi.org/10.36941/ajis-2021-0033>
- Schrading, N., Alm, C. O., Ptucha, R., & Homan, C. (2015). An Analysis of Domestic Abuse Discourse on Reddit. *Proceedings of the 2015 conference on empirical methods in natural language processing*: (pp. 2577–2583.)
- Segrest, M. (2021). *How to Speak Like a Sailor: 1500 Nautical Terms Every Mariner Should Know*. Alabama: Navigate Content.
- Stanlaw, J., Nobuko A., & Zdenek S. (2017). *Language, Culture, and Society: An Introduction to Linguistic Anthropology*. Westview Press.
- Varian, Hal R. (2000). Buying, Sharing and Renting Information Goods. *Journal of Industrial Economics* 48(4), 473–88.
- Wardhaugh, R., & Janet, M., F. (2021). *An Introduction to Sociolinguistics*. John Wiley & Sons.

