

Opening sentences of research papers

How academics defeat the blinking cursor

ABSTRACT

abstract

CCS CONCEPTS

• **xx**;

KEYWORDS

xx

1 INTRODUCTION

How do researchers start their papers? We wanted to understand how other researchers approach the opening sentence in their work. For this purpose, we studied three years of papers (379 papers in total) from USENIX Security, one of the top venues for security research. More pertinently, it is a venue with open access, easy-to-download full proceedings in a variety of friendly formats that made our job easier. We read every opening sentence from the main body of the paper (as opposed to the abstract). For this work, we drew inspiration from Grounded Theory [5]—we applied coding for each sentence. After numerous iterations several patterns started to emerge and as a result, we developed and refined a detailed categorization for opening sentences.

2 RELATED WORK

Different aspects of academic writing has been explored in the literature. King analyses opening sentences in medical research in his article [7]. He states that the opening sentence should draw readers' attention, and it should also be concise and clear about stating the main theme of the paper. He also gives examples from medical writing and explains different ways to simplify overcomplicated sentences by shortening them. Cameron *et al.* explains the struggles of the writing process and suggests strategies to help novice writers to overcome them [3]. Hartley presents a bilingual study in English and Spanish on research papers in psychology [6]. The paper focuses on improving different aspects of academic writing to increase readability. Biber *et al.* discuss the stereotypical characteristics of academic writing like complex grammar structures [2].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Conference'17, July 2017, Washington, DC, USA

© 2021 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

3 METHODOLOGY

3.1 Grounded theory

Grounded theory is a qualitative data analysis method [5]. The collected qualitative data (*e.g.*, interview transcripts) is sorted through qualitative coding. By performing coding, the aim is to come up with new high-level theories and concepts at the end of the process. Coding is an iterative process, as several rounds of coding can be performed to refine the categories. At the end of the process, a new theory that is based on the data is presented.

In this paper, we applied coding for each opening sentence. We created codes and labeled each sentence by these codes. We observed the relations among the codes to categorize them and we ended up with the categorization shown in Figure 1. We used coding to determine common patterns among opening sentences.

Grounded Theory is also used as a methodology in security and privacy research. It is used to study user mental models of cryptocurrency systems [8], how blockchain technology is perceived and how it is used [9], preferences for security warning types [4], the factors that influence software developers' motivation towards security [1], how users manage their online security posture [10], and how users manage their passwords [11].

4 DESCRIPTION OF EACH CATEGORY

Figure 1 shows the hierarchy chart of the categories where the area of the box they are in is proportional to how many times they were used. We provide a description of each category in this section.

4.1 Facts

4.1.1 Facts: Definition or Description. Many papers start with a straightforward definition of the subject of the paper. These tend to be neutral and like something you'd read in a glossary.

- “Malware sandboxes are automated dynamic analysis tools that execute samples in an isolated and instrumented environment.”
- “Secure two-party computation allows two parties to process their sensitive data in such a way that its privacy is protected.”
- “HMAC is a cryptographic authentication algorithm, the ‘Keyed-Hash Message Authentication Code,’ widely used in conjunction with the SHA-256 cryptographic hashing primitive.”

Similarly, papers might provide a description of what the subject of a sentence does or how it works. These are also neutral statements and like something you'd read in a textbook.

- “Traditionally, digital investigations have aimed to recover evidence of a cyber-crime or perform incident response via analysis of non-volatile storage.”

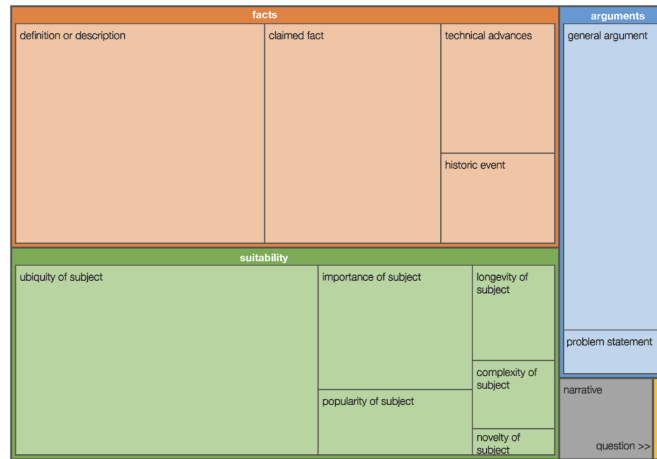


Figure 1: Overview of the categorization of opening sentences

- “Mobile social applications discover nearby users and provide services based on user activity (what the user is doing) and context (who and what is nearby).”
- “To reduce the memory footprint of a system, the system software shares identical memory pages between processes running on the system.”

4.1.2 Facts: Claimed Fact. Another neutral approach to an opening sentence is to provide a fact that is relevant to the subject of the paper. Later we will discuss arguments which are often expressed as if they are facts but are only debatably true. A claimed fact’s correctness should either be apparent or at least provable (i.e., falsifiable).

- “Users are often advised or required to choose passwords that comply with certain policies.”
- “Mobile apps frequently demand access to private information.”
- “For several decades, car keys have been used to physically secure vehicles.”
- “Some sentences use stronger and more vivid language but are still factually based.”
- “In spite of extensive industrial and academic efforts (e.g., [3, 41, 42]), distributed denial-of-service (DDoS) attacks continue to plague the Internet.”

4.1.3 Facts: Technical Advances. Many opening sentences lay out a technical advance in the subject of the sentence. This creates a window of opportunity for the researcher to later identify a novel research problem caused by the changing technology. It is common to see words like: evolve, become, transition, and improve.

- “Recent advances in cloud computing enable customers to outsource their computing tasks to the cloud service providers (CSPs).”
- “Browsers have evolved over recent years to mediate a wealth of user interactions with sensitive data.”
- “Since its beginning in the early nineties, the Web evolved from a mechanism to publish and link static documents into a sophisticated platform for distributed Web applications.”

4.1.4 Facts: Historic Events. A final neutral opening sentence will refer to some historic event.

- “In 1996, Wagner and Schneier performed an analysis of the SSL 3.0 protocol [67].”
- “In February 2011, a new Tor hidden service [16], called “Silk Road,” opened its doors.”
- “The Network Time Protocol (NTP) is one of the Internet’s oldest protocols, dating back to RFC 958 [15] published in 1985.”

In some cases, a paper opens with a “compound” sentence that makes reference to a historic event in one clause of the sentence, while having additional clauses of a different category. For example, the following sentence refers to a historic event as well as a technical advance.

- “Starting from Denning’s seminal work in 1986 [9], intrusion detection has evolved into a number of different approaches.”

4.2 Arguments

4.2.1 Arguments: General Argument. Many opening sentences issue a subjective argument that represents the authors’ opinion. Unlike a fact, it isn’t straightforward that the reader will accept it as true. While arguments are less neutral than facts, they can be more interesting and provocative, which can help draw the reader into the paper.

The arguments we categorize under “general arguments” do not fit elsewhere in our categorization system. As we go through more categories, we will see other more specific kinds of arguments.

- “It is a truth universally acknowledged, that password-based authentication on the web is insecure.”
- “The dismissal of human memory by the security community reached the point of parody long ago.”
- “In recent years, unwanted software has risen to the forefront of threats facing users.”
- “The phenomenal growth of Android devices brings in a vibrant application ecosystem.”

4.2.2 Arguments: Problem Statement. A special type of argument is a “problem statement” which uses the opening sentence to establish a problem or challenge to be solved.

- “A key challenge when running untrusted virtual machines is providing them with efficient and secure I/O.”
- “Determining the semantic similarity between two pieces of binary code is a central problem in a number of security settings.”
- “It is difficult to keep secrets during program execution.”

For some sentences, the problem is not stated explicitly but can be inferred from what is said. For example, the “pressure to respond” in the following sentence implies a problem.

- “As popular applications rely on personal, privacy-sensitive information about users, factors such as legal regulations, industry self-regulation, and a growing body of privacy-conscious users all pressure developers to respond to demands for privacy.”

4.3 Suitability

4.3.1 Suitability: Importance of subject. A large set of sentences make a special kind of argument: that the subject of the opening sentence is suitable or worthy of research. The exact reasons they are suitable fall into a few sub-categories: the subject is important, ubiquitous, complex, novel, popular with other researchers, or has been around a long time.

Many opening sentences state that their subject is important, with the implication that it is thus suitable for research.

- “Security has now become an important and real concern to connected and/or automated vehicles.”
- “Error handling is an important aspect of software development.”
- “SSL/TLS is, due to its enormous importance, a major target for attacks.”

Some sentences do not explicitly use the word “important” but find other ways to convey the same notion. For example, a concern or component might be described as essential or crucial or serious.

- “The threat of data theft in public and private clouds from insiders (e.g. curious administrators) is a serious concern.”
- “The same-origin policy (SOP) is a cornerstone of web security, guarding the web content of one domain from the access from another domain.”

4.3.2 Suitability: Ubiquity of subject. The most popular kind of opening sentence argues that a subject is suitable for research because it is ubiquitous and widely used.

- “Billions of users now depend on online services for sensitive communication.”
- “Embedded systems are omnipresent in our everyday life.”
- “Android is the major platform for mobile users and mobile app developers.”

4.3.3 Suitability: Popularity of subject. While the ubiquity of a subject corresponds to how widely it is used, a closely related variant points out that the subject has received a lot of attention. Often, this means attention from other researchers which lends credibility to the subject for further research.

- “Protecting the privacy of user data within mobile applications (apps for short) has always been at the spotlight of mobile security research.”
- “The black-market economy for purchasing Facebook likes, Twitter followers, and Yelp and Amazon reviews has been widely acknowledged in both industry and academia [6, 27, 37, 58, 59].”
- “Since the first widely-exploited buffer overflow used by the 1998 Morris worm [27], the prevention, exploitation, and mitigation of memory corruption vulnerabilities have occupied the time of security researchers and cybercriminals alike.”

4.3.4 Suitability: Longevity of subject. In this category, how long a subject has been around is the key component to why it is a suitable subject for study. In some cases, a specific duration is provided and in others, it is implied that the amount of time is significant.

- “Redaction of sensitive information from documents has been used since ancient times as a means of concealing and removing secrets from texts intended for public release.”
- “Since its beginning in the early nineties, the Web evolved from a mechanism to publish and link static documents into a sophisticated platform for distributed Web applications.”

4.3.5 Suitability: Complexity of subject. In this category, the complexity of the subject is highlighted, implying that the complexity creates new issues or requires further research. The complexity might be inherent to the subject itself. Or there might be a complex set of external factors to consider.

- “Today, large and complex software is built with many components integrated using APIs.”
- “The capabilities and limitations of disassembly are not always clearly defined or understood, making it difficult for researchers and reviewers to judge the practical feasibility of techniques based on it.”
- “As popular applications rely on personal, privacy-sensitive information about users, factors such as legal regulations, industry self-regulation, and a growing body of privacy-conscious users all pressure developers to respond to demands for privacy.”

4.3.6 Suitability: Novelty of subject. Finally, a degree of novelty is an important component in any research question so it is unsurprising that papers begin arguing novelty from their opening sentence. In this category, sentences focus on something that is new or emerging.

- “In the last few years, a new class of cyber attacks has emerged that is more targeted at individuals and organizations.”
- “Although the operating system (OS) kernel has always been an appealing target, until recently attackers focused mostly on the exploitation of vulnerabilities in server and client applications— which often run with administrative privileges—as they are (for the most part) less complex to analyze and easier to compromise.”
- “This sentence manages to appeal to both longevity and novelty by relating two subjects.”

- “While cryptocurrency has been studied since the 1980s [22, 25, 28], bitcoin is the first to see widespread adoption.”

4.4 Narrative

A potentially interesting way to draw a reader into a paper is by establishing a narrative: a scenario that gets the reader thinking about themselves or other people and what they might do.

- “Consider that you are a domain owner, holding a few domain names that you do not have a better use of.”
- “Consider the setting where a client owns a public input x , a server owns a private input w , and the client wishes to learn $z := F(x, w)$ for a program F known to both parties.”
- “Narratives might also set a scene, like the academic version of an “establishing shot” from films and TV.”
- “Our phones are always within reach and their location is mostly the same as our location.”
- “We live in a “big data” world.”
- “The battle for the living room is in full swing.”

4.5 Question

Making the reader curious is another good way to begin a paper, and this can be accomplished using a question. In our sample, this was underused with only one example.

- “Do programmers leave fingerprints in their source code?”

5 DISCUSSION

- “The widespread adoption of DEP, which ensures that all writable pages in memory are nonexecutable, has largely killed classic code injection attacks.” Do you have to be an expert in security&privacy to understand these first sentences? When you look at this example, even though you are not an expert (and you don’t know what DEP or code injection attack is) you can still understand the gist of it. One would know that attack has a bad connotation and DEP killing classic code injection attacks is a good thing. But what if the sentence was like this: “The widespread adoption of DEP, which ensures that all writable pages in memory are nonexecutable, has largely killed classic code injection.” Then, here we need the expertise. DEP killed classic code injection: Is it a desirable property or not?
- Color commentary: To add some colour, some papers use rhetoric to make a point. What makes these sentences more vivid is the rich choice of words, though they were not “purple prose”.
- Short sentences are striking and easy to remember. Long sentences can obviously convey more information or make a complex argument, but they might have to be read a few times to fully parse what they are saying. (King has a similar argument in [7]).
- arc of time: These sentences tend to appeal to the notion that a domain is important because of its longevity.
- Secondary code - biggest threat: In security, you can imagine that researchers want to tackle the biggest threats. Many use their opening sentence to position themselves as such.

6 CONCLUSION AND FUTURE WORK

As researchers, we often find ourselves staring at the blinking cursor, trying to come up with a witty and neat way to start our paper to engage the audience. The opening sentence has a responsibility because it is a way to be remembered, just like in the novels. Chances are high that you recognize many opening sentences of famous literary works even if you haven’t read them. Knowing how many times we have struggled with an opening sentence, we were curious how other researchers begin their papers. We know every researcher has their own style of telling the story of their research process.

We also discuss whether a similar classification can be used in other domains or the categories that we present in this paper are domain-specific. And if we can use the same categories, how the weights of these categories differ in other areas.

Future work maybe:

In other domains: categories can completely change or same categories can be used but their weights change.

Maybe in history, political science or economics a similar classification can be used?

But we think that we may need different categories if we do the same study in literature.

We also observe that the opening sentences that we coded are not too elaborate or ornate (this type of writing is defined as ‘purple prose’).

REFERENCES

- [1] Hala Assal and Sonia Chiasson. 2018. Motivations and amotivations for software security. In *Soups Workshop on Security Information Workers (WSIW)*. USENIX Association.
- [2] Douglas Biber and Bethany Gray. 2010. Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes* 9, 1 (2010), 2–20.
- [3] Jenny Cameron, Karen Nairn, and Jane Higgins. 2009. Demystifying academic writing: Reflections on emotions, know-how and academic identity. *Journal of Geography in Higher Education* 33, 2 (2009), 269–284.
- [4] Anastasia Danilova, Alena Naiakshina, and Matthew Smith. 2020. One size does not fit all: a grounded theory and online survey study of developer preferences for security warning types. In *2020 IEEE/ACM 42nd International Conference on Software Engineering (ICSE)*. IEEE, 136–148.
- [5] Barney G Glaser and Anselm L Strauss. 1967. The discovery of grounded theory; strategies for qualitative research. (1967).
- [6] James Hartley. 2012. New ways of making academic articles easier to read. *International Journal of Clinical and Health Psychology* 12, 1 (2012), 143–160.
- [7] Lester S King. 1967. The Opening Sentence. *JAMA* 202, 6 (1967), 535–536.
- [8] Alexandra Mai, Katharina Pfeffer, Matthias Gusenbauer, Edgar Weippl, and Katharina Krombholz. 2020. User Mental Models of Cryptocurrency Systems—A Grounded Theory Approach. (2020).
- [9] Scott Ruoti, Ben Kaiser, Arkady Yerukhimovich, Jeremy Clark, and Robert Cunningham. 2019. Blockchain technology: what is it good for? *Commun. ACM* 63, 1 (2019), 46–53.
- [10] Scott Ruoti, Tyler Monson, Justin Wu, Daniel Zappala, and Kent Seamons. 2017. Weighing context and trade-offs: How suburban adults selected their online security posture. In *Thirteenth Symposium on Usable Privacy and Security ({SOUPS} 2017)*. 211–228.
- [11] Elizabeth Stobert and Robert Biddle. 2014. The password life cycle: user behaviour in managing passwords. In *10th Symposium On Usable Privacy and Security ({SOUPS} 2014)*. 243–255.