Convergence of Truth Through Language Links in Historical Data – A Case Study on Wikipedia: A Semantic Analysis of Source Texts

Jonas Alexander Stöger

Work project carried out under the supervision of:

Advisor: Michael Kummer

Abstract

Our research examines whether numerical data on historical battles, such as casualty figures or troop strengths, converge over time across multiple languages on Wikipedia. Our analysis of annual revisions, reveals no trend of convergence, in most cases the numbers rarely change, with discrepancies persisting or even increasing over time.

The study also tests these patterns in underrepresented regions and simpler data categories outside of the battle context, like bridge length, while comparing Wikipedia figures and sentiment present in the articles with established reference works. Findings include persistent barriers, like editorial biases and resource disparities, that limit numerical consistency and reflect cultural and editing dynamics in global knowledge.

Collectively this thesis finds barriers that limit data alignment across languages and ultimately reveals that Wikipedia's goal of unifying knowledge globally faces unaddressed challenges.

Contents

Table of Figures

Table of Tables

1. Introduction

In today's globalized environment, Wikipedia stands as both a crucial knowledge resource and a reflection of global information practices. Available in over 300 language editions, it constitutes an unprecedented collective knowledge venture shaped by a multitude of cultural, linguistic, and historical backgrounds. Although this openness offers significant opportunities for democratizing information access, it also introduces complexities. Divergent cultural narratives, linguistic differences, and varying historical interpretations can create inconsistencies, particularly when comparing content across languages. This thesis contributes to broader discussions on the reliability of global knowledge-sharing platforms and the impact of cultural factors on collective memory.

The research within this work project focuses on numerical data related to historical battles, such as troop strengths, casualty figures, and outcomes, as an illustrative case for examining how contested historical information evolves on Wikipedia. By analyzing data from six major language editions – English, French, German, Portuguese, Spanish, and Italian – this study seeks to identify and understand patterns of convergence and divergence in reported figures over time.

To provide additional context, the thesis also investigates numerical data in less contentious domains. This includes metrics related to infrastructure (e.g., bridge lengths), scientific facts (e.g., melting points of elements), and geographic data (e.g., mountain heights), as outlined in the work of Hecht and Gergle (2010). Furthermore, it extends the analysis to underrepresented regions such as South America, Asia, and Africa to examine how unique cultural and historical contexts influence reported battle figures. In addition, the study explores the foundational references behind Wikipedia articles, including sources like Britannica and Encyclopædia Universalis, and assesses their early impact on shaping linguistic editions (Callahan & Herring, 2011). Through this lens, the thesis conducts both semantic and numeric analyses to elucidate how underlying sources and cultural biases are reflected in Wikipedia narratives (Pfeil, Zaphiris, & Ang, 2006).

Methodologically, the project employs a multi-layered approach to data collection and analysis. A custom-built pipeline integrates Wikipedia's API to retrieve historical article revisions, an HTML parser to extract structured data from infoboxes, and the Gemini AI model to preprocess unstructured textual content. This framework supports the systematic gathering of metrics (troop strengths, casualty counts, and battle outcomes) over extended temporal ranges for approximately 40-50 battles across six language editions.

Data preprocessing procedures addressed challenges such as inconsistent formatting, language-specific conventions, and missing values. A normalization phase ensured that data could be meaningfully compared across languages, while careful handling of missing and irregular values helped maintain dataset integrity. Statistical tools, including linear regression models, were used to detect trends in the alignment of numerical data over time. Convergence coefficients were calculated to quantify the extent to which different language editions approached similar values, and cross-language comparisons were employed to evaluate how editing frequency and interlanguage links influence data harmonization.

The implications of this research extend beyond Wikipedia. By illuminating the mechanisms that facilitate or obstruct convergence, this study provides insights applicable to other collaborative, multilingual platforms. It underscores the importance of transparent editorial guidelines and supportive technological infrastructures for achieving reliable and inclusive information systems. Although Wikipedia's scale and open-editing model are distinctive, the lessons learned from its challenges and achievements in harmonizing diverse contributions can inform strategies for improving content management across global digital knowledge ecosystems.

2. Background and Literature Review

This section explores how Wikipedia's multilingual environment shapes the construction and convergence of historical knowledge. We decided to focus on several intertwined subjects—Wikipedia's editorial ecosystem, multilingual content variations, philosophical theories of truth, and social theories of information convergence—because each offers a complementary lens on how knowledge is produced, validated, and potentially harmonized across diverse cultural contexts. Wikipedia's guiding principles (neutrality, verifiability, prohibition of original research) and collaborative model frame the platform's editorial dynamics. Multilingual disparities then illustrate how cultural and linguistic factors may shape historical narratives differently in each language edition. Philosophical theories of truth (coherence, consensus, and pragmatism), along with social frameworks (information diffusion and constructivism), provide deeper insight into how collective knowledge might converge despite lingering biases or conflicting sources. By tying these perspectives together and applying them to historically sensitive topics such as casualty figures, we gain a comprehensive understanding of both the challenges and mechanisms underpinning factual alignment in an open, global encyclopedia.

Building on these foundations, the review highlights tensions arising from cultural biases, political controversies, and editorial hierarchies. By reflecting on existing cross-language content studies, it becomes evident that Wikipedia is not a monolith but rather a rich tapestry of linked knowledge bases, with marked variations around contested historical events. Historical battles in particular serve as a revealing case study, spotlighting how collective negotiation, network effects, and linguistic barriers can impede or facilitate convergence. Finally, the role of language links and multilingual editors in bridging these gaps underscores the platform's evolving yet imperfect system for synchronizing content across different editions.

2.1. The Wikipedia Ecosystem

2.1.1 Introduction

Since its launch in 2001, Wikipedia has fundamentally altered the way information is compiled, shared, and accessed. As a free, collaboratively edited online encyclopedia, it empowers individuals worldwide to both consult and contribute knowledge. Today, Wikipedia hosts more than 300 language editions, each catering to diverse communities of readers and editors. These language-specific versions serve as primary reference points for a broad audience—ranging from students and educators to researchers, professionals, and the general public. The open-editing model, which enables virtually anyone to add, modify, or refine content, has democratized the production of knowledge. However, it has also introduced significant challenges related to maintaining accuracy, credibility, and internal consistency. This is particularly evident when dealing with sensitive or disputed content, such as casualty figures in historical battles, where cultural perspectives and national narratives may diverge (Jemielniak, 2014).

2.1.2 Wikipedia's Mission and Editorial Philosophy

At the core of Wikipedia's editorial framework are three guiding principles: neutrality, verifiability, and the prohibition of original research. These principles establish a foundational standard for contributors, ensuring that articles are grounded in credible external sources rather than personal opinions or unverified claims. Reagle (2010) characterizes the "neutral point of view" (NPOV) as more than a guideline; it is a core philosophical stance that supports the platform's credibility and public trust.

Neutrality plays a particularly critical role in articles addressing contentious subjects, including historical events and their associated casualty numbers. Differing accounts of such events may exist due to varying cultural interpretations and national narratives. Striving for NPOV encourages editors to present multiple perspectives fairly, allowing readers to encounter a balanced synthesis of competing viewpoints (Reagle, 2010).

The principle of verifiability further reinforces these standards. It mandates that all statements in Wikipedia articles be supported by reliable, authoritative sources. This focus on appropriate referencing is vital when confronting inconsistent casualty figures, as it helps editors scrutinize and reconcile divergent data across language editions and over time (Ford et al., 2013).

Finally, the prohibition of original research ensures that Wikipedia remains a tertiary source. Editors are expected to summarize and integrate existing knowledge rather than introduce new, untested assertions. This policy helps prevent the dissemination of misinformation and preserves Wikipedia's role as a credible repository of established facts (Konieczny, 2010).

2.1.3 The Collaborative Dynamics of Wikipedia

2.1.3.1 Crowdsourcing and Self-Correction Mechanisms

Wikipedia's success rests on its collaborative model, in which a global community of contributors continuously refines content. This "wisdom of the crowds" approach leverages collective intelligence to improve articles incrementally (Kittur & Kraut, 2008). In areas such as historical scholarship, where interpretations and data may evolve as new evidence emerges, this model allows Wikipedia to integrate updated information promptly.

A key strength of this system is its self-correcting capacity. Editors routinely monitor recent changes and evaluate the legitimacy of new edits, reverting or improving them as necessary. Over time, this iterative process helps maintain and enhance the accuracy of articles, including those reporting casualty figures, by encouraging editors to critically assess sources and align content with established editorial standards (Halfaker et al., 2009).

However, collaboration also has its tensions. Editors occasionally engage in “edit wars” when they repeatedly override each other’s revisions. Disputes over casualty numbers in historical battles, for example, may arise due to conflicting interpretations or national biases. Yet, these conflicts can ultimately lead to a more comprehensive and balanced presentation, as involved parties strive toward consensus through dialogue and negotiation (Keegan, Gergle & Contractor, 2011).

2.1.3.2 Influence of Editorial Hierarchies and Experienced Contributors

Although Wikipedia is open to all, an informal editorial hierarchy has emerged, guided by the experience, expertise, and demonstrated trustworthiness of certain contributors. Veteran editors and administrators play a pivotal role in enforcing policies, resolving content disputes, and maintaining editorial quality. Their involvement is especially valuable in complex or contested areas, such as historically sensitive articles with differing casualty accounts (Geiger & Ribes, 2010).

These experienced contributors often undertake quality assurance tasks, including verifying sources, improving article organization, and ensuring that content complies with established guidelines (Stvilia et al., 2008). Their oversight not only preserves the integrity of articles but also facilitates the convergence of accurate data across multiple language editions.

2.1.4 The Multilingual Foundation of Wikipedia

2.1.4.1 Cross-Language Content Variation and Consistency

Wikipedia’s multilingual nature allows it to represent diverse cultural perspectives. Each language edition operates with considerable autonomy, resulting in variations in content depth, scope, and emphasis (Hecht & Gergle, 2010). Consequently, articles on the same subject—such as a historical event—may differ significantly, including their casualty figures.

To mitigate these discrepancies, interlanguage links connect related articles across different language editions. These links enable editors and readers to compare content directly, identify inconsistencies, and transfer updates. When one language edition refines its casualty data based on new, reliable research, editors can propagate these improvements to other editions, promoting a more unified and accurate representation (Miquel-Ribé & Laniado, 2018).

2.1.4.2 Role of Multilingual Editors and Translation Tools

Multilingual editors are critical in bridging gaps between language editions. Their linguistic and cultural fluency allows them to translate content, verify sources in multiple languages, and ensure that important corrections and updates are integrated widely (Hecht et al., 2012). This capability is particularly important for achieving convergence on casualty figures, where data must be carefully cross-checked against international scholarship.

Additionally, translation tools and automated scripts (bots) support these efforts by performing routine updates and standardizations. While automated tools cannot fully replicate human judgment—especially in interpreting complex historical data—they assist in maintaining a baseline of consistency and accuracy across language editions (Massa & Scrinzi, 2012).

2.1.5 Challenges in Achieving Content Convergence

2.1.5.1 Cultural Biases and Systemic Disparities

Even with good intentions and useful tools, achieving perfect alignment of facts and figures across different language editions is far from guaranteed. Each language community comes with its own cultural lenses, which can subtly shape what's considered important, how data is framed, and which sources are trusted. Articles about historical battles, for instance, might reflect national pride, downplay certain events, or highlight specific interpretations. Variations in reported casualty numbers might simply stem from which historian or archive the local editors prefer (Pfeil et al., 2006).

Another factor is that not all language editions have the same human and material resources. Some editions are small, with too few editors to thoroughly monitor and update every article. That can leave outdated or less accurate info lingering there. This resource gap can slow down convergence, keeping certain discrepancies alive (Lewoniewski et al., 2017).

2.1.5.2 Misinformation and Verifiability Challenges

On top of cultural angles, Wikipedia's open-door policy makes it vulnerable to people who might add wrongful information, either as a prank or for more malicious reasons. Ensuring that all statements are tied to reliable sources is a constant struggle, especially when it comes to messy historical data that's debated among scholars. Editors have to carefully weigh sources, discuss them on talk pages, and sometimes revert or remove questionable additions (Luyt & Tan, 2010).

The only real defense is a strong editorial culture where everyone values proper sourcing, fact-checking, and thoughtful negotiation. Ford et al. (2013) suggest that reinforcing these reliable-sourcing habits is key to combating misinformation and nudging all the language versions toward consistent data over the long term.

2.1.6 The Dynamics of Convergence in Casualty Figures

2.1.6.1 Mechanisms Facilitating Convergence

In spite of the hurdles, you do see convergence happening—especially over longer timescales. Several mechanisms help:

- **Collaborative Editing & Consensus-Building:** Editors hash out differences on talk pages, weigh conflicting sources, and hopefully reach a stable consensus on figures (Viegas et al., 2004).

- **Cross-Language Information Transfer:** Thanks to multilingual editors and interlanguage links, improvements to one edition can cascade into others (Hecht et al., 2012).
- **Administrative Oversight:** Experienced editors and admins often guide discussions back to policy-based reasoning, insisting on verifiability and neutrality when numbers conflict (Geiger & Ribes, 2010).
- **Technological Tools:** Bots and scripts can handle smaller-scale data alignment tasks, like updating references or ensuring the same figure format, reducing noise and inconsistency (Massa & Scrinzi, 2012).

2.1.6.2 Factors Impeding Convergence

However, a few things make it hard to get everyone on the same page:

- **Conflicting Sources:** Different historians or archives present varying casualty numbers, making a clean consensus difficult (Luyt & Tan, 2010).
- **Cultural/National Biases:** Editors might lean toward figures that support their cultural narrative, resisting what others propose (Callahan & Herring, 2011).
- **Resource Limitations:** Smaller language editions might move slowly if they lack editors who can update or verify the numbers (Lewoniewski et al., 2017).

2.1.7 Implications for Knowledge Convergence Studies

Overall, Wikipedia provides a valuable setting in which to examine how dispersed information gradually—and at times unevenly—moves toward a more unified understanding. Historical battles and their reported casualty figures exemplify this process: initially divergent data points can, through continuous discourse, correction, and source verification, become more consistent over time. The platform's integrated talk pages, editorial guidelines, and multilingual connections all contribute to this intricate yet highly instructive dynamic of knowledge convergence.

Analyzing how facts evolve and align within Wikipedia articles offers researchers critical insights into collective intelligence and the global construction of knowledge. By understanding which factors support or impede the convergence process, one can inform strategies that improve both the reliability and coherence of Wikipedia and other collaborative platforms operating across linguistic and cultural boundaries. Studying the ways in which editors coordinate, negotiate, and ultimately agree upon certain figures provides a window into broader patterns of knowledge harmonization in today's increasingly interconnected world.

2.2. Studies on Content Comparison Across Languages

2.2.1. Introduction, significance of Multilingual Analysis

In an age where information flows rapidly online, Wikipedia has become a widely relied-upon, free source of knowledge. Since its creation in 2001, it has expanded into a vast platform containing more than 55 million articles in over 300 languages, each version shaped by volunteers who continually add, edit, and refine content.

This global scale and multilingual reach, however, also lead to noticeable differences across language editions. Each edition reflects the unique linguistic, cultural, and historical milieu of its contributors, influencing what topics receive attention, how content is structured, and which sources are deemed credible. In other words, these are not mere translations but distinct bodies of knowledge that emerge from local contexts and priorities.

Studying such variations is not simply a theoretical endeavor. Because people increasingly rely on digital resources for accurate information, it is important to ensure some level of consistency and fairness across linguistic boundaries. By exploring where language editions diverge, it becomes possible to spot biases, identify coverage gaps, and understand how shared knowledge can fragment into disparate narratives.

Insights gleaned from this analysis can then support practical interventions. Educators, librarians, and policymakers may use these findings to develop strategies that minimize inequalities. As more learners turn to Wikipedia as a starting point, striving for comparable quality and depth across languages is a meaningful and urgent goal.

2.2.2. Overview of Existing Research, Key Findings on Content Differences

A number of studies highlight the uneven nature of Wikipedia's multilingual landscape. Hecht and Gergle (2010) discovered that fewer than half the concepts covered by any two language editions overlap, pointing to considerable informational gaps. As a result, readers turning to one language edition may receive a significantly different knowledge set than those consulting another.

Hara, Shachaf, and Hew (2010) found that coverage of global events varies widely among language editions, often influenced by cultural familiarity or local relevance. Articles on natural disasters, for instance, tend to be more detailed in languages spoken in regions directly affected. Similarly, Callahan and Herring (2011) examined how cultural biases shape biographies of notable individuals, with different editions highlighting distinct attributes or achievements.

Massa and Scrinzi (2013) introduced "Manypedia," illustrating how the same topic can appear through different cultural lenses. Warncke-Wang et al. (2015) further showed that smaller

Wikipedias often face challenges meeting quality standards seen in larger editions, resulting in differences in both reliability and depth.

Taken together, this body of work suggests that Wikipedia's multilingual model produces not a single universal encyclopedia, but multiple culturally informed versions. Readers might thus remain within linguistic boundaries that limit their exposure to alternative viewpoints, potentially reinforcing pre-existing biases.

2.2.3. Methodologies in Prior Studies, Common Approaches

Researchers have adopted various methodologies to understand these disparities. Large-scale quantitative analyses often look at factors like article length, revision history, or link patterns. For example, Hecht and Gergle (2010) employed interlanguage link analysis to determine how much content different language editions actually share.

In contrast, qualitative methods probe the tone, framing, and narrative structure of articles. Pfeil, Zaphiris, and Ang (2006) analyzed cultural influences on collaborative writing, showing how differing communication styles affect what appears on the page. Some investigations (e.g., Bao et al. 2012) blend these approaches, providing a fuller picture that captures both structural patterns and subtle editorial choices.

However, these methods face challenges. Translation difficulties may obscure nuances or idiomatic expressions that carry cultural weight. Automated translation tools, while improving, still can miss context. Ethical considerations also arise when dealing with user-generated content that may be sensitive or politically charged. Researchers must proceed with care, ensuring that their findings acknowledge cultural complexity and avoid unintended misrepresentations.

2.2.4. Relevance to Historical Battles, the Impact on Historical Accuracy

Historical battles provide a particularly instructive example of how cultural perspective influences Wikipedia's multilingual content. Because these events often resonate deeply in national identities and collective memories, the way they are depicted can shape public understanding and reinforce certain narratives.

For instance, accounts of the Nanjing Massacre differ significantly between Chinese and Japanese Wikipedia editions, mirroring long-standing historical disputes and sensitivities. Likewise, the Falklands War (Guerra de las Malvinas) appears differently in English and Spanish editions. While English articles tend to emphasize British views, Spanish editions—especially those reflecting Argentine perspectives—focus on sovereignty claims and the conflict's human toll.

These divergences matter. A reader relying solely on one language edition may be unaware that a completely different framing exists elsewhere. Without comparing across languages, individuals risk absorbing a one-sided story. Historians, educators, and the wider public need to be aware of these variations if they hope to build a more balanced understanding of the past.

2.2.5. Gaps in the Literature, the Need for Focused Research

Although earlier studies have illuminated general disparities and variations in recent events, relatively few have zeroed in on historical battles. These episodes, charged with symbolic and cultural significance, may provide fresh insights into how Wikipedia's content evolves over time.

Most prior work offers static snapshots, not tracking how articles change as new research emerges, commemorations occur, or political environments shift. A long-term view would capture these dynamics, revealing how narratives stabilize, diverge, or converge with new editorial input.

In addition, we know less about the people behind the edits—who they are, what shapes their contributions, and why certain viewpoints gain prominence. Understanding these underlying factors could clarify the social and cultural mechanisms that produce content disparities.

Delving into these areas may show how Wikipedia influences collective memory in a digital era, prompting discussions on editorial policies, community engagement, and tools that help create a more balanced representation of historical events.

2.2.6. Contribution of the Current Study, Addressing the Gaps

The present study aims to address these limitations by examining selected historical battles through a longitudinal, cross-linguistic lens. By focusing on battles significant in multiple cultural settings, this research looks beyond one-time snapshots to see how content transforms over extended periods.

A combined quantitative and qualitative approach will be used. On the quantitative side, metrics such as article length, editing frequency, and source diversity will help identify trends and changes. Qualitative analysis will consider language choices, narrative framing, and the portrayal of key figures and events. Attention will also be paid to the backgrounds of those who contribute, offering clues as to why certain narratives prevail.

This perspective will help reveal how external factors, such as commemorations, political debates, or new academic findings, prompt shifts in Wikipedia's portrayal of historical battles. Beyond this specific case, the project hopes to refine methods for cross-language comparison and suggest approaches that can improve balance and accuracy in other areas of Wikipedia.

2.2.7. Conclusion

In essence, Wikipedia's multilingual nature is both a strength and a source of complexity. Historical battles, strongly tied to collective identity, underscore how language editions can produce distinct narratives, each aligned with particular cultural viewpoints. While existing research has highlighted many of these differences, it has yet to fully capture their evolution over time.

By focusing on historical battles and applying a longitudinal, multi-method framework, this study will deepen our understanding of how knowledge is formed and reformed across linguistic lines. Such insights have broader implications not only for improving Wikipedia's reliability and inclusiveness, but also for understanding how digital platforms shape public perceptions of the past.

2.3. Philosophical Foundations of Truth

2.3.1. Theories of truth

According to literature, there are several theories within the broad philosophical field of truth. All philosophical theories generally revolve around whether a specific statement or proposition is true, where derived narratives or stories may also be encompassed. Usually, one tries to obtain the meaning and application of phrases like "P is true" or "it is true that P," where P represents a statement or proposition (Rescher, 1973).

For this work, the most relevant are the concepts of coherence theory, consensus theory, and pragmatic theory.

2.3.1.1. Coherence theory

The first cannot be defined as a single concept, but rather consists of three doctrines – *nature of reality*, which stipulates that the truth is a coherent system; *definition of truth*, stating that the definition of truth has to be measured in how coherent the propositions are; and *criterion of truth*, that requires truth to clarify the coherence of propositions. Broadly speaking, the coherence theory rejects the need for foundational truths or certainties as basis for knowledge. Instead, under this framework, truth can be reached within the extralogical realm – without relying on any underlying factual certainty (Rescher, 1973). This concept does not achieve truthfulness by mirroring reality, but rather logically integrating ideas (Kirkham, 2001). Nonetheless, individual judgements remain only partially true and should be synthesizing them into an absolute whole, according to this theory, so that through coherence with this concatenated system, truth emerges (Bradley, 1914). Merely looking at isolated propositions by themselves is not enough, coherent context is necessary (Young, 2008).

The necessity of several systems of beliefs that are required in this framework represents a vulnerability, however, in that it is possible that one of such systems may be internally coherent but may be incompatible in combination with others. A set of false propositions could be internally consistent, which means that coherence may not always be a sufficient condition (Russel, 1910). To avoid that, it should be related to an external reality. This would also preclude circularity and subjectivism. If correspondent facts are missing, coherence cannot guarantee truth (Moore, 1901). As the internal set of beliefs plays a crucial role in this theory, empirical knowledge is often not attributed enough weight, whereas sensory experiences are accounted for (BonJour, 1985).

2.3.1.2. Consensus theory

The second, consensus theory, sees truth as consistency within a set of beliefs and that is determined by the agreement of a community of inquirers (Habermas, 1984). It hence emphasizes the social dimension of truth (Rescher, 1993). Essentially, propositions are perceived as true if all members of a given community accept them, following an open and rational discourse (Habermas, 1976). They are established through situations where participants engage among each other without facing any coercion, and subsequently being able to aim for a mutual understanding. Rational consensus achieved under ideal conditions is indicative of truth (Habermas, 1984). It is formed as a product of collective judgement, determined under conditions conducive to impartial and informed deliberation (Rescher, 1993).

Still, adhering to the consensus may not guarantee truthfulness of a proposition as entire communities can be mistaken (Lynch, 2001) and shared biases, or even misinformation which is vastly present in current times, may lead to false beliefs being accepted (Fumerton, 2002). Moreover, consensus is difficult to achieve in places where power dynamics are present (Guess, 1981), which is mostly the case for online platforms.

2.3.1.3. Pragmatic theory

The third one to be discussed here is the pragmatic theory where truth is based on practical consequences and utility by comparing propositions with plausible alternatives. Maximal utility serves as prime indicator of truthfulness and a proposition is considered true if the utility is sufficiently great, meaning that acceptance leads to consequences that outweigh those of rejection (Rescher, 1973). It is also described as the truth that is equivalent to justification or a mere "approximate truth" (Kirkham, 2001), proposing that truth is what is practical or beneficial to believe (James, 1907). Emphasizing action, experience and practicaly consequences (Peirce, 1878), the pragmatic theory serves as guidance to navigate experiences effectively. The meaning of this concept, therefore, lie in its practical effects (James, 1907). It is an instrumental tool for solving problems and facilitating human endeavors (Dewey, 1938).

The fact that under this framework, truth is linked to practical success, the objective nature of it can be undermined. Truth may be conflated with utility. As a result, the subjective understanding of what is truthful varies with individual and cultural preferences (Russel, 1910). Since that is the case, different practical outcomes for different groups run the risk of leading to conflicting truths (Talisse and Aikin, 2008), where the field of relativism comes in.

2.3.2. Application to information convergence

The above-mentioned theories can be applied to help analyse information convergence. As previously stated, information convergence, or in our case, the convergence of truth involves the amalgamation of data from multiple sources to form a unified perspective on a specific matter (Jenkins, 2006). The consensus theory hereby emphasizes the role of collective deliberation and shared understanding in establishing truth (Rescher, 1993). Social processes and collective agreement are essential (Habermas, 1984). The pragmatic theory, on the other hand, acknowledges that truth may change as ne information emerges, since it is the practical consequences of a belief that determine utility and truthfulness (James, 1907). The coherence theory tries to integrate new information into existing sets of belief while preserving coherence with the already existing ones (BonJour, 1985). The logical consistency when converging information is of utmost importance for this theory (Blanshard, 1939).

2.4.Information Convergence in Collaborative Environments

2.4.1. Knowledge construction

Collaborative platforms, such as Wikipedia, rely on collective efforts of users around the world to create, edit, and validate content (Kittur and Kraut, 2008). On these platforms, knowledge is constructed collaboratively through a dynamic process that is characterized by continuously updating, discussing, and building consensus among contributors (Jemielniak, 2014). The consensus theory perfectly embodies how Wikipedia operates as truth within articles emerges from the agreement of editors after open discussions (Reagle, 2010). This dialogue that is taking place on dedicated pages on the platform, the so-called “talk pages”, where improvements to articles and other pages are discussed (Wikipedia, XXX), leads to a common understanding and settles disputes among participants, which follows the theory's principle that truth stems from rational communication (Habermas, 1984).

Wikipedia then ties in with the coherence theory due to the fact that editors try to make sure that new information that is integrated into the platforms base of knowledge logically fits with already existing articles (Forte and Bruckman, 2008). The focus on consistency and alleviating contradictions to ensure that the information presented within a specific page remains coherent (Stvilia et al., 2005).

Lastly, the pragmatic point of view steers how content is contributed via usefulness of the information and its practical ability (Fallis, 2008). Relevance to readers plays a key role in this context.

2.4.2. Mechanisms facilitating convergence

Every platform that serves as a repository for information and has the claim to provide it in an accurate and truthful way, there have to be mechanisms in place that facilitate convergence towards such truth. Among a plethora of those, the mechanisms of information sharing, collective editing and consensus building are worth mentioning in the example of Wikipedia.

2.4.2.1. Information sharing

When it comes to effectively sharing information, convergence in collaborative environments is vital since editors constantly add new content, references, and perspectives to enhance existing work (Forte and Lampe, 2013). Open-source platforms like Wikipedia thrive on the willingness of users to contribute and disseminate information without restrictive barriers. One key aspect of this is the use of Creative Commons licenses, which allows users to use, modify and distribute content from the platform free of charge and without restrictions (Lih, 2009). This enables people from different backgrounds to collaborate and fosters a more inclusive and comprehensive knowledge base (Jemelniak, 2014).

Additionally, community guidelines have a large impact on how information-sharing is performed on the platform. Policies are set in place that ensure verifiability of provided information and encourage editors to reliably source and write reputable content that is truthful and accurate (Magnus, 2009). Contributors must adhere to certain quality standards when citing sources (Jemelniak, 2014), so that credibility and cross-verification can be preserved (Reagle, 2010).

2.4.2.2. Collective editing

The principle of collective editing is another pillar of Wikipedia's collaborative environment. As outlined before, dedicated "talk pages" give contributors spaces to exchange their views and perspectives to build consensus, which often involves negotiation and compromise (Forte et al., 2012). The Neutral Point of View policy, for instance, is employed in the platform to balance different viewpoints to create unbiased content (Reagle, 2010). While it may be challenging to integrate differing views into a unified whole, mediation and arbitration committees that are established help smoothening the sometimes-difficult process of collective editing by resolving disputes and facilitating resolution (Forte el al. 2012).

2.5. Multilingual Information Dynamics

2.5.1. Language as a barrier and a bridge

In a multilingual and multicultural context, language can both serve as a barrier and a bridge. The linguistic diversity of contributors can enrich the content but also presents challenges when it comes to accessibility or inclusivity (Danet and Herring, 2007). Language proficiency greatly affects the ability of editors to contribute effectively to the conversation (Pfeil et al., 2006). On Wikipedia, English serves as the primary language and hence dominates all pages present. This can marginalize readers who are more comfortable communicating in other languages (Danet and Herring, 2007). But not only the proficiency of a language itself may present a challenge on online platforms. Often, cultural nuances are embedded in language to further complicate interactions. While not known to proficient speakers, misinterpretations can arise from such cultural differences, which in turn lead to misunderstandings or even conflicts among contributors (Amant, 2007).

To mitigate these potential sources of misalignment, various strategies are put in place by online platforms. Multilingual interfaces and localized versions are worth mentioning, where users are able to access and contribute content in their native language (Hale, 2014). For Wikipedia, specifically, editions are available in more than 330 languages, which enables more people to participate in the discourse and profit from its broad knowledge base (Wikipedia Contributors, 2024).

2.5.2. Cross-language information flow and cultural influences

On the other hand, language can facilitate knowledge convergence and cross-language information flow is especially important in this regard (Hecht and Gergle, 2010). Multilingual editors are hence a crucial building block when it comes to cross-language information transfer and extending the existing knowledge repository (Hale, 2014). Because of their different background and perspectives, they can be seen as cultural intermediaries that adapt content to suit the linguistic and cultural context of various language communities that are present on the platform (Danet and Herring, 2007), which increases inclusivity and diversity.

It is already established that cultural context significantly influences how information is perceived and subsequently presented in collaborative environments (Pfeil et al., 2006). This is why variations on topics are to be found across different languages (Hecht and Gergle, 2010) and it is especially true for historical events where the cultural background of contributors may reflect national narratives and values (Callahan and Herring, 2011).Cross-cultural collaboration and dialogue that is encouraged on Wikipedia can help identify and resolve biased representations in such events (Hara et al., 2010).

2.6.The Role of Language Links in Wikipedia

In order to connect different language editions of the same topic on the platform, Wikipedia introduced the tool of language links. Fundamentally, they should facilitate information convergence and enhance user accessibility (Hecht and Gergle, 2010). By serving as a bridge element between linguistic communities, said links allow for the effective sharing and subsequent synchronization of content (Liu et al., 2018).

2.6.1. Functionality of language links

At their core, language links on Wikipedia are hyperlinks that connect articles on the same topic in different languages (Hecht et al., 2009). In the current layout of the webpage, they are to be found on the top right corner of the sidebar and essentially enable readers to switch between various language editions through two simple clicks or taps.

First and foremost, they were introduced to provide direct access to equivalent articles to not only retrieve information for multilingual users but also for those who seek alternative perspectives on a specific topic (Adafre and de Rijke, 2006), which may be very relevant for historical events that were traditionally very ambiguous in terms of factual information. By enabling this functionality, language barriers that may have existed before can be overcome and users who had previously been excluded from contributing to a topic are now able to do so.

The links themselves are maintained through a collaborative effort by editors who identify and manually link articles that cover the same topic across languages (Hecht and Gergle, 2010). This process has increased in terms of efficiency after the introduction of Wikidata, centralized knowledge base (Vrandecic, 2012). This repository stores language link information and enables automatic updates, which leads to a drop in redundancy in maintaining link across the various language editions by the editors (Liu et al., 2018).

Lastly, these links are also very helpful when mapping how interconnected the topics that are to be covered on the platform are. This increased structural organization of Wikipedia (Samoilenko et al., 2017) allows for discovery of content variations and the cultural differences that are present across the different editions (Callahan and Herring, 2011).

2.6.2. Facilitation of content synchronization

On top of these generated analytical insights, language links facilitate content synchronization as they enable readers to identify differences in content coverage between different versions, whether in terms of quality or other dimensions (Erdmann et al., 2009). By accessing language links, contributors can compare articles and transfer missing information between editions to enhance the knowledge presented (Adar et al., 2009). Potential gaps that are identified by an individual that exist due to varying contributor bases and available resources among different language editions can then be filled (Hale, 2014). This is especially relevant when up-to-date information is crucial for a certain event (Liu et al., 2018). In such instances, where rapidly evolving subjects require timely updates, synchronization of information is crucial (Adar et al., 2009) and language links are precisely meant to raise awareness among readers to integrate new developments into the existing knowledge base (Erdmann et al., 2009).

Still, full synchronization is difficult to achieve due to a variety of reasons, such as contributor availability, language proficiency, and cultural contexts (Hecht and Gergle, 2010). Moreover, article length and depth, as well as focus on certain aspects persist across different editions even after the introduction of language links (Callahan and Herring, 2011). However, Wikipedia is working on streamlining data and structuring it in a way, so that it can be used uniformly across various editions to help with consistency (Vrandecic, 2012).

2.7.Theoretical Models of Convergence:

For our work, information convergence can be defined as the process in which different pieces of information are synthesized and together from a coherent understanding or consensus on a specific subject (Floridi, 2011). In this section, we would like to highlight three frameworks that help understand how information convergence happens in collaborative environments, such as Wikipedia.

2.7.1. Information Diffusion Theory:

This theory examines how information is propagated through social networks over time (Rogers, 2003). It has been extended to various contexts, including the spreading of ideas, behaviours and information as such, even though it was originally developed to explain the adoption of ideas (Bakshy et al., 2012). Within the information diffusion theory, several concepts are integral.

The network structure, for example, describes the pattern of connections between nodes, or individuals, that influences the speed and to which extent information spreads within the network. In dense networks, many connections faciliated rapid diffusion. In sparse networks, by contrast, the process of information propagation is slower (Watts and Strogatz, 1998).

Thresholds and cascades are another characteristic of this theory, whereby each person has a threshold for adopting new information. If this new information exceeds the threshold, a cascade effect may occur, which in turn leads to widespread adoption. This is often due to multiple exposures from connected peers (Granovetter, 1978; Centola and Macy, 2007).

Influence and Adoption is the last concept we would like to mention in the context of information diffusion theory. Individuals in the network are influenced by their peers and may adopt information based on factors like social pressure, perceived utility, and personal relevance (Bandura, 1986).

On collaborative platforms such as Wikipedia, the information diffusion theory can help explain how edits, norms, and innovations spread among contributors (Yasseri et al., 2012). Several mechanisms are a part of this theory, content propagation being the first one. When a contributor makes a significant edit or adds new information, this change can propagate as other editors observe and incorporate similar updates in related articles or across language editions (Liu et al., 2018).

Experienced editors and administrators often hold central positions in the network. They exert greater influence on content and community norms than novice contributors. Their actions can set precedents that others follow, facilitating convergence of information (Zhang & Zhu, 2011). Through network effects, information is then spread more efficiently, which is enabled by the interconnectedness of contributors. Strong ties and frequent interactions on the platform enhance the diffusion of information (Kittur et al., 2009).

The theory of information diffusion can be leveraged to enhance information convergence in collaborative environments by designing effective networks, in that connections among editors are fostered, which encourages them to collaborate to facilitate faster and more cohesive spread of information (Centola, 2010), and by improving communication channels and tools to lower the barriers of interaction, hence promoting more effective information diffusion (Kittur et al., 2009).

2.7.2. Social constructivism:

The model of social constructivism stipulates that knowledge and understanding are constructed through social interactions and shared experiences (Vygotsky, 1978). It emphasizes that individuals develop cognitive functions and construe meaning in a collaborative manner with a cultural and social context (Berger and Luckmann, 1966). The role of social processes is hereby fundamental in shaping reality, which stands in stark contrast to individualistic views of knowledge acquisition (Gergen, 1999).

According to social constructivism, reality is not an objective entity but is co-created by individuals through language, communication, and interaction (Berger & Luckmann, 1966). Language serves as a primary tool for thought and is crucial in mediating social experiences that lead to knowledge construction (Vygotsky, 1978). The fact that learning and understanding are inherently social activities are also highlighted by this framework, influenced by cultural norms and shared practices (Lave & Wenger, 1991).

In the context of Wikipedia and collaborative environments, social constructivism provides a tool to understand how knowledge is collectively created and converged upon by a community of contributors (Bryant et al., 2005). Through social interactions, be it directly through discussions or indirectly through edits, editors build and refine content (Forte and Bruckman, 2006), which shows how individuals co-construct knowledge by negotiating meanings, resolving conflicts, and integrating diverse perspectives (Jemielniak, 2014).

The concept of a "community of practice" is particularly relevant here. Participants share a common interest and collectively advance their understanding through mutual engagement (Wenger, 1998). In Wikipedia specifically, editors form such communities around articles or topics, contributing their expertise and learning from one another (Forte & Bruckman, 2006). The platform's policies and guidelines emerge from collective negotiation, reflecting shared norms and values that guide content creation (Jemielniak, 2014). Facilitating communication and collaboration among contributors is essential for effective information convergence (Lave & Wenger, 1991).

In summary, the literature reveals Wikipedia's potential as a global, crowdsourced platform where diverse editors collaboratively shape historical knowledge. Yet, the very openness that fosters rapid information sharing also allows linguistic, cultural, and political biases to permeate its content—especially in multilingual or contested contexts. By examining the platform through various theoretical lenses, from coherence and consensus theories of truth to social constructivism and information diffusion models, researchers gain a deeper understanding of how knowledge can converge or diverge over time. Historical battles, with their charged narratives and data discrepancies, emerge as a particularly instructive setting for observing these processes. This perspective highlights the importance of cross-language comparisons, multilingual editorial oversight, and structural tools like language links in striving toward more consistent and reliable representations of historical events. Ultimately, these findings lay the groundwork for further investigation into the mechanisms that shape Wikipedia's evolving depiction of truth on a global scale.

3. Methodology

To accurately measure convergence and reduce the chance of statistical errors, we needed a solid and organized approach. This chapter walks through the main steps of our analysis, including how we gathered data, prepared it, selected the right metrics, and the statistical methods we used to assess convergence. Our goal was to create a clear and repeatable framework for the study, which we will also apply to the different avenues we explore outside of the battle context.

3.1. Data Collection

We aimed to build a dataset that tracks how battle-related information has changed over time across different languages. Specifically, for each battle, we collected one version of its Wikipedia page per year, starting from the first mention. This allowed us to follow how key metrics like troop strength and casualties evolved over time. The data we chose to track, composed of: troop numbers for each side, giving us an idea of the conflict's scale, and detailed casualty figures broken down into deaths, injuries, missing persons, captures, and total casualties. We chose these metrics because they are consistently reported and are the main numeric indicators for battles, that rely on different sources and could therefore converge over time with the introduction of language links. Additionally, having this detailed information lets us explore different ways convergence might occur in our analysis.

To ensure our analysis was thorough, we gathered data in six languages: English, French, German, Portuguese, Spanish, and Italian. These languages were selected because they represent major European powers that played significant roles in battles before World War I, which is the focus of our study. Furthermore, they are written in the Latin alphabet, which eased the process of collecting the data and reduced the risk of errors due to wrong translation.

We intentionally focused on battles before World War I. After this period, battles became much larger and more complex, which would have introduced many challenges to our analysis. For example, estimates of casualties and troop numbers often varied greatly between sides, sometimes differing by hundreds of thousands. Modern battles also include more detailed reporting on equipment losses like tanks, planes, and vehicles. These added complexities would have made our study too broad and introduced many inconsistencies in the data. By concentrating on pre-World War I battles, we kept the scope manageable and ensured our dataset remained reliable. Specifically we ensured consistency by only recording human numbers, filtering out data related to equipment like canons or animals like horses.

Our main goal was to see if adding interlanguage links on Wikipedia led to more consistent numerical data across different language editions. By collecting yearly revisions in multiple languages, we could examine whether these links helped standardize battle-related metrics through shared sources and cross-referencing. The data we collected forms the basis for assessing this trend toward convergence.

We primarily used the "Infobox" section of Wikipedia battle pages as the source for our data, see ***Error! Reference source not found.***. An infobox is a summary table, usually found at the top-right o f a page, that provides key information in a standardized format. For battles, the infobox is split into two parts, each showing metrics like troop strength and casualties for each side involved in the conflict, including the details like dead or injured which we aim to record. The consistent layout and format of infoboxes across different pages and languages made them a good choice for automated data collection. Exhibit 1 shows what a typical infobox may look like.

Belligerents	
France Sardinia	Austria
Commanders and leaders	
Napoleon III Victor Emmanuel II	Franz Joseph I
Strength	
82,935 infantry 9,162 cavalry 240 guns 37,174 infantry 1,562 cavalry 80 guns **Total**: 130,833[1] 320 guns	119,783 infantry 9,490 cavalry 429 guns **Total**: 129,273 429 guns[1][2]
Casualties and losses	
France: 3,887 killed *Including 117 officers* 8,530 wounded 1,518 missing[3] **Sardinia:** 691 killed *Including 49 officers* 3,572 wounded 1,258 missing[3] **Total**: c. 28,000	7,679 killed *Including 216 officers* 17,567 wounded 9,290 missing **Total**: c. 40,000[3]

Exhibit 1: Infobox example

Manually collecting this information was not feasible because of the large amount of data. Analyzing 40 to 50 battles, each with about 20 yearly revisions and six language versions, would result in thousands of data points. Manually extracting data from each infobox would take one to two hours per battle, making the total workload too high. Additionally, Wikipedia's dynamic nature means there are inconsistencies in formatting, language-specific conventions, or naming schemes, which would complicate manual efforts due to, for example, translation requirements.

To handle this, we decided to automate the process, allowing us to scale up in the future and also give us the opportunity to expand our information base beyond the battle context. However, directly scraping structured data from infoboxes was challenging. As is shown by Exhibit 1, the numerical metrics for troop strength and casualties were often stored as unstructured text rather than in separate columns, making automated parsing difficult. There were also many data inconsistencies: some infoboxes only included partial metrics (like the number of deaths but not injuries), while others provided ranges (e.g., "50,000–100,000"). When multiple nations were involved on one side, their contributions were listed separately instead of being combined, adding another layer of complexity. These issues were made worse by changes in formatting and naming conventions across different languages and over time.

To overcome these issues and build our automated Wikipedia scraper, we adopted a three-step approach:

As a first step, we used the Wikipedia API to collect yearly revisions of each battle's Wikipedia page. Utilizing the Wikipedia API, we were able to automatically find and store the yearly revisions for each language. These revisions were then saved by their URL in a structured dataframe that would be filled at a later stage with the numeric information scraped from the infobox.

Once the yearly revisions were obtained, we used an HTML parser to extract the raw infobox content. The raw data is then cleaned from any formatting and saved to be further processed in the next step. As explained previously, due to the dynamic nature of the infobox, we cannot directly process the raw text in a structured and scalable fashion.

In the final step, to overcome this challenge, the extracted raw data is processed using an AI processor. We chose the Gemini AI processor because it is freely available and accessible via an API. As large language models like Gemini or ChatGPT, are more powerful with English data, any non-English text is translated to English using Google's free translation API. The usage of an AI processor allows us to directly establish the same guideline as our manual scraping efforts through prompting, without having to investigate the data ourselves. The output of this processing step is a structured dictionary format, organizing the data into categories for each side of the conflict (e.g., Strength_A, Strength_B). The data returned can then easily be inserted into the columns of our previously empty data frame. The same procedure is then repeated for revision in the data frame and for each language specified.

During this process, we noticed that a modification to our approach was needed. While our initial plan was to collect as much data as possible (separating into Injured, Captured, Missing, etc), we realized that the inconsistency in reporting of these metrics made it impossible to analysis at a later stage. Specifically, in many cases these metrics were not reported at all or just summed up as a total number (for example "4000 Missing, Injured or Captured"). This left us with a large amount of NaN fields, which made comparisons across languages unfeasible. Hence, we decided to modify our approach by changing the data we wish to collect. We focused on Strength, which is always recorded consistently and the total number of casualties summed up, without separation in different categories. Additionally, we improved our handling of ranges (for example "6000-7000 dead") by recording both numbers as Lower- and Upper bound. This way we minimized the number of NaN values and handled ranges without modifying the original data, by for example calculating an average of the two numbers. As a result, our final data was stored in the columns: Strength_Lower_A, Strength_Upper_A, Total_Lower_A, Total_Upper_A, with the same repeated for the other side, as shown in *Table 1*.

Battle Name	**Date**	**lang**	**Strength Lower A**	**Strength Upper A**	**Total Upper A**	**…**	**Total Lower B**
Siege Paris	23.01.2009	de	6000	7000	1000		500

Table 1: Example Row of our Dataframe

To ensure the accuracy of the AI-processed data, we compared Gemini's results with data that had been manually scraped in previous tests. This comparison confirmed that the AI generally performed well, closely matching our manual extraction results. However, we did encounter some discrepancies due to variations in infobox formatting that Gemini was not able to handle. The handling of these inconsistencies are addressed in detail in the Data Processing chapter of this thesis. A detailed workflow of our automated bot is illustrated in ***Error! Reference source not f***

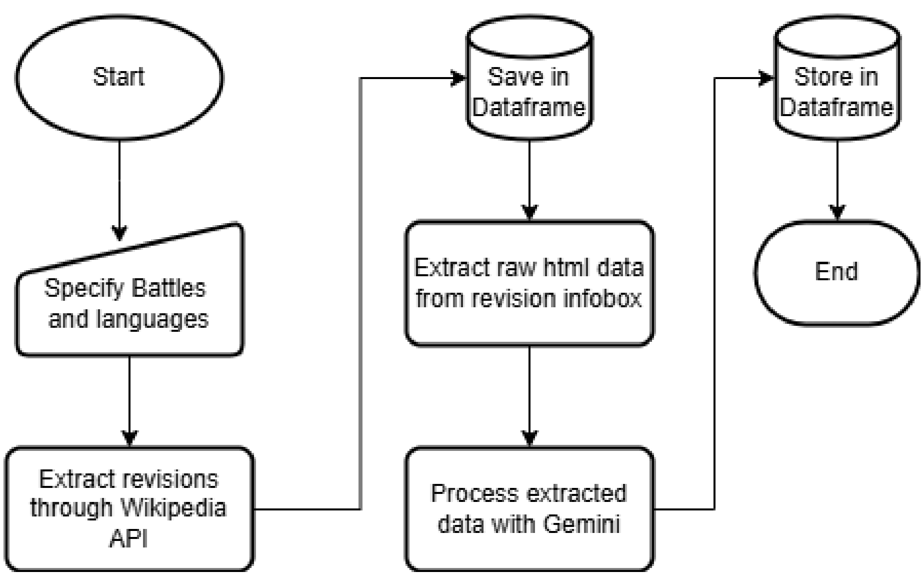

Figure 1: AI Processor Workflow

ound. which outlines the interactions between the Wikipedia API, the HTML parser, and the AI processor.

3.2. Data Cleaning

As we touched upon earlier, simply relying on AI-driven automated scraping doesn't guarantee perfect accuracy. That's why processing and thoroughly cleaning our data before moving forward is very important. We wanted to be sure our dataset was as accurate as possible, to achieve this we implemented the following techniques:

First, we dealt with missing values using a "backward fill" approach. If a data point was missing (NaN), we replaced it with the previously available value. If there was no earlier value, we removed the entire row. This step helped us fix issues caused by processing errors and also got rid of early revisions that didn't include any infobox data at all.

Next, we made sure all the data was in the same format. Different languages handle numbers differently—many European languages, like French or German, use a period to separate thousands instead of a comma. Python interprets this as a decimal point, which could cause the numbers to be largely skewed. To prevent that, we changed the separators so the numerical data would be stored correctly in our dataframe.

We also tackled sudden spikes in the numbers. Sometimes, due to how the data was processed, values could suddenly jump to very high levels. Given the historical context of battles, it's not impossible for certain figures (due to, for example, injuries now being included in total casualties) to surge over time. Still, we set a threshold of 1000%. If any number suddenly increased from one revision to another above that threshold, it was inspected and if needed, we corrected it with our backward fill method. We picked such a high threshold to make sure we weren't changing numbers that were actually correct.

Additionally, whenever we found values that differed by more than 500% from their English reference points, we inspected them manually. We wanted to see if the discrepancy was a real difference in reporting or just a mistake. We steered clear of automatic fixes here because we'd already noticed in our manual scraping attempts, that datapoints in different languages could sometimes vary significantly from other languages. The last thing we wanted was to overwrite good data just because it looked suspicious.

By following these steps—filling in missing values, standardizing formats, watching out for big spikes, and carefully reviewing suspicious entries—we ended up with a dataset that was much more reliable.

3.3. Metrics for Convergence

To statistically measure if our data converges, we developed and selected an approach that can easily be replicated in future expansions of our research.

The main metric we calculated was the relative difference of each datapoint to the reference point, see *Equation 1*. Specifically, for each recorded datapoint—such as troop strength or casualties—we calculated the relative difference between a given language's value and the corresponding English value. This was achieved using the formula:

$$Percentage\ Difference = \frac{Value - Reference\ Value}{Reference\ Value} \times 100$$

Equation 1: Relative Percentage Difference Equation

Choosing the relative difference and not the absolute was the obvious choice for us given the context. Absolute differences would have caused comparisons to be dominated by battles featuring extraordinarily large numbers, overshadowing the convergence patterns in smaller or more moderate battles. By normalizing the data in terms of proportional change rather than raw magnitude, we obtained a more balanced perspective. Adopting this approach also allowed us to skip the normalization steps we would have otherwise had to take, as this is our main metric which is already normalized against the reference value in its' calculation. As a result, every data point, whether from a large battle or a smaller skirmish, could be assessed on an equal term. Furthermore, after having computed the relative differences for each data point, we averaged these values per year for each metric and language. Visualizing these values over time allowed us to get a first feel for the data developments, see any spikes that we might have missed and estimate if convergence is happening or not. If the slopes of our trendlines decrease over time, we would be able to assume convergence, however without statistical assurance.

While visual inspection of trends is valuable and gives first impressions, it is also subjective. To move beyond qualitative observations, we introduced a quantitative approach to measure and confirm convergence. For each metric and language, we performed a linear regression of the yearly average percentage differences against time. The slope of this regression line served as a "convergence coefficient." A negative slope indicates that, as time progresses, the values reported in the given language edition are becoming more similar to those in English. On the other hand, a positive slope suggests that the differences are growing over time, implying divergence rather than convergence.

Just computing slopes is not enough; it is crucial to establish whether these slopes are statistically meaningful. To achieve this, we conducted one-sample t-tests on the slopes. The t-tests determine if the average convergence coefficient significantly differs from zero. A statistically significant negative slope would confirm that the observed trend toward convergence is unlikely to be the product of random variation. Similarly, a non-significant or positive slope would prompt us to reconsider the patterns or investigate potential sources of divergence. To gain a more complete understanding of convergence in each language, we then averaged these convergence coefficients across all metrics for each language. This allowed us to form a comprehensive picture of whether a particular language, considered in its entirety, tended to move closer to or further away from the English reference values over time. By aggregating across metrics, we minimized the risk of drawing conclusions based solely on isolated anomalies in one particular data category.

5. Findings

5.1. Results

Following the extensive data cleaning, we were left with 40 battles that were processed during our analysis. As stated, for each revision we compared the scraped data to the data of the reference language English. As a first insight we computed the average percentage difference in the various categories across languages, the results can be seen in *Table 2*.

Strength Lower A	Strength Upper A	Total Lower A	Total Upper A	Strength Lower B	Strength Upper B	Total Lower B	Total Upper B
14,6%	13,8%	29%	30%	16%	17%	26%	22%

Table 2: Average percentage difference per category across languages

At first glance, we can see that the lower and upper bounds of each category are fairly identical in the average percentage difference between them across languages. Only for the Side B values for the total casualties we see a 4% difference, which could be due to a few battles that are influencing these numbers, which isn't abnormal due to our smaller sample size. Furthermore, we can see that the average difference is much higher for both sides for the casualty values. This is to be expected due to the inconsistent reporting for this category. As mentioned before, it is common that there are large differences between the revisions in this category, some may include captured in the total, while some may not. For strength this is not the case, which is why we have much smaller differences on average in this category. As a next insight, we investigated each language separately by computing the average difference to English across all categories for each language. The results can be seen in *Table 3*.

German	Spanish	French	Italian	Portuguese	English
24%	23%	33%	28%	27%	0%

Table 3: Average percentage difference per language

The overall are fairly similar across all languages, with Spanish and German having the lowest average difference and French the highest. Of course, as English is the reference language, the difference to itself is 0%. This gives us an indication there might be a similar theme across languages and that there are a good number of differences in reported numbers between the English data. However, while the numbers show that reported values differ on average, it gives us no insights about convergence over time. For this we computed average percentages per year across all languages and all languages. The results can be seen in *Figure 2.*

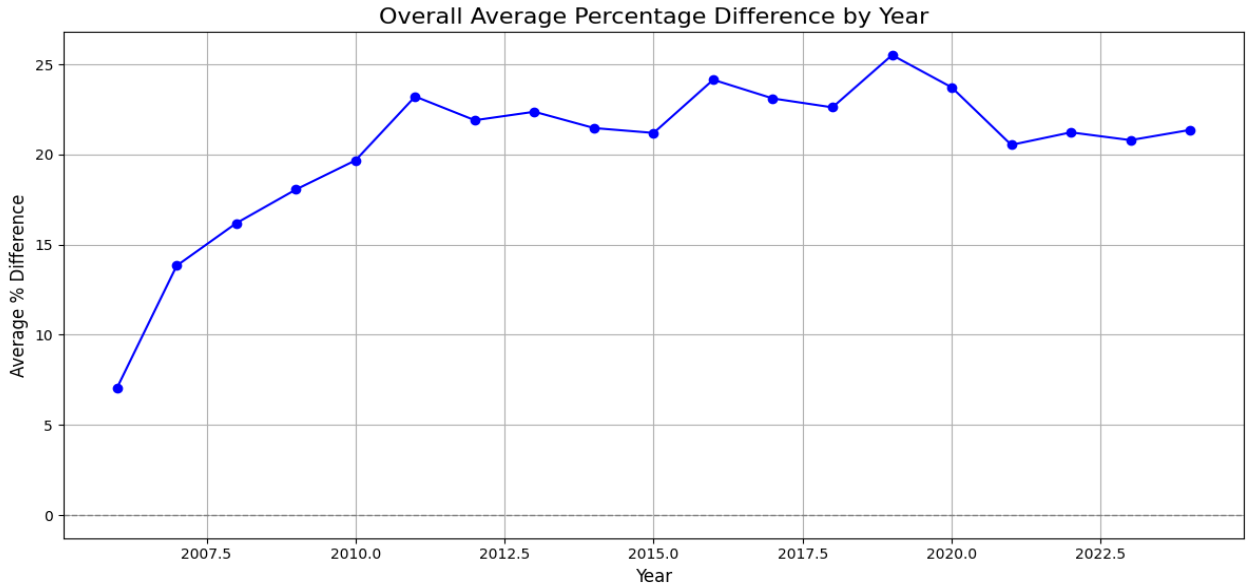

Figure 2: Overall Average Percentage Difference per Year

As evident, the average difference for all metrics and languages does not seem to decrease over time, which would signal convergence. It also does not seem to increase over time either, rather it stagnates over time. The lower difference at the beginning in earlier years can be attributed to lack of data available for many battles in the early years of Wikipedia or that early versions might have just been a direct one to one translation from another language. From the developments over time, we could draw the conclusion that instead of converging or diverging, the differences seem to stagnate, meaning that numbers are drawn from a source at one point of time and rarely change after that. This is interesting as it implies that languages seem to stick with their sources and do not revise or compare them with other languages. To gain a better understanding of these developments, we split the computation above into the various language groups, to see if the above holds across all languages. The results can be seen in *Figure 3.*

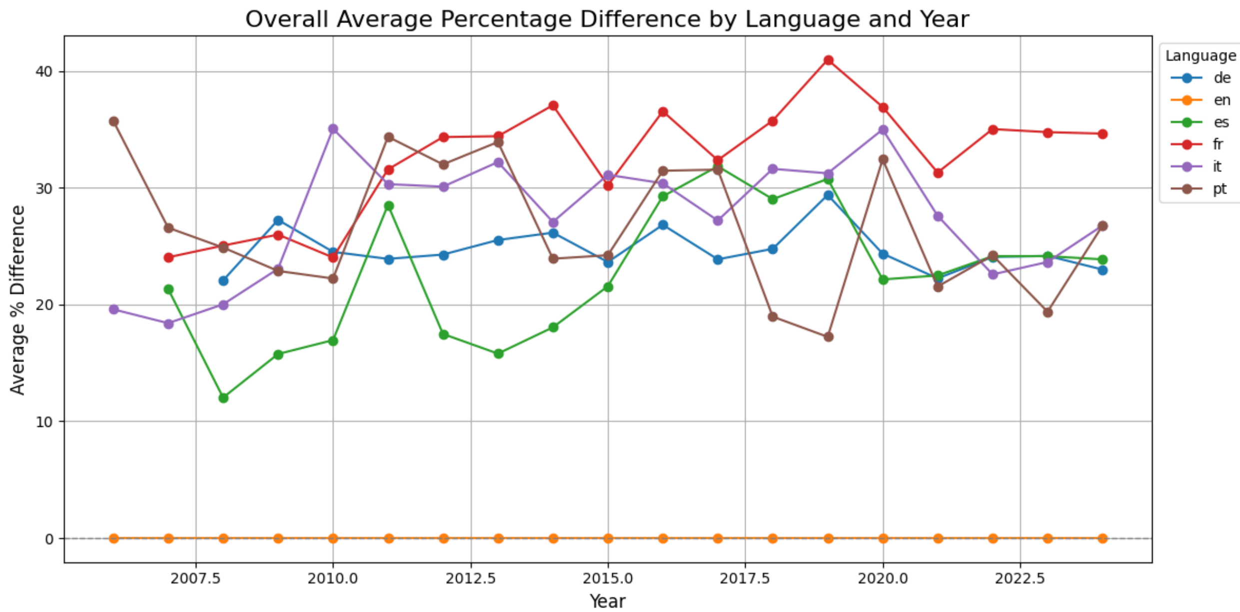

Figure 3: Overall Average Percentage Difference by Language and Year

The data confirms our previous suspicions, the trend for all languages moves fairly identically, with French having the highest average differences and German the lowest. Again, we are not able to see convergence or divergence for any language over time. The trends remain that the differences seem to stagnate around a certain point, indicating that languages stick to their sources and data. To validate these results, as explained in our methodology, we conducted a one sample t-test. The results of our t-test per language can be seen in *Table 4*.

Language	**t-statistic**	**p-value**	**Significant**
de	1.403276	0.203303	False
es	4.492156	0.002825	True
fr	2.688328	0.031161	True
it	2.214359	0.062395	False
pt	1.376840	0.210981	False

Table 4: T-test results of the original battle context

The results of our tests confirmed some of our suspicions but also revealed some surprises. First of all, for German, Italian and Portuguese we confirmed our theory that the trend of time is stagnation. The convergent coefficient (slopes of the trendlines) is statistically not different from zero, meaning we have no evidence to reject the null hypothesis that the coefficient is different from zero. On the other hand, for French and Spanish, we see a p-value <5%, meaning that we can have enough evidence to reject the null hypothesis, showing that statistically the convergence coefficient differs from zero. However, after looking at the coefficients resulting from our linear regression in *Table 5*, we see a positive slope for French and Spanish. This means that for these two languages, the data seems to diverge over time, contradicting our theory of convergence over time. Seeing these results, we found it unnecessary to include language links in our analysis, as they would only be relevant and interesting to analyze if we actually have convergence. These results do not surprise us, as during our manual scraping attempts, which were done to validate the results of our automatic bot, we saw rare movements in the battle metrics over time. On the contrary, we witnessed just small changes and rather as proven by our statistical tests and visualizations, the language's stayed with their sources over most of the revisions.

Language	**Average Slope**
de	0.414533
es	1.067906
fr	1.269892
it	1.021357
pt	0.614562

Table 5: Average Slope per language in the original battle context

5.2. Limitations

While these results provide valuable insights, that we will also explore and expand in the following parts of our thesis, they do come with certain limitations that we need to acknowledge. First of all, with the limitations we put on the languages and the battle timeframe, the sample size is relatively small. While we did conduct extensive cleaning, investigated outliers and corrected wrong data, some singular battles with this large of a sample size can still possibly skew our results. A larger sample size with 300+ battles would balance these out and allow for more generalizable results. Adding to this, there are still possible results from the battle scraper that are incorrect, which would also skew our results. A more robust scraping method with a stronger AI processor developed in the future, could strengthen the results. Furthermore, for some languages there is just more extensive work put into Wikipedia articles than others. Larger language communities like French, Spanish or English, are likely to have more revisions and data available than Italian for example. Finally in our scraping methodology, we are focusing only on the Infobox, if this not available or no casualty data is recorded in it, we have no data. The entire text offers a more complete picture and also offers more data on certain metrics. An expansion of the scraper that can include information from the entire text, would strengthen our results.

6. A Semantic Analysis of Source Texts

6.1. Motivation

The representation of historical events often varies across sources and languages, influenced by cultural perspectives, authorial intent, and the broader contexts in which those texts were produced. In particular, accounts of significant battles—whether from contemporary encyclopedias, memoirs, or modern Wikipedia entries—may differ not only in factual presentation but also in sentiment and emotional framing.

This project aims to shed light on how different types of sources (e.g., lexicons, encyclopedias, or early accounts) and languages (German, English, French) shape the portrayal of key historical battles. Specifically, I will:

- Compare the overall sentiment (positive, negative, neutral, and compound) across sources and languages.
- Examine stylistic differences, such as word count and average sentence length.
- Explore whether older, more contemporary narrative sources show heightened emotional language compared to early Wikipedia entries.

When selecting source texts, I placed particular emphasis on publications likely available to early Wikipedia contributors, as these works could plausibly have informed the initial Wiki articles. Meanwhile, the chosen battles were those for which I could reliably locate both older, possible source texts (in the original languages) and corresponding Wikipedia entries, thus minimizing distortions introduced by translation.

Research on historical texts has long noted the variability introduced by differing editorial conventions, cultural viewpoints, and even editorial aims. Encyclopedic entries may strive for neutrality yet still contain subtle subjective markers, while historic accounts or newspapers can be more emotionally charged, reflecting the immediacy of warfare experienced by writers at the time. To systematically capture these nuances, I employ sentiment analysis as a key methodological component. By quantitatively scoring texts for positivity, negativity, and overall sentiment, it becomes possible to compare how different sources and languages frame significant historical conflicts.

6.2. Methodology

6.2.1. Source Selection

Like already touched upon in the introduction, when looking for material to use as older (primary) sources, I focused on encyclopedias, lexicons, and historical accounts that were likely to have been widely accessible between 2000 and 2005, examples include well-regarded reference works such as Lexikon der deutschen Geschichte, Encyclopedia Britannica, and Brockhaus Brockhaus Enzyklopädie. The underlying reasoning was that contributors operating in Wikipedia's early years might have turned to these volumes for factual content and background. In addition to encyclopedias, I analysed historical accounts that frequently appeared in bibliographies on mid-20th-century warfare, selecting them either for their attention of detail or for the recognized authority they have among scholars.

Meanwhile, the modern (secondary) sources consisted of Wikipedia articles in German, English, and French. I retrieved the earliest versions that contained substantial information on each relevant battle, with the aim of identifying any influence from older materials and recognizing the incremental changes introduced by Wikipedia's continuous editing, expansions, and revisions.

To compile the source texts, I located print editions of these encyclopedias and books in local libraries. Because the exact editions required were unavailable in digital form, I resorted to scanning or photographing the pages and then performing Optical Character Recognition (OCR) to obtain machine-readable text, a tedious task. Additionally, an AI-assisted manual review was used to correct typical OCR errors, such as misread punctuation or character splits. Careful error correction at this stage was crucial to preserving textual integrity, as even minor distortions can significantly affect sentiment analysis results and the photographs often contained unwanted information from surrounding articles.

6.2.2. Battle Data

In order to avoid the artificial sentiment shifts that translations might introduce, all texts, both the older reference works and the corresponding Wikipedia articles, were examined in their original language. For example, when analyzing French sources, only the French encyclopedia text and the French Wikipedia article were used, ensuring that linguistic subtleties remained intact. This approach provided a clearer foundation for sentiment comparisons, free from biases that can creep in through translation.

With language consistency as a guiding principle, battles were then chosen to reflect a diverse range of historical periods and geopolitical contexts. Conflicts spanned the Middle Ages (such as the Battle of Hastings), the Napoleonic Wars (Waterloo, Austerlitz), and major 20th-century battles (Verdun, El Alamein, the Battle of the Bulge). Selecting battles in this way helped prevent any single era or national perspective, such as an overemphasis on World War II in German sources, from disproportionately influencing the entire dataset. Ultimately, the availability of authoritative older texts in the original language served as the decisive factor in determining which battles to include. Where multiple sources existed for the same conflict, the most reputable or widely cited references were prioritized, striking a balance between chronological breadth and textual reliability.

6.2.3. Sentiment Analysis

Given the variety in text lengths, from concise encyclopedic entries to much more expansive Wikipedia articles, I required a sentiment analysis tool both robust for short passages and sensitive to nuanced polarity in extended narratives. VADER (Valence Aware Dictionary for sEntiment Reasoning) proved especially well-suited for these requirements due to its features:

- **Rule-Based Lexicon**: A human-curated lexicon capturing both standard and colloquial sentiment-laden terms.
- **Sensitivity to Degree Modifiers**: It accounts for amplifiers (e.g., “very,” “extremely”) and negations, which can drastically alter measured sentiment.
- **Ease of Implementation**: Its Python interface allows for batch processing of texts in different languages (as long as recognizable sentiment terms or synonyms appear in the lexicon), which is important for this multilingual dataset.
- **Proven Reliability**: Although originally designed for social media, VADER has been validated on various text genres, including short historical documents, which are very close to what I am working with in the sentiment analysis. (Hutto & Gilbert, 2014).

While no automated tool is perfect, particularly when trying to analyze specialized historical vocabulary, VADER’s adaptability and transparent scoring system made it a strong choice for comparing older entries, sometimes written in a non-standard format and using period-specific vocabulary, to more detailed, modern Wikipedia texts.

6.2.4. Metrics

After cleaning each text, I used VADER to compute four key sentiment metrics:

- **Positive**: Proportion of words with a positive connotation.
- **Negative**: Proportion of words with a negative connotation.
- **Neutral**: Proportion of words deemed to lack strong emotional connotations.

- **Compound**: A normalized, aggregate sentiment measure ranging from -1 (extremely negative) to +1 (extremely positive).

To gain additional insight into each text's structure I also analyzed:

- **Total word count**
- **Number of sentences**
- **Average sentence length** (in words)

These metrics help clarify whether longer and more elaborate texts correlate with stronger negative or positive sentiment. For example, Wikipedia entries typically cover strategic context, casualty figures, and tactical decisions in greater depth, which could amplify negative sentiment when recounting heavy losses or brutal conditions.

6.3. Results

In this section, I compared sentiment metrics between older source texts and modern Wikipedia entries, organized by language (German, English, and French). Each subsection highlights one illustrative battle taken from the relevant dataset.

6.3.1. German Sources vs. German Wikipedia

Source	Total Words	Avg. Positive	Avg. Negative	Avg. Neutral	Compound Score
Der Zweite Weltkrieg - Ein Lexikon	49	0	0.161	0.839	**-0.277**
Wikipedia	87	0	0.179	0.821	**-0.391**

Table 6: Sentiment Analysis results from German sources for: Ardennenoffensive (1944–1945)

To illustrate a typical analysis for the German sources we are looking at the Ardennenoffensive. Both accounts emphasize Germany's intent to split the Allied front and reach Antwerp, the harsh winter conditions, and the eventual failure of the offensive. The older lexicon version is succinct yet still negatively toned, reflecting the losses incurred. The Wikipedia entry, despite being only somewhat longer (87 words), further underscores the scale of logistical problems and the significance of Bastogne, pushing the overall compound sentiment to -0.391, more negative than -0.277 in the lexicon source text.

6.3.2. English Sources vs. English Wikipedia

Source	Total Words	Avg. Positive	Avg. Negative	Avg. Neutral	Compound Score
Britannica	44	0	0.278	0.722	**-0.596**
Wikipedia	281	0.036	0.185	0.779	**-0.415**

Table 7: Sentiment Analysis results from English sources for: Gallipoli (Dardanelles) (1915–1916)

For the English sources we are looking at the Battle of Gallipoli: here, the older source is not only concise but also starkly negative. It highlights the failed campaign and especially the high Allied casualty count (214,000), pushing the compound score to -0.596. The Wikipedia article, while still clearly negative (-0.415), spreads the emphasis over strategic aims, Turkish resistance, and the eventual evacuation, "diluting" the immediate impact of the disaster. The result is a slightly less severe compound score than that of the older text.

6.3.3. French Sources vs. French Wikipedia

Source	Total Words	Avg. Positive	Avg. Negative	Avg. Neutral	Compound Score

Larousse	37	0	0	1	**0**
Wikipedia	449	0.022	0.017	0.961	**0.037**

Table 8: Sentiment Analysis results from French sources for: Bataille d'Austerlitz (1805)

The older French text describes Austerlitz in two short sentences, focusing on Napoleon's success without providing emotive language about casualties or the harsh realities of war. Consequently, VADER interprets it as perfectly neutral (compound 0.000), particularly with short lexicon entries, this was problem that occurred multiple times. The Wikipedia entry, with 449 words, includes more details on the scale of the victory, troop maneuvering, and losses on both sides. This broader coverage generates a slightly positive tilt (0.037), reflecting mentions of Napoleon's "brilliant tactics" or "decisive victory."

6.3.4. Summary

In comparing older sources and Wikipedia texts across German, English, and French, I was generally able to observe that both sets lean negative when describing conflicts, yet Wikipedia typically amplifies this negativity due to more detailed references to casualties, battlefield conditions, and strategic failures. In some of the sources, the conciseness and factual tone of the older source leads to near-neutral reading, while Wikipedia's lengthier explanations introduce stronger negative cues. Conversely, the older Gallipoli (7.3.2) excerpt is more negative than Wikipedia, indicating that very concise but stark mentions of high casualties can drive exceptionally negative scores.

Across languages, these patterns remain consistent: if an older source is negative, the corresponding Wikipedia article tends to be negative as well, often more so due to expanded context, while short, neutral-sounding older texts often see a shift to clearly negative or slightly positive in the Wikipedia version. Overall, the more comprehensive coverage in Wikipedia entries intensifies sentiment markers, whereas older encyclopedic or lexicon-style references tend to be short and register milder polarity.

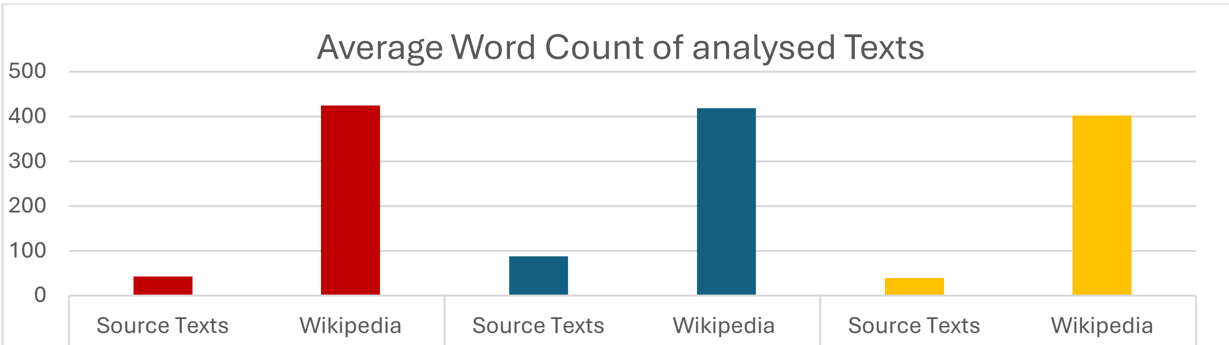

Figure 4: Average Wordcount across sources and languages

Figure 4 shows the stark differences in word count across languages and source types. On average, Wikipedia articles are substantially longer than source texts in all three languages (English, French, German). The most striking gap appears in German, where source texts barely exceed 40 words on average, compared to over 400 words for Wikipedia. Such variation aligns with the expectation that modern, collaboratively edited articles explore battles in greater detail, often including extensive context, casualty figures, and strategic analysis which then in turn often leads to a higher sentiment value of the text.

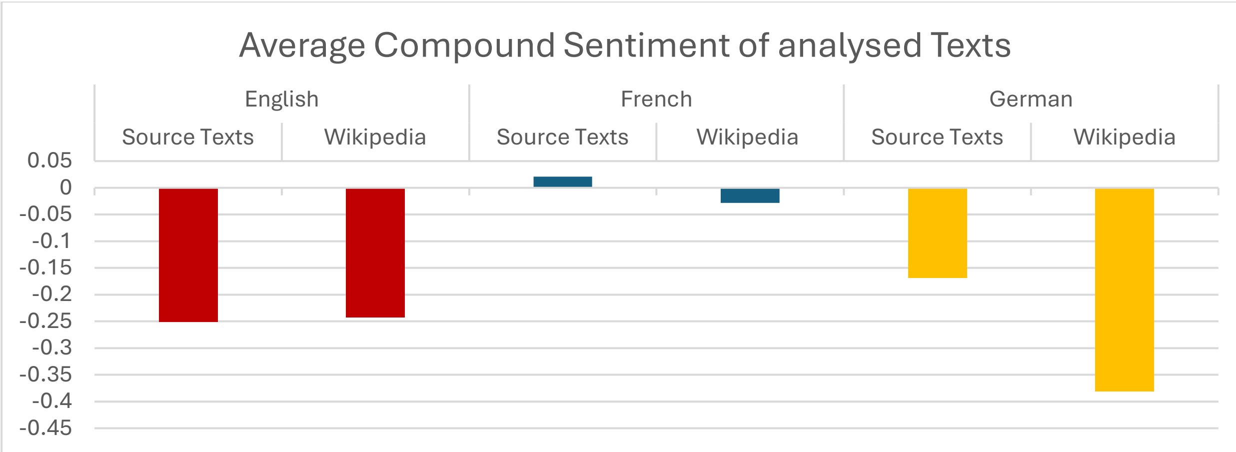

Figure 5: Average Compound Sentiment across sources and languages

Figure 5 illustrates the average compound sentiment. Here we see that German Wikipedia entries are the most negative overall, reflecting how expanded details (e.g., logistical difficulties or dramatic defeats) intensify negative sentiment compared to the older source texts. French source texts hover around neutral or slightly positive but shift to mildly negative in French Wikipedia. Meanwhile, English source texts and Wikipedia entries are comparatively close in negativity, albeit with shorter older texts occasionally becoming strongly negative (as in the Gallipoli excerpt). These trends confirm a broader pattern: more comprehensive coverage generally increases negative sentiment, likely due to mention of devastating events and human costs, though language-specific nuances and difference in the expression of emotion between the languages still shape the overall emotional tone.

6.4. Limitations

Although this research offers valuable insights, it is limited in several important ways. First, the range and depth of reference materials varied considerably among the three languages. While I had multiple German texts at my disposal to choose from, the English and French sources often relied on a single encyclopedia edition or sets of historic accounts, raising questions about whether

Wikipedia editors in those languages actually consulted those specific works, or if they turned instead to school textbooks or other widely used but less detailed publications. Moreover, older encyclopedias tend to provide only broad summaries rather than precise accounts of what events, limiting the availability of truly analyzable and comparable data.

A second issue concerns the length of many primary texts. Quite a few of the older entries, particularly in encyclopedias, are extremely short and contain little more than a paragraph of factual statements, sometimes only about 30 words. Since sentiment analysis performs best on fuller passages with a variety of lexical cues to work with, these condensed texts rarely offer sufficient context for accurately detecting nuanced emotional or narrative tones. As a result, direct comparisons between such source materials and the normally more elaborate Wikipedia articles can become skewed.

Additionally, there are the technical limitations of the tool used for the Sentiment Analysis: VADER. From a methodological standpoint, VADER matched this project's requirements well, yet it remains a rule-based tool that does not fully account for historical context or older writing styles. While more advanced NLP models like BERT could produce more granular insights, training them on specialized historical data would demand significant resources, computing power, and data. Additionally, the already mentioned disparity in length between concise encyclopedia excerpts and expansive Wikipedia narratives can distort negativity or neutrality scores of VADER, making direct comparisons less precise.

Lastly, collecting and processing data introduced logistical challenges. Manually photographing or scanning older sources and correcting OCR errors left room for accidental mistakes and the accidental capture of surrounding information that can distort the analysis, even though I tried to get rid of these distortions by manually reviewing the final text before analysis, mistakes can still slip in. A thoroughly digitized collection of encyclopedias would likely yield cleaner data. And though the selection of battles covers multiple centuries, practical access to robust source material ultimately limited both the scope and the scale of the study's findings.

6.5. Conclusion

The findings from this comparative sentiment analysis underscore the significant role that source type, language, and editorial style play in shaping the emotional tone and factual detail of historical battle accounts. The results indicate that older encyclopedic entries, often concise and factual, tend to register moderate or near-neutral sentiment, whereas early Wikipedia articles, with their more extensive coverage of casualties, strategic context, and eyewitness elements, frequently exhibit stronger negative polarity. This divergence is consistent across German, English, and French materials, although each language demonstrates subtle differences in phrasing and cultural framing.

By focusing on non-translated texts and carefully selecting widely accessible encyclopedias that predate or coincide with early Wikipedia development (circa 2000–2005), I tried to minimize translation artifacts and ensured realistic source relationships. The OCR process and subsequent text cleaning were crucial for capturing these reference materials accurately, highlighting the challenges inherent in working with older print publications.

Looking ahead, this approach could be expanded by incorporating other likely source texts for the analyzed Wikipedia articles like school history books or other encyclopedias. Furthermore, there is room for improvement by looking into additional languages, applying advanced NLP methods to capture more nuanced historical context, as well as looking at the same battles across different sources in different languages to analyse the sentiment the different parties had and potential reasoning behind it. The results presented here demonstrate both the utility of automated sentiment analysis in dissecting historiographical bias and the importance of not looking at quantitative findings of research without considering political and cultural context, e.g., the difference in emotional language used in different cultures. Through this dual lens, we can better understand how collective memory and editorial practices influence the narratives that shape our perception of pivotal military conflicts.

7. Outlook & Discussion

For many people Wikipedia has become a primary source of information and knowledge. People around the world collaborate on articles in their language to achieve Wikipedia's goal of making knowledge accessible and available for everyone. Our thesis has in many ways showed, how the promise of unified information from a single source still faces many challenges. Reflecting on our initial work in the original battle context, which formed the basis of our subsequent individual explorations, we found that even today, choosing a different language can lead to notably different representations of the same historical events. For us, this made a lot of sense when reflecting on the nature of battles. Historians cannot always be certain about the absolute facts, given that the records they rely on were typically produced by past governments or earlier historians. These sources often contain only estimates or figures that were altered to serve the political agendas or public relations aims of their time, making it difficult to obtain definitive information. Furthermore, there are conflicting sources in many cases, with one side reporting different numbers from the other. This means that Wikipedia editors need to choose the source they deem the most reliable, which could lead to subconscious biases. This is why our individual explorations were necessary and valuable, as they help to understand these observed dynamics. First of all, expanding the geographical scope of our data beyond Europe and analyzing the differences between continents and culture groups showed that there is still much we don't understand. Areas outside of Europe showed even less possible hints at convergence, which seems reasonable due to how Eurocentric a lot of source material is. Another aspect is that the further you expand into underrepresented culture groups, the less data is available. The potential for bias in the battle context is evident which paved the way for exploring convergence outside of that context, on categories that we deemed "factual". However again, we were faced with discrepancies. While much smaller on average than for battles, they still existed, with no sign of convergence over time. These findings made it evident for us, that we had to investigate the data on a more granular scale, comparing the Wikipedia entries with source text from encyclopedias. Again, here we concluded that Wikipedia faces challenges with data alignment. Only very few articles showed perfect alignment of the actual source text of the battle with the Wikipedia article, highlighting how editors rely on other sources. By comparing the linguistic variations via sentiment analysis of both the articles and the original source material, we found that the numerical data is not the only indicator of limited convergence. The emotional tone also differs significantly between Wikipedia entries and reference texts across languages, though a few similarities can be shown.

These findings underline the complexities of global knowledge sharing and the limitations of Wikipedia's open-editing framework in achieving consistency.

7.1. Limitations

Throughout this study, several challenges and limitations emerged that influenced the scope and reliability of our findings. One significant issue was data availability. Across all parts of the study, we encountered difficulties obtaining consistent information. This included gaps in source texts from encyclopedias, limited revisions for certain battles in less commonly edited languages, and inconsistencies in the availability of data for underrepresented regions. These limitations impacted the breadth and depth of our analysis.

Additionally, the dynamic and ever-changing nature of Wikipedia posed challenges to accuracy and consistency. As articles are frequently updated by contributors, capturing a stable dataset proved difficult. We constantly were faced with new formats to data representation, like changing names or variables in the infoboxes, which proved a challenge to the way we were collecting data.

The AI processor used for data scraping also introduced potential errors, highlighting the need for more advanced tools. A more robust AI approach could improve the precision and efficiency of future data extraction efforts, reducing the likelihood of mistakes and enabling more comprehensive datasets.

Another limitation was the sample size. To better account for outliers and ensure statistically reliable conclusions, future studies should expand the scope of their data collection. A larger sample size would help smooth inconsistencies and provide more generalizable results, especially in understanding patterns of convergence across languages and regions.

7.2. Future Research

This study has shed light on the challenges of achieving consistency across Wikipedia's multilingual content, but there is still much more to explore. One clear next step is to expand the dataset. By analyzing a larger number of battles, a longer timeframe, and a wider range of countries and languages, future research could offer a more complete picture of how Wikipedia evolves globally and over time.

It would also be valuable to look beyond infoboxes. While these provide clear and structured data, much of the story is told in the full text of articles. Examining these narrative sections could reveal even more about how events are framed differently across languages, showing us the nuances of interpretation and editorial choices.

Another direction is to include more encyclopedias and languages in the analysis. Comparing Wikipedia entries with a broader set of reference works could help us understand whether certain traditions or regions are more aligned or prone to discrepancies. This would also provide insight into how well Wikipedia reflects the diversity of perspectives it claims to unify.

To dig deeper into the causes of inconsistencies, future research could focus more narrowly on just two or three languages. This more concentrated approach would allow a closer examination of how cultural biases or editorial practices shape the information presented. By isolating specific factors, we might better understand why certain trends emerge and how to address them.

Lastly, trying new tools and methods could open up exciting possibilities. For example, advanced machine learning models or natural language processing techniques could help uncover subtle differences in tone, style, or sentiment across languages that might not be obvious through traditional methods. These tools could also improve how we measure alignment and divergence in multilingual content.

Bibliography

Adar, E., Skinner, M., & Weld, D. S. (2009). Information arbitrage across multi-lingual Wikipedia. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining* (pp. 94–103). ACM.

Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012). The role of social networks in information diffusion. In *Proceedings of the 21st International Conference on World Wide Web* (pp. 519–528). ACM.

Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory.* Prentice-Hall.

Bao, P., Hecht, B., Carton, S., Quaderi, M., Horn, M., & Gergle, D. (2012). Omnipedia: Bridging the Wikipedia Language Gap. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1075–1084).

Berger, P. L., & Luckmann, T. (1966). *The Social Construction of Reality: A Treatise in the Sociology of Knowledge.* Anchor Books.

Blanshard, B. (1939). *The Nature of Thought* (Vols. 1–2). George Allen & Unwin.

BonJour, L. (1985). *The Structure of Empirical Knowledge.* Harvard University Press.

Bradley, F. H. (1914). *Essays on Truth and Reality.* Oxford University Press.

Callahan, E. S., & Herring, S. C. (2011). Cultural bias in Wikipedia content on famous persons. *Journal of the American Society for Information Science and Technology, 62*(10), 1899–1915.

Centola, D. (2010). The spread of behavior in an online social network experiment. *Science, 329*(5996), 1194–1197.

Centola, D., & Macy, M. (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology, 113*(3), 702–734.

Danet, B., & Herring, S. C. (2007). *The Multilingual Internet: Language, Culture, and Communication Online.* Oxford University Press.

Dewey, J. (1938). *Logic: The Theory of Inquiry.* Holt, Rinehart and Winston.

Erdmann, M., Nakayama, K., Hara, T., & Nishio, S. (2009). An approach for extracting bilingual terminology from Wikipedia. In *Proceedings of the 13th International Conference on Database Systems for Advanced Applications* (pp. 380–392). Springer.

Floridi, L. (2011). *The Philosophy of Information.* Oxford University Press.

Ford, H., Sen, S., Musicant, D. R., & Miller, N. (2013). Getting to the Source: Where Does Wikipedia Get Its Information From? In *Proceedings of the 9th International Symposium on Open Collaboration* (Article No. 9:1–9:10).

Forte, A., & Bruckman, A. (2006). From Wikipedia to the classroom: Exploring online publication and learning. In *Proceedings of the 7th International Conference on Learning Sciences* (pp. 182–188). International Society of the Learning Sciences.

Forte, A., & Bruckman, A. (2008). Scaling Consensus: Increasing Decentralization in Wikipedia Governance. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences* (p. 157).

Forte, A., & Lampe, C. (2013). Defining, understanding, and supporting open collaboration: Lessons from the literature. *American Behavioral Scientist, 57*(5), 535–547.

Geiger, R. S., & Ribes, D. (2010). The work of sustaining order in Wikipedia: The banning of a vandal. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (pp. 117–126).

Gergen, K. J. (1999). *An Invitation to Social Construction.* SAGE Publications.

Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology, 83*(6), 1420–1443.

Habermas, J. (1976). *Communication and the Evolution of Society* (T. McCarthy, Trans.). Heinemann.

Habermas, J. (1984). *The Theory of Communicative Action: Reason and the Rationalization of Society* (Vol. 1, T. McCarthy, Trans.). Beacon Press.

Hale, S. A. (2014). Multilinguals and Wikipedia editing. In *Proceedings of the 2014 ACM Conference on Web Science* (pp. 99–108). ACM.

Halfaker, A., Kittur, A., Kraut, R. E., & Riedl, J. (2009). A jury of your peers: Quality, experience and ownership in Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration* (Article No. 15).

Hara, N., Shachaf, P., & Hew, K. F. (2010). Cross-cultural analysis of the Wikipedia community. *Journal of the American Society for Information Science and Technology, 61*(10), 2097–2108.

Hecht, B., & Gergle, D. (2010). The Tower of Babel meets Web 2.0: User-generated content and its applications in a multilingual context. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 291–300).

Hecht, B., Adams, B., & Riedl, J. (2012). The collaborative ecosystem of multilingual Wikipedia. In *Proceedings of the 2012 ACM Conference on Computer Supported Cooperative Work* (pp. 161–170).

Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International AAAI Conference on Weblogs and Social Media*.

James, W. (1907). *Pragmatism: A New Name for Some Old Ways of Thinking.* Longmans, Green, and Co.

Jemielniak, D. (2014). *Common Knowledge? An Ethnography of Wikipedia.* Stanford University Press.

Jenkins, H. (2006). *Convergence Culture: Where Old and New Media Collide.* New York University Press.

Keegan, B. C., Gergle, D., & Contractor, N. (2011). Hot off the wiki: Dynamics, practices, and structures in Wikipedia's coverage of the Tōhoku catastrophes. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration* (pp. 105–113).

Kirkham, R. L. (2001). *Theories of Truth – A Critical Introduction.* MIT (5th ed.).

Kittur, A., & Kraut, R. E. (2008). Harnessing the wisdom of crowds in Wikipedia: Quality through coordination. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work* (pp. 37–46).

Kittur, A., Pendleton, B. A., & Kraut, R. E. (2009). Herding the cats: The influence of groups in coordinating peer production. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration.* ACM.

Konieczny, P. (2010). Adhocratic governance in the Internet age: A case of Wikipedia. *Journal of Information Technology & Politics, 7*(4), 263–283.

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation.* Cambridge University Press.

Lewoniewski, W., Węcel, K., & Abramowicz, W. (2017). Multilingual ranking of Wikipedia articles with quality and popularity assessment in different language editions. *Online Information Review, 41*(4), 478–495.

Liu, S., Yan, X., Li, H., Zhao, T., & Wang, Y. (2018). Cross-lingual entity alignment via attribute-enhanced embedding. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 2486–2495). Association for Computational Linguistics.

Luyt, B., & Tan, D. (2010). Improving Wikipedia's accuracy: Is edit age a solution? *Journal of the American Society for Information Science and Technology, 61*(12), 2582–2593.

Massa, P., & Scrinzi, F. (2012). Wikipedia and the wisdom of crowds: A methodology to evaluate content quality. In *Proceedings of the 8th International Symposium on Wikis and Open Collaboration* (pp. 1–10).

Massa, P., & Scrinzi, F. (2013). Manypedia: Comparing language points of view of Wikipedia communities. In *Proceedings of the 9th International Symposium on Open Collaboration* (pp. 1–10).

Miquel-Ribé, M., & Laniado, D. (2018). Wikipedia Cultural Diversity Dataset: A cartography of Wikipedia and Wikimedia projects. In *Proceedings of the 14th International Symposium on Open Collaboration* (pp. 1–9).

Moore, G. E. (1901). Truth. *Mind, 10*(40), 446–450.

Peirce, C. S. (1878). How to Make Our Ideas Clear. *Popular Science Monthly, 12,* 286–302.

Pfeil, U., Zaphiris, P., & Ang, C. S. (2006). Cultural differences in collaborative authoring of Wikipedia. *Journal of Computer-Mediated Communication, 12*(1), 88–113.

Reagle, J. (2010). *Good Faith Collaboration: The Culture of Wikipedia.* MIT Press.

Rescher, N. (1993). *A System of Pragmatic Idealism, Volume II: The Validity of Values.* Princeton University Press.

Rescher, N. (1973). The Coherence Theory of Truth.

Rogers, E. M. (2003). *Diffusion of Innovations* (5th ed.). Free Press.

Russell, B. (1910). Pragmatism. *The Philosophical Review*, *19*(1), 1–13.

Samoilenko, A., Karimi, F., Edler, D., Kunze, T., & Strohmaier, M. (2017). Linguistic neighbourhoods: Explaining cultural borders on Wikipedia through language co-editing activity. In *Proceedings of the Eleventh International AAAI Conference on Web and Social Media* (pp. 217–226). AAAI Press.

Stvilia, B., Twidale, M. B., Smith, L. C., & Gasser, L. (2005). Assessing Information Quality of a Community-Based Encyclopedia. In *Proceedings of the International Conference on Information Quality* (pp. 442–454).

Stvilia, B., Twidale, M. B., Smith, L. C., & Gasser, L. (2008). Information quality work organization in Wikipedia. *Journal of the American Society for Information Science and Technology*, *59*(6), 983–1001.

Talisse, R. B., & Aikin, S. F. (2008). *Pragmatism: A Guide for the Perplexed.* Continuum.

Viegas, F. B., Wattenberg, M., & Dave, K. (2004). Studying cooperation and conflict between authors with history flow visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 575–582).

Vrandecic, D. (2012). Wikidata: A new platform for collaborative data collection. In *Proceedings of the 21st International Conference on World Wide Web* (pp. 1063–1064). ACM.

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes.* Harvard University Press.

Warncke-Wang, M., Cosley, D., & Riedl, J. (2015). Tell Me More: An Actionable Quality Model for Wikipedia. In *Proceedings of the 9th International AAAI Conference on Web and Social Media* (pp. 454–463).

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*(6684), 440–442.

Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity.* Cambridge University Press.

Wikipedia Contributors. (2024). List of Wikipedias. Retrieved from https://meta.wikimedia.org/wiki/List_of_Wikipedias

Yasseri, T., Sumi, R., & Kertész, J. (2012). Circadian patterns of Wikipedia editorial activity: A demographic analysis. *PLOS ONE, 7*(1), e30091.

Young, J. O. (2008). The Coherence Theory of Truth. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2008 ed.).

Zhang, X., & Zhu, F. (2011). Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *American Economic Review, 101*(4), 1601–1615.