FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Unravelling technology acceptance through text mining: An application to automated truck platooning

Rhaydrick Sandokhan Pinheiro Teixeira Tavares

Mestrado em Engenharia e Ciência de Dados

Supervisor: António Manuel Cabral Vieira Lobo

Co-Supervisor: Sérgio Pedro de Matos Duarte

Co-Supervisor: Vera Lúcia Miguéis Oliveira e Silva

October 29, 2024

Unravelling technology acceptance through text mining: An application to automated truck platooning

Rhaydrick Sandokhan Pinheiro Teixeira Tavares

Mestrado em Engenharia e Ciência de Dados

October 29, 2024

Abstract

The platooning technology facilitates the seamless coordination of two or more trucks as a connected convoy, using connectivity technology and automated driving support systems. While offering social-economic and environmental advantages, the transition to practical platooning systems necessitates advancements in infrastructure, in-vehicle technology, traffic legislation, and workforce capabilities. To comprehensively gauge the perspectives of key stakeholders influenced by this transformative shift, interviews were conducted with road regulators, truck drivers, and road operators. Leveraging Natural Language Processing (NLP) and LDA models, the study aims to analyze interview responses to ascertain if these diverse actors share similar views on critical aspects of truck platooning. Nevertheless, by employing NLP, the research seeks to uncover underlying patterns and commonalities in the opinions of road regulators, truck drivers, and road operators and the existing literature, shedding light on potential challenges and opportunities in the widespread adoption of platooning systems.

Keywords: Truck platooning, NLP, LDA, Decision-Making, Systems Approach

Acknowledgements

To my beloved wife and daughter, whose support and patience have been my pillars throughout this challenging journey. Your love and encouragement have carried me through countless late nights and stressful moments. Thank you for being my constant source of strength. To my family and friends, whose understanding and presence have meant the world to me. Your support was always felt, even when I couldn't be there in person. To my manager and company, who have provided me with an environment that fosters growth and innovation. Your support for my continued education and professional development has been invaluable. To my colleagues, who have been not only collaborators but also mentors and friends. Your guidance, teamwork, and willingness to share your knowledge have been instrumental in my academic achievements. I am truly fortunate to have worked alongside such talented individuals. Finally, to my supervisors, whose support, patience, and willingness to help have made this journey possible. Your guidance, suggestions, and belief in my abilities have been invaluable. Thank you for your unwavering support and dedication.

This work is financially supported by national funds through the FCT/MCTES (PIDDAC), under the project with DOI 10.54499/PTDC/ECI-TRA/4672/2020.

Rhaydrick Sandokhan

"Be curious, not judgmental."

Carl Sagan

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context

Sustainability and greener solutions are significant topics tackled every time we discuss innovative technology solutions to our society's problems. New technology faces the challenges of "dualism," functioning efficiently today while planning and innovating effectively for tomorrow, Katz and Allen (1985). Thus, the primary purpose of innovative technologies is to bring innovation and increase economic benefits while decreasing costs and social-environmental impacts.

In recent years, the automotive industry has attempted to be at the vanguard of this process by developing hybrid and fully electric vehicles and automated transport systems for freight and passengers. It is imperative to highlight that road transport is responsible for half of the freight transport ETRAC (2019). Thus, it is vital that we find a viable solution to answer the goals defined by the European Commission regarding safety and environmental concerns ETRAC (2022). Truck platooning is one of the proposed solutions since, at its core, it takes advantage of state-of-the-art communication and vehicle automation technologies while promising to reduce energy consumption and emissions.

Although truck platooning can transform the operations of logistics companies Ferrell et al. (2020), their workforce may perceive a potential loss of jobs and opportunities, as well as could resist and see more risks than benefits regarding the safety of vehicle automation Cunha et al. (2022).

Developed under the project **TRAIN**, this research aims to extract knowledge from the transcribed speeches of the actors involved in the deployment of vehicle automation in the context of truck platooning, as well as from a literature review previously conducted under the same project. Mainly, this work should identify the main topics referred to by those actors, their sentiments and arguments towards the benefits and risks of the technology, and the relationships between actors. Moreover, the research should allow us to contrast the insights from the transcribed speeches with those from the reviewed international literature. In the end, a comprehensive mapping of the main drivers and barriers to truck platooning technology acceptance and adoption at a large scale should be obtained from the different actor groups.

1.2 Motivation and Objectives

Platooning systems allow two or more trucks to operate in convoys like trains by establishing a virtual connection through connectivity and automated driving support technology. The trucks maintain a distance from each other PPMC (2021), and the initial truck is designated as the leader, taking on the responsibility of the driving task with support from Cooperative Adaptive Cruise Control (CACC) and system monitoring Willemsen et al. (2021). The subsequent trucks are referred to as followers, adjusting and responding to the leader's actions with the potential for little to no human interaction Castritius et al. (2020).

The motivation behind this analysis is rooted in the need to understand the potential impact of truck platooning systems in the Portuguese context. Introducing truck platooning systems will yield advantages, including improved road safety, reduced fuel consumption, and lower emissions. Nevertheless, the successful integration of this technology is contingent upon the endorsement and support of key stakeholders and all actors involved in this sector. Despite the benefits that truck platooning could provide, it is essential to understand users' requirements and acceptance, which is the key to ensuring a successful deployment in the real world. Some tests in more or less controlled environments have already been conducted, for instance, in Germany Castritius et al. (2020), but a full-scale implementation has yet to be carried out.

This research aims to understand the potential impact of platooning systems on various actors within the Portuguese context. It utilizes input from 11 focus groups and semi-structured interviews with key stakeholders, including truck drivers, logistics operators, and road authorities, conducted in previous project iterations. By analyzing this data through *Natural Language Processing* (NLP) techniques, the study seeks to extract valuable knowledge and compare it with existing international literature.

This study looks to address the following questions:

1. What are the impressions of different actors regarding truck platooning in the Portuguese context?

2. In what ways do these impressions diverge from or align with conclusions in international literature?

3. How can NLP methods be leveraged to widen our understanding of stakeholders' views and identify the potential benefits and challenges of implementing truck platooning systems?

The contributions of this work are multifaceted. First, it will provide a detailed analysis of stakeholders' perspectives, highlighting the benefits and challenges of platooning systems. Second, by applying NLP techniques, the study will offer a different approach to synthesizing qualitative data, making it easier to identify patterns and trends that might be overlooked through traditional methods. Third, the findings will contribute to the existing literature on platooning systems, offering specific insights into the Portuguese context, which can inform policymakers and industry stakeholders. Lastly, the study's outcomes can serve as a foundation for further research and development in autonomous vehicle technology and its integration into existing transportation systems.

1.3 Structure

This document is organized into 6 chapters. The present chapter briefly introduces this dissertation's motivation, objectives, and expected contributions. Chapter 2 presents concepts and contextualizes the problem, including a description of automated driving and vehicle communication protocols, introduces Truck Platooning and methodologies applied to assess the introduction of new technologies, and how we can use NLP as a tool as well. Chapter 3 describes the datasets and explains the first pre-processing step taken to standardize use and improve the results of the models in the first section. Chapter 4 presents the preprocessing stages performed in our datasets and the models to be applied. Chapter 5 presents and discusses the results obtained. Finally, chapter 6 closes this work with the main conclusions and present directions for future work.

Chapter 2

Literature Review

This chapter presents concepts and contextualizes the problem in light of the existing literature. The first section provides an overview of advances in automated driving and vehicle communications; the second part discusses truck platooning and some benefits and challenges; and the last section compiles relevant methodologies in introducing new technologies.

2.1 Advances in Automated Driving and Vehicle Communications

The automotive industry has progressed severely in automated driving technologies, vehicle communications, and safety protocols in the last decade. These advancements are stepping stones to the development of truck platooning systems. To some degree, automated driving technologies have revolutionized the automotive industry; in this scenario, several companies, i.e., Tesla and Uber, have developed advanced systems that depend on a synergy of sensor, machine learning, and artificial intelligence to assist and control vehicle with minimal human intervention (Katrakazas et al. (2015)). These systems comprehend a variety of features such as lane keeping, adaptive cruise control, and *advanced driver-assistance systems* (**ADAS**), which are core components to the automated driving concept (Litman (2020)).

To complement, *Vehicle-to-Everything* (**V2X**) protocols, which encompasses *Vehicle-to-Vehicle* (**V2V**) and *Vehicle-to-Infrastructure* (**V2I**) communication, Figure 2.1, has seen rapid development. These communications protocols provide an interface where vehicles can exchange information with each other and traffic infrastructure, amplifying the highway's awareness and enabling coordinated driving maneuvers.

Another important consideration is the impact of advancements in safety-related protocols and standards on the reliability and safety of automated driving systems. Many new vehicles now come standard with *advanced driver assistance systems* **(ADAS)** such as *automatic emergency braking* **(AEB)** and *electronic stability control* **(ESC)**. Additionally, policymakers and road regulators have implemented rigorous safety standards and testing protocols to ensure the safety of automated vehicles (Shladover (2018)).

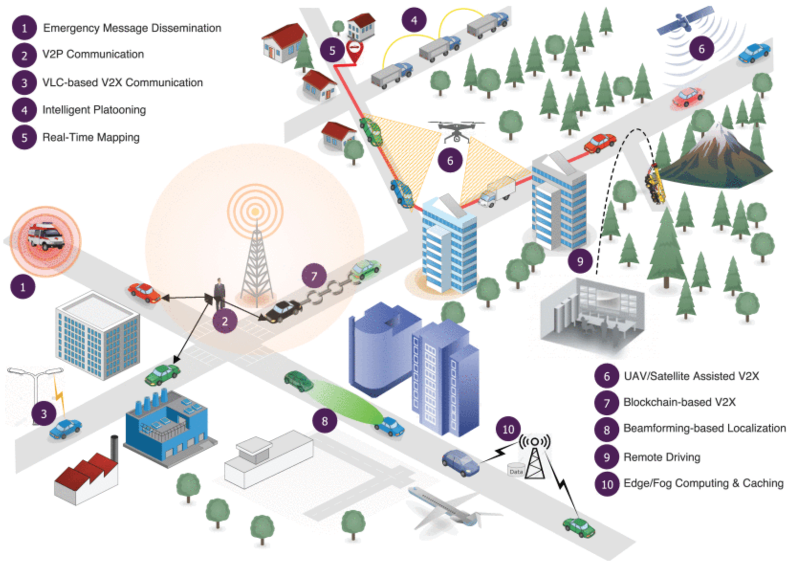

Figure 2.1: Overview of V2X communications Noor-A-Rahim et al. (2022).

Therefore, when discussing automated driving and its advancements, it is essential to understand that there are different levels of automation, ranging from L0 to L5 on the *SAE scale* (Society of Automotive Engineers), as noted in SAE International (2018). On this scale, L0 represents no automation, where the driver is always entirely controlled by the vehicle. For example, most vehicles already belong to L1, taking advantage of single-drive assistance systems, either on steering or acceleration/deceleration. L2, with **ADAS**, provides an interface that allows the driver to take off the hands from the steering wheel and also its feet from the pedals, pushing this responsibility to *dynamic driving task* (**DDT**), although, continuous monitoring of the road is required. In contrast, L3 systems can handle all aspects of driving in certain conditions and issue a *takeover request* (**TOR**) when it encounters situations it cannot manage. The driver must be available to take control but does not need to monitor the road constantly while the system is active. L4 provides full automation in specific driving conditions or environments, allowing the driver to disengage from the driving task, with the system capable of requesting human intervention if necessary, whereas L5 enables full automation under all driving conditions and environments, allowing the vehicle to perform all driving tasks autonomously without human intervention (Soares et al.).

Consequently, by introducing these advancements, we are laying the groundwork for a more in-depth conversation on truck platooning, which leverages these fundamental technologies to establish a transportation system that is both more efficient and safer.

2.2 Truck Platooning

Truck platooning (**TP**) is an emerging transportation technology where a group of trucks travels closely together in a convoy, emulating a "short train," using **ADAS** and vehicle communication protocols. This innovation aims to enhance fuel efficiency, reduce emissions, improve road safety, and optimize logistics operations (Bhoopalam et al. (2018); Lioris et al. (2017)).

Typically, in a **TP** setup, multiple trucks follow a "lead" truck in a tightly coordinated formation. The "lead" dictates the convoy's speed and direction, while the following trucks, equipped with technologies like *Cooperative Adaptive Cruise Control* (**CACC**), preserve a predefined and precise distance from one another and the lead truck. These distances are dynamically adjusted based on real-time data exchanges facilitated by **V2X** communication systems, ensuring synchronized responses to the "lead" truck's maneuvers (Figure 2.2).

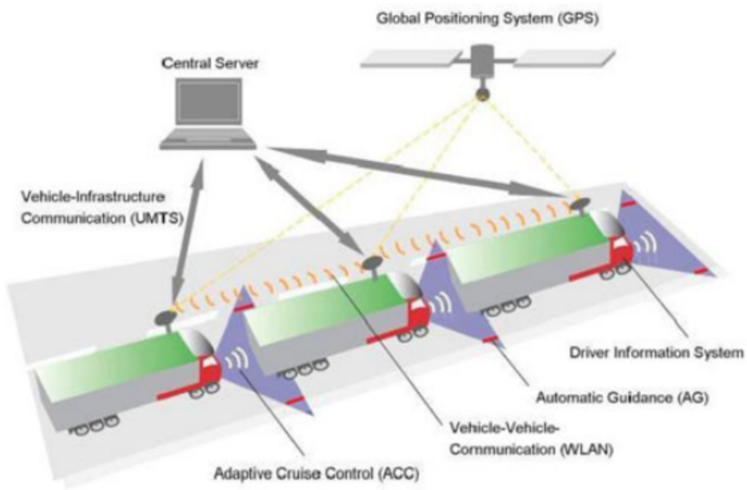

Figure 2.2: The KONVOI PLATOON SYSTEM Tsugawa et al. (2016)

Integrating advanced automated driving technologies and **V2X** communication minimizes inter-vehicle gaps and enhances overall safety and operational efficiency. For highly automated trucks (ETRAC (2019)), the gap between vehicles can be reduced to as low as 0.2 seconds, which at 80 km/h represents a distance of approximately 7 meters. This capability allows for closer coordination and responsiveness among vehicles in the platoon, contributing to substantial savings in fuel consumption and significant reductions in greenhouse gas emissions, particularly benefiting following vehicles due to improved aerodynamic drag characteristics (Bhoopalam et al. (2018); Kaiser et al. (2022)).

Hence, vehicles with this technology can operate with minimal human intervention, maintaining safe distances from each other, as described in PPMC (2021). The lead truck primarily drives the platoon in this setup, supported by **CACC** and continuous system monitoring Willemsen et al. (2021). However, in the first implementation stage, the leading position and its driver are expected to manually drive the vehicle, establishing a reference for other vehicles to follow. The following trucks in this convoy, referred to as followers, adjust and respond to the leader's actions, with little to no human interaction potential; it is worth noting that, still, its drivers are expected to continuously monitor their vehicles while in automated mode and be ready to take control over, when required.

Recent research, such as that by Simoes et al. (2022), highlights evolving considerations in truck platooning deployment. Their findings underscore drivers' shifting roles and responsibilities within platoons, from increased alertness demands for lead drivers to potential impacts on situational awareness among followers. Moreover, the adaptation of drivers' workspace to accommodate platooning systems offers opportunities for improved workflow and working conditions, including issues related to driver adaptation, regulatory compliance, and the need for infrastructure upgrades to support **V2X** communication and automated driving systems Fröhlich et al. (2018).

The potential benefits of truck platooning are numerous, including improved road safety, reduced fuel consumption, lower emissions, and the shortage of truck drive professionals (Bhoopalam et al. (2018); Kaiser et al. (2022); Shladover et al. (2015), Castritius et al. (2020)). However, the successful integration of this technology is contingent upon the endorsement and support of critical stakeholders, including truck drivers, logistics operators, road authorities, and road operators. Surveys conducted by the *European Transport Advisory Council* (**ETAC**) indicate a growing interest and cautious optimism among industry decision-makers towards embracing such transformative technologies for their potential to revolutionize freight transport efficiency and safety (ETRAC (2022)).

The deployment of Truck Platooning presents a promising yet complex landscape of benefits and challenges. Although it holds potential for fuel and greenhouse gas reductions, particularly for following vehicles benefiting from drag aerodynamics, challenges arise regarding the distribution of responsibilities between leading and following positions. Drivers should face distinct tasks and potential risks, from increased responsibility in leading positions to impacting alertness and situational awareness, Simoes et al. (2022). Notwithstanding, the adaptation of drivers' workplaces to accommodate platooning systems offers opportunities for improved workflow and working conditions but may conflict with existing regulations, Fröhlich et al. (2018).

Additionally, the technology's success hinges on acceptance from industry decision-makers, front-line users, and the general public (ETRAC (2019); ETRAC (2022)). As directives push for more sustainable and safe transport solutions, further studies on technology acceptance are imperative. Ultimately, the adoption and usage of truck platooning will depend on navigating these complexities and balancing its potential advantages and implementation challenges Pająk and Cyplik (2020).

2.3 Methodologies Applied to the Introduction of new technologies

Introducing new technologies has been a part of the human experience and is vital for economic growth. This phenomenon has revolutionized how businesses operate, the interaction between people and nations, and how societies function. Despite the importance of technological progress, developers face difficulties in identifying and measuring the initial creation of new technologies and the subsequent diffusion and impact of these new technologies. Particularly, one way of assessing and measuring the introductions of new technologies predominantly relies on patent statistics, which are broadly available across countries, industries, and time (Griliches (1990); Hall et al. (2001);Nagaoka et al. (2010)). Although one could argue that other methods could be used to assess the impact of new technologies, it is natural to view these innovations in different ways, such as the impact on the environment, the impact on society, the impact on the economy, and time perspectives.

Qualitative methodologies comprehensively understand the potential impact of new technologies and their adoption, user experience, and organizational dynamics (Saghafian et al. (2021), El-Haddadeh (2020)). Additionally to qualitative methods, using quantitative methods can provide a new layer of insights and, thus, a comprehensive understanding of the potential impact of new technologies (Oliveira and Martins (2010), Ghobakhloo et al. (2011), Chatterjee et al. (2021)). Hence, researchers might capture stakeholders' subjective perceptions, attitudes, and behaviors, enriching the analysis and providing valuable context for the decision-making process.

Halicka (2017), mention the importance of methods such as patent analysis and statistical and bibliometric analysis, which are crucial in evaluating the impact in the innovation landscape. Meanwhile, patent analysis enables businesses to assess the novelty and uniqueness of their ideas, providing valuable insights into the intellectual property landscape and competitive positioning. Statistical analysis, on the other hand, empowers organizations to analyze large datasets, uncover patterns, and derive actionable insights to inform decision-making processes.

Nonetheless, the introduction of new technologies in transport has been studied through different perspectives. The most adopted frameworks are the *Technology Acceptance Model* (**TAM**) (Davis (1985), Davis (1989)), the *Theory of Planned Behaviour* (**TPB**) (Ajzen (1991)) and *Unified Theory of Acceptance and Use of Technology* (**UTAUT**) (Venkatesh et al. (2003)). These models evaluate the preference and the actual use of technology by representing the components of behavioral preference and the influence relationships between these elements, alongside other mediating factors grounded in the psychology field.

The Technology Acceptance Model (TAM), developed by Davis (1989), suggests that the perceived ease of use and perceived usefulness are the main factors influencing technology acceptance (Figure 2.3). This model is particularly applicable in the field of transportation as it helps in evaluating how drivers perceive the effectiveness of new support systems, such as lane departure warnings and automatic braking systems, in improving their driving efficiency and safety. Understanding these perceptions can lead to more effective implementation strategies that are in line with user expectations and can improve overall adoption rates.

The Theory of Reasoned Action (TRA), formulated by Ajzen and Fishbein (1975), emphasizes the role of attitudes and subjective norms in shaping behavioral intentions. When applied to transport technology, TRA can help understand how drivers' attitudes towards new technologies and the influence of significant others (such as family, friends, or authority figures) shape their willingness to use driver support systems. This approach underscores the importance of addressing both individual attitudes and the broader social context in promoting the adoption of transport technologies.

As proposed by Ajzen (1991), the Theory of Planned Behavior (TPB) extends the Theory of

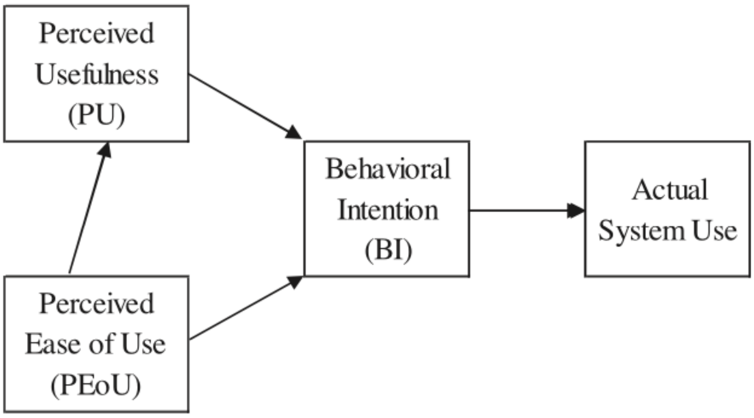

Figure 2.3: Technology Acceptance Model Davis (1989).

Reasoned Action (TRA) by including perceived behavioral control as an additional factor influencing behavioral intentions and actual behavior. In the context of driver support systems, TPB can elucidate how drivers' perceived control over using these technologies, influenced by their past experiences and the availability of support, affects their acceptance and usage, Figure 2.4. For instance, if drivers feel confident using a new speed adaptation system, they are more likely to adopt it. This model also highlights the importance of subjective norms aligning with UTAUT's social influence construct, further validating its applicability in transport technology studies.

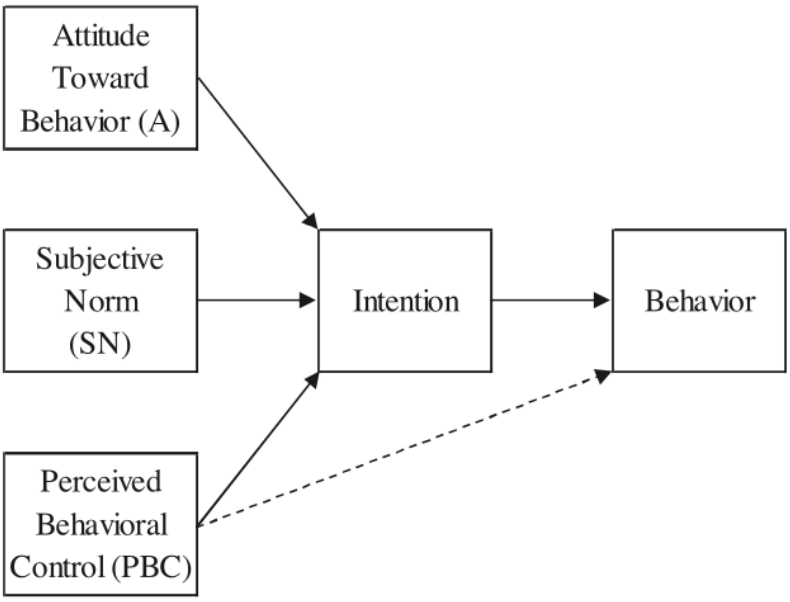

Figure 2.4: Theory of Planned Behavior Ajzen (1991).

The Unified Theory of Acceptance and Use of Technology (UTAUT) has been effectively applied to study the acceptance of technology within the transport industry, particularly in the context of driver support systems. Venkatesh et al. (2003) consolidated the eight most significant models of individual acceptance into the comprehensive UTAUT model, see Figure 2.5. According to Adell (2009), driver support systems such as Intelligent Speed Adaptation (ISA) and Forward Collision Warning (FCW) are crucial for enhancing traffic safety. However, their success largely depends on drivers' acceptance and use of these technologies. UTAUT provides a comprehensive framework for understanding this acceptance by considering factors like performance expectancy, effort expectancy, social influence, and facilitating conditions.

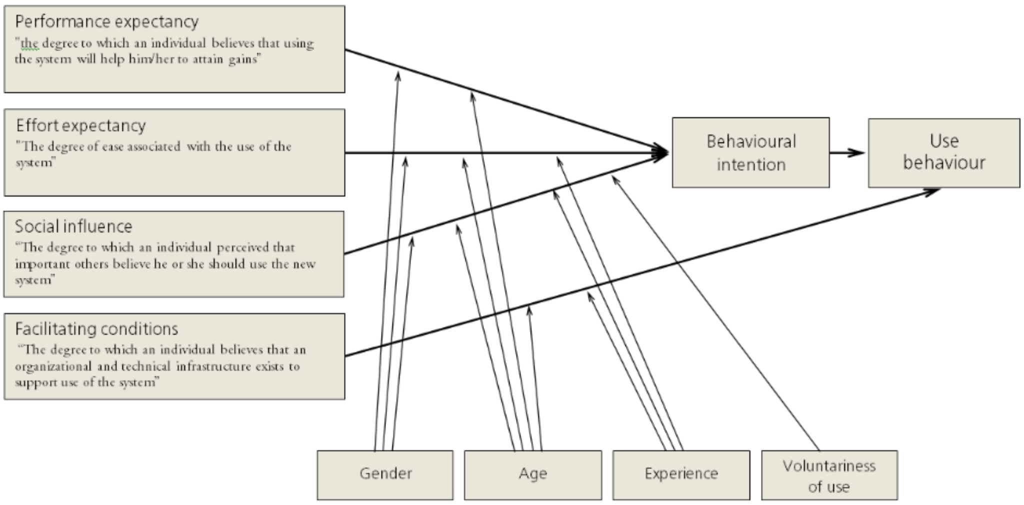

Figure 2.5: The Unified Theory of Acceptance and Use of Technology Venkatesh et al. (2003).

Statistical methods play a crucial role in validating the output of previously described frameworks. We have structural equation modeling (SEM), hierarchical regression, and cluster analysis among the most utilized methods. SEM is especially used in analysis where the variables are latent, which provides a way to validate the complexity of the framework of technology acceptance (Bagozzi and Yi (2012)).

Other works have applied hierarchical regression to research the potential influence of contextual variables, i.e., technology confidence or safety, that could influence the intent to use or adopt it (Venkatesh et al. (2003)). Additionally, cluster analysis allows the development of various personalized strategies to engage, promote, and accept new technologies more efficiently (Hair et al. (2010)).

In summary, as discussed by Geels (2012), introducing new technologies should be evaluated more sustainably. The transition to a more sustainable society should be viewed as a process of co-evolution involving technologies, infrastructures, markets, and user practices. This process is influenced by the interaction of various actors, including governments, companies, and society. Integrating multiple frameworks like **UTAUT**, **TAM**, **TPB**, and **TRA** provides a robust framework for understanding the complexities of technology acceptance, ensuring that innovations are effectively adopted and utilized for maximum societal benefit.

2.4 Natural Language Processing Applied to the Introduction of New Technologies

In recent decades, *Text Mining* and *Natural Language Processing* (**NLP**) fields have gained significant importance due to the exponential growth of digital information. Text mining involves extracting insights and knowledge from large datasets, using techniques from data mining, machine learning, and statistics, Manning and Schütze (1999). In contrast, **NLP** aims to enable computers to understand, interpret, and generate human language meaningfully.

The popularization of the internet, digital artifacts (text, audio, video, and pictures), and advancements in artificial intelligence and its sub-fields have contributed to a rise in the volume of text data available. This increase is not only in quantity but also in the variety and velocity at which this data is produced. Consequently, the challenge and necessity of effectively processing and analyzing this data have become paramount Martinez and Herrera (2020).

Moreover, the miniaturization of processors and advancements in computer technologies still play a pivotal role in developing and enhancing **NLP** techniques. These technological advancements have significantly expanded computational power, driving the processing of large datasets and complex algorithms that were previously infeasible. Likewise, the increased processing capabilities have facilitated the development of sophisticated **NLP** models and algorithms, allowing for more accurate and efficient analysis of textual data.

NLP methods provide several sophisticated techniques that are valuable in contrasting stakeholders' perceptions from interviews with the existing literature. One key method is sentiment analysis, which involves assessing the sentiments expressed in textual data Jurafsky and Martin (2020). By analyzing interview transcriptions, one can gauge all actors' attitudes, opinions, and concerns regarding truck platooning. This initial analysis can then be contrasted with the existing literature to comprehensively map the main drivers and barriers for truck platooning technology acceptance and adoption on a large scale, thus obtaining insights from different actor groups.

Additionally, methods such as topic modeling, which is a type of statistical model for discovering the abstract "topics" that occur in a collection of documents Blei (2012), can be used to identify the main topics referred to by stakeholders, their sentiments and/or arguments towards the benefits and risks of the technology, and the relationships between actors.

As noted by Mihalcea and Tarau (2004), summarization techniques can condense lengthy transcriptions and academic articles into concise summaries, highlighting key points and facilitating comparative analysis. By systematically comparing summarized stakeholder perspectives with synthesized summaries of relevant literature, one can identify overarching themes, discrepancies, and areas requiring further investigation.

Several key studies have demonstrated the effectiveness of text mining in understanding technology acceptance. For instance, Chen et al. (2019) used text mining to analyze scientific literature on Convolutional Neural Networks, uncovering trends and key research areas. Similarly, Manning and Schütze (1999) discussed foundational techniques in statistical NLP that are crucial for text mining applications.

NLP methods have impacted several industries, including healthcare. Zhou et al. (2022) presents a panoply of different applications in this context, such as clinical practice, where NLP is used to assist doctors in communicating with patients in other languages. Public health, where these techniques could identify populations with higher health risk factors, could improve population screening, among others.

In the context of automated truck platooning, text mining can help identify the main barriers to acceptance, such as safety concerns, regulatory issues, and economic implications. It can also highlight the benefits perceived by different stakeholders, such as improved efficiency, cost savings, and environmental benefits. Despite its potential, text mining has not been extensively used in this area.

In summary, **NLP** offers several tools and methodologies valuable for contrasting stakeholder perceptions with the existing literature in truck platooning. Leveraging these techniques can extract insights into the alignment or disparity between stakeholder perspectives and scholarly discourse. This integrated approach enhances our understanding of the multifaceted considerations surrounding truck platooning, informing future research, policy-making, and implementation strategies.

2.5 Summary

In this chapter, we covered the basic concepts and put the issue of truck platooning into context. We started by looking at the improvement in automated driving and vehicle communications, emphasizing the importance of these technologies in developing truck platooning systems. We then discussed the benefits and challenges of truck platooning, highlighting its potential to improve fuel efficiency, reduce emissions, and enhance road safety. We also discussed ways to assess technology acceptance, including **TAM**, **TPB**, and **UTAUT** methodologies, and the benefits of using Natural Language Processing (NLP) to analyze stakeholder opinions and academic discussions to evaluate technology acceptance from other lenses. This review establishes a theoretical basis for understanding the factors that affect the adoption and implementation of truck platooning technologies.

Chapter 3

Data

This chapter describes the dataset used in our experimental setups and explains the essential steps required to preprocess the data to create the final dataset for our analysis.

Our dataset is derived from a comprehensive series of interviews conducted with various stakeholders, including truck drivers, logistics operators, road authorities, and road operators. These interviews provide invaluable insights into the practical and operational aspects of truck platooning. Additionally, we have incorporated articles from the scientific community to assess and validate our findings, ensuring a robust and well-rounded analysis.

We conducted a preprocessing stage to transform these diverse sources of information into a suitable input dataset. This involved converting interview transcripts and scientific articles into a standardized format, ensuring consistency and coherence across the dataset.

This section outlines the data collection process and the first preprocessing stage employed in our initial dataset, which later will provide standardized datasets on which our analysis will built.

3.1 Datasets

For the purpose of this study, four datasets were developed to encompass stakeholders from the Portuguese Truck Platooning context. As part of the TRAIN project, different actors were called to participate through focus groups and interviews to gather data that supported the design of the actor-network map. In total, 11 focus groups and semi-structured interviews were conducted (Table 3.1).

As stated in Duarte et al. (2023), the focus groups included diverse stakeholders, such as truck drivers, logistics operators, road authorities, and road operators. Given the interest in safety and working conditions, drivers had the highest representation, with 30 male participants. Practitioners noted a need for more women in freight transport driving roles than in passenger transport. Logistics companies were represented by 16 assistants and managers responsible for tasks impacted by technological changes in driving activities. Additionally, two regulating entities and three road operators facilitated implementation discussions.

Neto et al. (2024) detailed the procedure used in the interviews with the truck drivers' focus groups. The sessions occurred on December 12, 2022, January 6, and January 25, 2023, in a meeting room on the companies' premises. With a duration of 1h30 minutes each, one of them was carried out in the morning (10:30 am - 12:00 pm) and the two others in the afternoon (2:30 pm - 4:00 pm). On each Focus Group session, one moderator was directing the session, and two assistants were present to collect relevant or more detailed information, completing the required transcriptions. Each session aimed to address four main topics:

1. **Own Representations of Automated Driving Systems**: Understanding the participants' perceptions and attitudes towards automated driving systems.

2. **Impacts of Truck Platooning Implementation on Own Activity**: Exploring how the introduction of truck platooning would affect the participants' own work and operations.

3. **Impacts of Truck Platooning Implementation on Others' Activity**: Considering the broader implications of truck platooning on other stakeholders and the overall logistics ecosystem.

4. **Expectations of Road Freight Transport for the Near Future**: Discussing future projections and expectations for road freight transport in the context of technological advancements.

Table 3.1: Focus group and interviews distribution by groups of actors. Duarte et al. (2023)

Actors	**Focus group / interviews (quantity)**
Drivers	focus group (4)
Logistic companies	focus group (2)
Road operators	focus group (2), interview (1)
Regulators	interview (2)

To ensure the reliability and validity of the collected data, the moderators remained neutral during these interactions, consciously avoiding any bias that could influence the participants' responses. Nevertheless, the moderators showed a short video describing the truck platooning concept and operation [1] and clarified participants' doubts regarding the study (Neto et al. (2024)). The discussions with the logistics companies, road operators, and regulators have generically revolved around the same four main topics and followed similar procedures, with slight adaptations related to the different roles of the interviewees.

3.1.1 Scientific Community

In addition to the interview datasets, we leveraged the literature review by Lourenço et al. (2024), also conducted under the **TRAIN project**, to gain deeper insights into the international scientific

[1] https://www.youtube.com/watch?v=X7vziDnNXEY

community's perspectives on truck platooning. Lourenço et al. (2024) reviewed existent literature to identify and analyze how truck platooning technology is represented in previous studies and to uncover acceptance levels among various stakeholder groups. This literature review involved searching titles, abstracts, and keywords in the Scopus, Web of Science, and TRID databases, with search results retrieved on July 28th, 2022.

As a result of this literature review, 35 studies presented that experience played an important role in the representations developed by participants. Thus, these studies provided diverse perspectives and were grouped by type of platooning experience included, namely:

1. **Discussions with No Platooning Experiences**: Papers that explore theoretical and speculative aspects of truck platooning without practical implementation.
2. **Simulated Experiences**: Research based on simulations that model the potential impacts and benefits of truck platooning under various scenarios.
3. **On-the-Road Experiences**: Studies that report on real-world implementations and trials of truck platooning, offering empirical data and insights.

Lourenço et al. (2024), has observed that the papers in the first category (18 studies) rely on self-reported measures, such as interviews and surveys. However, literature search and observation methodologies are employed to understand the different perspectives. Also, the analysis and approaches differ considerably, yet others provide models or scenario projections based on quantitative analyses.

The studies involving simulated platooning experiences included 11 studies, one using a paper prototyping setup and the rest being conducted in a driving simulator. These studies explored human-machine interaction concepts and tested specific platooning driving scenarios by assessing users' preferences and performance in a simulated environment. Thus, several physiological, behavioral, and self-report metrics were employed to measure drivers' state, driving behavior, and acceptance.

In contrast to the previous categories, the final category examined truck driver acceptance of platooning following a real-world experience driving in a platoon. This group shifted from analyzing general driving behavior and distance metrics to understanding the drivers' direct experience. Here, interviews and self-report questionnaires were used to evaluate how drivers felt about driving in a platoon and their overall acceptance of the technology.

This blend of firsthand stakeholder input and rigorous academic research forms the basis of our dataset, enabling a thorough and multi-faceted analysis of the factors influencing truck platooning deployment and its potential impacts on the road freight transport sector.

3.2 Dataset Transcription

Our data comprises two sets: (1) audio and video files from various stakeholders in the Portuguese scenario and (2) academic papers that provide insights into the scientific community's perspectives in different setups. The datasets were available in different formats, including audio and video files from focus group sessions and PDF files from scientific papers. Thus, to ensure compatibility with the two distinct models developed, we performed a transcription of the interviews (*drivers, logistic companies, and road regulators and operators*) and standardized the second set.

The first set consists of audio and video recordings from focus group sessions involving essential stakeholders in the Portuguese context. The interviews were conducted in Portuguese and transcribed for analysis. Initially, manual transcription was considered and performed in some interviews; however, this approach was impractical due to the large volume and length of the recordings.

The selected files were transcribed in two stages by different transcribers. The transcriptions of the six focus groups with the Drivers and Logistics Operators were made manually prior to this research without any specific tool (Neto et al. (2024)).

In the second stage, performed under this dissertation work, we used GoodTape. However, we also manually inspected the transcriptions to correct any occasional mistakes in transcribing Portuguese terms and sentences. This step was important in assuring the transcriptions' accuracy and dependability, ensuring they were suitable for input into our models.

In the second set, to standardize the PDF files of the scientific papers for analysis, we implemented a Python script to scrape the textual data and save it as plain text files. However, due to digital rights protections on 10 of the papers, we could only successfully scrape textual data from 25 papers. These text files form the basis for our analysis of the scientific literature on truck platooning.

Normalizing the various data sources was necessary during our data transcription and preparation process. Transcriptions from stakeholder interviews were converted into uniform text files to ensure consistency in format and structure. Similarly, the text files obtained from scraped scientific papers were organized and formatted consistently to simplify their integration into our analytical pipeline.

We prepare our datasets to establish a solid basis for the experimental study, ensuring that our dataset have the proper format and structure to provide good inputs to our model.

3.3 Summary

In this chapter, we outlined the dataset used for our experimental setups, detailing the data collection and preprocessing steps to create a final dataset for analysis. Our dataset originates from interviews with stakeholders such as truck drivers, logistics operators, road authorities, and road operators, alongside scientific articles validating our findings. We described the development of four datasets, focus groups with key actors, and a literature review by Lourenço et al. (2024) for a comprehensive understanding of truck platooning.

The datasets comprised audio, video files, and academic papers, all transcribed and standardized for analysis. This process ensured compatibility and consistency, establishing a solid basis for our experimental study and facilitating the integration of diverse data sources into our analytical pipeline.

Chapter 4

Methods

The analysis of technology acceptance is a complex field that could significantly benefit from text mining techniques, especially when dealing with datasets derived for interview transcriptions and textual data from documents. Using NLP methods and topic modeling, such as Latent Dirichlet Allocation (LDA), allows for an understanding of individuals' and groups' perceptions and attitudes toward emerging technologies.

This chapter will present the methods used to analyze the acceptance of truck platooning from a dataset of interview transcripts. The data preprocessing procedures, including tokenization and stopword removal, and the applied analysis strategies, such as NLP and LDA, will be discussed.

4.1 Data Preprocessing

Data preprocessing is a crucial step in NLP, transforming unstructured text into a structured format suitable for analysis and modeling. This process is essential for enhancing the quality and consistency of the data, ultimately improving the performance of NLP models.

In our project, we aim to preprocess interview datasets to ensure both datasets' privacy, confidentiality, and data quality. The preprocessing pipeline involves various techniques to extract and prepare text data effectively. Initially, we anonymized all personnel mentioned in the interview datasets to protect privacy using simple anonymization techniques.

Subsequent steps focus on refining the content of the datasets, transforming raw data into a clean, consistent, and relevant format. Basic text cleaning includes converting text to lowercase and removing punctuation and non-alphanumeric values. Advanced techniques such as tokenization, stopword removal, normalization (stemming and lemmatization), part-of-speech (POS) tagging, named entity recognition (NER), and handling out-of-vocabulary (OOV) words could be applied to refine the text further.

Leveraging programming languages like Python and R is beneficial due to their robust environments and supportive communities. In R, one may utilize several packages from the *tidyverse* collection, along with specialized packages such as *udpipe*, *tidytext*, and *quanteda*. These tools enable efficient preprocessing and data transformation into a format suitable for further analysis and modeling.

Our preprocessing pipeline includes the following tasks:

1. **Text Cleaning**: Removing words that start or end with numerals, words with fewer than three characters, all punctuation, non-alphanumeric characters, and hyphens between words.

2. **Tokenization**: Splitting text into individual tokens or words for easier analysis.

3. **Stopword Removal**: Eliminating common stopwords in both Portuguese and English to focus on meaningful words.

4. **Normalization**: Applying stemming and lemmatization to reduce words to their base forms.

In summary, by cleaning and transforming the text, we aimed to create a high-quality dataset that enhances the performance of our NLP models. The specific implementation steps, challenges encountered, and the impact of preprocessing on our data quality are detailed in the following sections.

4.1.1 Tokenization

The exact definition of a token is subject to variation and interpretation, as highlighted by Jurafsky and Martin (2021). Different linguists provide manifold explanations based on the study domain and the corpora application. One widely accepted definition of tokenization, described by Schütze (1993), is segmenting text into alphanumeric sequences, separated by spaces, which may include hyphens and apostrophes but exclude punctuation marks.

Jurafsky and Martin (2021) state that numerous challenges arise in tokenizing a corpus. One primary issue is the composition of tokens, where words are frequently followed by punctuation without an intervening space, complicating the tokenization process.

Another significant challenge is the phenomenon known as haplology, where the same punctuation mark has different meanings depending on the context. For instance, the period ('.') can denote the end of a sentence or a decimal point, such as in monetary values (e.g., USD 100.00). Similarly, the hyphen ('-') can indicate a compound word (e.g., "long-term") or syllable separation, which is prevalent in Portuguese. In Portuguese, hyphens also join object pronouns to the nouns they refer to, such as in the phrase "bem-vindo" (welcome), Table 4.1.

Table 4.1: Examples of Word and Sentence Tokenization

Original Text	**Word Tokenization**	**Sentence Tokenization**
Autonomous and connected vehicles have promising potential to significantly impact almost all aspects.	`['Autonomous', 'and', 'connected', 'vehicles', 'have', 'promising', 'potential', 'to', 'significantly', 'impact', 'almost', 'all', 'aspects', '.']`	`['Autonomous and connected vehicles have promising potential to significantly impact almost all aspects.', 'Cooperative adaptive cruise control is an extension of adaptive cruise control that incorporates dedicated short-range communications to enable wireless vehicle-to-vehicle communications.']`

To address these issues, we utilized the `unnest_tokens` function from the `tidytext` package in R. This function includes algorithms designed to recognize patterns such as parenthesized expressions and sub-string divisions. This approach, different from traditional tokenization, does not rely solely on spaces or punctuation but adds a layer that utilizes context-independent criteria to ensure precise and efficient word tokenization as described by Silge and Robinson (2016).

4.1.2 Stop-words Removal

Stopwords are words that do not carry significant meaning and are usually filtered out during the initial text analysis phase. According to Manning et al. (2008), these words are very common in a language but provide little helpful information for text mining. Words like articles, conjunctions, and prepositions frequently appear in various documents but do little to distinguish content or aid in information retrieval. Therefore, we performed a stop-word removal process to eliminate these terms from our corpus, ensuring that only the most relevant and informative words remain for analysis.

As Moens (2001) recommended, we used stopword removal in our study to eliminate words that do not significantly contribute to the classification of legal texts. To assist with this process, we utilized the stopwords and quanteda packages, which contain a standard set of Portuguese and English stopwords. This method enabled us to identify and remove these words from our corpus efficiently.

An additional benefit of using stopwords is that they help reduce the size of the final vocabulary, thereby improving algorithm performance. As stated by Manning and Schütze (1999), using a simplified list of words can cut the size of the inverted index in half.

We show an example of an application where we remove stopwords in sentences in Tables 4.2 and 4.3

Table 4.2: Sentences before and after stopword removal - in English.

Original Sentences	**Altered Sentences**
Autonomous and connected vehicles have promising potential to impact almost all aspects dramatically.	Autonomous connected vehicles promising potential dramatically impact almost all aspects.
Cooperative adaptive cruise control is an extension of adaptive cruise control by incorporating dedicated short-range communications to enable wireless vehicle-to-vehicle communications.	Cooperative adaptive cruise control extension adaptive cruise control incorporating dedicated short-range communications enable wireless vehicle-to-vehicle communications.

Table 4.3: Sentences before and after stopword removal—in Portuguese.

Original Sentences	**Altered Sentences**
A regulação e a abertura também dos operadores em começarem a fazer as coisas em conjunto.	Regulação abertura operadores começarem fazer coisas conjunto.
Acho que também não avançou muito a comunicação de infraestrutura para veículos neste caso se calhar para os produtores também não têm, porque cada um tem os seus protocolos.	Acho avançou comunicação infraestrutura veículos neste caso se calhar produtores têm, cada um seus protocolos.

In conclusion, eliminating stopwords is a key text analysis preprocessing step. By removing words with high frequency but low information content, we reduced the dimensionality of our dataset without sacrificing its vital substance. This optimization increased computing efficiency and simplified the process of identifying significant connections and trends in the documents.

4.1.3 Normalization

One essential text preparation technique is normalization. Normalization methods generally lower a word's inflectional and occasionally derivatively related forms to a common base form.

There are two methods in natural language processing for breaking down words into their base or root forms, and each has a distinct outcome. Removing suffixes from words allows them to be shortened to their root form or stemmed. For instance, "run" can be used to replace the terms "running," "ran," and "runner." However, stemming frequently yields stems that might not even be words. On the other hand, lemmatization considers a word's context and part of speech and reduces it to its primary or dictionary form (lemma). For instance, "running" turns into "run," "ran" turns into "run," while "runner" remains "runner." Tables 4.4 and 4.5.

Words	Stemming	Lemmatization
Improve	Improv	Improve
Improving		
Improved		
Improvement		

Table 4.4: Example of words, their stemms, and lemmas in English. Adaptation Jacob Murel (2023)

Words	Stemming	Lemmatization
Melhorar	Melhor	Melhorar
Melhorando		
Melhorado		
Melhoria		

Table 4.5: Example of words, their stemms, and lemmas (Portuguese). Adaptation Jacob Murel (2023)

In the text normalization process, lemmatization and stemming are both valuable strategies. The NLP task's particular needs will determine which of them to use. Lemmatization, for example, is usually the better choice for tasks that demand high linguistic accuracy and where word context is essential. When computing efficiency is more important than accuracy, stemming is useful.

Furthermore, the language being processed might have an impact on the decision. Because the morphological structures of other languages—like English—are comparatively simpler, stemming works better in these cases. More complex morphological languages, such as Portuguese or Finnish, might profit more from lemmatization (Balakrishnan and Lloyd-Yemoh, 2014).

In conclusion, knowing the benefits and limitations of lemmatization and stemming allows for improved decision-making across the text preparation pipeline, improving the performance of NLP models and applications. The trade-offs between computing efficiency and linguistic correctness can be carefully considered when choosing the proper technique, in our case, lemmatization, resulting in text data that is better prepared for additional analysis and modeling.

4.2 Latent Dirichlet Allocation (LDA)

LDA is a generative probabilistic model commonly used for topic modeling in NLP. The goal of LDA models, which function as three-level hierarchical Bayesian models, is to group unlabeled documents into representative and cohesive topics (Blei et al. (2003)). LDA models assume that a corpus, or group of documents, contains latent themes or topics, each comprising terms that occur in multiple documents (Chu et al. (2020)).

There are three essential steps to generate a LDA model. First, the total number of topics must be determined. Next, each term in every document must have a subject assigned to it. Next, until convergence is reached, fresh subjects are interactively allocated to each term using a sampling process. The final step involves creating a topic-word co-occurrence frequency matrix, which serves as the finalized LDA model and summarises the learned associations between topics and words.

As a result, LDA generates probabilities for words within themes and topics inside documents. These probabilities produce distributions that allow us to identify the most representative phrases for each topic and determine the probability that documents align with specific topics (Karami et al. (2020)). Furthermore, LDA enables the extraction of prominent topics throughout an entire corpus, providing insights into the dominant themes that pervade the collection of texts.

Finally, several studies have proved the success of LDA in extracting key topics from large textual datasets. For example, Blei et al. (2003) proposed the LDA model and showed how it could be used to identify scientific themes. Griffiths and Steyvers (2004) used LDA to model the evolution of subjects in academic writing.

4.3 Parameter tunning: Optimal number of topics

It is essential to consider that the dimension of the corpus impacts the number of optimal topics in a topic modeling, which affects the model's ability to generalize. The literature usually defines the number of topics between 2 and 100 or an interval within this range (Karami et al. (2020). However, other studies define different intervals but always in the interval set by Karami et al. (2020); for instance, Chu et al. (2020) consider a range between 20 to 80 topics.

The interval used in this work ranged from 2 to 20 topics. A model was implemented for each number of topics between this interval, and its performance was then analyzed using four metrics: *Griffiths2004*, *CaoJuan2009*, *Arun2010*, and *Deveaud2014*. To determine the optimal number of topics, *Griffiths2004* and *Deveaud2014* should be maximized, while *CaoJuan2009* and *Arun2010* should be minimized (Griffiths and Steyvers (2004); Cao et al. (2009); Arun et al. (2010); Deveaud et al. (2014)).

This process used the R environment's "*ldatuning*" library. Nevertheless, it is worth noting that each metric measures different things and could provide a distinct number of topics. However, when choosing the number of topics to adopt, it is important to consider the interpretability of the results.

4.4 Frequency Analysis

Frequency analysis examines the frequency of words or phrases in a text corpus to determine the most popular terms. This method is essential in text mining and NLP for determining the importance and relevance of terms within a document or group of documents. It gives a direct method for identifying patterns, trends, and significant phrases.

This method is used for several applications, including content analysis, sentiment analysis, and information retrieval. Researchers can use phrase frequency analysis to detect dominating topics, measure the sentiment of a text, and improve search engine efficacy. For example, in sentiment analysis, the frequency of positive and negative words might represent the overall sentiment of a text or a collection of documents.

The required steps in frequency analysis are (1) preprocessing text data, including tokenization, stopword removal, and stemming/lemmatization; (2) counting the frequency of each term in the text corpus; and (3) analyzing frequency distributions to locate frequent phrases.

Many studies have used frequency analysis to extract meaning from textual data. For example, by highlighting frequently used phrases, Manning et al. (2008) improved information retrieval systems through frequency analysis [5]. Turney (2002) conducted a study using machine learning methods with frequency analysis to achieve sentiment classification.

4.5 Summary

This chapter covers many methods and techniques, focusing on preprocessing, frequency analysis, and LDA. These preprocessing methods, including tokenization, stopword removal, stemming, and lemmatization, will be used to analyze our dataset in the following phases.

Chapter 5

Results and Discussion

The first section begins with an overview of the dataset sizes before and after the preprocessing stage. It is followed by presenting the base LDA model and hyperparameter tuning processes. Finally, the chapter concludes with the results from the frequency analysis and topic modeling, providing insights into the thematic structure of the datasets.

5.1 Data Preprocessing

As previously described, we have five datasets: four from several stakeholders in the Portuguese context (truck drivers, logistics companies, road authorities, and road operators) and one derived from the literature review. Thus, as described in chapter 4, our pipeline is composed of different stages, i.e., text cleaning, normalization, and tokenization, and all of those processes are performed to obtain a dataset that is in the right format to be input into our models.

However, this process significantly reduces our dataset, as demonstrated in Table 5.1. After preprocessing, the size of datasets (the number of tokens) is reduced by around 45%, even more, if only unique tokens are considered.

We apply the *lemma* rather than the *stem* of the words in both English and Portuguese corpora to increase interpretability and get reliable insights. For each corpus, we investigate, present its relevant characteristics, and then develop the LDA model to extract its main topics.

Table 5.1: Corpus sizes before and after preprocessing

Corpus	**Number of Tokens**		
	Before Preprocessing	**After Preprocessing**	**Unique Tokens**
Road operators and regulators	23357	11039	4505
Drivers	24355	10407	3745
Logistic companies	11436	5036	1973
Scientific literature	138358	60881	17446

The next stage involves applying hyperparameter tuning to determine the optimal number of topics in each data set. This stage is essential to guarantee that the topic models accurately capture the underlying themes within each dataset, thereby delivering meaningful insights into the textual data. By optimizing the number of topics, we can enhance the model's performance and ensure that it effectively represents the diverse perspectives and nuances present in the data from each stakeholder group.

5.2 Results

To further understand our corpora collection, we performed frequency analysis, which involved extracting the overall distribution of word frequency, selecting the most common terms, and expressing them in a word cloud. This gave us a summary of the corpus characteristics, such as the texts' vocabulary diversity.

5.2.1 Frequency Analysis

Figure 5.1 illustrates the most frequent words and the number of times they appear in the scientific literature. Analyzing the frequency of tokens such as "truck," "driver," and "platoon" reflects the central subject matter of the literature, which revolves around vehicles and driving behaviors, particularly in the context of trucking and logistics, suggesting that the scientific community focus on its practical and operational aspects and how this technology affects the road environment. Thus, we perceived the weight academia put on operational challenges for truck platooning.

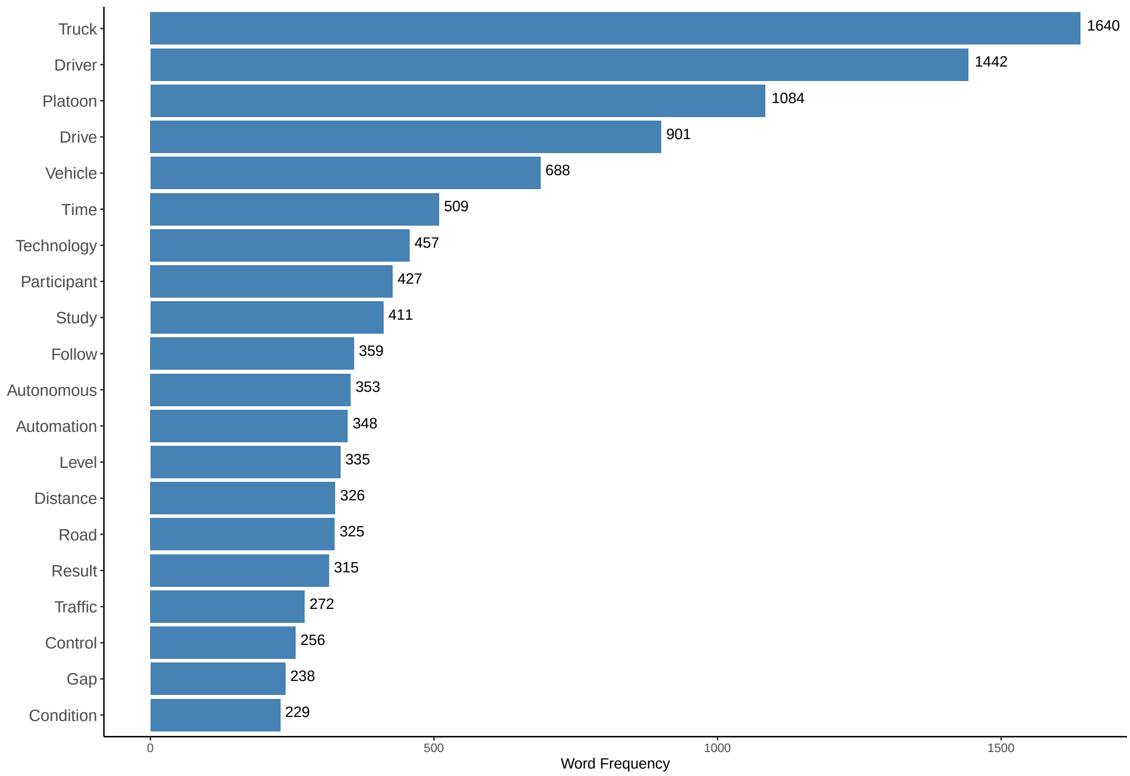

Figure 5.1: Frequency of the 20 most frequent words (scientific literature)

The logistics companies prioritize operational aspects and business-related topics, such as "*cliente*," "*empresa*," "*condução*," "*camião*," and "*motorista*." Additionally, there is an evident emphasis on strategic planning and tactical decisions related to logistics and transportation. The large number of operational and action-oriented terminologies suggests discussions of decision-making centered on practical activities and strategies involving several stakeholders (Figure 5.2).

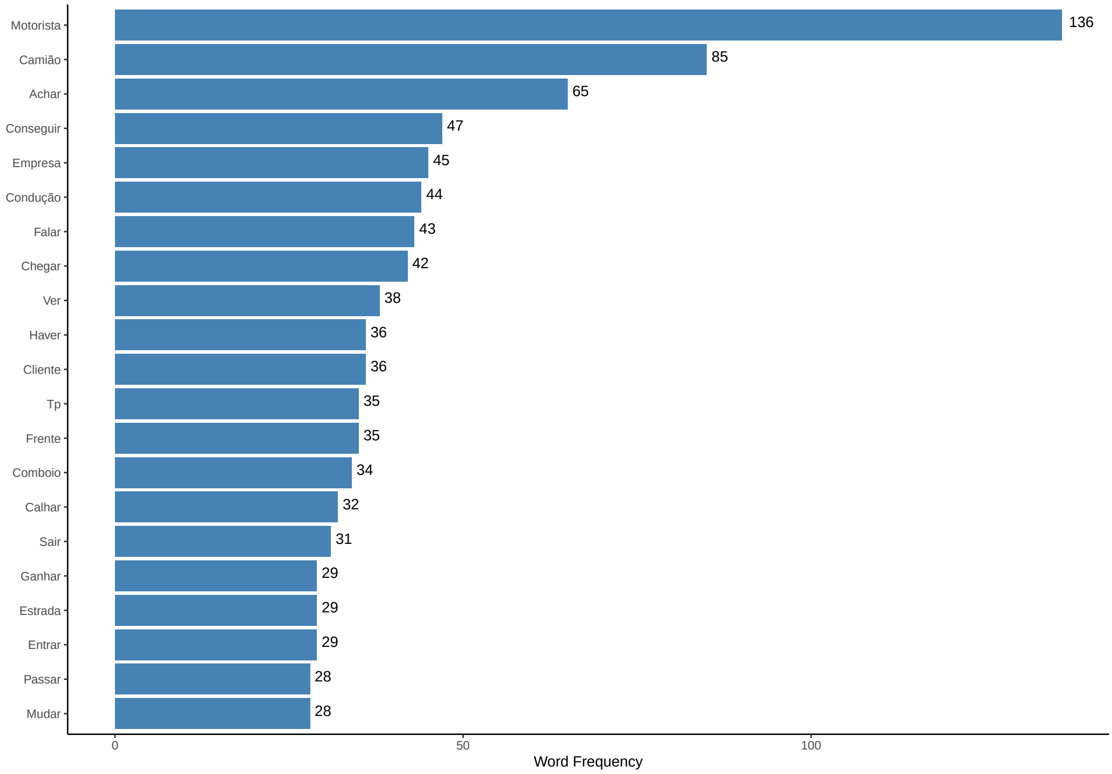

Figure 5.2: Frequency of the 20 most frequent words (logistics companies)

The drivers' corpus contains a high frequency of operational and action-oriented tokens, indicating that conversations center on real-world examples, firsthand experiences, and the operational environment in which drivers operate (Figure 5.3). This covers dealing with businesses, overseeing services, moving cargo, and operating vehicles—all essential daily tasks for drivers. Moreover, it highlights physical and situational contexts within which drivers operate, including road conditions, unexpected events, specific cases or incidents, and positional awareness.

The prominence of terms related to specific cases, technological advancements, and infrastructure projects indicates a comprehensive approach to managing and regulating road use and development. The mention of safety and operational terms underscores the regulators' focus on ensuring roadways' safe and efficient operation. Additionally, the inclusion of truck platooning highlights the emerging regulatory challenges and opportunities associated with advanced transportation technologies (Figure 5.4).

The 50 most frequent words are shown in word clouds (5.5; 5.7; 5.6; 5.8). In each word cloud, words were grouped by frequency level, and each group had its own color. The most frequent

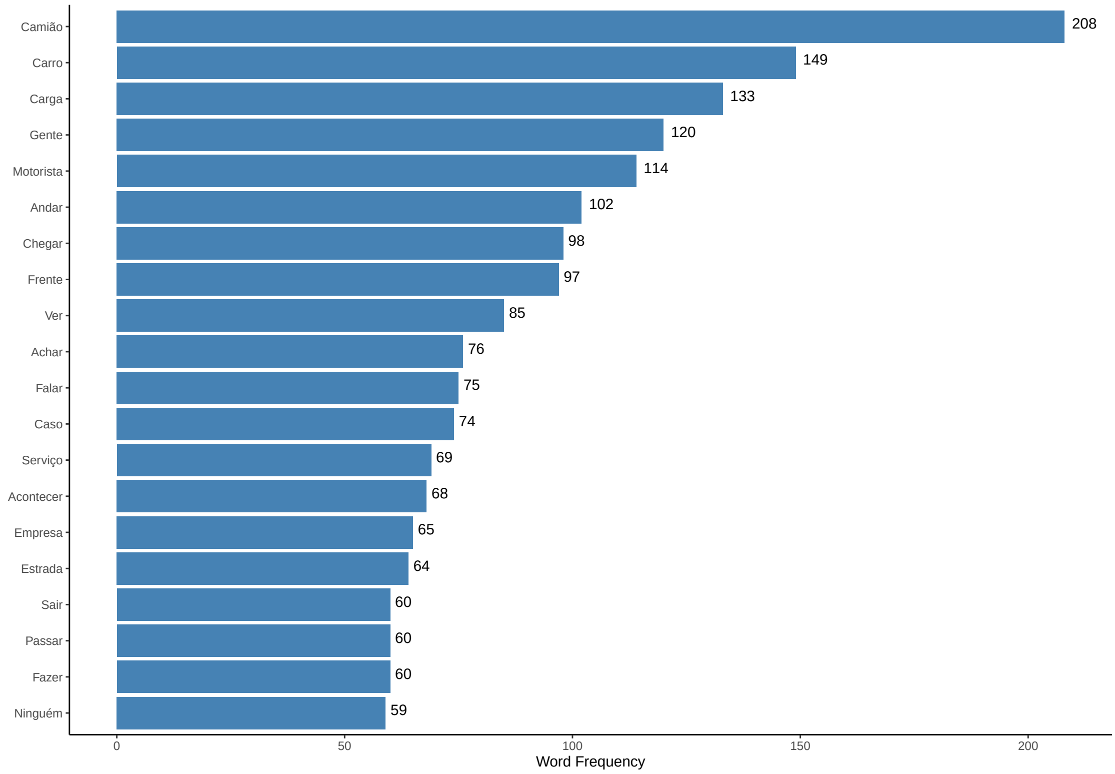

Figure 5.3: Frequency of the 20 most frequent words (drivers)

word populates the middle of each word cloud, and the less frequent words are closest to the board of the cloud.

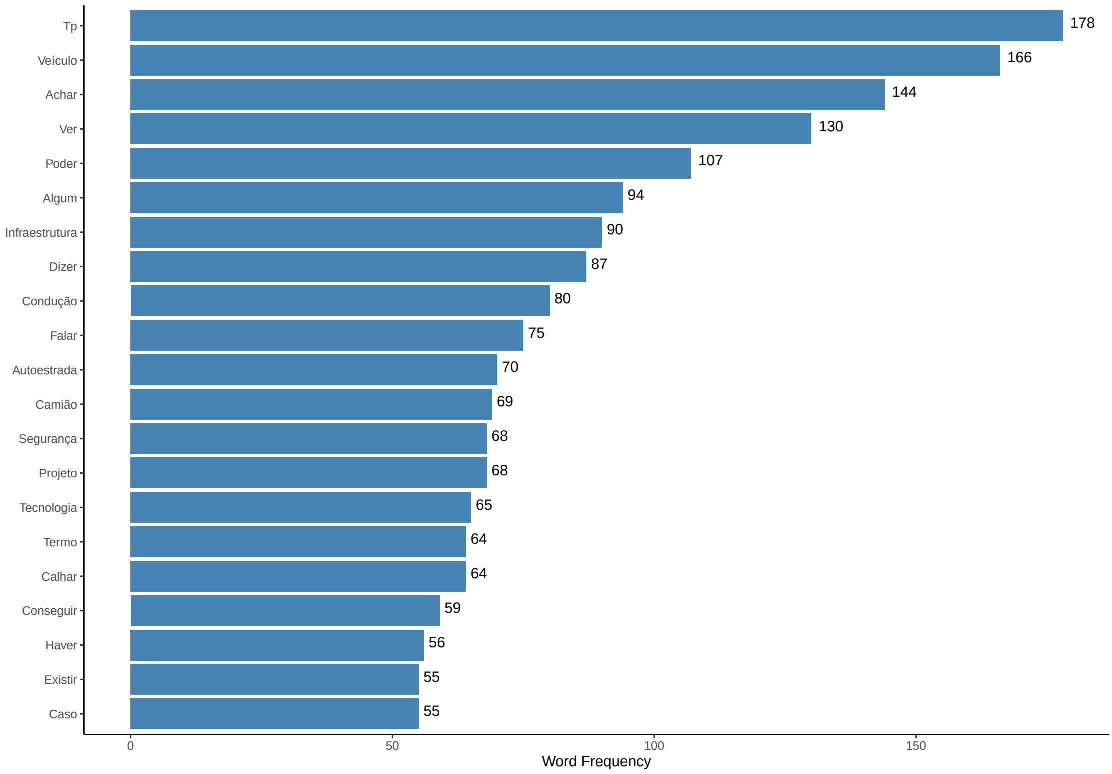

Figure 5.4: Frequency of the 20 most frequent words (road operators and regulators)

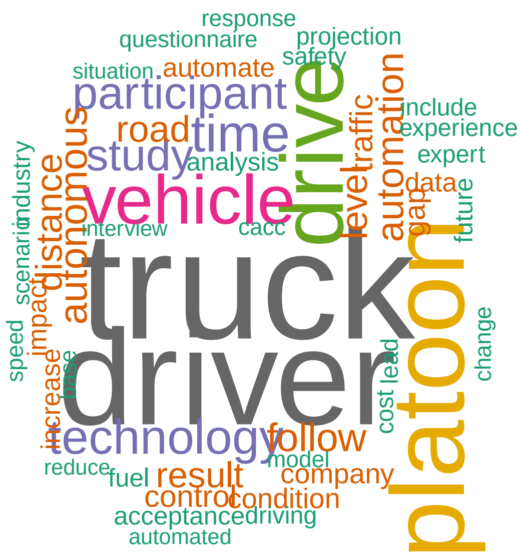

Figure 5.5: Word cloud of the 50 most frequent words (scientific literature)

Figure 5.6: Word cloud of the 50 most frequent words (drivers)

Figure 5.7: Word cloud of the 50 most frequent words (logistic companies)

Figure 5.8: Word cloud of the 50 most frequent words (road operators and regulators)

Overall, it is possible to observe that all actors converge on two crucial points: the need for operational efficiency and safety. Still, the scientific community and road regulators emphasize advanced technologies and safety measures. At the same time, decision-makers and drivers focus on practical and operational aspects of transportation, which one can draw from the frequency of topics.

5.2.2 Number of topics

As mentioned before, multiple metrics were utilized to determine the ideal number of topics to extract from every corpus; hence, an array of potential topics was required to be defined, and the appropriate *LDA models* were implemented. A range of 2 to 20 was considered, meaning 19 different models were generated and compared according to the four different metrics (4.3).

Figure 5.13 presents the different number of topics suggested by each metric regarding each corpus. Respectively, *Griffiths2004* and *Deveaud2014* suggested different but similar ranges of values, while *CaoJuan2009* and *Arun2010* were more consistent in the number of topics 5.2.

Table 5.2: Optimal Number of Topics per Metric and Corpus

	(Scientific literature	Drivers	Logistic companies	Regulators and Operators
Griffiths2004	16 Topics	5 Topics	6 Topics	6 Topics
CaoJuan2009	17-18 Topics	6-10 Topics	6 Topics	7 Topics
Arun2010	17-18 Topics	6-10 Topics	6 Topics	7 Topics
Deveaud2014	17 Topics	6-10 Topics	6 Topics	6-10 Topics

Readability and interpretability are crucial to consider while analyzing LDA results. Therefore, reducing the number of topics could make explaining the available data more simple and effective (Chu et al. (2020)). In this regard, the number of topics suggested by the Deveaud2014 metrics was more appropriate. This metric, as explained in 4.3, maximizes the information divergence of all topic pairs, which maximizes the differences between topics (Sutherland and Kiatkawsin (2020)) and is a common choice to define the number of topics in LDA applications. Hence, the LDA model was developed considering different settings regarding each corpus: Scientific literature (17 topics), Logistic companies (6 topics), Regulators and Operators (7 topics), and Drivers (10 Topics)

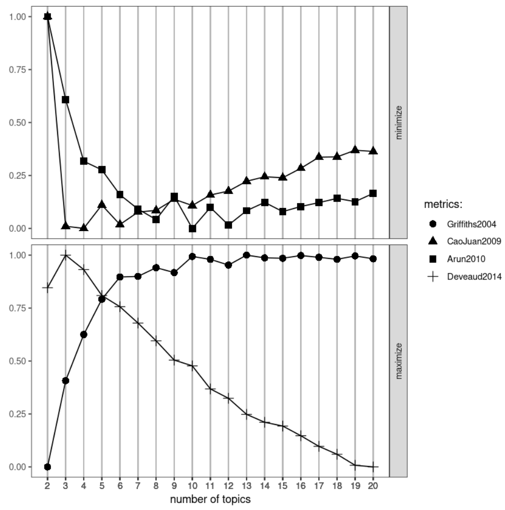

Figure 5.9: Logistic companies (Optimal Number of Topics)

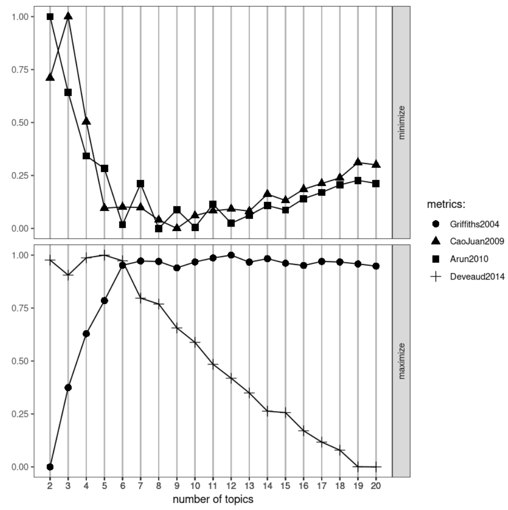

Figure 5.10: Drivers (Optimal Number of Topics)

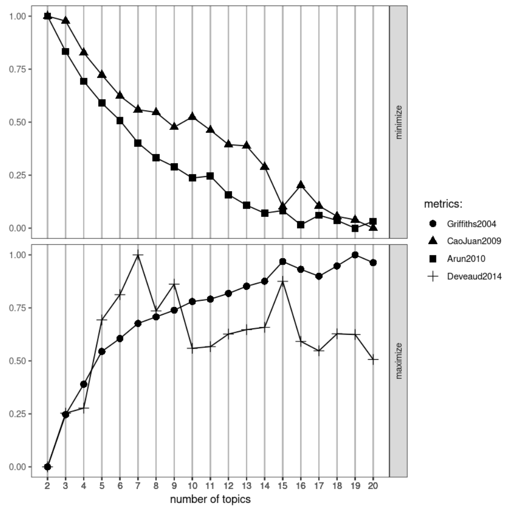

Figure 5.11: Scientific literature (Optimal Number of Topics)

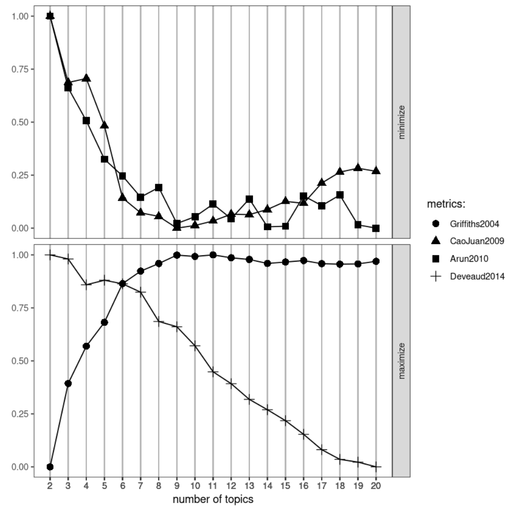

Figure 5.12: Regulators and Operators (Optimal Number of Topics)

Figure 5.13: Optimal number of Topics per corpus

5.2.3 Topics

The four models offered information to assist us in understanding the most relevant themes of truck platooning. For instance, it presented the most relevant words for each topic, providing an easy platform to distinctly identify what topic they related to since the LDA model does not perform tasks for naming the group of topics.

For each model, the optimal number of topics was selected as presented in Figure 5.13, calculated using the metrics explained in section 4.3 on chapter 4. Although naming 40 different topics is an excellent starting point for understanding the themes around Truck Platooning from all the actors, analyzing these 40 identified topics provides an understanding of this domain's key concerns and focus areas. By examining both the convergence and divergence of these topics, we can derive meaningful insights into each stakeholder group's collective and individual priorities while guiding future strategies and interventions (Figure 5.14).

	Topics	Top 10 words
Logistic companies	1. Driver Actions and Perceptions	motorista, achar, frente, camião, haver, falar, ver, muito, mudar, dizer
	2. Driver Situations and Client Interactions	motorista, achar, calhar, falar, conseguir, haver, porto, frente, cliente, coisa
	3. Driving and Company Operations	motorista, condução, empresa, camião, falar, mudar, ganhar, sair, andar, veículo
	4. Road Conditions and Driver Feedback	achar, camião, frente, dizer, mudar, certo, haver, estrada, ficar, cola
	5. Vehicle Logistics and Client Interactions	motorista, camião, achar, conseguir, via, chegar, cliente, horário, mudar, falar
	6. Company Coordination and Client Relations	camião, motorista, comboio, empresa, conseguir, cliente, horário, chegar, muito, carro
Drivers	1. Driver Experiences and Cargo Handling	motorista, camião, carro, gente, carga, achar, calhar, acontecer, haver, fazer
	2. Driver Sensors and Situations	motorista, sensor, gente, depender, muito, chegar, acontecer, carga, situação, pessoa
	3. Company Operations and Vehicle Service	carro, empresa, frente, serviço, trás, gente, passar, camião, ninguém, veículo
	4. Customer Service and Driving Conditions	gente, cliente, andar, empresa, carro, serviço, exatamente, chegar, dormir, problema
	5. Cargo Management and Navigation	camião, carro, carga, radar, sair, andar, frente, gente, motorista, falar
	6. Driving Safety and Incidents	camião, carro, chegar, condução, acontecer, caso, sair, distância, segurança, travar
	7. Cargo and Route Planning	camião, frente, carga, motorista, passar, caso, parar, chegar, metro, achar
	8. Loading Operations and Route Events	carregar, camião, falar, carro, gente, chegar, acontecer, frente, dever, andar
	9. Driver Schedules and Client Services	andar, gente, chegar, motorista, horário, cliente, serviço, loja, caso, ninguém
	10. Sensor Data and Vehicle Maintenance	sensor, falar, levar, carga, chegar, avariar, motorista, haver, casa, tirar
Regulators	1. Vehicle Communication and Infrastructure	calhar, ver, achar, veículo, vantagem, algum, via, comunicação, poder, acontecer
	2. Project Management and Traffic Impact	poder, impacto, ter, projeto, fazer, algum, tráfego, portagem, veículo, camião
	3. Vehicle Circulation and Infrastructure	veículo, falar, circulação, infraestruturo, piloto, comunicação, projeto, ter, real, acontecer
	4. Driving Regulations and Safety Standards	poder, condução, veículo, condutor, segurança, convenção, nível, empresa, internacional, regular
	5. Traffic Management and Technology	tráfego, veículo, algum, gestão, segurança, tecnologia, comboio, perceber, ter, falar
	6. Operational Scenarios and Infrastructure	achar, veículo, algum, existir, empresa, cenário, infraestruturo, condutor, pensar, camião
	7. Regulatory Discussions and Vehicle Use	dizer, condução, veículo, achar, poder, calhar, camião, claramente, algum, efetivamente
Scientific literature	1. Automotive Projections and Studies	projection, expert, study, automotive, delphi, logistic, truck, development, estimate, include
	2. Autonomous Vehicle Technology	driver, participant, vehicle, technology, autonomous, logistic, agvs, study, impact, drive
	3. Truck Platooning and Logistics	platooning, truck, road, driver, interviewee, platoon, carrier, drive, route, freight
	4. Driving Dynamics and Speed Control	truck, cacc, gap, driver, time, drive, road, vehicle, control, speed
	5. Traffic and Workload Management	konvoi, drive, traffic, driver, vehicle, participant, distance, workload, project, acceptance
	6. Fuel Efficiency and Emissions	truck, platooning, fuel, vehicle, scenario, emission, platoon, reduction, cost, saving
	7. Technology Adoption and Company Models	truck, level, company, adoption, technology, firm, variable, model, cluster, scenario
	8. Driver Experience and Platooning	drive, driver, platoon, truck, vehicle, gap, distance, follow, experience, traffic
	9. Cost Analysis and Autonomous Scenarios	truck, autonomous, scenario, cost, level, drive, driver, chain, supply, factor
	10. Automation and Technology Integration	truck, driver, automation, drive, autonomous, automate, technology, company, comment, study
	11. Platooning and Maneuverability	driver, platooning, truck, vehicle, platoon, drive, merge, distance, speed, maneuver
	1. Brake and Distance Control	truck, driver, time, drive, brake, condition, experiment, control, distance, task
	13. Driver Intentions and Technology Acceptance	truck, platoon, vehicle, drive, intention, acceptance, technology, highway, cut, driver
	14. Fuel Savings and Data Management	truck, follow, fuel, distance, datp, offset, lateral, saving, platoon, datum
	15. Driver Interaction and Display	concept, drive, driver, participant, display, input, takeover, score, situation, truck
	16. Trust and Automation Response	automation, automate, condition, driver, drive, baseline, trust, response, truck, study
	17. Industry Projections and Supplier Integration	truck, expert, autonomous, industry, future, projection, supplier, oems, business, delphi

Figure 5.14: Topic summary

One of the findings from this analysis is the convergence of topics related to driver operations and experiences. Both logistic companies and drivers exhibit a significant focus on operational and experiential aspects, as evidenced by topics like "Driver Actions and Perceptions," "Driving Safety and Incidents," and "Driver Experiences and Cargo Handling." This convergence indicates a shared interest in improving driver conditions and enhancing safety measures. The focus on these areas means that both groups realize the importance of addressing the daily practical challenges drivers face, thereby highlighting the need for policies and technologies that support drivers' operational efficiency and safety.

Another area of convergence is vehicle and technology management, which regulators and the scientific literature have highlighted. Common themes include “Vehicle Communication and Infrastructure,” “Driving Regulations and Safety Standards,” and “Technology Adoption and Company Models.” This alignment reflects a collective emphasis on ensuring vehicle technologies are safely and effectively integrated into transportation systems. The shared emphasis on these topics indicates a demand for steady innovation, the evolution of regulatory frameworks, and joint efforts between regulatory bodies and researchers to advance technological adoption.

Similarly, logistics and infrastructure emerge as essential themes among drivers and the scientific literature. To promote transportation efficiency, topics like "Cargo Management and Navigation" and "Loading Operations and Route Events" emphasize the importance of solid logistical frameworks and infrastructure. This confluence emphasizes the importance of infrastructure and logistics in maintaining seamless and efficient transportation operations.

Although there are certain areas of agreement, some significant differences in the subjects covered reflect their distinct priorities and concerns. For instance, drivers concentrate on precise operational details, such as "Driver Sensors and Situations" and "Cargo and Route Planning." These topics are more granular and operationally focused, highlighting drivers’ immediate and practical challenges.

Road regulators and operators, on the other hand, focus on more general systemic problems. Subjects like "Driving Regulations and Safety Standards" and "Project Management and Traffic Impact" highlight their broad responsibility for monitoring and guaranteeing adherence throughout the transportation network. This emphasis on systemic issues highlights the regulators’ duty to uphold transportation networks’ sustainability, efficiency, and safety through legislation and regulation.

In contrast, the scientific literature explores cutting-edge and futuristic subjects. Topics such as "Industry Projections and Supplier Integration" and "Autonomous Vehicle Technology" emphasize research and development for prospective breakthroughs. This futuristic view emphasizes how scientific literature drives technological advances and influences future transportation directions.

Additionally, decision-makers have unique themes like "Company Coordination and Client Relations," which highlight relational and strategic elements. These topics indicate their interest in more comprehensive organizational and client-facing tactics, emphasizing simplifying business operations and supporting satisfying client interactions.

While we applied the optimal number of topics, it is visually noticeable that some topics are still similar. Thus, we recalculated the similarity of tokens in each topic, and by performing this process, the tokens from each group did not change. However, the number of topics was reduced in this interaction for three of the four groups: drivers (10 to 5 topics), scientific literature (17 to 15 topics), and regulators (7 to 5 topics).

5.2.4 Similarity estimation

The cosine similarity measure was used to quantify the similarity between the information extracted from the Portuguese context and the Scientific literature datasets. Our approach is twofold: (1) to quantify the similarity of information within the corpora from the Portuguese context, we categorized our data into three groups: drivers, logistic companies, and regulators (road operators and legislators). (2) To contrast the Portuguese context with the scientific literature, we took advantage of Lourenço et al. (2024) findings, which clustered the scientific literature into three groups: Truck Drivers, General Public, and decision-makers (comprehend road operators, legislators, logistic companies). This methodology allows for the evaluation and contrast of the different perspectives within the datasets.

The cosine similarity scale extends from -1 to 1, where -1 denotes the total opposite orientation, 0 is orthogonality (no similarity), and 1 is the same orientation. The cosine similarity values for the three group pairs are displayed in Figure 5.15, each providing a perspective on the dynamics of their relationships demonstrating the similarity and how aligned the perspectives from different groups are.

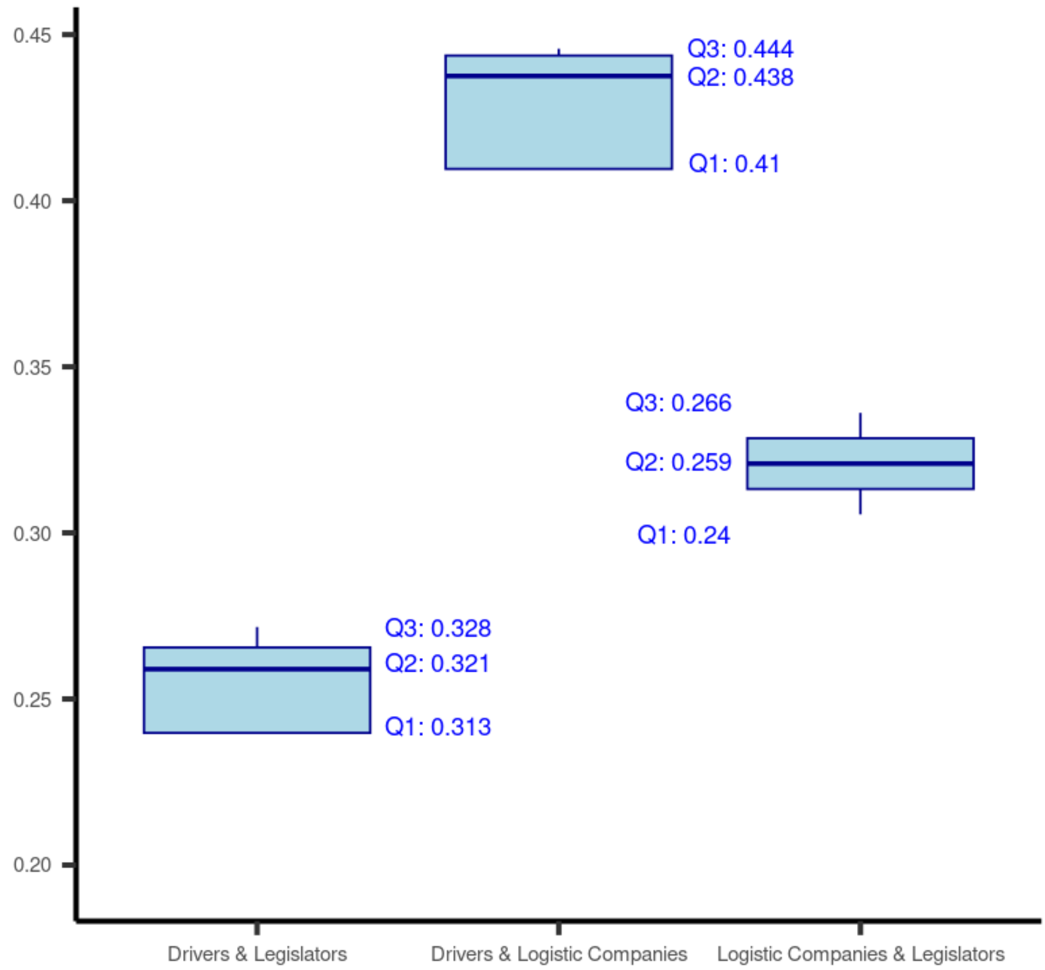

Figure 5.15: Cosine Similarity Portuguese Context

The first cluster pair indicates moderate similarity. The cosine similarity values for this group are closely grouped around the median, suggesting a high degree of agreement or consistency among the similarity scores for this group. Since the similarity values do not change much, the narrow *interquartile range* (**IQR**) indicates that legislators and logistic companies have relatively constant opinions or tendencies. This moderate proximity indicates that these two groups are reasonably aligned, which could lead to more seamless policy-making and collaborative processes.

Drivers and Decision Makers have the highest similarity of the three couples evaluated. This group's median similarity rating is significantly higher, implying a high degree of agreement between decision-makers and drivers. The values demonstrate a fairly high similarity, even if the range is wider than in the first group. This higher median similarity suggests a greater degree of agreement between drivers and decision-makers, which could prove beneficial when implementing policies or choices that directly affect both groups. Their opinions are strongly aligned, which may facilitate better collaboration and communication. On the other hand, the Drivers and Legislators cluster displays the lowest similarity values, highlighting the substantial contrasts between the two groups. This low similarity implies that drivers and legislators have different perspectives or parameters, which could make it difficult to implement policies and comply with regulations. These values draw attention to the distinctions between these groups, which suggests that attempts to align interests and objectives and implement specific interventions may be necessary to close the gap between them.

Since the cosine of similarity provided a measure on how the different actors are similar, one other point that we tackled was the overall differences across them, and using *ANOVA* on the corpus regarding the stakeholders from the Portuguese context, we find the following: The *ANOVA* test results indicate a statistically significant difference in the similarity means between the groups ($F = 17.71$, $p = 0.0018$).

Nevertheless, the cosine similarity analysis between Portuguese and international drivers shows a moderate level of alignment, with a median similarity score of 0.0827. This underlines that while drivers in various situations may have similar behaviors and perspectives, there are also evident geographical differences. The regular pattern of similarity values shown in Figure 5.16, the narrow IQR, and the absence of outliers suggest that these data are reliable and indicate broader trends in driving behavior and laws.

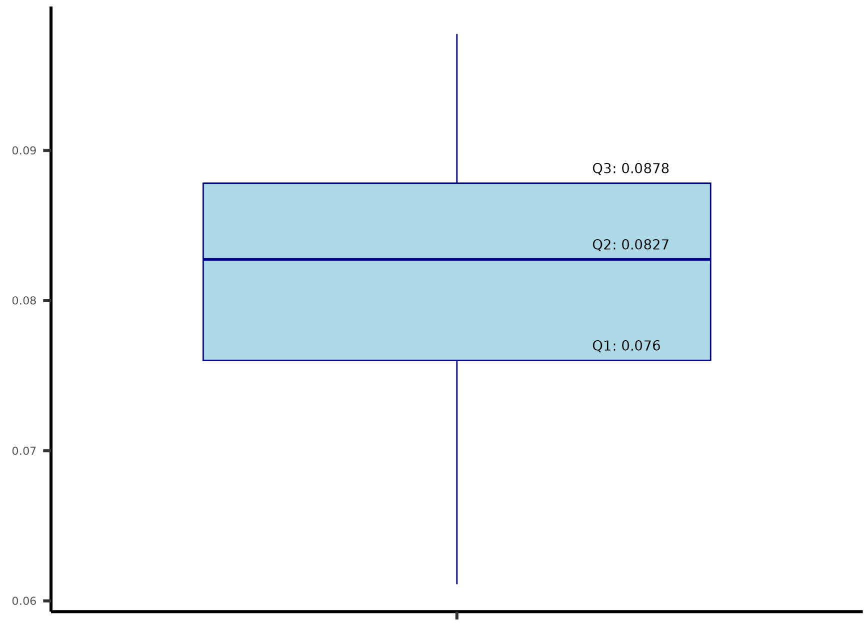

Figure 5.16: Cosine Similarity Drivers (Portuguese and International context)

Since the numbers are low, almost near zero (orthogonality), we can infer some alignment between drivers from the Portuguese context and scientific literature; naturally, this low alignment is due to the nature of each work. The scientific literature revolves around the search for innovations, while the drivers are more concerned with the daily challenges; additionally, it reflects the lack of knowledge of a technology they did not have contact with.

There is a moderate similarity between other stakeholders in the Portuguese context and their foreign counterparts, according to the median similarity value of 0.0871. The similarity values are rather closely clustered around the median, as shown in the interquartile range (IQR), which ranges from 0.0642 to 0.0917. This suggests a consistent pattern of alignment (Figure 5.17).

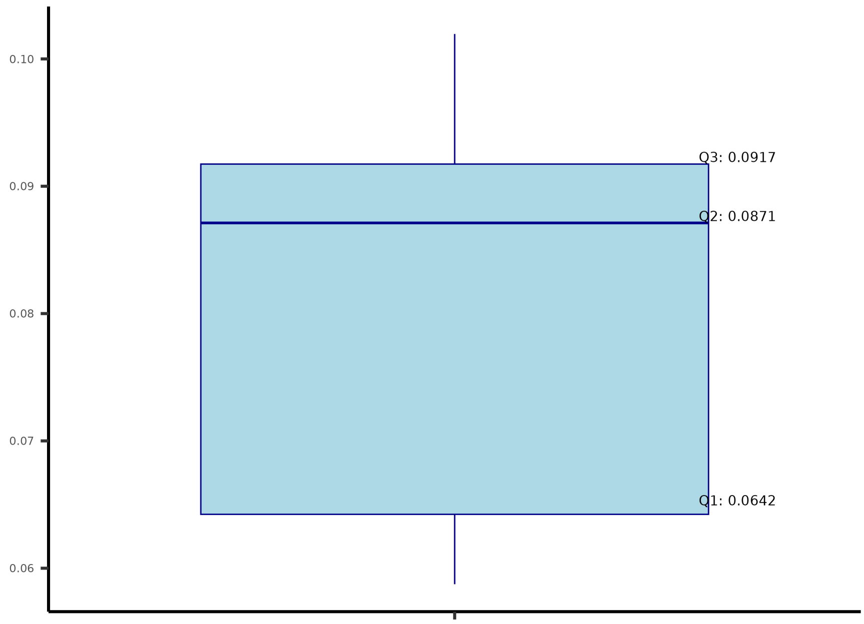

Figure 5.17: Cosine Similarity Legislators and Decision Makers (Portuguese and International context)

There is a range of congruence levels, from modest to high, according to the cosine similarity study across different stakeholder groups. Yet, there are also some significant areas of divergence. Overall, Drivers and Logistics companies, Logistics companies, and Legislators show higher degrees of similarity, indicating strong alignment that is favorable to productive collaboration. On the other hand, there is the least similarity between drivers and legislators, emphasizing the necessity of focused efforts to bring them together.

In Portuguese and international contexts, drivers and other stakeholders exhibit modest alignment, with consistent similarity values expressing common behaviors and values. These findings highlight the significance of understanding the shared and distinct elements of stakeholder relationships in various situations.

Thus, we used *Wilcoxon* to calculate the significant statistical differences when we contrasted the drivers and other stakeholders in international and Portuguese scenarios. For the drivers, the result is not statistically significant, indicating no significant difference in cosine similarity between the two groups, with a *p-value* of 0.6017 and a *Wilcoxon* score of 180. The result is not statistically significant for other stakeholders, indicating no difference in cosine similarity between the two groups, with a *p-value* of 0.0598 and a *Wilcoxon* score of 1062.

As Lourenço et al. (2024) described, there is a diversity of perspectives and themes; however, we can infer that there needs to be more dissemination and public debate around TP, which affects all parts' synergy and engagement.

In general, these understandings are essential for creating focused plans, regulations, and actions that consider any organization's unique dynamics and alignments. Stakeholders may promote more efficient coordination, communication, and collaboration across groups and areas by utilizing these commonalities and addressing the disparities, ultimately increasing the efforts' overall efficacy.

5.3 Summary

The four examined corpora each have intrinsic characteristics that provide distinct perspectives on TP. While drivers and logistic companies have their thoughts rooted in operational challenges and issues in their current reality, the scientific literature and regulators look forward to the future, especially pushing for advancements in technology and regulations. To some degree, the different perspectives clarify that we have different realities that should be addressed.

The topics extracted from different actors in truck platooning have highlighted significant insights and nuanced distinctions regarding their concerns and priorities. The discussion of the convergence and divergence of these topics provides a comprehensive understanding of how each group perceives and addresses issues within truck platooning, providing numerous directions and potential policy implementations.

One of the findings from this analysis is the convergence of topics related to driver operations and experiences. Logistic companies and drivers exhibit a significant focus on operational and experiential aspects. This convergence suggests a shared interest in improving driver conditions and enhancing safety measures, underscoring the need for policies and technologies supporting drivers' operational efficiency and safety.

The scientific literature and authorities have emphasized the importance of vehicle and technology management as another crucial area of convergence. Regular topics are intended to ensure that vehicle technologies are correctly and effectively incorporated into transportation systems. This alignment suggests that to promote technological adoption, continuous innovation, the creation of regulatory frameworks, and cooperative efforts between regulatory authorities and researchers are required.

Moreover, drivers and the scientific literature identify logistics and infrastructure as critical concerns. To promote transportation efficiency, topics like "Cargo Management and Navigation" and "Loading Operations and Route Events" highlight the importance of solid logistical frameworks and infrastructure. This convergence emphasizes the importance of infrastructure and logistics in maintaining fluid and effective transportation operations.

The examination of these 40 subjects yields several important conclusions. First, all stakeholder groups share a joint emphasis on operational effectiveness and safety. The overall acknowledgment of safety and efficiency as crucial goals implies that these elements should be prioritized in any changes or advancements made in the transportation sector.

Second, the convergence of logistics and technology management problems suggests that different stakeholders must work together. Concurrently, the scientific literature, decision-makers, regulators, and drivers can create solutions that address systemic and practical issues, resulting in more sustainable and efficient transportation networks.

The topic deviation also underscores each stakeholder group's diverse perspectives and priorities. While drivers are primarily concerned with immediate operational challenges, regulators focus on more general compliance and systemic issues, and the scientific literature explores future specialized advancements. Comprehending these priorities is essential for conceiving targeted interventions that address each group's needs and concerns.

According to the cosine similarity, congruence levels exist, from modest to high, across different stakeholder groups. However, there are also some significant areas of divergence. Overall, drivers, decision-makers, and legislators show higher degrees of similarity, indicating strong alignment favorable to productive collaboration. On the other hand, there is the slightest similarity between drivers and legislators, emphasizing the necessity of focused efforts to bring them together.

These understandings are essential for creating focused plans, regulations, and actions considering any organization's unique dynamics and alignments. Stakeholders may promote more efficient coordination, communication, and collaboration across groups and areas by utilizing these commonalities and addressing the disparities, ultimately increasing the efforts' overall efficacy.

These topics precisely map the primary concerns and focus areas within the transportation sector. This information is valuable for steering the development of targeted procedures, policies, and inventions that enhance transportation systems' overall efficiency, safety, and sustainability. By identifying both the convergence and divergence of topics, we can ensure that the needs of all actors are addressed, leading to a more cohesive and practical approach to transportation management.

Chapter 6

Conclusion and Future Directions

This thesis aimed to explore the dynamics of truck platooning by applying LDA models to various datasets representing different stakeholders: truck drivers, logistics operators, road operators and regulators, and the scientific literature. The primary goal was to identify the predominant themes and topics within these groups to better understand their perspectives and concerns and if, in any case, they have an intersection.

The analysis revealed significant insights into the different themes among the stakeholders. Operational efficiency and security are shared priorities. Logistic companies and drivers alike have paid substantial attention to practical characteristics and driver experiences, reflecting a common goal of improving everyday valuable features. The scientific community and regulators underlined the need for robust regulatory frameworks and continuous technical innovation related to vehicle and technology management.

The differences in subjects, however, also highlighted specific problems that each group encountered. Drivers concentrated on specific, practical difficulties, while regulators dealt with more general systemic problems. The scientific community focused on future breakthroughs by investigating innovative and futuristic topics. On the other hand, decision-makers acknowledged the importance of strategic and relational elements, emphasizing the necessity of efficient business operations and strong client relationships.

These results highlight the complexity of truck platooning as a multifaceted area that requires a comprehensive approach to strategy and policy development. The subjects identified serve as a concise outline of the main issues and areas of focus, directing the development of specific strategies, policies, and innovations that improve the overall effectiveness, sustainability, and safety of truck platooning systems.

Even though this study examines truck platooning, some shortcomings and areas could be researched. The LDA model's static nature, which ignores temporal changes in the data, is one of its main drawbacks. Dynamic topic modeling could be helpful in future research to capture the themes' temporal progression. This strategy would enable more responsive and adaptable policy-making by offering insights into how stakeholder concerns and priorities change.

The preprocessing of these diverse datasets involved several stages, such as text cleaning, normalization, and tokenization, significantly reducing the size of the corpora by around 45%. The application of lemmatization, as opposed to stemming, enhanced the interpretability of the text, ensuring that words were analyzed in their base forms. This process laid a solid foundation for the subsequent LDA modeling.

Given their varying sizes, hyperparameter tuning was crucial in determining each dataset's optimal number of topics. Metrics such as *Griffiths2004*, *CaoJuan2009*, Arun2010, and *Deveaud2014* were employed to evaluate model performance, with a range of 2 to 20 topics considered. The optimal number of topics was chosen based on maximizing and minimizing these metrics, ensuring a balance between detail and interoperability.

In addition, there is potential for expanding the range of datasets utilized in the present study. Including more extensive and varied corpora—including international datasets—would provide a more comprehensive awareness of different perspectives on truck platooning worldwide. Cross-cultural analysis may highlight particular geographical issues and their solutions, resulting in more broadly applicable tactics.

Advanced methods, such as large language models (LLMs), can significantly enhance the analysis. Large volumes of text data can be processed by these models more accurately and contextually. Using LLMs for topic modeling may enhance the detected subjects' coherence and granularity. These models benefit specialist fields like truck platooning because they may be adjusted to specific domains. LDA models could be combined with NLP techniques like named entity identification and sentiment analysis to provide a more complex analysis. Sentiment analysis would help identify stakeholders' views and opinions regarding different facets of truck platooning. Named entity recognition may also recognize critical entities in the text and their connections. Through these methods, the analysis would gain more depth and provide deeper insights into the stakeholders' viewpoints.

Lastly, involving stakeholders through participatory research could guarantee that the developed strategies and policies are grounded in practical realities. Workshops, focus groups, and surveys could be utilized to gather direct feedback from stakeholders, refining the models and strategies based on their inputs. This participatory approach would enhance the relevance and usefulness of the proposed solutions.

In conclusion, while this study has provided insights into the themes and concerns of various stakeholders in truck platooning, there is vast potential for further research and development. By pursuing these future directions and addressing the limitations, we can continue improving truck platooning systems' efficiency, safety, and sustainability, ultimately contributing to advancing transportation technologies and practices.

References

Emeli Adell. *Driver experience and acceptance of driver support systems–a case of speed adaptation*. PhD thesis, Lund University, 2009.

Icek Ajzen. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2):179–211, 1991.

Icek Ajzen and Martin Fishbein. An attitude-behavior relationship: The theory of reasoned action. *Psychological Bulletin*, 81(2):88–118, 1975.

R Arun, V Suresh, CE Veni Madhavan, and MN Murthy. Finding topics in documents: a probabilistic approach. In *Proceedings of the 2010 European conference on Machine learning and knowledge discovery in databases: Part III*, pages 174–186. Springer, 2010.

Richard P Bagozzi and Youjae Yi. Specification, evaluation, and interpretation of structural equation models. *Journal of the Academy of Marketing Science*, 40(1):8–34, 2012.

Vimala Balakrishnan and Ethel Lloyd-Yemoh. Stemming and lemmatization: A comparison of retrieval performances. *International Journal of Advanced Computer Science and Applications*, 3(2):193–200, 2014.

Anil K Bhoopalam, Niels Agatz, and Rob Zuidwijk. Planning of truck platoons: A literature review and directions for future research. *Transportation Research Part B: Methodological*, 107:212–228, 2018. doi: 10.1016/j.trb.2017.10.016.

David M. Blei. Probabilistic topic models. *Commun. ACM*, 55(4):77–84, apr 2012. ISSN 0001-0782. doi: 10.1145/2133806.2133826. URL https://doi.org/10.1145/2133806.2133826.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003. ISSN 1532-4435. URL http://dl.acm.org/citation.cfm?id=944919.944937.

Juan Cao, Tian Xia, Jintao Li, Yong Zhang, and Sheng Tang. A density-based method for adaptive lda model selection. In *Proceedings of the 18th international conference on Information and knowledge management*, pages 51–60, 2009.

Sarah Maria Castritius, Heiko Hecht, Johanna Möller, Christoph J. Dietz, Patric Schubert, Christoph Bernhard, Simone Morvilius, Christian T. Haas, and Sabine Hammer. Acceptance of truck platooning by professional drivers on german highways. a mixed methods approach. *Applied Ergonomics*, 85, 5 2020. ISSN 18729126. doi: 10.1016/j.apergo.2019.103042.

Subhomoy Chatterjee, Nripendra P Rana, Yogesh K Dwivedi, and Abdulmohsen M Baabdullah. Understanding ai adoption in manufacturing and production firms using an integrated tam-toe model. *Technological Forecasting and Social Change*, 170:120880, 2021.

Qinyu Chen, Yi Chen, and Ying Chen. Text mining for uncovering technology intelligence: A case on convolutional neural network. *Technological Forecasting and Social Change*, 146:785–796, 2019.

Chia-Yu Chu, Kwanwoo Park, and Geoffrey E Kremer. A global supply chain risk management framework: An application of text-mining to identify region-specific supply chain risks. *Advanced Engineering Informatics*, 2020. doi: 10.1016/j.aei.2020.101053.

Liliana Cunha, Daniel Silva, Daniela Monteiro, Sara Ferreira, António Lobo, António Couto, Anabela Simões, and Catarina Neto. *Who Really Wants Automated Vehicles? Determinant Factors of Acceptability Profiles in Portugal*. 2022. ISBN 9781792389887. URL https://openaccess.cms-conferences.org/#/publications/book/978-1-7923-8988-7.

Fred D Davis. *A technology acceptance model for empirically testing new end-user information systems: Theory and results*. PhD thesis, Massachusetts Institute of Technology, 1985.

Fred D Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3):319–340, 1989.

Romain Deveaud, Eric SanJuan, and Patrice Bellot. Accurate and effective latent concept modeling for ad hoc information retrieval. *Document numérique*, 17(1):61–84, 2014.

Sérgio Pedro Duarte, Liliana Cunha, Luciano Moreira, Sara Ferreira, and António Lobo. Requirements and expectations for truck platooning - a multidisciplinary perspective. In *Proceedings of the AHFE 2023 International Conference on Human Factors in Transportation*. AHFE Open Access, 2023. doi: 10.54941/ahfe1003072.

Ramzi El-Haddadeh. Digital innovation dynamics influence on organisational adoption: The case of cloud computing services. *Information Systems Frontiers*, 22:985–999, 2020. doi: 10.1007/s10796-019-09912-2.

ETRAC. Long distance freight transport a roadmap for system integration of road transport status: final for publication, 2019.

ETRAC. Connected, cooperative and automated mobility roadmap status: final for publication, 2022. URL www.ertrac.org.

William Ferrell, Kimberly Ellis, Phil Kaminsky, and Chase Rainwater. Horizontal collaboration: opportunities for improved logistics planning. *International Journal of Production Research*, 58(14):4267–4284, 2020. doi: 10.1080/00207543.2019.1651457. URL https://doi.org/10.1080/00207543.2019.1651457.

Peter Fröhlich, Alexandra Sackl, Sandra Trösterer, Alexander Meschtscherjakov, Luke Diamond, and Manfred Tscheligi. Acceptance factors for future workplaces in highly automated trucks. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 129–136, 2018. doi: 10.1145/3239060.3240446.

Frank W. Geels. A socio-technical analysis of low-carbon transitions: introducing the multi-level perspective into transport studies. *Journal of Transport Geography*, 24:471–482, 2012. ISSN 0966-6923. doi: https://doi.org/10.1016/j.jtrangeo.2012.01.021. URL https://www.sciencedirect.com/science/article/pii/S0966692312000269. Special Section on Theoretical Perspectives on Climate Change Mitigation in Transport.

Morteza Ghobakhloo, Daniel Arias-Aranda, and Jose Benitez-Amado. Adoption of e-commerce applications in smes. *Industrial Management & Data Systems*, 111(8):1238–1269, 2011.

Thomas L Griffiths and Mark Steyvers. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(suppl 1):5228–5235, 2004.

Zvi Griliches. *Patent Statistics as Economic Indicators: A Survey, Part II*. National Bureau of Economic Research, 1990.

Joseph F Hair, William C Black, Barry J Babin, and Rolph E Anderson. *Multivariate Data Analysis*. Pearson Prentice Hall, 2010.

Katarzyna Halicka. Main concepts of technology analysis in the light of the literature on the subject. *Procedia Engineering*, 182:291–298, 2017.

Bronwyn H Hall, Adam B Jaffe, and Manuel Trajtenberg. The nber patent citation data file: Lessons, insights and methodological tools, 2001.

Eda Kavlakoglu Jacob Murel. What are stemming and lemmatization? https://www.ibm.com/topics/stemming-lemmatization, 2023. Accessed: 2024-07-13.

Daniel Jurafsky and James H. Martin. *Speech and Language Processing*. Pearson, 3rd edition, 2020.

Daniel Jurafsky and James H. Martin. *Speech and Language Processing*. Pearson, 2021.

Stefan Kaiser, Matthias Winkelbauer, Elmar Wannenmacher, Philipp Blass, and Hüseyin Atasayar. Road safety issues related to truck platooning deployment. pages 175–186, 2022. doi: 10.1007/978-3-030-88682-0_13.

Amir Karami, Megan Lundy, Farren Webb, and Yogesh K Dwivedi. Twitter and research: A systematic literature review through text mining. *IEEE Access*, 2020. doi: 10.1109/ACCESS.2020.2983656.

Christos Katrakazas, Mohammed Quddus, Wei-Hua Chen, and Lipika Deka. Real-time motion planning methods for autonomous on-road driving: State-of-the-art and future research directions. *Transportation Research Part C: Emerging Technologies*, 60:416–442, 2015.

Ralph Katz and Thomas J Allen. Organizational issues in the introduction of new technologies. In *The management of productivity and technology in manufacturing*, pages 275–300. 1985.

John Lioris, Ramtin Pedarsani, Faruk Y. Tascikaraoglu, and Pravin Varaiya. Platoons of connected vehicles can double throughput in urban roads. *Transportation Research Part C: Emerging Technologies*, 77:292–305, 2017.

Todd Litman. Autonomous vehicle implementation predictions: Implications for transport planning. *International Journal of Transportation Science and Technology*, 9(3):208–218, 2020.

Vladimiro Lourenço, Sérgio Pedro Duarte, António Lobo, Sara Ferreira, and Liliana Cunha. Representations of truck platooning acceptance of truck drivers, decision-makers, and general public: A systematic review. *Transportation Research Part F: Traffic Psychology and Behaviour*, 104:359–373, 2024. ISSN 1369-8478. doi: https://doi.org/10.1016/j.trf.2024.06.008. URL https://www.sciencedirect.com/science/article/pii/S1369847824001414.

Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, Massachusetts, 1999. URL http://nlp.stanford.edu/fsnlp/.

Christopher D Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

Luis Martinez and Francisco Herrera. Text mining and natural language processing: Synergies and divergence. *Information Fusion*, 64:3–5, 2020.

Rada Mihalcea and Paul Tarau. Textrank: Bringing order into text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 404–411, 2004.

Marie-Francine Moens. *Automatic Indexing and Abstracting of Document Texts*. Springer, 2001.

Sadao Nagaoka, Kazuyuki Motohashi, and Akira Goto. Chapter 25 - patent statistics as an innovation indicator. In Bronwyn H. Hall and Nathan Rosenberg, editors, *Handbook of the Economics of Innovation, Volume 2*, volume 2 of *Handbook of the Economics of Innovation*, pages 1083–1127. North-Holland, 2010. doi: https://doi.org/10.1016/S0169-7218(10)02009-5. URL https://www.sciencedirect.com/science/article/pii/S0169721810020095.

Catarina Neto, Anabela Simoes, Liliana Cunha, Sérgio Pedro Duarte, and António Lobo. Qualitative data collection to identify truck drivers' attitudes toward a transition to platooning systems. *Accident Analysis Prevention*, 195:107405, 2024. ISSN 0001-4575. doi: https://doi.org/10.1016/j.aap.2023.107405. URL https://www.sciencedirect.com/science/article/pii/S0001457523004529.

Md. Noor-A-Rahim, Zilong Liu, Haeyoung Lee, Mohammad Omar Khyam, Jianhua He, Dirk Pesch, Klaus Moessner, Walid Saad, and H. Vincent Poor. 6g for vehicle-to-everything (v2x) communications: Enabling technologies, challenges, and opportunities. *Proceedings of the IEEE*, 110(6):712–734, 2022. doi: 10.1109/JPROC.2022.3173031.

Tiago Oliveira and Maria Fraga Martins. Firms patterns of e-business adoption: evidence for the european union-27. *Electronic Journal of Information Systems Evaluation*, 13(1):47–56, 2010.

Marcin Pająk and Piotr Cyplik. Truck platooning in the context of sustainable development's targets defined in european union's strategies. *Logforum*, 16(2):311–321, 2020. doi: 10.17270/J.LOG.2020.411.

PPMC. Truck platooning trial in the netherlands with ecotwin. http://www.ppmc-transport.org/truck-platooning-trial-in-the-netherlands-with-ecotwin, 2021. Accessed: 2024-01-26.

SAE International. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. SAE Standard J3016_201806, SAE International, 2018. URL https://www.sae.org/standards/content/j3016_201806.

Mina Saghafian, Karin Laumann, and Martin Rasmussen Skogstad. Stagewise overview of issues influencing organizational technology adoption and use. *Frontiers in Psychology*, 12, 2021. ISSN 1664-1078. doi: 10.3389/fpsyg.2021.630145. URL https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2021.630145.

Hinrich Schütze. Word space. *Advances in Neural Information Processing Systems*, 5, 1993.

Steven E. Shladover. Connected and automated vehicle systems: Introduction and overview. *Journal of Intelligent Transportation Systems*, 22(3):190–200, 2018.

Steven E Shladover, Christopher Nowakowski, Xiao-Yun Lu, and Robert Ferlis. Cooperative adaptive cruise control: Definitions and operating concepts. *Transportation Research Record*, 2489:145–152, 2015. doi: 10.3141/2489-17.

Julia Silge and David Robinson. *Text Mining with R: A Tidy Approach*. O'Reilly Media, 2016.

Ana Simoes, António Lobo, Sérgio Ferreira, Carlos Rodrigues, José P Tavares, Ana Couto, Luís Cunha, and Céu Neto. Mapping risks and requirements for truck platooning: A human-centred approach. *Lecture Notes in Networks and Systems*, 319:514–522, 2022. doi: 10.1007/978-3-030-85540-6_65.

S. Soares, A. Lobo, S. Ferreira, et al. Takeover performance evaluation using driving simulation: a systematic review and meta-analysis. *European Transport Research Review*, 13.

I. Sutherland and K. Kiatkawsin. Determinants of guest experience in airbnb: A topic modeling approach using lda. *Sustainability (Switzerland)*, 2020. doi: 10.3390/SU12083402.

Sadayuki Tsugawa, Sabina Jeschke, and Steven E. Shladover. A review of truck platooning projects for energy savings. *IEEE Transactions on Intelligent Vehicles*, 1(1):68–77, 2016.

Peter D Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 417–424. Association for Computational Linguistics, 2002.

Viswanath Venkatesh, Michael G Morris, Gordon B Davis, and Fred D Davis. User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3):425–478, 2003.

Dehlia Willemsen, Antoine Schmeitz, Edoardo Mascalchi, Rudolf Huisman, Valerio Liga, Roman Alieiev, Hans Nordin, and Wen Xu. European commission ensemble enabling safe multi-brand platooning for europe deliverable no. d2.3: V2 platooning use cases, scenario definition and platooning levels. Dissemination level: Public, 2021. URL https://ec.europa.eu/programmes/horizon2020/ENSEMBLE.

Binggui Zhou, Guanghua Yang, Zheng Shi, and Shaodan Ma. Natural language processing for smart healthcare. *IEEE Reviews in Biomedical Engineering*, 2022.