

Babeş-Bolyai University
Faculty of Political, Administrative and Communication Sciences
Institute for Doctoral Studies
Doctoral School of Political and Communication Sciences
Communication Sciences Program

PHD THESIS

ABSTRACT

The Archetypal Myths of War News Discourse: The Discourse-Mythocritical Model and The New York Times' Macro Mythological-Discursive Construction of International Armed Conflicts

Scientific coordinator: Professor Andreea-Alina Mogoş, PhD.

PhD Candidate: Teodora-Elena Grapă

Copyright © 2026 by Teodora-Elena Grapă. All Rights Reserved.

Table of contents

Chapter I: Introduction.....	1
The Research Problem: The Research Engine.....	1
Theoretical, Conceptual, Analytical, Methodological and Empirical Solutions.....	2
Chapter II: Literature Review and Theoretical Framework.....	3
Chapter III: The Discourse-Mythocritical Model.....	4
Chapter IV: Research Design.....	5
Theoretical Thesis and the Bonus Pilot Study: Purpose, Objectives, Research Questions	5
Methodology	6
Limitations, Reproducibility and Replicability.....	8
Chapter V: Analysis and Results	9
Chapter VI: Discussion and Conclusions. Original Contributions	10
Bibliography	14

Chapter I: Introduction

The Research Problem: The Research Engine

Chapter I of the present doctoral thesis offers an introduction (**section 1.1.**) and an overview of all the thesis chapters (**section 1.2.**), and briefly explains the particularities of the research purpose, focus, the main research question, objectives, and the process of developing and refining the multi-faceted original contributions of the study: the Discourse-Mythocritical Model and the 30 archetypal myths, the novel methods of honing in on narrative discourse and the pilot study testing and showcasing the application of the model.

Since the initial purpose of the present research was to determine archetypal myths constructing news media on various events and actors, and since the research process requires the naming and detailed exploration of justifiably relevant theoretical and analytical frameworks, the initial research plan quickly became obstructed by lack of answers that are pertinent to communication sciences and the study of discourse. As such, the identification of major theoretical and conceptual gaps led to a down-the-rabbit-hole multidisciplinary search for the definition of archetype, beyond media studies and the often-favoured psychoanalysis-relevant definition, into ancient, medieval and modern philosophy, logic, cultural anthropology, cultural studies, and back into cognitive and socio-cognitive linguistics. The process proved fruitful in the sense of supporting the development of original definitions put forward through the present thesis, which set the necessary theoretical and conceptual foundations facilitating the precise operationalization of archetypal myths as analytical categories within communication sciences research.

Since important analytical and methodological gaps were also identified, the research quest was defined by a long process of testing (multiple analytical approaches and methods, various text analysis software), learning (computational linguistics, using programming languages and the relevant tools) and more intensive testing, aimed at developing and refining original solutions to the problem.

Therefore, the present thesis takes on a three-part challenge meant to account for major theoretical, conceptual, analytical and methodological gaps in the study of archetypal myths within the field of communication sciences: (1) the communication-centred definition of archetypes and archetypal myths, (2) the conceptualisation of a set of narrative schemata and linguistic indicators

per archetypal myth and (3) the operationalisation and application of analytical categories employing quantitative methods, not just qualitative methods. As the main purpose of the present study is the underscoring and definition of an archetypal myth-discursive logic that is communication-centred and operational in the pursuit of analysing the archetypal dimensions of discourse, especially focused on discourse on conflict, the main research question (which is introduced and justified in Chapter I and further detailed in Chapter IV) that guides the entire study is preoccupied with the specifics of the narrative-discursive construction of archetypal myths.

Theoretical, Conceptual, Analytical, Methodological and Empirical Solutions

The Discourse-Mythocritical Model (DMM) put forward by the present research is designed to address the overarching RQ; hence, the thesis advances communication theory and analytical approaches by conceptualizing and cementing archetype and archetypal myth in the communication science field. Furthermore, since this is first and foremost a theoretical thesis, its core objectives (which are resolved in detail in Chapter III) include the in-depth exploration of the novel proposal of the linguistic indicators of myth, together with the definition of a conflict discourse-relevant comprehensive list of archetypal myths, their narrative schemata and linguistic indicators, 30 in total. The archetypal myths are discussed in detail in Chapter III, including through exemplification of their logic on a wide variety of texts (from historical text, various news sources, to music lyrics).

The purpose of the additional pilot study is, as with any pilot study, to fill an *empirical gap*, meaning to test new theory, new methodology and/or applicability of new analytical models, all of which are relevant to the present thesis. This is secondary to the main research, hence secondary research questions are formulated around the following objectives, which are introduced in Chapter I and detailed in Chapter IV (Research Design).

To fulfil the larger purpose and meet all objectives of the present research, the thesis essentially offers (1) a conceptual and analytical framework that accounts for various nuances in mythopoetic logic, including the (2) in-depth conceptualisation of 30 archetypal myths, all detailed in Chapter III, (3) an innovative research design (detailed in Chapter IV) including novel methods of

determining narrative-discursive patterns and a pilot study (4) meant to show the applicability of the theoretical, conceptual, analytical and methodological facets of the thesis' main proposal (the DMM) through an analysis of New York Times' news on two international armed conflicts (Russia-Ukraine and Israel-Palestine). Overall, the original contributions of the present study to communication sciences are varied, with theoretical, analytical, and methodological implications, which are highlighted throughout the thesis.

Chapter II: Literature Review and Theoretical Framework

Chapter II is dedicated to reviewing the relevant literature and theories, therefore the first **subchapter (2.1)** defines and explains the mechanisms surrounding media discourse, mythopoesis and mediated conflict, by first focusing, in **section 2.1.1.**, on media logic and the spectacularization of society, media events, news values and media frames, as well as audience frames.

The rest of **section 2.1.** explains paradigms, theories, concepts and notions such as myth, the mytheme and the structuralist perspective, mythopoetics, ideology, narratology, discourse and narrative, as well as the particularities of news mediated war, including a review of previous research on the two analysed conflicts and The New York Times' role in the global scene of news production. The last **section (2.1.4.)** also explains the rationale behind the choice of NYT as a case study and the positive implications of focusing on "hard news" reports to demonstrate the applicability of the model, namely its high replicability potential.

The second **subchapter of Chapter II (2.2)** is the product of the process of reviewing literature that answers the question "what is the archetype?", tracing the history of the concept from ancient philosophical thought to contemporary media studies, before reviewing literature relevant to media research and archetypal analysis (in **section 2.2.4.**), and defining major research gaps, including the lack of a relevant communication-centred conceptualisation of archetype and archetypal myths and an operational discourse-centred analytical framework.

Chapter III: The Discourse-Mythocritical Model

Chapter III introduces and outlines the main novelty of the present research, the Discourse-Mythocritical Model, which contributes to knowledge on myth criticism with a comprehensive account of the archetypal myths of war and conflict, their structuring narrative schemata and especially, the linguistic indicators of archetypal myths. The proposed list of archetypal myths is discussed in detail in this chapter, including through exemplification of their logic on a wide variety of texts (from historical text, various news sources, to music lyrics).

Subchapter 3.1. explains the conceptual facet of the model and the proposed definition of the concept of archetype and archetypal myths, including an in-depth differentiation and exemplification of archetypal myths and other myths. **Section 3.2.** explains the analytical facet of the proposed approach, focusing on the relevant analytical units and variables. To facilitate operationalisation of the model, **section 3.3.** (pp. 47-85) defines the relevant archetypal myths in detail, starting with the logic of valorisation of archetypal roles in **section 3.3.1**, and continuing with the in-depth conceptualisation of 30 archetypal myths, their dominant and, where relevant, auxiliary narrative schemata and their linguistic indicators (**the rest of subchapter 3.3.: sections 3.3.2.-3.3.15**).

The entire **section 3.3.** is a major contribution of the present study, as the comprehensive list and detailed conceptualization of the archetypal myths put forward here includes original elements such as the linguistic indicators of myth (meaning *all* linguistic indicators for all 30 archetypal myths are a novel contribution of the thesis), and in most cases, narrative schemata too, as much previous research proposes an understanding of archetypes based on symbolic and/or psychoanalytical themes and schemata, not enough (and/or not a vast enough list of archetypes) on narrative schemata. Furthermore, the thesis also proposes some original (not defined in/not proposed by previous literature) archetypal myths (**sections 3.3.4., 3.3.5., 3.3.6., 3.3.11., 3.3.14., 3.3.15.**), as well as a wide variety of reconfigured myths (renamed and reconceptualised in light of potential contemporary relevance, or in order to account for and eliminate potential limitations of gender-based naming of archetypes).

Chapter IV: Research Design

Chapter IV explains the research purpose (**section 4.1.**), objectives and research questions (**section 4.2.**), details and justifies the relevant methodological steps and tools (**section 4.3.**), the limitations of the study (**section 4.4.**) and discusses the reproducibility of the pilot study and replicability of the Discourse-Mythological Model and the novel analysis methods (**section 4.5.**).

Theoretical Thesis and the Bonus Pilot Study: Purpose, Objectives, Research Questions

Since the contribution of the present study is multifaceted, as this is first and foremost, a theoretical thesis, the main purpose is the underscoring and definition of a macro-mythopoetic discursive logic that is communication centred and operational in the pursuit of analysing the archetypal dimension of discourse. Hence, the main question (**RQ1**) is answered throughout Chapter III through (1) synthesis and reconfiguration of previous relevant research and paradigms and (2) original conceptualisation of the Discourse-Mythocritical Model, including the definitions of archetype, archetypal myth, the terms macro myth and micro myth, and the conceptualisation of 30 archetypal myths (together with their defining narrative schemata and linguistic indicators) and their exemplification.

The theoretical proposal and the development of the relevant concepts are further complemented by the investigation of mythological patterns in the form of empirical analysis of news discourse, in a pilot study meant to show the application of the theoretical and methodological proposal. Unlike research meant to fill population/geographical/data/context gaps (i.e., lack of evidence from particular data sources and contexts based on existing, tested and established theory, for instance, in news media discourse analysis, missing timeframes or missing comparisons between a variety of news sources framing a recent particular issue/media event), a pilot study's purpose is to test the feasibility of new theory, or conceptual/analytical frameworks or novel methods, all of which are relevant here. While empirical research solving a population/geographical research gap often focuses on large data samples (e.g., longer timeframe, more data sources), a pilot study focuses on smaller data sets, as the aim is to test and show the feasibility of a previously untested theory/methods and reach saturation (meaning focusing on the minimum sample size that proves the applicability of proposed novelties).

In the case of the present empirical analysis, as with many such pilot studies, the process of developing, testing and proving the applicability of the model was a laborious process requiring significant resources and time, as it included (1) the development of an analysis method of determining the relevant analytical variables, using computational linguistics tools and committing to a long trial-and-error process, and (2) the triangulation of methods for each analysis step (two or more different methods of arriving at results, to ensure and prove validity of the model). This also further explains pilot studies' focus on small data sets. However, in the case of the present study, the data sample is considerably larger than what would be necessary and sufficient for testing, as the required saturation is reached after (1) analysis of one month (timeframe) of news texts referencing a single media event, and (2) after analysis of less than 100 sentences and around 12 large sentences focusing on no more than 3 actors. Therefore, the present empirical study's contribution goes well beyond the aims of preliminary research, with exhaustive testing on larger samples and comprehensive in-depth analysis of not just one, but two media events focused on considerably longer timeframes (six months and two years, respectively): over 180.000 sentences and 161 large sentences in 3582 news articles. To tackle the empirical study, which is secondary to the main theoretical research, secondary research questions and objectives are defined, detailed and justified in **section 4.2.2.**

Methodology

Section 4.3. tackling the proposed methodology is further split into sections delving into the data collection and sampling strategy, as well as the rationale behind the focus on the two events (**section 4.3.1.**), and comprehensive detailed account of methods, tools and subsamples per research question (**sections 4.3.2. – 4.3.5.**). The detailed, time-expensive and laborious approach to data collection, sampling, and filtering is relevant due to the purpose of the study, which includes extensive testing and developing quantitative and qualitative methods, a process that becomes efficient and productive with clean data and a precise approach to sampling. To that end, the filtering process is also accompanied by two other important processes: (1) manual checking and supplementation of any missing data and (2) manually cleaning every single article text, headline and lead. All the data (complete initial, raw data, as well as the filtered and cleaned data) is included in the supplementary materials.

To answer the research questions, each corpus is examined using two different approaches that are triangulated, to demonstrate that the application of the conceptual and analytical model is possible: qualitative and quantitative. To approach the text through qualitative analysis, the present research proposes and applies a multi-step process including (1) Supervised Machine Learning to classify texts (a multi-label binary class task) for an initial clustering and only then (2) applying multi-level mythological-discursive analysis to sub-samples (per cluster), i.e., syntagmatic and paradigmatic analysis focusing on the grammar of narrative and actants, further determining archetypal myths based on the analytical categories defined in Chapter III. Furthermore, the qualitative analysis includes manual data annotation, data labelling, as well as identification and manual collection of the linguistic indicators of the coded archetypal myths for each sentence and each macro myth. The particularities of steps, methods, illustrative examples, tools, libraries, and the rationale behind the proposed approach are elaborated in **sections 4.3.2., 4.3.3.** (which are relevant to **RQ2**) and **4.3.4.** (relevant to **RQ3**), and every analysis file and output, from the qualitative mythological-discursive analysis, manual text annotation and linguistic indicators manual collection, to the relevant projects and code (using Python programming language and the PyCharm IDE) pre-processing data, fine-tuning, testing and evaluating the text classification model, BERT (Devlin et al., 2018; Kumar & Patro, 2020) are all included in the supplementary files for full reproducibility of the pilot study.

To approach the text through quantitative analysis determining the macro mythology of the relevant frequent actors/issues referenced within the news media events (**RQ4**, detailed in **section 4.3.5.**), the research proposes a novel quantitative method developed using computational linguistics tools, which is applied to the entire corpora, and separately, also to the qualitatively annotated data, resulting in the definition and individual visualisation (semantic co-occurrence networks) of a total 161 large sentences containing automatically annotated data. All 161 semantic co-occurrence networks (data visualisations), together with each analysis output applying the proposed method resulting in 550 output files, are included in the supplementary files folders.

The quantitative approach is developed and implemented in a series of steps using PyCharm (PyCharm 2025.2, 2025), a Python IDE (integrated development environment), making this another essential contribution of the present study, in the form of original code (which further research can use to apply the quantitative method to a variety of texts). Using various Python data

frame-relevant and Natural Language Processing libraries (thoroughly detailed and explained in **section 4.3.5.**), especially the spaCy library and its dependency parser (Honnibal et al., 2020), the defined functions facilitate data annotation upon identification of relevant terms designating carefully-defined and easily generalizable analytical variables. To develop, test and apply the method, the automated analysis is accompanied by qualitative validation of results (which is why every co-occurring pair of terms is accompanied by its source sentence in the output). The PyCharm projects (Jupyter Notebooks) including data visualisation codes (complete with seed numbers for full reproducibility) are included in the supplementary materials folders. To fully answer the last research question, a last qualitative analysis step is applied, i.e., a paradigmatic reading of each large sentence visualisation, which applies both deductive and inductive reasoning (detailed within the same section).

Limitations, Reproducibility and Replicability

Section 4.4. details the relevant limitations of the research. For instance, a limitation of the quantitative approach is that it does not include coreference resolution (Mitkov, 2022), a complex task that requires extensive additional testing to implement, hence the lack of incorporation here. An additional limitation of the novel quantitative method is that it could prove computationally expensive (RAM, processing time), so future research would require additional resources to apply the same functions to bigger datasets; and considering its wide generalisability to a variety of texts and research purposes (since it is built on stable analytical variables, i.e., grammatical roles, not niche dictionaries), to successfully achieve results using the same defined functions, the data sample texts need to be clean (without artifacts from automated data collection) in order to obtain accurate results. To account for any potential lack of accuracy, the present study included a laborious process of manually parsing and where needed, cleaning of each sentence in the entire corpora.

As **section 4.5.** details (reproducibility and replicability), every step and sub step of the quantitative analysis can be rerun with the expectation of producing the same results, as the research adheres to transparency regarding all levels and provides supplementary materials (approx. 550, including all the data, the qualitative analyses, the automated analysis output files and data visualizations, and the Jupyter Notebooks with all the codes and project requirements). As the present analysis is the result of a single coder, the lack of possibility of providing an inter-coder agreement (multiple

coders) is fully compensated through methods-triangulation and careful validation of all methods, from data collection to analysis and data visualisation. Therefore, regarding the reproducibility of the qualitative steps, as the quantitative validation of the qualitative data shows, it is also expected that attempts at reproducing the analysis would yield similar results overall, as the well-defined categories and well documented analysis (including all the steps and the sub steps) facilitate the process of testing and reproducing the same analysis.

In terms of replicability, the entire set of 30 analytical can be deductively applied to a large variety of texts and adapted to various languages through translation. With potential for replicability in analysis of a wide variety of texts, the analytical variables, the logic of the Discourse Mythocritical Model, as well as the novel methods, are therefore expected to be easily applicable to new data, since the model, the archetypal myths, the quantitative and qualitative facets of the methodology are thoroughly tested, explained and exemplified. As the pilot study does not seek, nor claims specific results' generalizability (since it is not meant to fill a population/geographical/context/data research gap), considering its purpose and focus, i.e., to fill an *empirical gap*, namely to test theory, the model and the proposed methods, the successful demonstration of the feasibility of the theoretical proposal, the defined model and methods demonstrate the high potential for replicability of the model and the methods. Otherwise put, as the present research advances a new conceptual and analytical model with theoretical implications, as well as a new methodology, the pilot study does not advance analysis results (the way studies preoccupied with geographical/populations gaps do), it *uses* the results to exemplify and demonstrate the validity of the model. Therefore, it is a mark of the efficiency, value and replicability of the proposed model and methods to expect different findings (in terms of the combinations of prevalent macro myths found per corpus, the list of frequent actors and their precise descriptors and actions, etc.) and a variety of combinations of mythologies when applying the same conceptual, analytical and methodological approach to new data.

Chapter V: Analysis and Results

Chapter V presents the analysis and findings of the empirical study, with the purpose of highlighting the application of the proposed conceptual and analytical model. The chapter is divided in three subchapters. The first subchapter, **segment 5.1.** focuses on the overall New York

Times' news reports analysis of archetypal myths of war and their linguistic indicators, as identified in the analysed data, thus exemplifying the operationalisation of analytical categories and the proposed model, and meeting the third objective of this research (the first two objectives are met by developing the Discourse-Mythocritical Model in Chapter III): the identification of the linguistic indicators and narrative-discursive patterns structuring whichever macro myths can be found within the analysed corpus.

Hence, **section 5.1.1.** presents the analysis and results of the patterns in mythopoetic logic of both qualitatively analysed and manually annotated corpora (**RQ2, RQ3**), also examined through triangulation of methods by applying, on the manually annotated data, the quantitative approach put forward. Furthermore, the segment details the problem definition patterns found in the qualitative data. A total of 23 myths designating actors and issues were identified and explored in the presentation of analysis results. Moreover, **segments 5.1.2. – 5.1.16** (pp. 125-169) focus on an extensive account of the top 15 most frequently represented archetypal myths. Ergo, the rest of **subchapter 5.1. sections** include an exhaustive account of the narrative and discursive logic of each myth.

To further show the application of the model in terms of identifying actors, issues, their mythological construction and therefore meet the final objective of the thesis (relevant to **RQ4**), the other two subchapters of **Chapter V, segments 5.2 and 5.3** (pp. 169-217) are dedicated to the specifics of the macro mythological construction of each of the two news events and the most prominent actors and issues.

Chapter VI: Discussion and Conclusions. Original Contributions

The final **Chapter (VI)** discusses the findings of the empirical study and the conclusions of the research, as well as the implications and contributions of the present study. **Section 6.1.** discusses the findings of the research and how The Discourse-Mythocritical Model is designed to facilitate the critical study and the identification of the macro mythological construction of conflict media events. The macro-structure aspect of archetypes is reflected in the findings of linguistic indicators of macro frames, determined through focus on narrative schemata and discursive patterns which are visible at a corpus level, not just individual article level.

Overall, the discussion of findings in **section 6.1.** (1) contextualizes the findings of the pilot study, (2) compares and contrasts the mythological-discursive construction of actors/issues within the two different media events (the Russia-Ukraine and Israel-Palestine conflicts) by the New York Times, but also highlights (3) the aspects of the study that are congruent with previous research on war journalism and the two conflicts, and (4) the aspects of the findings that contribute to knowledge on the multi-faceted mythological construction of the relevant news actors/issues for the analysed timeframe. Furthermore, (5) the section discusses the applicability and validity of the conceptual and analytical model, as well as the novel methodological approach, in light of evidence from the pilot study.

Section 6.2. of Chapter VI delves into the conclusions of the entire research, including the conclusions of the pilot study findings, and the implications and contributions of the research, as well as recommendations for future research. Essentially, the present research proposes a model that operationalizes the process of determining the particularities of the macro mythological construction of media events, thus furthering previous approaches by designing a more holistic model that accounts for various nuances in mythopoetic logic. The Discourse-Mythocritical Model is designed to address the overarching research question and support the process of identifying the particularities of the archetypal construction of discourse; hence, the research conceptualizes and cements archetype and archetypal myth in the communication science field. Future research could employ the Discourse-Mythocritical Model especially in analysis of conflict reporting, political discourse on wars and crises, populist political communication, but also a variety of other discourses which frame characters, political/social actors and issues through the logic of opposition and conflict, including social media, literature and historical texts.

The Discourse-Mythocritical Model, the proposed and thoroughly defined conflict-relevant archetypal frames and their linguistic indicators, the research design which triangulates manual and quantitative methods and innovatively leverages computational linguistics tools, and the operationalization of the model and the proposed analytical categories through case studies on news media of two frequently mediated international armed conflicts by a worldwide popular news outlet, The New York Times, constitute this study's multi-level multi-faceted contributions to communication sciences: theoretical, conceptual, methodological, as well as empirical.

More specifically, as the thesis proposes a reconceptualization of archetype and consequently, archetypal myths, the research offers a novel **conceptual** alternative in understanding and defining archetypes, which has important **theoretical** implications. Another **conceptual contribution** of this study, meant to facilitate understanding of the particularities of the archetypal myth compared to other myths, is the proposal of the two terms: *macro myth* and *micro myth*, and the in-depth exploration and conceptual differentiation of the archetypal myth (*macro ideology*) from other myths (*micro ideologies*). A major **theoretical and conceptual contribution** of this study is the comprehensive list and detailed conceptualization of the 30 archetypal myths put forward (Chapter III), including through exemplification of their logic on a wide variety of texts (from historical text, various news sources, to music lyrics). Furthermore, the thesis also proposes some original (not defined in/not proposed by previous literature) archetypal myths, as well as some reconfigured frames (renamed and reconceptualised in light of potential contemporary relevance).

The pilot study, i.e., the bonus research complementing the theoretical research, is designed to address the remaining empirical gap and test the model, hence, the contribution of the entire research includes comprehensive testing and a detailed demonstration through analysis of two media events, which is the **empirical contribution** of the thesis. Therefore, the theoretical proposal is further validated by the pilot study, which shows how the analytical variables and the paradigms (the archetypal myths) can be operationalised.

Furthermore, in terms of **analytical and methodological contributions**, the study furthers previous approaches by also advancing a more reliable step-by-step **research design** including a **methodological contribution**: the operationalisation of analytical variables through qualitative analysis of statements that, once read paradigmatically (applying the paradigms defined in Chapter III: the archetypal myths), facilitate the identification of the mythological construction of events. The same principles are applied in developing the quantitative method of tackling corpora. By facilitating the identification of narrative-discursive patterns based on syntactic dependencies, the study forwards a **novel quantitative method** of determining actants, which contributes to advancing Critical Discourse Analysis approaches to narrative discourse.

And finally, since the pilot study's data sample is larger (in terms of data sample and timeframes) than what would be sufficient for demonstrating the feasibility of the model and thus for closing the relevant empirical gap, the pilot study's contribution goes well beyond the aims of preliminary

research, with exhaustive testing and comprehensive in-depth analysis of not just one, but two media events. As such, beyond the main theoretical, conceptual, analytical, methodological and empirical contributions of the research discussed so far, an additional **empirical contribution** is seen in the results of the analysis, in the identification of the macro-mythological-discursive construction of the two media events by the analysed news source. Future research can employ the model and the proposed methods to determine the mythological particularities of not just the news discourse on the two conflicts (for different timeframes or different news sources), but also different media events, as well as a variety of other discourses.

Bibliography

This bibliographic list contains all references discussed within this abstract, and therefore represents only a part of the full PhD thesis' reference list.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 1(Mlm)*, 4171–4186.

Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). *spaCy: Industrial-strength Natural Language Processing in Python*. <https://doi.org/10.5281/zenodo.1212303>

Kumar, P. V, & Patro, T. (2020). *Bert Model with 0.845 Accuracy*. Kaggle.Com.
<https://www.kaggle.com/code/vpkprasanna/bert-model-with-0-845-accuracy/notebook>

Mitkov, R. (2022). *The Oxford Handbook of Computational Linguistics 2nd edition*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199573691.001.0001>

PyCharm 2025.2. (2025). *PyCharm Documentation*. JetBrains.Com.
<https://www.jetbrains.com/help/pycharm/2025.2/getting-started.html>