

Medical Context Variation: A source of impairment for Event classification

Aman Sinha^{1,2,3} Marianne Clausel¹ Mathieu Constant² Xavier Coubez³

¹IECL, University of Lorraine, France

²ATILF, CNRS, Université de Lorraine, Nancy, France

³Institut Strauss, Strasbourg, France

Correspondence: aman.sinha@univ-lorraine.fr

Abstract

The variation in writing style encapsulates nuanced characteristics, which are often exploited for author or demographic identification. In the medical domain, language models are frequently deployed to capture relevant information from unstructured or complex data, such as clinical notes that often include patients' medical histories. Such data is largely free-form and unstructured, obtained through diverse clinician–patient interactions. In this work, we present a case study investigating whether variations in clinicians' writing styles can lead to differences in medical context understanding capabilities for pre-trained language models (PLMs) on five downstream tasks related to medical event classification. Our findings indicate that variation in writing style, characterized by linguistic features, can indeed lead to suboptimal performance in deployed NLP models. Furthermore, we explore linguistically guided counterfactual reasoning to mitigate the impact of writing-style variation, suggesting that LLM-based stylistic normalization is effective for this purpose.

1 Introduction

Clinical narrative plays important role in determining the history of the patient. It helps clinicians in answering prognostic inquiries and is essential for identifying the confounding between disease progression and interventions (Pham et al., 2017). The documentation writing practices are designed to be chronological and divided into identifiable sections, such as patient's prior medical context, physical examination, possible diagnosis and actions (Bansler et al., 2016). Therefore, the narrative found in clinical notes can depend on the writing style of the clinical staff. Furthermore, the variation in writing can be a source of communicating incomplete, ambiguous or at times wrong information.

For eg., In the sentence – *The infusion rate was then increased and blood was taken 4,8,12 and*

20 minutes after the new target concentration had been achieved. The writer were interested in how quickly the concentration reached its new steady-state after the rate of infusion had been increased; they took blood 4, 8, 12 and 20 minutes after taking the action that sought to produce a new target concentration, that is, after altering the rate of infusion. The investigators could not know the concentration of the drug in the blood until after analysis some days later. A recommended alternative (Goodman and Edwards, 2014) for the text is as follows – *The infusion rate was then increased and blood was taken 4, 8, 12 and 20 minutes later, or (although it is implicit and not strictly necessary) ... after the new target concentration had been set.* This sentence clarifies the order of the actions and the events happening.

English for Medical Purposes (EMP) is a key focus in medical education, with research emphasizing the importance of collocation acquisition by EFL learners (Džuganová, 2019). Medical writing often suffers from issues such as overly complex sentences, excessive passive voice, weak verb usage, and medical jargon, all of which hinder clarity and comprehension (Goodman and Edwards, 2014). Given that free-text clinical notes make up about 80% of clinical data, models trained on such text are likely influenced by stylistic and demographic variation. While large-scale training may offer robustness to writing style differences, this assumption has not been systematically evaluated for language models.

In this work, we present a case study focusing on linguistic variation, in particular writing style variation, that is found in the clinical notes for the task of extracting contextual information related to medicinal changes. The primary objective is to investigate *whether the characteristics of writing style in clinical notes has subsequent impact on language models* for event classification tasks. Our key contributions in this paper are as follows:

- (a) We provide a thorough linguistic feature analysis to find empirical evidence for stylistic variation in clinical notes has impact on capability of pre-trained language models.
- (b) We show that linguistically-guided counterfactual reasoning can be used to mitigate the impact of writing style variation on LMs.
- (c) Additionally, we study the generated counterfactuals for medical relevance for data augmentation.

2 Background

Style Variation Style variation is a form of linguistic variation has been studied for describing and explaining variation in linguistic form across society (Nguyen et al., 2021). Stylistic Vagueness or variation is very well acknowledged in medical writing (Mitchell, 2007) and is advised to avoid as much as possible in medical communication. Although, the existence of any linguistic variation can be attributed to demographics of international doctors across the world, majority of whom have English as their second-language (Verma et al., 2016). Linguistic variation between user and the content of clinical notes already has studied for medical dialog systems (Llanos et al., 2016). However, the impact of the writing style variation has not been studied thoroughly for language models for medical context understanding.

Medical Contextual Understanding Clinical notes are narrative and longitudinal, capturing a patient’s medical history, assessments, and care plans across time. These texts vary widely in structure and writing style depending on the provider, institution, and note type (Pivovarov and Elhadad, 2015). Events such as medication changes are often embedded in nuanced context — temporal cues, attribution, and intent — that must be accurately captured for tasks like medication reconciliation (Cadwallader et al., 2013). While clinical event extraction has received significant attention, the effect of writing-style variation on the medical contextual understanding of language models remains underexplored (Xu et al., 2022).

Counterfactual Explanations We consider counterfactual explanation (Wachter et al., 2017; Sinha et al., 2023) as a mean to verify and potentially mitigate the impact the different linguistic feature that contribute to writing style variation in clinical notes on medical language models’ capabilities for contextual medical event

extraction task. Various works that exist inline with this approach such as POLYJUICE (Wu et al., 2021), MICE (Ross et al., 2021) which provide a pipeline for counterfactual generation and evaluation. However, there is a lack of exploration into mitigating medical writing style variation, which our work addresses.

3 Methodology

Dataset CMED (Mahajan et al., 2022) dataset is a collection of clinical notes that are associated to the patients who were admitted for the Type II Diabetes (T2D). The dataset consists of total 500 clinical notes from two different hospitals (See Appendix A for more details). Each clinical note was annotated by a team of three annotators led by a physician for medication events extraction and context annotations.

Contextual Event Extraction The contextual medical event extraction was introduced in n2c2 challenge¹. It involved two-fold tasks, firstly, it requires to identify medicine for which any *medication change*² is being discussed in the clinical note and secondly, different contextual information associated the medication change has to be extracted. For any medication change that is identified as *disposition*, the model is expected to extract five different contextual information regarding the Action, Negation, Temporality, Certainty and Actor that are associated to for each identified medication change (Mahajan et al., 2022). See Appendix A for more details.

Linguistic Characterization of Clinical Narrative We utilize BEA linguistic feature extractor tool (Lee and Lee, 2023) in order to extract linguistic features from any segmented medical sentence of the clinical notes. The tool facilitates to extract 220 linguistic feature per sentence (See Table 19,20). These features cover various linguistic characteristics on sentence-level such as surface, syntax, lexico-semantics, and discourse. We utilized SciSpacy³ instead of general domain modules for processing the segmented sentences.

Models We consider four pre-trained language models as follows: BERT (Devlin et al., 2018),

¹<https://n2c2.dbmi.hms.harvard.edu/>

²Medication change, is defined as “any discussion about a medicine change” and is termed to as *disposition*, for a given patient mentioned in the clinical note.

³`en_core_sci_sm`

Linguistic Feature (<i>lf</i>)	Medical Sentence	Modified Sentence	BEA(<i>s</i>)	BEA(<i>s_l</i>)
#unique_nouns	–check chem 7 given recent restart BP meds	–check chemistry 7 given recent restart blood pressure medications	3	5
simple_verbs_variation	5 x2 and lopressor 12.	5 multiplied by 2 and lopressor 12.	0	1
#unique_adverbs	Pain was relieved by Morphine (4, 4, 2 mg).	Pain was completely relieved by Morphine (4, 4, 2 mg).	0	1
root_verbs_variation	8. Neurontin 200 mg b.i.d. (she was told she could discontinue	Neurontin 200 mg b.i.d. (she was informed she could cease taking it).	1.414	1.732
#named_entities_person	We will have her start the Norvasc and Cozaar, have her return in two weeks for followup of blood pressure.	We will have Jane start the Norvasc and Cozaar, have Jane return in two weeks for followup of blood pressure.	0	2
simple_type_token_ratio	She was treated with IV lopressor 2.	She received IV lopressor twice .	1.125	1.2
avg_syllables_per_word	A1c off it 3/94 8.	A1c off it three over ninety-four eight .	0.667	1.25

Table 1: Samples showing linguistic features modification. BEA(..) denotes linguistic feature counter.

<i>lf</i>	<i>D</i>	<i>D(l)</i>	Action	Actor	Certainty	Negation	Temporality
#unique_nouns	2.94	3.55	neg.	pos.	pos.	pos.*	pos.
simple_verbs_variation	0.89	0.95	pos.*	neg.	neg.	pos.*	neg.
#unique_adverbs	0.60	1.04	pos.*	pos.	pos.	pos.*	pos.
root_verbs_variation	1.23	1.34	neg.*	pos.	pos.	neg.*	pos.
#named_entities_person	0.27	0.65	/	pos.	/	/	/
simple_type_token_ratio	1.10	1.13	neg.*	pos.*	pos.	neg.*	pos.
avg_syllables_per_word	1.41	1.67	pos.*	pos.	neg.*	pos.*	neg.

Table 2: Mean shift for selected linguistic features and their correlation impact (pos./neg.) on context categories. * denotes non-significant, / denotes no correlation. *D*: original dataset and *D(l)*: modified dataset.

ModernBERT (Warner et al., 2024), ClinicalBERT (Huang et al., 2019) and PubmedBERT (Gu et al., 2021). For each of these BERT-like models, we train classification model using the segmented sentences containing the target medicine as an input against each of the five context category labels for 20 epochs with early stopping. Details for reproducibility in Appendix C.

4 Experimental Setup

Each clinical note is segmented at the sentence level, and then sentences containing a medication change annotated with a *disposition* label are retained. We train the different LMs to obtain predictions on the development and test sets. Additionally, we extract 220 linguistic features using BEA linguistic feature extraction library for selected disposition based sentences. To test our hypothesis, we first define sample-level *correctness* by comparing ensemble predictions of all the language models with gold labels. Then, with the obtained ensemble predictions, we examine correlations between linguistic features and correctness via the OLS method (Kuchibhotla et al., 2019) across the five context categories.

Based on obtained correlation analysis, we select a subset of features and list them with their expected impact in Table 2. Next, we use GPT-

<p>You are an assistant supporting a linguistic study. Your task is to rewrite the given sentence by modifying the following linguistic feature: [feature]. Return only the modified sentence, without additional text or explanation.</p> <p>Sentence: "sentence"</p>
--

Table 3: Prompt used for generating sentence explanations.

3.5-Turbo to modify validation and test segmented sentences to alter linguistic features for counterfactual evaluation using the prompt (See Table 3). We provide examples of modified sentence with respect to a set of linguistic features in Table 1). Finally, the modified sentences are used for inference with the trained LMs to verify the if the counterfactual are able to change the predictions or not as expected.

5 Results & Discussion

Table 4 shows the percentage of linguistic features (*lf*) that were significantly correlated with prediction mistakes in dev and test set across the five context tasks (See details in Table 12 - 18). We consider seven most common linguistic features for which we obtain *llm* modification across the five context tasks. Table 1 presents examples of linguistic modifications generated via GPT-3.5-Turbo for each of the selected linguistic feature. Additionally, we measure and provide the overall feature shift shown in Table 2 as modified dataset compared to the original average *lf* distribution. Table 5 presents the detailed results on original and modified dataset for test set. We also report broader linguistic categorization for each selected features based on (Lee and Lee, 2023) documentation in Table 11.

Finding I: Writing style variation impacts models’ ability to understand clinical narratives.

CONTEXT	Action	Actor	Certainty	Negation	Temporality
DEV	3.64	42.27	27.73	12.27	67.73
TEST	11.82	37.73	21.36	7.73	9.55

Table 4: % of linguistic features that are significantly ($\rho < 0.05$) correlated with prediction mistakes.

Among the five context tasks, *Actor*, *Certainty*, and *Temporality* are most affected by variation in linguistic features (see Table 4). In the test set, *Negation* appears to be least influenced by linguistic features, whereas in the development set, *Action* showed the least impact, indicating inconsistency across datasets. As shown in Table 11, lexico-semantic and syntactic feature families account for the majority of significant correlations with model prediction correctness, suggesting that models may be relying on surface stylistic cues rather than meaningful medical context distinctions. The exception is observed for *Actor* and *Negation*, where discourse-related linguistic features play a dominant role. Overall, these findings reinforce that writing style variation substantially shapes how models interpret and process clinical narratives.

Models	Action	Actor	Certainty	Negation	Temporality
BERT	51.84	44.70	52.80	49.54	48.30
ModernBERT	43.55	35.83	42.54	63.90	53.58
BioClinicalBERT	49.13	50.38	48.77	49.54	50.12
PubmedBERT	71.48	49.95	53.65	49.54	39.37
ENSEMBLE	52.50	50.20	50.20	49.50	48.50
#unique_nouns	-0.7	+1.0	-1.9	± 0.0	+1.5
simple_verbs_variation	+2.1	-6.3	+5.2	± 0.0	+0.3
#unique_adverbs	-1.0	+0.9	-1.9	± 0.0	-1.7
root_verbs_variation	-0.7	-3.0	+0.6	± 0.0	-1.3
#named_entities_person	-0.3	-2.2	± 0.0	± 0.0	+0.1
simple_type_token_ratio	-0.2	-6.8	+3.8	± 0.0	+1.5
avg_syllables_per_word	-2.7	-4.1	-1.2	± 0.0	+2.6

Table 5: Ensemble context classification results (Macro F1) under counterfactual feature shift on the test set. Green indicates expected outcomes, red indicates unexpected outcomes, and gray denotes excluded cases.

Finding II: Linguistically-guided counterfactuals show potential to mitigate impact of writing style variation. Compared to ensemble baseline performance across all the five tasks, counterfactual modified dataset leads to both gains and losses depending on the linguistic feature are modified (See Table 5). Complete results are in the Appendix B. Concretely, for overall $\sim 60\%$ cases the modified dataset via counterfactuals leads to the expected behavior. *Negation* remains relatively invariant across all feature shifts, however for *Actor*, *Certainty* and *Temporality*, we obtain mixed behavior. *Actor*, *Cer-*

	Modification Type	Example
LOW	Underspecification	This is rather a simple infection , and may be treated with Ancef. This is rather a simple cellulitis, and may be treated with Ancef.
	Overspecification	Reinitiated Lasix promptly . Reinitiated Lasix.
MODERATE	Entity	She was started on Xenical by two doctors . She was started on Xenical by Internal Medicine.
	Coreference	insulin dosing per Dr. Nettles. Dr. Nettles has been eating and drinking insulin dosing per Dr. Nettles. She has been eating and drinking
HIGH	Content Violation	No John with Mary given possible Sarah bleed. No heparin with HD given possible GI bleed.
	Numeric	2 liters of IV fluid, came up to 90s/50s. 5 liters of IV fluid, came up to 90s/50s.

Table 6: Different LLM generated counterfactuals categorized based on medical risk.

	if	LOW	MODERATE	HIGH
#unique_nouns		79.6	10.2	10.2
simple_verbs_variation		65.3	14.3	20.4
#unique_adverbs		89.8	8.2	2.0
root_verbs_variation		98.0	0.0	2.0
#named_entities_person		65.3	14.3	20.4
simple_type_token_ratio		95.9	2.0	2.0
avg_syllables_per_word		93.9	6.1	0.0

Table 7: Categorization of Counterfactual Explanation based on clinical relevance risk.

tainty are the most affected showing significant performance change under *simple_verbs_variation* and *simple_type_token_ratio* respectively. Conversely, *Action* and *Negation* seems to experience least changes on when subjected to linguistic feature shift. These results suggest that further confirms that linguistic variations can both degrade and enhance contextual understanding suggesting a more non trivial phenomena.

Finding III: Linguistic-guided counterfactuals are double-edged sword.

The use of LLM to generate counterfactuals by default is prone to produce hallucinations. Therefore, we analyse the modified sentences (see Table 1) for hallucination to understand the clinical relevance distortion⁴. We consider 50 segmented sentences that are modified with respect to seven linguistic features via GPT-3.5-Turbo. We manually annotate each of the modified sentence ($50 \times 7 = 350$) for the level of potential medical relevance distortion risk into three categories: low, moderate, and high (see Table 6). For LOW risk, we consider modifications that involve under-specification of any linguistic unit i.e., plausible but unsupported modifiers. For MODER-

⁴Medical relevance distortion denotes to what extent the counterfactual deviates from original sentence on clinical ground.

ATE risk, we consider modification of clinically relevant content while remaining linguistically plausible. For HIGH risk, we consider violations of transformation constraints.

Among the seven features we utilized for generating counterfactuals, based on 50 instances, we observed that *named_entities_nouns*, *unique_nouns* leads to high risk counterfactuals whereas *root_verbs_variation* leads to low risk counterfactuals. We present in distribution in Table 7. While HIGH risk modifications, due to distortion of medical relevance, can mislead the models to deviate from intended downstream tasks. LOW and MODERATE level modifications can be useful to augment the model to overcome class imbalance or lack of clinical annotated data.

6 Conclusion

In this study, we investigated the impact of writing style variation on medical context classification task performance of pre-trained language models. Our findings indicate that stylistic variation in clinical narratives can lead to difference in medical context interpretation of language models. Our study highlights the impact of considering writing style heterogeneity, suggesting that incorporating stylistic normalization via counterfactual explanation generation could enhance model robustness. While the counterfactuals explanation can often require clinical validation, otherwise, it can positively contribute to address the lack of clinical annotated data.

Limitations

While the study highlights the impact of writing style variation on language models in medical domain, we acknowledge the following limitations to the best of our knowledge.

- Firstly, the study focuses primarily on linguistic features and does not account for other potential sources of variation, such as differences in clinical knowledge or documentation practices.
- Secondly, our case study examines the effect on selected downstream tasks from correlation angle, and therefore findings may not generalize to all possible confounders.
- And finally, our study focuses on pre-trained language models considering practical deploy-

ment settings and encourage future studies to generalize the finding of LLMs.

Acknowledgments

We would like to extend our thanks to the anonymous reviewers. Additionally, we would also like to express thanks to Cristina Garcia Holgado and Timothée Mickus for the valuable discussion on the initial version of this work.

References

- Jørgen Bansler, Erling Havn, Kjeld Schmidt, Troels Mønsted, Helen Petersen, and Jesper Svendsen. 2016. [Cooperative epistemic work in medical practice: An analysis of physicians' clinical notes](#). *Computer Supported Cooperative Work*, 25.
- Justin Cadwallader, Kenneth Spry, Justin Morea, AL Russ, Jon Duke, and Michael Weiner. 2013. Design of a medication reconciliation application. *Applied clinical informatics*, 4(01):110–125.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. [BERT: pre-training of deep bidirectional transformers for language understanding](#). *CoRR*, abs/1810.04805.
- Božena Džuganová. 2019. Medical language-a unique linguistic phenomenon: Engleski. *Jahr–European Journal of Bioethics*, 10(1):129–145.
- Neville W Goodman and Martin B Edwards. 2014. *Medical writing: a prescription for clarity*. Cambridge University Press.
- Yu Gu, Robert Tinn, Hao Cheng, Michael Lucas, Naoto Usuyama, Xiaodong Liu, Tristan Naumann, Jianfeng Gao, and Hoifung Poon. 2021. [Domain-specific language model pretraining for biomedical natural language processing](#). *ACM Trans. Comput. Healthcare*, 3(1).
- Kexin Huang, Jaan Altosaar, and Rajesh Ranganath. 2019. Clinicalbert: Modeling clinical notes and predicting hospital readmission. *arXiv preprint arXiv:1904.05342*.
- Arun K Kuchibhotla, Lawrence D Brown, Andreas Buja, and Junhui Cai. 2019. All of linear regression. *arXiv preprint arXiv:1910.06386*.
- Bruce W. Lee and Jason Lee. 2023. [LFTK: Handcrafted features in computational linguistics](#). In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, pages 1–19, Toronto, Canada. Association for Computational Linguistics.
- Leonardo Campillos Llanos, Dhouha Bouamor, Pierre Zweigenbaum, and Sophie Rosset. 2016. Managing linguistic and terminological variation in a medical

dialogue system. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3167–3173.

Diwakar Mahajan, Jennifer J. Liang, and Ching-Huei Tsou. 2022. [Toward Understanding Clinical Context of Medication Change Events in Clinical Narratives](#). *AMIA Annual Symposium Proceedings*, 2021:833–842.

SLM Mitchell. 2007. Medical writing: A prescription for clarity.

Dong Nguyen, Laura Rosseel, and Jack Grieve. 2021. On learning and representing social meaning in nlp: a sociolinguistic perspective. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human language technologies*, pages 603–612.

Trang Pham, Truyen Tran, Dinh Phung, and Svetha Venkatesh. 2017. Predicting healthcare trajectories from medical records: A deep learning approach. *Journal of biomedical informatics*, 69:218–229.

Rimma Pivovarov and Noémie Elhadad. 2015. Automated methods for the summarization of electronic health records. *Journal of the American Medical Informatics Association*, 22(5):938–947.

Alexis Ross, Ana Marasović, and Matthew Peters. 2021. [Explaining NLP models via minimal contrastive editing \(MiCE\)](#). In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3840–3852, Online. Association for Computational Linguistics.

Aman Sinha, Cristina Garcia Holgado, Marianne Clausel, Mathieu Constant, and Xavier Coubez. 2023. Can llms be used to understand clinical notes better? *5èmes journées du Groupement de Recherche CNRS «Linguistique Informatique, Formelle et de Terrain» LIFT 2023*, page 94.

Anju Verma, Ann Griffin, Jane Dacre, and Andrew Elder. 2016. Exploring cultural and linguistic influences on clinical communication skills: a qualitative study of international medical graduates. *BMC Medical Education*, 16(1):162.

Sandra Wachter, Brent Mittelstadt, and Chris Russell. 2017. Counterfactual explanations without opening the black box: Automated decisions and the gdpr. *Harv. JL & Tech.*, 31:841.

Benjamin Warner, Antoine Chaffin, Benjamin Clavié, Orion Weller, Oskar Hallström, Said Taghadouini, Alexis Gallagher, Raja Biswas, Faisal Ladhak, Tom Aarsen, and 1 others. 2024. Smarter, better, faster, longer: A modern bidirectional encoder for fast, memory efficient, and long context finetuning and inference. *arXiv preprint arXiv:2412.13663*.

Tongshuang Wu, Marco Tulio Ribeiro, Jeffrey Heer, and Daniel Weld. 2021. [Polyjuice: Generating counterfactuals for explaining, evaluating, and improving](#)

[models](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6707–6723, Online. Association for Computational Linguistics.

Zhuo Xu, Yue Wang, Lu Bai, and Lixin Cui. 2022. Writing style aware document-level event extraction. *arXiv preprint arXiv:2201.03188*.

A More Information for CMED Dataset

The CMED dataset contains clinical notes that came from the Partners HealthCare Electronic Medical Records (EMR). EMR at Partners HealthCare comprises a platform shared by two large academic tertiary hospitals – Massachusetts General Hospital (MGH) and Brigham and Women’s Hospital (BWH), USA. The five context categories description are below:

Action	What is the change discussed?
Actor	Who initiated the change?
Certainty	How likely is this change to have occurred / will occur?
Negation	Is the change being discussed negated?
Temporality	When is this change intended to occur?

Table 8: Description of Medical contextual event classification tasks.

The distribution of classes across the different context tasks for train set is shown in Figure 1.

B Detailed Test Results

Models	Action	Actor	Certainty	Negation	Temporality
BERT	51.84	44.70	52.80	49.54	48.30
ModernBERT	43.55	35.83	42.54	63.90	53.58
BioClinicalBERT	49.13	50.38	48.77	49.54	50.12
PubmedBERT	71.48	49.95	53.65	49.54	39.37
ENSEMBLE	52.50	50.20	50.20	49.50	48.50
#unique_nouns	51.80	51.20	48.30	49.50	50.00
simple_verbs_variation	54.60	43.90	55.40	49.50	48.80
#unique_adverbs	51.50	51.10	49.50	49.50	46.80
root_verbs_variation	51.80	47.20	50.80	49.50	47.20
#named_entities_person	52.20	48.00	50.20	49.50	48.60
simple_type_token_ratio	52.30	434.00	54.00	49.50	50.00
avg_syllables_per_word	49.80	46.10	49.00	49.50	51.10

Table 9: Ensemble context classification results (Macro F1) under counterfactual feature shift on the test set (Complete Results).

Table 9 presents the experiment results in detail. The top section contains the five context categories across all four PLMs, along with the ENSEMBLE results. The bottom section comprises ensemble results corresponding to linguistic features listed in each row when the models are trained with a

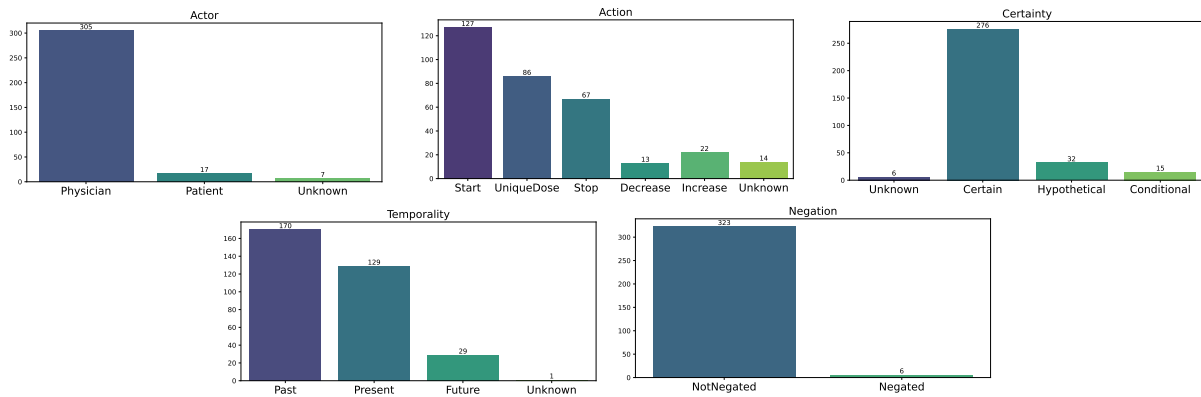


Figure 1: Distribution of context labels across CME2 trainset.

dataset with modified linguistic features. For instance, compared to the ENSEMBLE, when the model is trained with the dataset with modified #unique_nouns feature, the performance of the ENSEMBLE changes in an expected manner (See table 2) for Action, Actor, Temporality, highlighted by green, and unexpected manner for Certainty.

C Reproducibility

All the model were trained on the same hyperparameter setting.

Batch Size	4
Learning Rate	1e-5
Loss Function	Focal Loss
EPOCHS	20
Early stopping	True
Patience	3
MAX LEN	128

Table 10: Hyperparameter Setup.

D Linguistic Features

Table 19-20 list all the linguistic features that can be extracted from BEA (Lee and Lee, 2023) extractor tool.

Action	Actor	Certainty	Negation	Temporality
avg_brybaert_age_of_acquisition_of_words (avgworddiff, lex-sem)	#unique_verbs (partofspeech, syntax)	corrected_verbs_variation (lexicalvariation, lex-sem)	simple_adpositions_variation (avgworddiff, lex-sem)	#unique_punctuations (partofspeech, syntax)
corrected_adverbs_variation (lexicalvariation, lex-sem)	#named_entities_cardinal (entity, discourse)	root_verbs_variation (lexicalvariation, lex-sem)	#named_entities_product (entity, discourse)	simple_type_token_ratio_no_lemma (typetokenratio, lex-sem)
root_adverbs_variation (avgworddiff, lex-sem)	avg_adverbs_per_word (partofspeech, syntax)	avg_particles_per_word (partofspeech, syntax)	smog_index (readformula, surface)	simple_type_token_ratio (typetokenratio, lex-sem)
#unique_nouns (partofspeech, syntax)	#named_entities_loc (entity, discourse)	avg_adjectives_per_word (partofspeech, syntax)	avg_named_entities_loc_per_word (entity, discourse)	simple_punctuation_variation (lexicalvariation, lex-sem)
simple_numerals_variation (lexicalvariation, lex-sem)	avg_named_entities_norp_per_sent (entity, discourse)	#unique_nouns (avgworddiff, lex-sem)	avg_named_entities_time_per_sent (entity, discourse)	#unique_nouns (partofspeech, syntax)

Table 11: Top 5 significant linguistic features for each context category in the order of decreasing effect size.

Action	Coef.	Std.Err.	t	P> t	[0.025	0.975]
average_brybaert_age_of_acquisition_of_words_pe...	-1.332967	0.528185	-2.523674	1.388597e-02	-2.386399	-0.279535
corrected_adverbs_variation	5175.404330	2283.831956	2.266106	2.653914e-02	620.445122	9730.363538
root_adverbs_variation	-5201.900492	2296.742110	-2.264904	2.661663e-02	-9782.608191	-621.192792
total_number_of_unique_nouns	-13.631678	6.105775	-2.232588	2.877686e-02	-25.809262	-1.454095
simple_numerals_variation	-12.722929	5.826538	-2.183617	3.234429e-02	-24.343593	-1.102265
total_kuperman_age_of_acquisition_of_words	37.321799	18.411220	2.027123	4.645964e-02	0.601778	74.041819
average_kuperman_age_of_acquisition_of_words_per...	-36.646022	18.083067	-2.026538	4.652102e-02	-72.711562	-0.580481

Table 12: Action

Actor	Coef.	Std.Err.	t	P> t	[0.025	0.975]
total_number_of_unique_verbs	-6.603329e+00	8.827628e-01	-7.480298	1.650764e-10	-8.363944e+00	-4.842714e+00
total_number_of_named_entities_cardinal	1.423989e-08	2.126493e-09	6.696419	4.446148e-09	9.998731e-09	1.848104e-08
average_number_of_adverbs_per_word	1.320402e+00	2.292887e-01	5.758689	2.077356e-07	8.631005e-01	1.777704e+00
total_number_of_named_entities_loc	2.985107e-08	5.366359e-09	5.562631	4.546192e-07	1.914821e-08	4.055394e-08
average_number_of_named_entities_norp_per_sentence	3.550854e-08	6.615868e-09	5.367178	9.835470e-07	2.231361e-08	4.870347e-08
average_number_of_named_entities_ordinal_per_se...	-1.419000e-08	2.679636e-09	-5.295494	1.302100e-06	-1.953436e-08	-8.845632e-09
total_number_of_named_entities_date	7.829471e-08	1.498465e-08	5.224995	1.713522e-06	4.840878e-08	1.081806e-07
simple_verbs_variation	-3.609825e+00	7.089304e-01	-5.091932	2.866104e-06	-5.023743e+00	-2.195908e+00
total_number_of_unique_adverbs	7.335495e+00	1.457505e+00	5.032912	3.594641e-06	4.428592e+00	1.024240e+01
total_number_of_named_entities_money	6.785820e-08	1.353844e-08	5.012262	3.890109e-06	4.085664e-08	9.485977e-08
simple_coordinating_conjunctions_variation	9.094260e+00	1.823606e+00	4.986966	4.284578e-06	5.457193e+00	1.273133e+01
simple_determiners_variation	-3.627148e+00	7.319451e-01	-4.955492	4.830352e-06	-5.086966e+00	-2.167329e+00
average_number_of_named_entities_date_per_word	6.510582e-08	1.318313e-08	4.938572	5.151270e-06	3.881290e-08	9.139874e-08
simple_adverbs_variation	9.767410e+00	1.989599e+00	4.909235	5.757781e-06	5.799280e+00	1.373554e+01
total_number_of_named_entities_norp	-3.300574e-07	6.728979e-08	-4.905014	5.850624e-06	-4.642626e-07	-1.958521e-07
root_adverbs_variation	-2.730079e+03	5.568992e+02	-4.902286	5.911379e-06	-3.840780e+03	-1.619379e+03
average_number_of_named_entities_law_per_word	-8.650389e-09	1.764684e-09	-4.901949	5.918936e-06	-1.216994e-08	-5.130839e-09
corrected_adverbs_variation	2.712329e+03	5.537689e+02	4.897944	6.009383e-06	1.607872e+03	3.816786e+03
average_number_of_coordinating_conjunctions_per...	9.809528e-01	2.006043e-01	4.889989	6.193042e-06	5.808601e-01	1.381045e+00
average_number_of_syllables_per_word	2.316189e+02	4.836869e+01	4.788613	9.072603e-06	1.351506e+02	3.280872e+02
total_number_of_named_entities_ordinal	-2.596594e-08	5.438161e-09	-4.774765	9.555796e-06	-3.681201e-08	-1.511987e-08
average_number_of_auxiliaries_per_word	1.251121e+00	2.645758e-01	4.728781	1.134728e-05	7.234414e-01	1.778801e+00
average_number_of_pronouns_per_word	1.311673e+00	2.800749e-01	4.683294	1.343990e-05	7.530815e-01	1.870265e+00
average_number_of_nouns_per_word	2.257798e+00	4.860383e-01	4.645308	1.547154e-05	1.288425e+00	3.227171e+00
total_number_of_unique_nouns	6.716098e+00	1.480489e+00	4.536405	2.309635e-05	3.763356e+00	9.668840e+00
average_number_of_named_entities_event_per_word	2.472929e-08	5.466195e-09	4.524040	2.416458e-05	1.382730e-08	3.563127e-08
total_number_of_unique_determiners	-3.375619e+00	7.541151e-01	-4.476264	2.876153e-05	-4.879654e+00	-1.871584e+00
root_verbs_variation	1.868781e+03	4.193529e+02	4.456344	3.091963e-05	1.032408e+03	2.705154e+03
corrected_verbs_variation	-1.860470e+03	4.178374e+02	-4.452618	3.134043e-05	-2.693820e+03	-1.027120e+03
average_number_of_named_entities_language_per_word	4.124240e-08	9.411679e-09	4.382045	4.044268e-05	2.247140e-08	6.001340e-08
total_number_of_named_entities_language	-4.495325e-08	1.068470e-08	-4.207255	7.539254e-05	-6.626321e-08	-2.364329e-08
average_number_of_adpositions_per_word	1.292568e+00	3.124311e-01	4.137131	9.644439e-05	6.694440e-01	1.915692e+00
root_auxiliaries_variation	7.020887e+02	1.715549e+02	4.092502	1.126829e-04	3.599333e+02	1.044244e+03
average_number_of_named_entities_law_per_sentence	-1.571381e-08	3.840805e-09	-4.091279	1.131628e-04	-2.337405e-08	-8.053563e-09
colemans_liau_index	3.698407e+00	9.062749e-01	4.080889	1.173224e-04	1.890899e+00	5.505916e+00
corrected_auxiliaries_variation	-6.962070e+02	1.707454e+02	-4.077808	1.185835e-04	-1.036808e+03	-3.557260e+02
average_number_of_named_entities_org_per_sentence	1.349471e-08	3.326202e-09	4.057091	1.274113e-04	6.860804e-09	2.012861e-08
average_number_of_named_entities_date_per_sentence	1.978967e-08	4.912671e-09	4.028292	1.407396e-04	9.991661e-09	2.958769e-08
total_number_of_named_entities_time	7.983593e-08	2.011475e-08	3.969023	1.725146e-04	3.971832e-08	1.199535e-07
average_number_of_named_entities_fac_per_sentence	1.480192e-08	3.734692e-09	3.963358	1.758897e-04	7.353312e-09	2.225053e-08
average_number_of_named_entities_loc_per_sentence	1.433150e-08	3.616952e-09	3.962314	1.765185e-04	7.117716e-09	2.154528e-08
average_number_of_named_entities_language_per_s...	3.009464e-08	7.654272e-09	3.931745	1.959175e-04	1.482868e-08	4.536061e-08
average_number_of_spaces_per_word	1.032595e+00	2.675687e-01	3.859178	2.504993e-04	4.989463e-01	1.566244e+00
corrected_adjectives_variation	4.139136e+03	1.074507e+03	3.852126	2.565202e-04	1.996099e+03	6.282173e+03
root_adjectives_variation	-4.143990e+03	1.076025e+03	-3.851202	2.573188e-04	-6.290054e+03	-1.997926e+03
smog_index	-1.740051e-01	2.018465e-01	-3.834623	2.720674e-04	-1.176575e+00	-3.714350e-01
total_number_of_words_more_than_three_syllables	1.110019e+00	2.919974e-01	3.801468	3.040247e-04	5.276483e-01	1.692389e+00
average_number_of_verbs_per_word	1.098956e+00	2.953297e-01	3.721116	3.970383e-04	5.099396e-01	1.687973e+00
root_coordinating_conjunctions_variation	-5.918521e+03	1.630449e+03	-3.629995	5.353012e-04	-9.170348e+03	-2.666693e+03
corrected_coordinating_conjunctions_variation	5.907623e+03	1.629720e+03	3.624931	5.441962e-04	2.657249e+03	9.157997e+03
average_number_of_proper_nouns_per_word	1.291358e+00	3.613331e-01	3.573872	6.421061e-04	5.707021e-01	2.012015e+00
average_number_of_named_entities_quantity_per_s...	-6.742536e-09	1.893705e-09	-3.560499	6.703914e-04	-1.051941e-08	-2.965659e-09
average_number_of_subordinating_conjunctions_pe...	4.302424e-01	1.223792e-01	3.515651	7.741333e-04	1.861648e-01	6.743199e-01
average_number_of_named_entities_cardinal_per_s...	-2.051229e-08	5.878661e-09	-3.489279	8.420618e-04	-3.223691e-08	-8.787671e-09
average_number_of_named_entities_time_per_word	2.397853e-08	7.173665e-09	3.342577	1.335283e-03	9.671104e-09	3.828595e-08
reading_time_for_slow_readers	2.543545e+01	7.808767e+00	3.257294	1.736213e-03	9.861352e+00	4.100954e+01
average_number_of_named_entities_ordinal_per_word	-2.398542e-08	7.495725e-09	-3.199880	2.067104e-03	-3.893517e-08	-9.035665e-09
simple_nouns_variation	4.572277e-01	1.447439e-01	3.158873	2.338662e-03	1.685451e-01	7.459103e-01
total_number_of_unique_pronouns	-5.933595e+00	1.886614e+00	-3.145102	2.437108e-03	-9.696329e+00	-2.170861e+00
simple_pronouns_variation	-5.690443e+00	1.838407e+00	-3.095312	2.826292e-03	-9.357029e+00	-2.023856e+00
average_number_of_numerals_per_word	9.211165e-01	3.016055e-01	3.054044	3.192013e-03	3.195834e-01	1.522650e+00
average_number_of_adjectives_per_word	9.039703e-01	2.996039e-01	3.017218	3.555111e-03	3.064291e-01	1.501511e+00
average_number_of_named_entities_norp_per_word	3.013695e-08	1.038449e-08	2.902112	4.952210e-03	9.425735e-09	5.084816e-08
root_particles_variation	1.355049e+01	4.699383e+00	2.883462	5.221432e-03	4.177868e+00	2.292311e+01
corrected_particles_variation	1.369325e+01	4.750674e+00	2.882380	5.237457e-03	4.218327e+00	2.316817e+01
root_adpositions_variation	2.611851e+02	9.240457e+01	2.826539	6.128892e-03	7.689001e+01	4.454802e+02
total_number_of_named_entities_org	7.365727e-08	2.613092e-08	2.818778	6.263278e-03	2.154080e-08	1.257737e-07
corrected_adpositions_variation	-2.592618e+02	9.275231e+01	-2.795206	6.688263e-03	-4.442505e+02	-7.427314e+01
simple_particles_variation	-1.393290e+01	5.031651e+00	-2.769051	7.190673e-03	-2.396821e+01	-3.897588e+00
root_determiners_variation	8.658698e+02	3.227626e+02	2.682683	9.105436e-03	2.221401e+02	1.509599e+03
total_number_of_unique_numerals	4.208560e+00	1.576313e+00	2.669876	9.426012e-03	1.064703e+00	7.352417e+00

Table 13: Actor - Part1

Actor	Coef.	Std.Err.	t	P> t	[0.025	0.975]
corrected_determiners_variation	-8.581236e+02	3.219952e+02	-2.665020	9.550229e-03	-1.500323e+03	-2.159244e+02
average_number_of_named_entities_gpe_per_word	1.416567e-08	5.513924e-09	2.569072	1.233065e-02	3.168492e-09	2.516284e-08
total_number_of_unique_particles	-1.333033e+01	5.194925e+00	-2.566030	1.242971e-02	-2.369128e+01	-2.969382e+00
average_number_of_named_entities_gpe_per_sentence	-1.223607e-08	4.854813e-09	-2.520399	1.400475e-02	-2.191869e-08	-2.553447e-09
total_number_of_named_entities_fac	-1.487456e-07	6.189584e-08	-2.403160	1.890746e-02	-2.721930e-07	-2.529826e-08
total_number_of_named_entities_law	7.439520e-08	3.105679e-08	2.395456	1.927782e-02	1.245437e-08	1.363360e-07
simple_adjectives_variation	3.805237e+00	1.590028e+00	2.393189	1.938805e-02	6.340270e-01	6.976447e+00
total_number_of_named_entities_quantity	5.019254e-09	2.127752e-09	2.358947	2.112259e-02	7.755860e-10	9.262923e-09
average_number_of_named_entities_percent_per_se...	-4.051988e-09	1.740027e-09	-2.328693	2.276858e-02	-7.522361e-09	-5.816140e-10
average_number_of_named_entities_person_per_sen...	6.094086e-09	2.686051e-09	2.268790	2.636675e-02	7.369257e-10	1.145125e-08
root_type_token_ratio	-9.129534e+01	4.082464e+01	-2.236281	2.852244e-02	-1.727175e+02	-9.873173e+00
root_type_token_ratio_no_lemma	-9.129534e+01	4.082464e+01	-2.236281	2.852244e-02	-1.727175e+02	-9.873173e+00
corrected_type_token_ratio	8.965713e+01	4.088803e+01	2.192748	3.165116e-02	8.108534e+00	1.712057e+02
corrected_type_token_ratio_no_lemma	8.965713e+01	4.088803e+01	2.192748	3.165116e-02	8.108534e+00	1.712057e+02
total_number_of_named_entities_gpe	-1.436278e-08	6.634177e-09	-2.164968	3.380136e-02	-2.759423e-08	-1.131333e-09
simple_auxiliaries_variation	-2.849675e+00	1.345381e+00	-2.118117	3.771737e-02	-5.532954e+00	-1.663967e-01
average_number_of_named_entities_person_per_word	-2.567186e-08	1.213546e-08	-2.115442	3.795238e-02	-4.987526e-08	-1.468454e-09
average_number_of_named_entities_percent_per_word	1.101978e-08	5.301360e-09	2.078670	4.131448e-02	4.465514e-10	2.159301e-08
average_number_of_particles_per_word	3.744410e-01	1.805495e-01	2.073897	4.176934e-02	1.434636e-02	7.345356e-01
simple_adpositions_variation	-9.511057e-01	4.714977e-01	-2.017201	4.751080e-02	-1.891478e+00	-1.073309e-02
average_number_of_characters_per_word	-2.612976e+02	1.305870e+02	-2.000947	4.927703e-02	-5.217451e+02	-8.500675e-01

Table 14: Actor - Part2

Certainty	Coef.	Std.Err.	t	P> t	[0.025	0.975]
corrected_verbs_variation	-3.212384e+03	7.606418e+02	-4.223254	7.125228e-05	-4.729436e+03	-1.695331e+03
root_verbs_variation	3.221559e+03	7.634007e+02	4.220011	7.207333e-05	1.699005e+03	4.744114e+03
average_number_of_particles_per_word	1.256040e+00	3.286769e-01	3.821503	2.843090e-04	6.005144e-01	1.911565e+00
average_number_of_adjectives_per_word	1.975107e+00	5.454067e-01	3.621348	5.505765e-04	8.873279e-01	3.062887e+00
total_number_of_unique_nouns	9.644532e+00	2.695120e+00	3.578517	6.325497e-04	4.269285e+00	1.501978e+01
average_number_of_proper_nouns_per_word	3.132926e+00	8.847966e-01	3.540843	7.141205e-04	1.368255e+00	4.897597e+00
average_number_of_spaces_per_word	1.704043e+00	4.870889e-01	3.498423	8.178939e-04	7.325748e-01	2.675511e+00
simple_verbs_variation	-4.506679e+00	1.290555e+00	-3.492047	8.346754e-04	-7.080610e+00	-1.932748e+00
average_number_of_coordinating_conjunctions_per...	1.272030e+00	3.651852e-01	3.483246	8.583776e-04	5.436909e-01	2.000369e+00
average_number_of_proper_nouns_per_word	2.275784e+00	6.577801e-01	3.459795	9.246828e-04	9.638833e-01	3.587685e+00
average_number_of_numerals_per_word	1.845811e+00	5.490503e-01	3.361825	1.257749e-03	7.507647e-01	2.940857e+00
total_number_of_words_more_than_three_syllables	1.747913e+00	5.315596e-01	3.288274	1.579013e-03	6.877512e-01	2.808075e+00
total_number_of_unique_verbs	-5.212617e+00	1.607004e+00	-3.243686	1.809805e-03	-8.417686e+00	-2.007549e+00
average_number_of_pronouns_per_word	1.622757e+00	5.098556e-01	3.182777	2.176561e-03	6.058818e-01	2.639632e+00
average_number_of_verbs_per_word	1.678832e+00	5.376258e-01	3.122677	2.605744e-03	6.065711e-01	2.751093e+00
total_number_of_named_entities_gpe	-3.761732e-08	1.207703e-08	-3.114784	2.667656e-03	-6.170419e-08	-1.353045e-08
average_number_of_adpositions_per_word	1.745591e+00	5.687575e-01	3.069130	3.053488e-03	6.112399e-01	2.879942e+00
smog_index	-1.118901e+00	3.674465e-01	-3.045073	3.277147e-03	-1.851750e+00	-3.860523e-01
simple_determiners_variation	-4.040180e+00	1.332452e+00	-3.032140	3.403587e-03	-6.697671e+00	-1.382689e+00
total_number_of_unique_numerals	8.620557e+00	2.869560e+00	3.004139	3.693049e-03	2.897399e+00	1.434371e+01
coleman_liaw_index	4.932388e+00	1.649806e+00	2.989677	3.851343e-03	1.641953e+00	8.222822e+00
total_number_of_named_entities_quantity	-1.147253e-08	3.873415e-09	-2.961865	4.173613e-03	-1.919781e-08	-3.747248e-09
total_number_of_unique_adverbs	7.468378e+00	2.653280e+00	2.814772	6.333705e-03	2.176579e+00	1.276018e+01
bilogarithmic_type_token_ratio	-6.978865e-01	2.541982e-01	-2.745443	7.673647e-03	-1.204869e+00	-1.909043e-01
bilogarithmic_type_token_ratio_no_lemma	-6.978865e-01	2.541982e-01	-2.745443	7.673652e-03	-1.204869e+00	-1.909042e-01
total_kuperman_age_of_acquisition_of_words	-2.174240e+01	8.126806e+00	-2.675392	9.286694e-03	-3.795080e+01	-5.533992e+00
average_number_of_named_entities_cardinal_per_word	1.447710e-08	5.514504e-09	2.625278	1.062418e-02	3.478774e-09	2.547544e-08
simple_type_token_ratio	1.541694e+00	6.013573e-01	2.563691	1.250639e-02	3.423247e-01	2.741063e+00
simple_type_token_ratio_no_lemma	1.541694e+00	6.013573e-01	2.563691	1.250639e-02	3.423247e-01	2.741063e+00
average_kuperman_age_of_acquisition_of_words_per...	2.037079e+01	7.981958e+00	2.552104	1.289251e-02	4.451275e+00	3.629030e+01
simple_numerals_variation	6.483988e+00	2.571864e+00	2.521124	1.397837e-02	1.354567e+00	1.161341e+01
average_number_of_named_entities_product_per_se...	-1.031647e-08	4.109783e-09	-2.510223	1.437966e-02	-1.851318e-08	-2.119768e-09
total_number_of_unique_determiners	-3.325580e+00	1.372811e+00	-2.422461	1.800730e-02	-6.063565e+00	-5.875961e-01
average_number_of_adverbs_per_word	9.997150e-01	4.174031e-01	2.395083	1.929594e-02	1.672309e-01	1.832199e+00
total_number_of_named_entities_cardinal	-9.126049e-09	3.871122e-09	-2.357468	2.120049e-02	-1.684676e-08	-1.405338e-09
flesch_kincaid_grade_level	3.737562e+03	1.594155e+03	2.344541	2.189264e-02	5.581205e+02	6.917003e+03
flesch_kincaid_reading_ease	2.999031e+03	1.286925e+03	2.330386	2.267354e-02	4.323410e+02	5.565722e+03
reading_time_for_fast_readers	-1.356564e+01	5.852540e+00	-2.317907	2.338238e-02	-2.523817e+01	-1.893119e+00
average_number_of_named_entities_money_per_word	2.653509e-08	1.160470e-08	2.286581	2.524896e-02	3.390247e-09	4.967994e-08
root_type_token_ratio_no_lemma	-1.652804e+02	7.431822e+01	-2.223955	2.937941e-02	-3.135034e+02	-1.705738e+01
root_type_token_ratio	-1.652804e+02	7.431822e+01	-2.223955	2.937941e-02	-3.135034e+02	-1.705738e+01
total_number_of_unique_particles	2.098733e+01	9.456974e+00	2.219244	2.971291e-02	2.125994e+00	3.984867e+01
total_number_of_unique_auxiliaries	-5.758467e+00	2.625123e+00	-2.193599	3.158723e-02	-1.099411e+01	-5.228236e-01
average_number_of_subordinating_conjunctions_pe...	4.876051e-01	2.227822e-01	2.188708	3.195621e-02	4.328008e-02	9.319301e-01
total_number_of_unique_coordinating_conjunctions	4.736511e+00	2.166110e+00	2.186644	3.211308e-02	4.163406e-01	9.056682e+00
corrected_type_token_ratio_no_lemma	1.619719e+02	7.443361e+01	2.176059	3.292807e-02	1.351877e+01	3.104251e+02
corrected_type_token_ratio	1.619719e+02	7.443361e+01	2.176059	3.292807e-02	1.351877e+01	3.104251e+02
average_number_of_proper_nouns_per_sentence	-6.589614e+01	3.039285e+01	-2.168146	3.354909e-02	-1.265128e+02	-5.279508e+00
total_number_of_named_entities_ordinal	2.134446e-08	9.899768e-09	2.156057	3.451764e-02	1.599997e-09	4.108893e-08
total_number_of_unique_proper_nouns	3.251199e+01	1.523059e+01	2.134651	3.629253e-02	2.135541e+00	6.288845e+01
total_number_of_proper_nouns	3.251199e+01	1.523059e+01	2.134651	3.629253e-02	2.135541e+00	6.288845e+01
root_particles_variation	-1.801303e+01	8.554877e+00	-2.105586	3.882938e-02	-3.507519e+01	-9.508670e-01
corrected_particles_variation	-1.820879e+01	8.648248e+00	-2.105489	3.883811e-02	-3.545718e+01	-9.604052e-01
root_determiners_variation	1.223518e+03	5.875652e+02	2.082353	4.096658e-02	5.165614e+01	2.395380e+03
average_number_of_named_entities_law_per_word	6.668953e-09	3.212475e-09	2.075955	4.157272e-02	2.618729e-10	1.307603e-08
corrected_determiners_variation	-1.214813e+03	5.861683e+02	-2.072465	4.190660e-02	-2.383889e+03	-4.573735e+01
average_number_of_verbs_per_sentence	-5.335525e+01	2.651491e+01	-2.012273	4.804043e-02	-1.062376e+02	-4.729229e-01
average_number_of_auxiliaries_per_sentence	-6.196236e+01	3.080374e+01	-2.011520	4.812178e-02	-1.233985e+02	-5.262309e-01
average_number_of_adpositions_per_sentence	-8.579195e+01	4.280723e+01	-2.004146	4.892499e-02	-1.711683e+02	-4.156157e-01
automated_readability_index	-6.747181e+02	3.371693e+02	-2.001125	4.925731e-02	-1.347181e+03	-2.255082e+00

Table 15: Certainty

Negation	Coef.	Std.Err.	t	P> t	[0.025	0.975]
simple_adpositions_variation	-1.719098e+00	6.155434e-01	-2.792814	6.732834e-03	-2.946761e+00	-4.914355e-01
total_number_of_named_entities_product	4.244266e-08	1.733488e-08	2.448397	1.685814e-02	7.869341e-09	7.701598e-08
smog_index	-6.244617e-01	2.635119e-01	-2.369767	2.056014e-02	-1.150020e+00	-9.890378e-02
average_number_of_named_entities_loc_per_word	5.714464e-08	2.488083e-08	2.296733	2.463010e-02	7.521382e-09	1.067679e-07
average_number_of_named_entities_time_per_sentence	-1.229495e-08	5.436076e-09	-2.261733	2.682211e-02	-2.313686e-08	-1.453039e-09
total_number_of_named_entities_percent	9.312301e-08	4.136844e-08	2.251064	2.752364e-02	1.061625e-08	1.756298e-07
total_number_of_named_entities_law	9.120490e-08	4.054485e-08	2.249482	2.762905e-02	1.034075e-08	1.720690e-07
average_number_of_named_entities_cardinal_per_s...	-1.719405e-08	7.674631e-09	-2.240375	2.824268e-02	-3.250062e-08	-1.887482e-09
average_number_of_named_entities_gpe_per_sentence	-1.393150e-08	6.337991e-09	-2.198094	3.125139e-02	-2.657222e-08	-1.290775e-09
average_number_of_named_entities_org_per_word	4.446542e-08	2.024354e-08	2.196524	3.136830e-02	4.090961e-09	8.483988e-08
average_number_of_named_entities_money_per_sent...	2.024129e-08	9.350754e-09	2.164669	3.382518e-02	1.591797e-09	3.889078e-08
total_number_of_named_entities_fac	-1.743072e-07	8.080543e-08	-2.157123	3.443124e-02	-3.354686e-07	-1.314590e-08
total_number_of_named_entities_event	6.852808e-08	3.183012e-08	2.152932	3.477191e-02	5.044914e-09	1.320112e-07
average_number_of_named_entities_art_per_sentence	2.453337e-08	1.147241e-08	2.138468	3.597032e-02	1.652384e-09	4.741436e-08
average_number_of_named_entities_loc_per_sentence	1.000902e-08	4.721955e-09	2.119678	3.758082e-02	5.913809e-10	1.942667e-08
average_number_of_named_entities_gpe_per_word	1.519277e-08	7.198464e-09	2.110557	3.838489e-02	8.358857e-10	2.954965e-08
average_number_of_named_entities_art_per_word	-5.169465e-08	2.462123e-08	-2.099597	3.937085e-02	-1.008001e-07	-2.589165e-09
simple_coordinating_conjunctions_variation	4.958396e+00	2.380729e+00	2.082721	4.093188e-02	2.101811e-01	9.706611e+00
average_number_of_named_entities_percent_per_word	1.438223e-08	6.920960e-09	2.078069	4.137159e-02	5.788102e-10	2.818565e-08
average_number_of_named_entities_fac_per_word	-3.128243e-08	1.506596e-08	-2.076365	4.153365e-02	-6.133055e-08	-1.234316e-09
root_coordinating_conjunctions_variation	-4.413945e+03	2.128561e+03	-2.073675	4.179054e-02	-8.659227e+03	-1.686630e+02
corrected_coordinating_conjunctions_variation	4.409099e+03	2.127610e+03	2.072325	4.192006e-02	1.657140e+02	8.652483e+03
average_number_of_named_entities_event_per_sent...	1.840485e-08	8.899373e-09	2.068107	4.232679e-02	6.556136e-10	3.615409e-08
average_number_of_named_entities_person_per_word	-3.237656e-08	1.584292e-08	-2.043598	4.475826e-02	-6.397426e-08	-7.788546e-10
average_number_of_named_entities_language_per_word	2.490433e-08	1.228701e-08	2.026883	4.648480e-02	3.986602e-10	4.941000e-08
total_number_of_named_entities_person	8.564630e-08	4.238720e-08	2.020570	4.715160e-02	1.107697e-09	1.701849e-07

Table 16: Negation

Temporality	Coef.	Std.Err.	t	P> t	[0.025	0.975]
total_number_of_unique_punctuations	-8.488106e+00	1.458012e+00	-5.821699	1.612211e-07	-1.139602e+01	-5.580193e+00
simple_type_token_ratio_no_lemma	4.004287e+00	6.977944e-01	5.738491	2.252798e-07	2.612580e+00	5.395994e+00
simple_type_token_ratio	4.004287e+00	6.977943e-01	5.738491	2.252798e-07	2.612580e+00	5.395994e+00
simple_punctuations_variation	-5.202325e+00	9.263226e-01	-5.616105	3.674957e-07	-7.049817e+00	-3.354833e+00
total_number_of_unique_nouns	1.714003e+01	3.127325e+00	5.480731	6.289046e-07	1.090277e+01	2.337728e+01
total_number_of_unique_particles	5.991561e+01	1.097355e+01	5.460003	6.825611e-07	3.802956e+01	8.180166e+01
simple_auxiliaries_variation	-1.550611e+01	2.841929e+00	-5.456191	6.929087e-07	-2.117416e+01	-9.838059e+00
corrected_particles_variation	-5.234579e+01	1.003513e+01	-5.216255	1.772690e-06	-7.236023e+01	-3.233136e+01
root_particles_variation	-5.177689e+01	9.926785e+00	-5.215876	1.775294e-06	-7.157523e+01	-3.197854e+01
average_number_of_punctuations_per_word	-3.707239e+00	7.182851e-01	-5.161235	2.193892e-06	-5.139813e+00	-2.274664e+00
simple_particles_variation	5.464536e+01	1.062865e+01	5.141324	2.369345e-06	3.344717e+01	7.584354e+01
total_number_of_unique_auxiliaries	-1.535464e+01	3.046103e+00	-5.040749	3.488348e-06	-2.142990e+01	-9.279380e+00
coleman_liau_index	9.580976e+00	1.914378e+00	5.004745	4.003460e-06	5.762868e+00	1.339908e+01
average_number_of_adpositions_per_word	3.269788e+00	6.599667e-01	4.954474	4.849078e-06	1.953526e+00	4.586050e+00
total_number_of_words_more_than_three_syllables	2.960606e+00	6.168035e-01	4.799918	8.695859e-06	1.730431e+00	4.190782e+00
simple_verbs_variation	-7.145826e+00	1.497516e+00	-4.771787	9.662941e-06	-1.013253e+01	-4.159125e+00
average_number_of_named_entities_quantity_per_word	-4.121338e-08	8.651680e-09	-4.763627	9.962572e-06	-5.846861e-08	-1.395814e-08
average_number_of_interjections_per_word	1.792648e+00	3.835171e-01	4.674232	1.389958e-05	1.027747e+00	2.557549e+00
corrected_nouns_variation	-9.959689e+02	2.159900e+02	-4.611182	1.754958e-05	-1.426747e+03	-5.651905e+02
root_nouns_variation	9.975494e+02	2.166140e+02	4.605194	1.794119e-05	5.655264e+02	1.429572e+03
simple_proper_nouns_variation	9.242736e+00	2.047195e+00	4.514829	2.499143e-05	5.59734e+00	1.332574e+01
average_number_of_named_entities_art_per_word	1.797830e-07	3.983808e-08	4.512842	2.517333e-05	1.003284e-07	2.592375e-07
average_number_of_named_entities_percent_per_word	-5.033151e-08	1.119838e-08	-4.494536	2.691110e-05	-7.266598e-08	-2.799705e-08
total_number_of_unique_adpositions	1.096727e+01	2.450401e+00	4.475704	2.882018e-05	6.080100e+00	1.585444e+01
total_number_of_words_more_than_two_syllables	4.532242e+00	1.019734e+00	4.444535	3.227238e-05	2.498447e+00	6.566036e+00
root_proper_nouns_variation	-3.887525e+03	8.832845e+02	-4.401215	3.774369e-05	-5.649180e+03	-2.125870e+03
corrected_proper_nouns_variation	3.870173e+03	8.797788e+02	4.399030	3.804234e-05	2.115510e+03	5.624837e+03
root_verbs_variation	3.864364e+03	8.858240e+02	4.362452	4.339410e-05	2.097644e+03	5.631085e+03
corrected_verbs_variation	-3.847501e+03	8.826227e+02	-4.359168	4.390870e-05	-5.607836e+03	-2.087165e+03
average_number_of_named_entities_fac_per_word	1.061814e-07	2.437730e-08	4.355750	4.445068e-05	5.756244e-08	1.548004e-07
root_punctuations_variation	7.581011e+02	1.749593e+02	4.333015	4.822404e-05	4.091559e+02	1.107046e+03
total_number_of_named_entities_event	-2.224247e-07	5.150234e-08	-4.318730	5.075161e-05	-3.251429e-07	-1.197066e-07
corrected_punctuations_variation	-7.458231e+02	1.732774e+02	-4.304215	5.345123e-05	-1.091414e+03	-4.002322e+02
total_number_of_named_entities_product	-1.185646e-07	2.804849e-08	-4.227129	7.028313e-05	-1.745055e-07	-6.262363e-08
average_number_of_named_entities_event_per_sent...	-6.009121e-08	1.439952e-08	-4.173139	8.501091e-05	-8.881016e-08	-3.137272e-08
average_subtlex_us_zipf_of_words_per_word	-3.918422e+00	9.524894e-01	-4.113878	1.046008e-04	-5.818106e+00	-2.018745e+00
average_number_of_numerals_per_word	2.587804e+00	6.370992e-01	4.061854	1.253275e-04	1.317149e+00	3.858458e+00
average_number_of_auxiliaries_per_word	2.262518e+00	5.588793e-01	4.048313	1.313395e-04	1.147869e+00	3.377168e+00
gunning_fog_index	-2.912632e+00	7.260163e-01	-4.011800	1.489660e-04	-4.360626e+00	-1.464638e+00
simple_nouns_variation	-1.215635e+00	3.057511e-01	-3.975896	1.685040e-04	-1.825436e+00	-6.058333e-01
total_kuperman_age_of_acquisition_of_words	-3.737562e+01	9.430067e+00	-3.963452	1.758329e-04	-5.618330e+01	-1.856795e+01
total_number_of_unique_spaces	5.749141e+01	1.456223e+01	3.947981	1.853725e-04	2.844796e+01	8.653487e+01
total_number_of_named_entities_percent	-2.633546e-07	6.693571e-08	-3.934441	1.941272e-04	-1.298537e-07	-6.334667e-07
average_number_of_named_entities_org_per_word	-1.288588e-07	3.275481e-08	-3.934041	1.943917e-04	-1.941862e-07	-6.353138e-08
average_number_of_named_entities_loc_per_word	-1.580689e-07	4.025814e-08	-3.926385	1.995235e-04	-2.383613e-07	-7.777661e-08
root_adjectives_variation	-8.902584e+03	2.272952e+03	-3.916750	2.061655e-04	-1.343584e+04	-4.369325e+03
corrected_adjectives_variation	8.889159e+03	2.269745e+03	3.916369	2.064328e-04	4.362295e+03	1.341602e+04
average_number_of_nouns_per_word	4.013621e+00	1.026688e+00	3.909291	2.114528e-04	1.965957e+00	6.061285e+00
average_number_of_named_entities_time_per_sentence	3.425673e-08	8.795779e-09	3.894679	2.221910e-04	1.671410e-08	5.179936e-08
corrected_spaces_variation	-5.365895e+01	1.386591e+01	-3.869848	2.416467e-04	-8.131363e+01	-2.600427e+01
simple_spaces_variation	5.451684e+01	1.428988e+01	3.815067	2.905048e-04	2.601658e+01	8.301710e+01
average_kuperman_age_of_acquisition_of_words_per...	3.524573e+01	9.261990e+00	3.805416	3.000387e-04	1.677327e+01	5.371818e+01
reading_time_for_fast_readers	-2.574417e+01	6.791087e+00	-3.790877	3.149690e-04	-3.928857e+01	-1.219978e+01
root_spaces_variation	-4.460722e+01	1.180521e+01	-3.778605	3.281203e-04	-6.815197e+01	-2.106248e+01
average_number_of_named_entities_money_per_sent...	-5.695573e-08	1.512988e-08	-3.764454	3.439378e-04	-8.713132e-08	-2.678014e-08
average_number_of_named_entities_person_per_word	9.622665e-08	2.563444e-08	3.753803	3.563216e-04	4.510036e-08	1.473529e-07
total_brysaert_age_of_acquisition_of_words	2.551302e+01	6.819914e+00	3.740959	3.718210e-04	1.191113e+01	3.911491e+01
total_number_of_named_entities_law	-2.418963e-07	6.560311e-08	-3.687269	4.438607e-04	-3.727376e-07	-1.110550e-07
average_number_of_named_entities_art_per_sentence	-6.825043e-08	1.856279e-08	-3.676733	4.594789e-04	-1.052727e-07	-3.122811e-08
total_number_of_unique_words	-2.855957e+01	7.842706e+00	-3.641545	5.155262e-04	-4.420135e+01	-1.291778e+01
total_number_of_named_entities_fac	4.710148e-07	1.307463e-07	3.602511	5.852995e-04	2.102496e-07	7.317800e-07
average_number_of_named_entities_percent_per_se...	-1.321096e-08	3.675562e-09	-3.594270	6.011350e-04	-2.054164e-08	-5.880286e-09
total_number_of_named_entities_art	1.704759e-07	4.760417e-08	3.581112	6.272691e-04	7.553235e-08	2.654194e-07
average_brysaert_age_of_acquisition_of_words_pe...	-2.492547e+01	7.055126e+00	-3.532959	7.324102e-04	-3.899647e+01	-1.085447e+01
average_number_of_named_entities_product_per_word	9.089588e-08	2.590158e-08	3.509279	7.900521e-04	3.923681e-08	1.425550e-07
total_number_of_sentences	2.287950e+01	6.519742e+00	3.509265	7.900892e-04	9.876285e+00	3.588272e+01
average_number_of_verbs_per_word	2.158172e+00	6.238425e-01	3.459483	9.255977e-04	9.139578e-01	3.402387e+00
average_number_of_named_entities_cardinal_per_s...	4.226246e-08	1.241785e-08	3.403365	1.104648e-03	1.749585e-08	6.702907e-08
simple_adpositions_variation	3.327248e+00	9.959727e-01	3.340702	1.343072e-03	1.340843e+00	5.313653e+00
average_number_of_named_entities_person_per_sen...	1.894099e-08	5.673906e-09	3.338263	1.353269e-03	7.624741e-09	3.025724e-08
average_number_of_particles_per_word	1.264886e+00	3.813854e-01	3.316555	1.447285e-03	5.042365e-01	2.025535e+00

Table 17: Temporality-Part1

Temporality	Coef.	Std.Err.	t	P> t	[0.025	0.975]
simple_pronouns_variation	-1.276292e+01	3.883376e+00	-3.286553	1.587379e-03	-2.050807e+01	-5.017773e+00
average_kuperman_age_of_acquisition_of_words_per...	9.428420e-01	2.882325e-01	3.271116	1.664341e-03	3.679803e-01	1.517704e+00
average_number_of_named_entities_language_per_word	-6.366250e-08	1.988085e-08	-3.202202	2.052643e-03	-1.033136e-07	-2.401140e-08
bilogarithmic_type_token_ratio	-9.381622e-01	2.949628e-01	-3.180611	2.190802e-03	-1.526447e+00	-3.498773e-01
bilogarithmic_type_token_ratio_no_lemma	-9.381621e-01	2.949628e-01	-3.180611	2.190803e-03	-1.526447e+00	-3.498773e-01
average_number_of_named_entities_gpe_per_sentence	3.259444e-08	1.025511e-08	3.178361	2.205693e-03	1.214127e-08	5.304762e-08
average_number_of_named_entities_norp_per_word	6.964400e-08	2.193577e-08	3.174905	2.228745e-03	2.589448e-08	1.133935e-07
average_number_of_characters_per_sentence	-1.453636e+02	4.604424e+01	-3.157042	2.351536e-03	-2.371960e+02	-5.353126e+01
total_number_of_named_entities	-2.994569e+01	9.508760e+00	-3.149274	2.406884e-03	-4.891031e+01	-1.098107e+01
smog_index	-1.334089e+00	4.263723e-01	-3.128930	2.557660e-03	-2.184462e+00	-4.837163e-01
total_number_of_characters	1.355410e+02	4.332840e+01	3.128224	2.563044e-03	4.912519e+01	2.219567e+02
average_number_of_stop_words_per_word	2.581518e+00	8.283695e-01	3.116384	2.654992e-03	9.293867e-01	4.233648e+00
total_number_of_named_entities_gpe	-4.354429e-08	1.401377e-08	-3.107251	2.728013e-03	-7.149386e-08	-1.559471e-08
average_number_of_named_entities_per_sentence	2.983344e+01	9.603292e+00	3.106585	2.733412e-03	1.068028e+01	4.898660e+01
root_coordinating_conjunctions_variation	1.058921e+04	3.444094e+03	3.074600	3.004668e-03	3.720180e+03	1.745824e+04
average_number_of_named_entities_quantity_per_s...	-1.229481e-08	4.000186e-09	-3.073559	3.013899e-03	-2.027293e-08	-4.316689e-09
corrected_coordinating_conjunctions_variation	-1.057527e+04	3.442554e+03	-3.071925	3.028455e-03	-1.744123e+04	-3.709309e+03
simple_coordinating_conjunctions_variation	-1.181247e+01	3.852111e+00	-3.066492	3.077308e-03	-1.949526e+01	-4.129673e+00
corrected_adpositions_variation	-5.967113e+02	1.959262e+02	-3.045592	3.272164e-03	-9.874738e+02	-2.059487e+02
average_number_of_spaces_per_word	1.706844e+00	5.652012e-01	3.019887	3.527553e-03	5.795856e-01	2.834102e+00
root_adpositions_variation	5.875629e+02	1.951917e+02	3.010184	3.628681e-03	1.982654e+02	9.768603e+02
root_pronouns_variation	1.358380e+03	4.603614e+02	2.950682	4.310101e-03	4.402180e+02	2.276542e+03
corrected_pronouns_variation	-1.332799e+03	4.538840e+02	-2.936432	4.490009e-03	-2.238043e+03	-4.275563e+02
average_number_of_named_entities_time_per_word	-4.445738e-08	1.515336e-08	-2.933830	4.523594e-03	-7.467980e-08	-1.423496e-08
average_number_of_stop_words_per_sentence	-1.660031e+02	5.700572e+01	-2.912043	4.814133e-03	-2.796974e+02	-5.230877e+01
average_number_of_syllables_per_word	-2.972141e+02	1.021721e+02	-2.908957	4.856657e-03	-5.009898e+02	-9.343833e+01
uber_type_token_ratio	-3.692130e-01	1.272081e-01	-2.902434	4.947675e-03	-6.229215e-01	-1.155045e-01
uber_type_token_ratio_no_lemma	-3.692127e-01	1.272080e-01	-2.902432	4.947696e-03	-6.229211e-01	-1.155043e-01
average_number_of_subordinating_conjunctions_pe...	7.495776e-01	2.585088e-01	2.899615	4.987489e-03	2.339965e-01	1.265155e+00
average_number_of_named_entities_per_word	8.542054e-01	2.951073e-01	2.894558	5.059662e-03	2.656324e-01	1.442778e+00
average_number_of_interjections_per_sentence	-6.440576e+00	2.232746e+00	-2.884599	5.204637e-03	-1.089365e+01	-1.987505e+00
total_number_of_punctuations	3.316925e+01	1.154945e+01	2.871933	5.394545e-03	1.013460e+01	5.620390e+01
total_number_of_punctuations	3.316925e+01	1.154945e+01	2.871933	5.394546e-03	1.013460e+01	5.620390e+01
average_number_of_subordinating_conjunctions_pe...	-1.982583e+01	6.934693e+00	-2.858934	5.596079e-03	-3.365664e+01	-5.995022e+00
simple_adjectives_variation	9.458067e+00	3.358710e+00	2.815982	6.312358e-03	2.759332e+00	1.615680e+01
corrected_numerals_variation	2.361105e+03	8.449339e+02	2.794426	6.702768e-03	6.759375e+02	4.066273e+03
root_numerals_variation	-2.369576e+03	8.488371e+02	-2.791555	6.756402e-03	-4.062528e+03	-6.766232e+02
average_number_of_named_entities_gpe_per_word	-3.240184e-08	1.164739e-08	-2.781897	6.939742e-03	-5.563182e-08	-9.171852e-09
total_number_of_stop_words	1.489059e+02	5.454004e+01	2.730213	8.000787e-03	4.012924e+01	2.576826e+02
total_number_of_named_entities_person	-1.822952e-07	6.858410e-08	-2.657981	9.732926e-03	-3.190819e-07	-4.550856e-08
average_number_of_punctuations_per_sentence	-5.759106e+01	2.174557e+01	-2.648404	9.986602e-03	-1.009612e+02	-1.422089e+01
total_number_of_named_entities_quantity	-1.189930e-08	4.494578e-09	-2.647478	1.001144e-02	-2.086345e-08	-2.935145e-09
simple_adverbs_variation	-1.106562e+01	4.202748e+00	-2.632950	1.040869e-02	-1.944774e+01	-2.683507e+00
average_number_of_named_entities_date_per_sentence	-2.714601e-08	1.037733e-08	-2.615896	1.089321e-02	-4.784293e-08	-6.449081e-09
average_number_of_named_entities_loc_per_sentence	-1.948994e-08	7.640303e-09	-2.550938	1.293195e-02	-3.472805e-08	-4.251837e-09
total_number_of_unique_adjectives	5.061019e+01	1.986889e+01	2.547207	1.305892e-02	1.098293e+01	9.023744e+01
total_number_of_spaces	-3.129192e+01	1.242049e+01	-2.519378	1.404195e-02	-5.606381e+01	-6.520029e+00
average_number_of_named_entities_language_per_s...	-4.039236e-08	1.616857e-08	-2.498202	1.483419e-02	-7.263956e-08	-8.145154e-09
root_adverbs_variation	2.935009e+03	1.176371e+03	2.494968	1.495863e-02	5.888104e+02	5.281208e+03
total_number_of_unique_pronouns	-9.932665e+00	3.985208e+00	-2.492383	1.505879e-02	-1.788091e+01	-1.984419e+00
corrected_adverbs_variation	-2.913601e+03	1.169759e+03	-2.490770	1.512159e-02	-5.246611e+03	-6.850910e+02
total_number_of_adverbs	-4.744528e+01	1.920919e+01	-2.469926	1.595468e-02	-8.575681e+01	-9.133750e+00
total_number_of_named_entities_loc	-2.761345e-08	1.133568e-08	-2.435977	1.739997e-02	-5.022175e-08	-5.005156e-09
total_number_of_unique_determiners	3.816127e+00	1.592962e+00	2.395617	1.927005e-02	6.390641e-01	6.993190e+00
total_number_of_subordinating_conjunctions	1.573261e+01	6.600784e+00	2.383445	1.986821e-02	2.567758e+00	2.889746e+01
total_subtlex_us_zipf_of_words	5.537261e+01	2.323442e+01	2.383214	1.987971e-02	9.033015e+00	1.017122e+02
total_number_of_syllables	-1.370393e+02	5.794875e+01	-2.364836	2.081476e-02	-2.526144e+02	-2.146416e+01
average_number_of_named_entities_ordinal_per_word	3.711067e-08	1.583367e-08	2.343783	2.193388e-02	5.531423e-09	6.868993e-08
automated_readability_index	-9.018830e+02	3.912397e+02	-2.305193	2.412469e-02	-1.682186e+03	-1.215800e+02
total_number_of_named_entities_date	-7.290189e-08	3.165296e-08	-2.303162	2.424519e-02	-1.360317e-07	-9.772050e-09
average_number_of_characters_per_word	6.271808e+02	2.758466e+02	2.273658	2.605667e-02	7.702202e+01	1.177340e+03
average_number_of_adjectives_per_word	1.435990e+00	6.328712e-01	2.269008	2.635280e-02	1.737679e-01	2.698212e+00
total_number_of_named_entities_language	-5.113484e-08	2.256992e-08	-2.265619	2.657052e-02	-9.614914e-08	-6.120547e-09
total_number_of_pronouns	-1.088058e+02	4.855724e+01	-2.240773	2.821561e-02	-2.056501e+02	-1.196139e+01
total_number_of_determiners	-1.064012e+02	4.761202e+01	-2.234755	2.862733e-02	-2.013604e+02	-1.144201e+01
total_number_of_named_entities_time	-9.358431e-08	4.248959e-08	-2.202523	3.092355e-02	-1.783271e-07	-8.841489e-09
average_number_of_named_entities_event_per_word	2.519426e-08	1.154657e-08	2.181970	3.247077e-02	2.165356e-09	4.822316e-08
average_number_of_adverbs_per_word	1.048742e+00	4.843402e-01	2.165301	3.377489e-02	8.275605e-02	2.014728e+00
average_number_of_pronouns_per_sentence	9.992538e+01	4.657076e+01	2.145668	3.536939e-02	7.042922e+00	1.928078e+02
average_number_of_syllables_per_sentence	1.298597e+02	6.068938e+01	2.139743	3.586327e-02	8.818525e+00	2.509008e+02
average_number_of_determiners_per_sentence	1.010542e+02	4.728980e+01	2.136913	3.610127e-02	6.737672e+00	1.953707e+02
simple_determiners_variation	3.261927e+00	1.546131e+00	2.109735	3.845809e-02	1.782658e-01	6.345588e+00
average_number_of_coordinating_conjunctions_per...	5.408573e+01	2.591138e+01	2.087335	4.049987e-02	2.407114e+00	1.057643e+02
total_number_of_nouns	-1.968856e+02	9.514501e+01	-2.069322	4.220928e-02	-3.866464e+02	-7.124892e+00
total_number_of_adjectives	-1.159769e+02	5.607170e+01	-2.068368	4.230152e-02	-2.278084e+02	-4.145414e+00
total_number_of_coordinating_conjunctions	-5.389970e+01	2.619666e+01	-2.057503	4.336443e-02	-1.061473e+02	-1.652108e+00
average_number_of_pronouns_per_word	1.197110e+00	5.916190e-01	2.023447	4.684669e-02	1.716281e-02	2.377056e+00

Table 18: Temporality-Part2

Key	Feature-name	Key	Feature-name
0	t_word	total_number_of_words	55 a_n_ent_language_pw
1	t_stopword	total_number_of_stop_words	56 a_n_ent_date_pw
2	t_punct	total_number_of_punctuations	57 a_n_ent_time_pw
3	t_syll	total_number_of_syllables	58 a_n_ent_percent_pw
4	t_syll2	total_number_of_words_more_than_two_syllables	59 a_n_ent_money_pw
5	t_syll3	total_number_of_words_more_than_three_syllables	60 a_n_ent_quantity_pw
6	t_uword	total_number_of_unique_words	61 a_n_ent_ordinal_pw
7	t_sent	total_number_of_sentences	62 a_n_ent_cardinal_pw
8	t_char	total_number_of_characters	63 a_n_ent_ps
9	a_word_ps	average_number_of_words_per_sentence	64 a_n_ent_person_ps
10	a_char_ps	average_number_of_characters_per_sentence	65 a_n_ent_norp_ps
11	a_char_pw	average_number_of_characters_per_word	66 a_n_ent_fac_ps
12	a_syll_ps	average_number_of_syllables_per_sentence	67 a_n_ent_org_ps
13	a_syll_pw	average_number_of_syllables_per_word	68 a_n_ent_gpe_ps
14	a_stopword_ps	average_number_of_stop_words_per_sentence	69 a_n_ent_loc_ps
15	a_stopword_pw	average_number_of_stop_words_per_word	70 a_n_ent_product_ps
16	t_kup	total_kuperman_age_of_acquisition_of_words	71 a_n_ent_event_ps
17	t_bry	total_brysaert_age_of_acquisition_of_words	72 a_n_ent_art_ps
18	t_subtlex_us_zipf	total_subtlex_us_zipf_of_words	73 a_n_ent_law_ps
19	a_kup_pw	average_kuperman_age_of_acquisition_of_words_per...	74 a_n_ent_language_ps
20	a_bry_pw	average_brysaert_age_of_acquisition_of_words_per...	75 a_n_ent_date_ps
21	a_kup_ps	average_kuperman_age_of_acquisition_of_words_per...	76 a_n_ent_time_ps
22	a_bry_ps	average_brysaert_age_of_acquisition_of_words_per...	77 a_n_ent_percent_ps
23	a_subtlex_us_zipf_pw	average_subtlex_us_zipf_of_words_per_word	78 a_n_ent_money_ps
24	a_subtlex_us_zipf_ps	average_subtlex_us_zipf_of_words_per_sentence	79 a_n_ent_quantity_ps
25	t_n_ent	total_number_of_named_entities	80 a_n_ent_ordinal_ps
26	t_n_ent_person	total_number_of_named_entities_person	81 a_n_ent_cardinal_ps
27	t_n_ent_norp	total_number_of_named_entities_norp	82 simp_adj_var
28	t_n_ent_fac	total_number_of_named_entities_fac	83 simp_adp_var
29	t_n_ent_org	total_number_of_named_entities_org	84 simp_adv_var
30	t_n_ent_gpe	total_number_of_named_entities_gpe	85 simp_aux_var
31	t_n_ent_loc	total_number_of_named_entities_loc	86 simp_cconj_var
32	t_n_ent_product	total_number_of_named_entities_product	87 simp_det_var
33	t_n_ent_event	total_number_of_named_entities_event	88 simp_intj_var
34	t_n_ent_art	total_number_of_named_entities_art	89 simp_noun_var
35	t_n_ent_law	total_number_of_named_entities_law	90 simp_num_var
36	t_n_ent_language	total_number_of_named_entities_language	91 simp_part_var
37	t_n_ent_date	total_number_of_named_entities_date	92 simp_pron_var
38	t_n_ent_time	total_number_of_named_entities_time	93 simp_propn_var
39	t_n_ent_percent	total_number_of_named_entities_percent	94 simp_punct_var
40	t_n_ent_money	total_number_of_named_entities_money	95 simp_sconj_var
41	t_n_ent_quantity	total_number_of_named_entities_quantity	96 simp_sym_var
42	t_n_ent_ordinal	total_number_of_named_entities_ordinal	97 simp_verb_var
43	t_n_ent_cardinal	total_number_of_named_entities_cardinal	98 simp_space_var
44	a_n_ent_pw	average_number_of_named_entities_per_word	99 root_adj_var
45	a_n_ent_person_pw	average_number_of_named_entities_person_per_word	100 root_adp_var
46	a_n_ent_norp_pw	average_number_of_named_entities_norp_per_word	101 root_adv_var
47	a_n_ent_fac_pw	average_number_of_named_entities_fac_per_word	102 root_aux_var
48	a_n_ent_org_pw	average_number_of_named_entities_org_per_word	103 root_cconj_var
49	a_n_ent_gpe_pw	average_number_of_named_entities_gpe_per_word	104 root_det_var
50	a_n_ent_loc_pw	average_number_of_named_entities_loc_per_word	105 root_intj_var
51	a_n_ent_product_pw	average_number_of_named_entities_product_per_word	106 root_noun_var
52	a_n_ent_event_pw	average_number_of_named_entities_event_per_word	107 root_num_var
53	a_n_ent_art_pw	average_number_of_named_entities_art_per_word	108 root_part_var
54	a_n_ent_law_pw	average_number_of_named_entities_law_per_word	109 root_pron_var

Table 19: Linguistic Features (0-109)

Key	Feature-name	Key	Feature-name
110	root_propn_var	root_proper_nouns_variation	total_number_of_unique_interjections
111	root_punct_var	root_punctuations_variation	total_number_of_unique_nouns
112	root_sconj_var	root_subordinating_conjunctions_variation	total_number_of_unique_numerals
113	root_sym_var	root_symbols_variation	total_number_of_unique_particles
114	root_verb_var	root_verbs_variation	total_number_of_unique_pronouns
115	root_space_var	root_spaces_variation	total_number_of_unique_proper_nouns
116	corr_adj_var	corrected_adjectives_variation	total_number_of_unique_punctuations
117	corr_adp_var	corrected_adpositions_variation	total_number_of_unique_subordinating_conjunctions
118	corr_adv_var	corrected_adverbs_variation	total_number_of_unique_symbols
119	corr_aux_var	corrected_auxiliaries_variation	total_number_of_unique_verbs
120	corr_cconj_var	corrected_coordinating_conjunctions_variation	total_number_of_unique_spaces
121	corr_det_var	corrected_determiners_variation	average_number_of_adjectives_per_word
122	corr_intj_var	corrected_interjections_variation	average_number_of_adpositions_per_word
123	corr_noun_var	corrected_nouns_variation	average_number_of_adverbs_per_word
124	corr_num_var	corrected_numerals_variation	average_number_of_auxiliaries_per_word
125	corr_part_var	corrected_particles_variation	average_number_of_coordinating_conjunctions_pe...
126	corr_propn_var	corrected_pronouns_variation	average_number_of_determiners_per_word
127	corr_propn_var	corrected_proper_nouns_variation	average_number_of_interjections_per_word
128	corr_punct_var	corrected_punctuations_variation	average_number_of_nouns_per_word
129	corr_sconj_var	corrected_subordinating_conjunctions_variation	average_number_of_numerals_per_word
130	corr_sym_var	corrected_symbols_variation	average_number_of_particles_per_word
131	corr_verb_var	corrected_verbs_variation	average_number_of_pronouns_per_word
132	corr_space_var	corrected_spaces_variation	average_number_of_proper_nouns_per_word
133	simp_ttr	simple_type_token_ratio	average_number_of_punctuations_per_word
134	root_ttr	root_type_token_ratio	average_number_of_subordinating_conjunctions_p...
135	corr_ttr	corrected_type_token_ratio	average_number_of_symbols_per_word
136	bilog_ttr	bilogarithmic_type_token_ratio	average_number_of_verbs_per_word
137	uber_ttr	uber_type_token_ratio	average_number_of_spaces_per_word
138	simp_ttr_no_lem	simple_type_token_ratio_no_lemma	average_number_of_adjectives_per_sentence
139	root_ttr_no_lem	root_type_token_ratio_no_lemma	average_number_of_adpositions_per_sentence
140	corr_ttr_no_lem	corrected_type_token_ratio_no_lemma	average_number_of_adverbs_per_sentence
141	bilog_ttr_no_lem	bilogarithmic_type_token_ratio_no_lemma	average_number_of_auxiliaries_per_sentence
142	uber_ttr_no_lem	uber_type_token_ratio_no_lemma	average_number_of_coordinating_conjunctions_pe...
143	n_adj	total_number_of_adjectives	average_number_of_determiners_per_sentence
144	n_adp	total_number_of_adpositions	average_number_of_interjections_per_sentence
145	n_adv	total_number_of_adverbs	average_number_of_nouns_per_sentence
146	n_aux	total_number_of_auxiliaries	average_number_of_numerals_per_sentence
147	n_cconj	total_number_of_coordinating_conjunctions	average_number_of_particles_per_sentence
148	n_det	total_number_of_determiners	average_number_of_pronouns_per_sentence
149	n_intj	total_number_of_interjections	average_number_of_proper_nouns_per_sentence
150	n_noun	total_number_of_nouns	average_number_of_punctuations_per_sentence
151	n_num	total_number_of_numerals	average_number_of_subordinating_conjunctions_p...
152	n_part	total_number_of_particles	average_number_of_symbols_per_sentence
153	n_propn	total_number_of_pronouns	average_number_of_verbs_per_sentence
154	n_propn	total_number_of_proper_nouns	average_number_of_spaces_per_sentence
155	n_punct	total_number_of_punctuations	flesch_kincaid_reading_ease
156	n_sconj	total_number_of_subordinating_conjunctions	flesch_kincaid_grade_level
157	n_sym	total_number_of_symbols	gunning_fog_index
158	n_verb	total_number_of_verbs	smog_index
159	n_space	total_number_of_spaces	coleman_liau_index
160	n_uadj	total_number_of_unique_adjectives	automated_readability_index
161	n_uadp	total_number_of_unique_adpositions	reading_time_for_fast_readers
162	n_uadv	total_number_of_unique_adverbs	reading_time_for_average_readers
163	n_uaux	total_number_of_unique_auxiliaries	reading_time_for_slow_readers
164	n_uconj	total_number_of_unique_coordinating_conjunctions	
165	n_udet	total_number_of_unique_determiners	
166	n_uintj		
167	n_unoun		
168	n_unum		
169	n_upart		
170	n_upron		
171	n_uproprn		
172	n_upunct		
173	n_usconj		
174	n_usym		
175	n_uverb		
176	n_uspace		
177	a_adj_pw		
178	a_adp_pw		
179	a_adv_pw		
180	a_aux_pw		
181	a_cconj_pw		
182	a_det_pw		
183	a_intj_pw		
184	a_noun_pw		
185	a_num_pw		
186	a_part_pw		
187	a_propn_pw		
188	a_propn_pw		
189	a_punct_pw		
190	a_sconj_pw		
191	a_sym_pw		
192	a_verb_pw		
193	a_space_pw		
194	a_adj_ps		
195	a_adp_ps		
196	a_adv_ps		
197	a_aux_ps		
198	a_cconj_ps		
199	a_det_ps		
200	a_intj_ps		
201	a_noun_ps		
202	a_num_ps		
203	a_part_ps		
204	a_propn_ps		
205	a_propn_ps		
206	a_punct_ps		
207	a_sconj_ps		
208	a_sym_ps		
209	a_verb_ps		
210	a_space_ps		
211	fkre		
212	fkgl		
213	fogi		
214	smog		
215	cole		
216	auto		
217	rt_fast		
218	rt_average		
219	rt_slow		

Table 20: Linguistic Features (110-219)