Text vs. Transcription: A Study of Differences Between the Writing and Speeches of U.S. Presidents

Mina Rajaei Moghadam
Northern Illinois University
mina.rajaei.moghadam@niu.edu

Mosab Rezaei
Northern Illinois University
mosab.rezaei@niu.edu

Gülşat Aygen
Northern Illinois University
gaygen@niu.edu

Reva Freedman
Northern Illinois University
rfreedman@niu.edu

Abstract

Even after many years of research, answering the question of the differences between spoken and written text remains open. This paper aims to study syntactic features that can serve as distinguishing factors. To do so, we focus on the transcribed speeches and written books of United States presidents. We conducted two experiments to analyze high-level syntactic features. In the first experiment, we examine these features while controlling for the effect of sentence length. In the second experiment, we compare the high-level syntactic features with low-level ones. The results indicate that adding high-level syntactic features enhances model performance, particularly in longer sentences. Moreover, the importance of the prepositional phrases in a sentence increases with sentence length. We also find that these longer sentences with more prepositional phrases are more likely to appear in speeches than in written books by U.S. presidents.

1 Introduction

Scholars across various fields have sought to answer what makes writing different from speaking. The answers range from the notion that there is no fundamental difference to the belief that they are entirely distinct domains. These investigations lead to in-depth explorations with different approaches and perspectives, depending on the population or the system under study. For example, some scholars look for answers to support non-native speakers during the language acquisition process, while others attempt to measure the cognitive load through the sound or word production process. Likewise, some try to enhance our ability to program machines and unlock new insight into the differences between spoken and written text.

Rajaei Moghadam et al. (2024) investigated the difference between speaking and writing, focusing on morphological, lexical, and syntactic features at both sentence and chunk levels. They showed the superiority of BERT (Devlin et al., 2018) as well as the importance of sentence length, percentage of nouns, percentage of verbs, and depth of the parse tree.

In this paper, we expand the corpus and focus on high-level syntactic features to examine their effectiveness in distinguishing the transcriptions of speeches and written books by United States presidents. We analyze linguistically inspired features instead of simply counting categories.

Throughout the paper, the text is analyzed in different sentence lengths, categorized as short, medium, and long. We use CoreNLP (Manning et al., 2014) as a state-of-the-art to parse the sentences. Pinto et al. (2016) provide a comparison of several NLP toolkits, including NLTK, OpenNLP, and Stanford CoreNLP on both formal and informal texts and conclude that depending on the task and text types, the toolkits perform differently.

Regarding text similarity measurement, Wang and Dong (2020) recommend using a combination of techniques and models for higher accuracy, concluding that no single method works best for all similarity measurement tasks in NLP. For this investigation, we use Support Vector Machines (SVM), Decision Tree (DT), Random Forest (RF), and BERT as our machine-learning classification models to answer the following questions:

RQ1: How do syntactic features impact detection performance in sentences with different lengths?

RQ2: Which syntactic features improve the model performance in distinguishing written sentences from transcribed spoken sentences?

In the literature review, we will explore how other scholars have approached these questions. The methodology will be detailed in the procedure section. Following that, we will present the re-

Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities, pages 352–361
November 16, 2024.

sults and discussion, draw conclusions, outline the limitations, and finally offer future research. The artifacts of this work are published online.[1]

2 Related Work

In this section, we explore related findings in linguistics, cognitive science, and computer science.

2.1 Linguistic and Syntactic Analysis

In comparing writing and speaking, Blankenship (1962) found more general similarity than dissimilarity in sentence patterns with little variation in sentence length, along with mood indicators like imperative. Additionally, Blankenship found more passive constructions in writing. Working with a group of psychology students, Drieman (1962) realized that, compared to speaking, writing has shorter text with more diverse vocabulary and more multisyllabic and longer words. Later, DeVito (1967) analyzed samples of spoken and written language from university professors and concluded that spoken language relies more on verbs and adverbs, while written language uses more nouns and adjectives. These differences were further quantified by analyzing ratios between parts of speech, showing that speech uses fewer qualifiers than written language. O'Donnell (1974) used samples by the same male adult and examined syntactic features like gerunds, passive constructions, and attributive adjectives, which can be found more in written language. Einhorn (1978) also kept the subject and content similar, as she believed this would help understanding the effect of mode of communication. She worked with the writing and recorded speeches of ten famous men, and even though speeches were edited for publication, she still found many differences. For example, they contained more personal references, both singular and plural.

Through multiple attempts to understand why writing and speaking differ, Biber (1986b) identifies three key parameters, interactive vs. edited text, abstract vs. situated content, and reported vs. immediate style, that underlie textual variation in English. Additionally, Biber (1986a) believes that for such studies a comprehensive approach can capture the existing complexity between these two modalities. However, Biber and Grey (2011), in contrast to the conventional view, showed that both conversation and academic writing are grammatically complex, though the sources of complexity are different. In writing, sentences are compressed due to more use of phrasal expressions, including prepositional phrases as post-modifiers.

2.2 Linguistic and Contextual Influence

Akinnaso (1982) believes these differences are rooted in the objectives of the speaker and writer as well as the communicative and situational context. Chafe (1979), while underscoring the matter of context, identifies integration and involvement as two key distinctions between spoken and written language, meaning that writing is more integrated due to its coherent structure while speaking has a higher involvement rate as speakers are more engaged with the audience. Redeker (1984) believes the four categories of involvement, integration, detachment, and fragmentation work better for such distinctions and similarly refers to speaking as a mode with higher involvement that contains more self-reference items. Poole and Field (1976) highlight that oral language has simpler structures with more adverbial elaboration, which reflects the immediate and personal nature of spoken communication and recalls the importance of communicative context.

2.3 Cognitive Science

Early studies like Woolbert (1922) argue that despite some similarities between these two modes of communication, they are fundamentally different. Woolbert categorizes three processes in writing, thought, language, and typography, while identifying four processes in speaking including thought, language, voice, and action. Therefore, Woolbert counts both production mediums as means to manifest thought. Olson (1996) sees writing as a gateway for studying language and understanding the relationship between writing and cognition.

Liu (2023) reviews the distinction between production, perception, and form. During the production process, voice quality in speaking and, equally, punctuation in writing convey the meaning. Perception deals with the immediacy of feedback. Regarding form, Liu identifies differences in three main areas: lexical richness, grammar, and structure. Trying to explore the possibility of language measurement, Fairbanks (1944) and Mann (1944) worked on two groups: freshman students and individuals with schizophrenia. Both employ methods such as type-token ratios and grammatical analysis, including examining prepositions and conjunctions. The most notable difference, as Fairbanks men-

[1] https://github.com/mosabrezaei/Text-vs.-Transcription

tions, is the increased use of personal pronouns by patients with schizophrenia.

In another study, Rezaii (2022) works on productions of individuals with primary progressive aphasia and finds that familiarity with terms and topics decreases the cognitive load, thus easing the production process by either use of complex syntax structure with simple vocabulary or the opposite. Cleland and Pickering (2006) notice the use of syntactic priming, meaning that the speaker tends to reuse previously used syntactic structure. Their results show that during the production process, syntax is accessed the same way in both speaking and writing, suggesting a similar underlying cognitive mechanism.

DeVito (1966) highlights that writing has greater verbal diversity than speaking due to differences in the encoding process, including time constraints and pressure on the speaker when uttering a sentence, which is the reason speakers use more familiar and shorter words. Such differences show the distinct cognitive and linguistic demands on individuals through the production process. Likewise, Chafe and Tannen (1987) look at structural differences, cognitive implications, and social functions. They similarly refer to the immediate and context-dependent characteristics of speaking as opposed to writing. Gray and Biber (2013) analyze lexical frames in academic prose and conversational English. Their study shows that writing tends to use more grammatical structure and function word-based frames, while in conversation, fixed, verb-based frames are more common, which also, like DeVito, reflects the immediate and interactive aspects of spoken communication.

2.4 Computer Science

While Biber (2020) invites more investment in phonetic and phonological corpora to help in studies of speaking, Pangtay-Chang (2009) shows that text-based computer-mediated communication is becoming similar to what we produce in oral communication.

Understanding differences in writing and speaking will serve other areas of research like human-human-computer interactions. With this focus, Akhtiamov et al. (2017) analyzed speech through acoustical, syntactical, and lexical lenses. Ultimately, their study suggests a greater reliance on conversational context rather than acoustic cues. Similarly, Balagopalan et al. (2020) use NLP to detect symptoms of Alzheimer's disease (AD), which impacts both the content and acoustics of spontaneous speech. This study reveals that fine-tuned BERT models outperform traditional feature-based methods in detecting cognitive impairments associated with AD.

Such exploration will also be useful for related research in stylistics. Blankenship (1962) concluded that the formation of a syntactic structure is a matter of individual style; therefore the medium of delivery, whether writing or speaking, has minimal influence. Kurzynski (2023), through an analysis of perplexity, systematicity, and characteristic words of Mao Zedong, introduces these metrics as helpful ones to understand Mao's writing style. In another study, Freedman (2017) employs syntactic and bag-of-words approaches to distinguish different sections of the book of Isaiah. Also, Freedman and Krieghbaum (2014) used features like prepositional phrases along with machine learning techniques to investigate student responses. Expanding on stylistics, Khalid and Srinivasan (2020) used 262 stylistic features to analyze style across nine online communities to explore the importance of style in these communities rather than individual style. They found higher accuracy in style-based prediction as opposed to content-based predictions of community membership, particularly in smaller data sets.

Rajaei Moghadam et al. (2024) study syntactic and non-syntactic features to identify the most important ones for detecting spoken and written textual data. However, their study did not examine high-level constructs like prepositional phrases. Katre (2019), with a discourse analysis approach, used NLTK and Matplotlib to process a large corpus of political speeches to create visual tools like lexical dispersion plots, time-series plots, word clouds, and bar graphs.

Berriche and Larabi-Marie-Sainte (2024) examine writing style differences between human and ChatGPT-generated content. They employed classical classifiers and ensemble methods, training them with over 30 stylometric features. They extracted lexical and syntactic features including the frequency of conjunctions, pronouns, and prepositions. Through multiple experiments, they concluded that the ensemble learning classifiers outperformed the classical classifiers. Regarding style generation, Montfort et al. (2021) focus on generating narrative style (not the plot) with referring ex-

pressions. In other words, they explore how changing the referring expressions can model different literary styles. By keeping all other influential elements in the discourse constant and changing only reference conventions, they emphasize the use of nouns and noun phrases for generating different writing or narration styles.

3 Procedure

In this section, we describe the dataset and the extracted features.

3.1 Dataset

As outlined in the future work section of Rajaei Moghadam et al. (2024), we aimed to extend the number of extracted sentences. Therefore, in this study, we have expanded the corpus volume, which now contains 41,306 sentences, comprising 20,654 spoken samples and 20,652 written samples, compared to the earlier dataset of 13,600 spoken and 13,600 written samples.

We obtained transcriptions of spoken language from Miller Center of Public Affairs University of Virginia (2022), which covers transcriptions from George Washington to the present time. For the writing samples, we used ten complete books written by presidents, three of which we obtained from Project Gutenberg (n.d.).

To ensure the accuracy of calculations, all the pages that were not part of the main content were removed. Furthermore, multiple whitespaces were changed into single whitespaces. For sentence extraction, we used the *nltk* library (Bird et al., 2009), while CoreNLP (version 4.5.7) was employed for tokenization and word counting.

3.2 New Features

In the exploration of what exactly makes writing and speaking different, there is no single definite answer. Therefore, in addition to utilizing some of the features from Rajaei Moghadam et al. (2024), we will examine the following six features:

- Pronoun and noun phrases in the subject
- Passive and active sentences
- Comparative and superlative
- Imperative structures
- Conjunction phrases
- Prepositional phrases

3.2.1 Pronoun and Noun Phrase in Subject

We examined syntactic subjects to determine whether they were occupied by noun phrases (NP) or pronouns (PRN). Rajaei Moghadam et al. (2024) counted noun phrases and personal pronouns as separate features. In this paper, we only consider these two elements in the subject position. Such analysis deepens our understanding of nominal construction and sentence complexity in both modalities.

According to de Marneffe and Manning (2008), a nominal subject (*nsubj*) refers to a noun phrase that is the syntactic subject of a clause. Here, we use a combination of the parse tree and the Enhanced Dependency subsystem of Stanford CoreNLP to identify nominal subjects and their referents with higher accuracy.

3.2.2 Passive and Active Sentences

According to Aygen (2016), the active voice is the typical form in which the subject of the sentence is the agent. To do this, PassivePy package (Sepehri et al., 2023) in the SpaCy library (Honnibal et al., 2020) enables us to compute active, agentless passive, and agentive passive forms.

3.2.3 Comparative and Superlative

The comparative form is used to compare two sets of entities, whereas the superlative form compares more than two sets of entities or groups (Aygen, 2016). We extracted comparative and superlative structures with JJR, JJS, RBR, and RBS tags from the dataset using Stanford CoreNLP. This extraction includes irregular forms, such as "good", "well", and "best", in addition to those that end with "-er" and "-est" or contain indicators like "more" and "most".

3.2.4 Imperative

The imperative mood is used in direct requests or commands. According to Aygen (2016), imperatives do not have tense or aspect markers and have an implied subject (you). Therefore, this analysis focuses on structures without a stated subject and verbs without tense or aspect modifiers such as gerunds. To achieve this, we use StanfordCoreNLP to extract only sentences that begin with a VB tag.

It is common to find fragments and informal questions in spoken language that start with the base form of the verb, such as "Want fries?", which could be counted as an imperative structure. To address this, we examined the role of punctuation and ultimately decided to only consider sentences

that start with a verb and end with a period or exclamation mark.

In other cases, phrases like "Sleep well, gentlemen" may structurally appear as imperatives but are not interpreted as commands. Similarly, some proverbs and idiomatic expressions use the base form of the verb without an explicit subject, much like imperatives, e.g., "Hit the nail on the head". Although these cases offer interesting avenues for further analysis, they fall outside the scope of this research since our primary focus is syntactic analysis.

3.2.5 Conjunction Phrases

Conjunctions are connector words that link two words, phrases, clauses, or sentences (Zokirjon kizi, 2023). While Rajaei Moghadam et al. (2024) focused on the percentage of coordinators, this study investigates conjunction phrases with the assistance of the Standford CoreNLP parse tree. The work by de Marneffe and Manning (2008) notes that the parser does not account for symmetrical relations, meaning that we do not observe two conjunction phrase tags (CONJP) in cases like correlative conjunctions, e.g., "not only...but also...". Based on this, we check only for the presence of a single CONJP tag in the parse tree.

3.2.6 Prepositional Phrases

Among the different parts of speech, prepositions are considered as function words and prepositional phrases as grammatical units that act as connectors, typically with noun phrases. They precede or follow other phrases or elements in a sentence to create another phrase or constituent. According to Benelhadj (2015), prepositional phrases exhibit varying levels of structural complexity and generally cannot be understood without considering other elements of the sentence.

For this feature, we first extract and calculate the percentage of each sentence occupied by prepositional phrases (PP) using the Stanford CoreNLP tags. Then, we calculate the percentages of words with PP tags that modify verbs vs. nouns or other parts of speech. In this calculation, words in nested PPs are labeled according to their closest parent.

3.3 Pre-existing Features

In this paper we continue to use the features in Rajaei Moghadam et al. (2024). The features are listed below by category.

Morphological aspects:
- Average syllables per word
- Average words per sentence
- Average characters per word

Lexical aspects of sentences:
- Number of words in a sentence
- Percentage of POS
- Percentage of personal pronouns

Syntactical aspects:
- Percentage of subordinate clauses
- Depth of parse tree
- Percentage of noun phrases
- Average length of noun phrases
- Yes/no questions
- Direct wh-questions

4 Experiments

In this section, the two experiments that we conducted will be described. The first focuses on evaluating syntactic features in sentences with different lengths, and the second analyzes both low- and high-level syntactic features.

4.1 First Experiment

In the first experiment, we evaluated the impact of sentence length on model performance. Given the important role of length (Rajaei Moghadam et al., 2024), we used this insight to minimize model dependency on sentence length. The goal was to determine whether the models performed better when trained on the entire dataset or when focused on specific sentence lengths.

We divided our dataset into three categories based on sentence length: sentences with 18 or fewer words were classified as short, those with more than 18 and up to 37 words as medium, and those with more than 37 words as long. The boundary numbers that define short, medium, and long sentences were determined based on the data distribution to ensure a sufficient number of samples in each category. Then we trained each model on each section, utilizing both syntactic and non-syntactic features.

4.2 Second Experiment

In the second experiment, we shift our focus from sentence length to features. In this experiment, we evaluate the effectiveness of the combination of features in sections 3.2 and 3.3. We ran four

models, SVM, DT, RF, and BERT, on the new feature set. The comparison of these results with the results from Rajaei Moghadam et al. (2024) will determine whether combining features can improve model accuracy. It is important to note that in both experiments the BERT model examines words in sequence but does not utilize any of these features, while the other models are used to analyze the features and their role in distinguishing spoken from written language.

5 Results and Discussion

The results of our first experiment, shown in Figure 1, indicate that accuracy across all models is lower for short sentences compared to longer ones. This suggests two key points.

First, the extracted syntactic features are more informative in longer sentences, and their rare occurrence in short sentences leads to lower performance across all models. Second, despite having more short samples in the dataset, the selected features performed better on longer sentences. This implies that there is likely to be higher accuracy in a larger dataset containing more long sentences.

It should also be noted that the similarity between the results for short sentences and the overall dataset is due to the large number of short sentences, which biases the models' performance. Furthermore, the greater similarity between the results for medium and short sentences, compared to medium and long sentences, is due to the closer boundary numbers for short and medium sentences. Although BERT does not have explicit access to the linguistic features, we note that it performs better than any of the other models.

As shown in Table 1, although the length is the most important feature, its importance decreases as the sentence length increases. On the other hand, as sentences get longer the importance of prepositional phrases (PP, PP_NP, PP_VP) increases significantly.

Looking at RQ2, Table 2 and Figure 2 complement each other. We compare accuracy metrics for spoken vs. written sentences in each model. Table 2 shows that models trained with high-level syntactic features alone tend to have slightly lower accuracy, partially because length is not included in the high-level features and partially because the high-level feature set contains fewer features. Moreover, as shown in the last column, combining all features improves the performance of the models, with Random Forest outperforming the others.

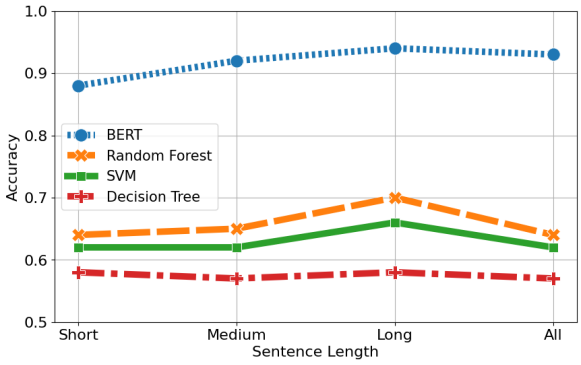

Figure 1: The performance of the models on sentences with different lengths.

Figure 2: Comparing the performance of models trained with low-level and high-level features.

Figure 2 compares the F1 scores of the SVM, RF, and BERT models for low-level and high-level features. The figure shows improvement in almost all performance metrics. Notably, the significant improvement in the SVM performance for the speaking class is particularly striking. This improvement may be attributed to several factors, such as an increased sample size, the use of different versions of CoreNLP, better feature normalization, and improved feature extraction methods. On the other hand, there is a slight reduction in performance for the writing class in the SVM model, which could indicate that the models are becoming more stable and less biased.

Table 3 presents a comparison between trained models with all syntactic features as well as the BERT model, which corresponds to the "All Features" columns in Table 2. When comparing Table 3 with a similar table in Rajaei Moghadam et al. (2024), we observe an overall improvement in model performance.

Notably, as sentences become longer, the perfor-

Table 1: The report of the six most important extracted features across four different sentence lengths. The numbers indicate the percentage of importance for each feature.

	Rank	Short		Medium		Long		All	
DT	1	Length	0.158	Length	0.115	**PP**	**0.089**	Length	0.117
	2	Verb	0.096	Noun	0.075	Length	0.086	Noun	0.084
	3	Noun	0.096	**PP**	**0.070**	**PP_VP**	**0.073**	Verb	0.077
	4	Words	0.075	Subord	0.069	**PP_NP**	**0.070**	Words	0.067
	5	Adverb	0.058	Verb	0.068	Words	0.066	**PP**	**0.063**
	6	**PP**	**0.054**	**PP_NP**	**0.064**	Verb	0.065	D_Tree	0.061
RF	1	Length	0.122	Length	0.095	**PP**	**0.093**	Length	0.094
	2	Verb	0.097	Noun	0.072	Length	0.087	Noun	0.075
	3	Noun	0.084	**PP**	**0.070**	**PP_NP**	**0.074**	Verb	0.075
	4	Words	0.065	**PP_NP**	**0.068**	**PP_VP**	**0.068**	Words	0.069
	5	D_Tree	0.059	Verb	0.065	Noun	0.066	**PP**	**0.064**
	6	**PP**	**0.052**	Subord	0.059	Subord	0.066	**PP_NP**	**0.061**

Table 2: The performance of the models with different levels of features. The low-level feature data comes from Rajaei Moghadam et al. (2024).

		Low-level Features		High-level Features		All Features	
		Precision	Recall	Precision	Recall	Precision	Recall
SVM	Spoken	0.59	0.67	0.57	0.57	0.61	0.65
	Written	0.63	0.54	0.58	0.58	0.64	0.59
DT	Spoken	0.56	0.56	0.55	0.49	0.56	0.57
	Written	0.57	0.58	0.55	0.60	0.58	0.57
RF	Spoken	0.62	0.69	0.58	0.52	0.63	0.67
	Written	0.66	0.58	0.58	0.64	0.66	0.62

mance of our syntactic features improves, probably because longer sentences provide more information, which enables the models to more accurately distinguish between speech and writing.

Figure 3 shows the feature importance ranking of the merged set of features. Note that high-level features like conjunction phrases (CONJP) and imperatives show less influence. This is possibly due to the fact that these features rarely appear in sentences in our dataset. For instance, the CONJP feature appears in only about 2 percent of all sentences.

As shown in Table 1 and Figure 3, two features stand out in distinguishing written text from the transcribed spoken text: length for all sentences and the percentage of PP for long sentences. The results show that longer sentences and a higher percentage of prepositional phrases appear more frequently in speech than in written books. In other words, U.S. presidents tend to use longer sentences and more prepositional phrases in their speeches than in their books. We conducted a statistical analysis and visualized the distribution of each of these features to better understand the relationship between these features and the classes of written and spoken texts. As expected, the t-tests in Table 4 show large absolute values and extremely small p-values for both features, indicating significant differences between the two classes. On the other hand, the small negative correlation values in Table 4 and the slight differences in class distribution as shown in Figure 4 indicate that increasing the percentages of these features decreases the probability of labeling a sentence as writing.

6 Conclusions

In this study, we analyze low-level and high-level syntactic features to identify the differences between the speeches and written books of presidents of the United States. We conducted two experiments to achieve these goals.

In the first experiment, sentences were divided into three categories: short, medium, and long. We found that, despite having fewer samples for long

Table 3: Comparing the results of trained models with all syntactic features and the BERT model.

	Labels	Precision	Recall	F1
SVM	Spoken	61%	65%	63%
	Written	64%	59%	61%
DT	Spoken	56%	58%	57%
	Written	58%	56%	57%
RF	Spoken	63%	67%	65%
	Written	66%	62%	63%
BERT	Spoken	92%	93%	**93%**
	Written	93%	92%	**93%**

Table 4: Statistical test on the length for all sentences and PP for long sentences.

	Length (all)	PP (long)
t-statistic	12.00	16.43
p-value	3.979×10^{-33}	9.029×10^{-60}
correlation	-0.05	-0.17

sentences, accuracy improves across all models. Increasing the sentence length also raised the importance ranking of prepositional phrases. Furthermore, the most significant features identified are sentence length, verb percentages, noun percentages, and prepositional phrases.

In the second experiment, we added a new set of syntactic features to morphological, lexical, and other syntactic features. The results showed that combining both groups of features improves model performance. Furthermore, sentence length and prepositional phrases emerged as the two important features in distinguishing the textual data of U.S. presidents. Based on our analysis, U.S. presidents are more likely to use prepositional phrases and longer sentences in their speeches than in their books.

7 Limitations

Although the dataset is balanced, we encountered some imbalanced features that appear rarely in sentences. For instance, there were only 351 imperative sentences, which account for less than 1 percent of all sentences. This limitation could affect future work in identifying effective features for this task.

Another limitation is the number of long sentences. By increasing the number of long sentences, or balancing with that of short sentences, we might observe higher model performance and and allow for the extraction of more accurate patterns.

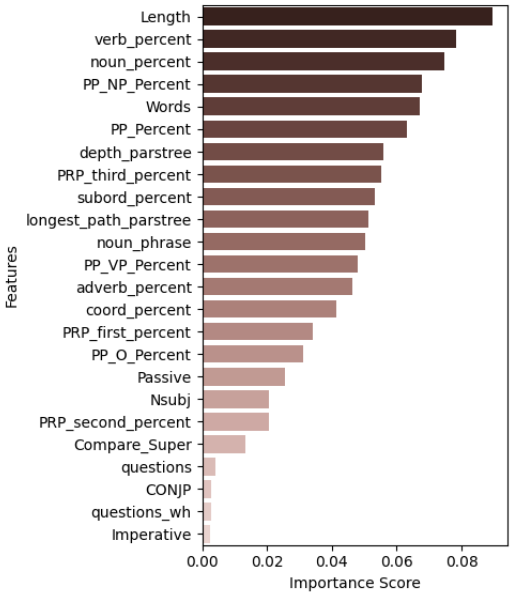

Figure 3: The importance of features in Random Forest model.

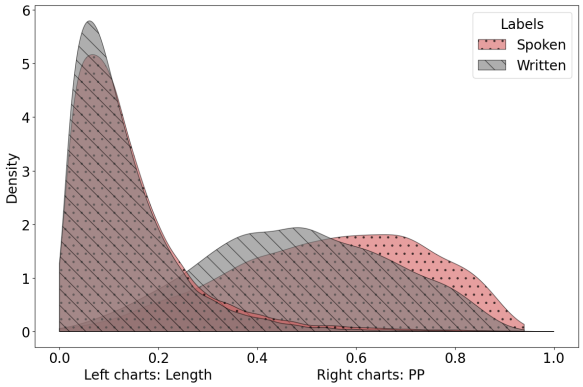

Figure 4: Distribution of the length and PP for each class using kernel density estimation (KDE).

8 Future Work

Explaining the differences between transcribed spoken and written text is an open area of research, with each study revealing more possible directions for future work. For example, our study demonstrates the weak performance of models on short sentences. For future work, new features need to be introduced and extracted in order to improve the model performance on short sentences.

We are going to perform a deeper analysis of the relationship between sentence length and its impact on both low- and high-level features. We will shift from a categorical approach to a regression-based analysis of sentence length. This means that instead

of categorizing sentences into three groups (short, medium, long), we will analyze the effects across the full range of sentence lengths.

Based on the importance of prepositional phrases, we plan to expand our analysis and study nested prepositional phrases. Additionally, we aim to apply deeper analysis to the different types of prepositional phrases introduced in this study, such as those modifying verb phrases or noun phrases.

References

Oleg Akhtiamov, Maxim Sidorov, Alexey A Karpov, and Wolfgang Minker. 2017. Speech and text analysis for multimodal addressee detection in human-human-computer interaction. In *Interspeech*, pages 2521–2525.

F Niyi Akinnaso. 1982. On the differences between spoken and written language. *Language and Speech*, 25(2):97–125.

Gulsat Aygen. 2016. *English Grammar: A Descriptive Linguistic Approach*, third edition. Kendall Hunt.

Aparna Balagopalan, Benjamin Eyre, Frank Rudzicz, and Jekaterina Novikova. 2020. To BERT or not to BERT: Comparing speech and language-based approaches for Alzheimer's disease detection. *arXiv preprint arXiv:2008.01551*.

Fatma Benelhadj. 2015. Prepositional phrases across disciplines and research genres: A syntactic and semantic approach. Doctoral dissertation, Department of English, University of SFAX.

Lamia Berriche and Souad Larabi-Marie-Sainte. 2024. Unveiling ChatGPT text using writing style. *Heliyon*, 10(12):e32976.

Douglas Biber. 1986a. On the investigation of spoken/written differences 1. *Studia Linguistica*, 40(1):1–21.

Douglas Biber. 1986b. Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language*, 62(2):384–414.

Douglas Biber. 2020. Corpus analysis of spoken discourse. *Pronunciation in Second Language Learning and Teaching Proceedings*, 11(1).

Douglas Biber and Bethany Grey. 2011. Is conversation more grammatically complex than academic writing? In *Grammatik und Korpora 2009: Dritte Internationale Konferenz*, pages 47–61.

Steven Bird, Edward Loper, and Ewan Klein. 2009. *Natural Language Processing with Python*. O'Reilly.

Jane Blankenship. 1962. A linguistic analysis of oral and written style. *Quarterly Journal of Speech*, 48(4):419–422.

Wallace Chafe. 1979. Integration and involvement in spoken and written language. In *2nd Congress of the International Association for Semiotic Studies*, pages 195–215.

Wallace Chafe and Deborah Tannen. 1987. The relation between written and spoken language. *Annual Review of Anthropology*, 16(1):383–407.

Alexandra A Cleland and Martin J Pickering. 2006. Do writing and speaking employ the same syntactic representations? *Journal of Memory and Language*, 54(2):185–198.

Marie-Catherine de Marneffe and Christopher D. Manning. 2008. *Stanford Typed Dependencies Manual*. Revised for the Stanford Parser v. 3.7.0 in September 2016.

Joseph A DeVito. 1966. The encoding of speech and writing. *Communication Education*, 15(1):55–60.

Joseph A DeVito. 1967. A linguistic analysis of spoken and written language. *Communication Studies*, 18(1):81–85.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Gerard HJ Drieman. 1962. Differences between written and spoken language: An exploratory study. *Acta Psychologica*, 20:36–57.

Lois Einhorn. 1978. Oral and written style: An examination of differences. *Southern Journal of Communication*, 43(3):302–311.

Helen Fairbanks. 1944. II. The quantitative differentiation of samples of spoken language. *Psychological Monographs*, 56(2):17–38.

Reva Freedman. 2017. Can natural language processing help identify the author(s) of the book of Isaiah? In *30th International FLAIRS Conference*, pages 297–300.

Reva Freedman and Douglas Krieghbaum. 2014. Effects of rewriting essays on linguistic measures of complexity. In *25th Annual Meeting of the Society for Text and Discourse*.

Bethany Gray and Douglas Biber. 2013. Lexical frames in academic prose and conversation. *International journal of Corpus Linguistics*, 18(1):109–136.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python. doi: 10.5281/zenodo.1212303.

Paritosh D Katre. 2019. NLP based text analytics and visualization of political speeches. *International Journal of Recent Technology and Engineering*, 8(3):8574–8579.

Osama Khalid and Padmini Srinivasan. 2020. Style matters! Investigating linguistic style in online communities. *Proceedings of the International AAAI Conference on Web and Social Media*, 14(1):360–369.

Maciej Kurzynski. 2023. The stylometry of Maoism: Quantifying the language of Mao Zedong. In *Proceedings of the Joint 3rd International Conference on Natural Language Processing for Digital Humanities and 8th International Workshop on Computational Linguistics for Uralic Languages*, pages 76–81.

Yijun Liu. 2023. Differences between spoken and written English. *Communications in Humanities Research*, 3:757–761.

Mary Bachman Mann. 1944. III. The quantitative differentiation of samples of written language. *Psychological Monographs*, 56(2):39–74.

Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60.

Miller Center of Public Affairs University of Virginia. 2022. Presidential speeches: Downloadable data. Accessed: 2022-03-17, Available at https://data.millercenter.org.

Nick Montfort, Ardalan SadeghiKivi, Joanne Yuan, and Alan Y Zhu. 2021. Using referring expression generation to model literary style. In *Proceedings of the Workshop on Natural Language Processing for Digital Humanities*, pages 64–74.

Roy C O'Donnell. 1974. Syntactic differences between speech and writing. *American Speech*, 49(1/2):102–110.

David R Olson. 1996. Towards a psychology of literacy: On the relations between speech and writing. *Cognition*, 60(1):83–104.

Yolanda Pangtay-Chang. 2009. IM conversations in Spanish: Written or oral discourse? *Illinois Language and Linguistics Society 1 (ILLS)*.

Alexandre Pinto, Hugo Gonçalo Oliveira, and Ana Oliveira Alves. 2016. Comparing the performance of different NLP toolkits in formal and social media text. In *5th Symposium on Languages, Applications and Technologies (SLATE' 16)(2016)*. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

Millicent E Poole and TW Field. 1976. A comparison of oral and written code elaboration. *Language and Speech*, 19(4):305–312.

Project Gutenberg. n.d. Project Gutenberg. Retrieved February 21, 2016, from https://www.gutenberg.org.

Mina Rajaei Moghadam, Mosab Rezaei, Miguel Williams, Gülşat Aygen, and Reva Freedman. 2024. Investigating lexical and syntactic differences in written and spoken English corpora. *Proceedings of the 37th International FLAIRS Conference*.

Gisela Redeker. 1984. On differences between spoken and written language. *Discourse Processes*, 7(1):43–55.

Neguine Rezaii. 2022. The syntax-lexicon tradeoff in writing. *arXiv preprint arXiv:2206.12485*.

Amir Sepehri, Mitra Sadat Mirshafiee, and David M Markowitz. 2023. PassivePy: A tool to automatically identify passive voice in big text data. *Journal of Consumer Psychology*, 33(4):714–727.

Jiapeng Wang and Yihong Dong. 2020. Measurement of text similarity: A survey. *Information*, 11(9):421.

Charles H Woolbert. 1922. Speaking and writing—A study of differences. *Quarterly Journal of Speech*, 8(3):271–285.

Tukhtasinova Zarina Zokirjon kizi. 2023. Conjunctions in English. *Modern Science and Research*, 2(9):29–35.