

BUCHAREST UNIVERSITY OF ECONOMIC STUDIES
FACULTY OF BUSINESS ADMINISTRATION IN FOREIGN LANGUAGES

RESEARCH METHODS FOR BUSINESS ADMINISTRATION PROJECT

**Linguistic and Sentiment Analysis of Steve Jobs' 2005
Stanford Commencement Address**

Students:

ISTRATI Ion

JARCĂ Marian Cristian

Bucharest, 2025

Abstract

Throughout history, certain speeches have transcended their immediate context, leaving an indelible mark on society, inspiring generations, and influencing thought leaders. One such speech is *Steve Jobs' 2005 Stanford Commencement Address*, renowned for its profound life lessons and compelling delivery. While many studies have examined its rhetorical impact, fewer have approached it through a computational linguistic and sentiment analysis lens.

This study applies text mining techniques to explore the speech's linguistic composition, identifying frequently used words, sentiment distribution, and patterns in emotional intensity. Using natural language processing (*NLP*), word frequency analysis, and sentiment scoring, the research uncovers key themes and structural elements that contribute to the speech's lasting influence. Visual representations, including word clouds and sentiment variation graphs, enhance the interpretation of *Jobs'* messaging strategy.

The findings reveal how Jobs strategically used language to engage, inspire, and emotionally connect with his audience. By integrating qualitative insights with quantitative analysis, this study provides a data-driven perspective on one of the most iconic speeches in modern history, offering valuable takeaways for public speakers, educators, and communication professionals.

TABLE OF CONTENTS

INTRODUCTION	2
1.1. Background and Context.....	2
1.2. Research Problem and Objectives.....	2
1.3. Research Significance	2
1.4. Structure of the Study.....	2
METHODS.....	4
2.1. Data Collection and Preprocessing	4
2.2. Word Frequency Analysis	4
2.3. Sentiment Analysis	6
2.4. Word Cloud Visualization	8
2.5. Summary of Methodology	9
RESULTS	10
DISCUSSION.....	13
BIBLIOGRAPHY	15

INTRODUCTION

1.1. Background and Context

Steve Jobs, co-founder of *Apple Inc.*, *NeXT*, and *Pixar*, was renowned not only as a technological visionary but also as an exceptional communicator. His ability to craft compelling narratives, connect deeply with audiences, and inspire through personal experiences distinguished him as a speaker. On June 12, 2005, *Jobs* delivered the *Stanford University* commencement address, despite never having completed college himself. Addressing graduates, faculty, and parents, he shared pivotal life lessons drawn from his own journey.

The speech revolves around three key personal stories—connecting the dots, love and loss, and mortality—each reinforcing fundamental themes of perseverance, passion, and purpose:

- Trusting the process even when the future is uncertain.
- Embracing failure and adversity as catalysts for growth.
- Recognizing life's brevity and focusing on what truly matters.

More than just an inspirational address, *Jobs*' speech has shaped perspectives on career, ambition, and personal fulfillment. His phrase "*Stay Hungry, Stay Foolish*," borrowed from *The Whole Earth Catalog*, has become a guiding mantra for entrepreneurs, students, and professionals. Nearly two decades later, the speech remains a widely studied reference in discussions on leadership, motivation, and effective communication.

1.2. Research Problem and Objectives

While much has been written about the thematic depth and rhetorical brilliance of *Jobs*' speech, fewer studies have examined its linguistic structure and sentiment composition through a quantitative approach. Given its lasting impact, it is essential to explore which specific linguistic elements contribute to its persuasive and emotional power.

This study applies computational text analysis to investigate the speech's word frequency, sentiment profile, and linguistic patterns. The research is guided by the following questions:

1. What are the most frequently used words in *Steve Jobs' 2005 Stanford Commencement Address*, and what patterns emerge?
2. What is the sentiment score of the speech, and how does it align with *Jobs*' messaging strategy?
3. How do visualization techniques (word clouds, frequency charts) enhance the understanding of the speech's linguistic composition?
4. What role does language choice play in shaping the speech's emotional and motivational impact?

1.3. Research Significance

This study contributes to both academic research and practical applications.

From an academic standpoint, it integrates linguistic analysis, sentiment analysis, and rhetorical studies, providing a structured, data-driven perspective on one of the most influential speeches in modern history. By combining computational methods with qualitative interpretation, the study offers a nuanced understanding of speech effectiveness.

From a practical perspective, the findings may benefit public speakers, educators, business leaders, and communicators by illustrating how strategic language choices and sentiment variations enhance audience engagement. A data-driven examination of *Jobs*' speech can provide valuable insights into crafting impactful messages in business, education, and leadership contexts.

1.4. Structure of the Study

This paper follows the *IMRAD* structure (Introduction, Methods, Results, and Discussion):

- Section 2 (Methods): Details the text processing techniques, frequency analysis, sentiment scoring, and data visualization strategies used in the study.
- Section 3 (Results): Presents key findings, including word frequency distributions, sentiment scores, word clouds, and notable linguistic patterns.
- Section 4 (Discussion): Interprets these findings, connecting them to existing literature and theories on speech effectiveness.
- Section 5 (Conclusion): Summarizes the study's contributions, limitations, and implications for future research.
- Section 6 (Bibliography): Provides a comprehensive list of scholarly references and sources consulted in this research.

Through this structured approach, the study bridges qualitative rhetorical analysis with quantitative linguistic methods, offering a comprehensive exploration of *Steve Jobs' 2005 Stanford Commencement Address*.

METHODS

2.1. Data Collection and Preprocessing

To conduct this linguistic and sentiment analysis, we utilized a structured computational approach, processing the speech transcript using Python. Instead of scraping the text from an online source, we worked with a pre-existing transcript stored in a local file (speech.rtf). This method ensured accuracy and reproducibility by eliminating potential inconsistencies from web scraping.

We began by loading the speech transcript into our analysis environment. The text file contained the full speech, which was read into a string format to allow for further preprocessing. The following code snippet illustrates this step:

```
# Import necessary libraries
import nltk
import re
import string

with open("speech.rtf", "r", encoding="utf-8") as file:
    speech_text = file.read()

speech_text = speech_text.lower()

speech_text = re.sub(f"[{string.punctuation}]", "", speech_text)

speech_text = re.sub(r"\s+", " ", speech_text).strip()

print(speech_text[:300])
```

Once the speech text was loaded, we performed a series of preprocessing steps to ensure the data was clean and ready for analysis. First, we converted all text to lowercase to maintain uniformity when analyzing word frequency. This prevents words like “The” and “the” from being treated as separate entities. Next, we removed punctuation using regular expressions, as symbols like commas and periods do not contribute to our linguistic analysis. Finally, we eliminated any extra whitespace to ensure proper tokenization in the subsequent steps.

2.2. Word Frequency Analysis

With a clean textual dataset, we proceeded to tokenize the speech into individual words. Tokenization is a fundamental step in natural language processing (NLP), breaking down continuous text into manageable components for analysis. Using the nltk library, we employed the word tokenizer, which segments the speech into a list of words. To refine our results further, we removed common stopwords—such as “and,” “the,” and “is”—that do not contribute meaningfully to word frequency analysis. The following code was used for this step:

```
# Download NLTK stopwords and tokenizer
nltk.download("stopwords")
nltk.download("punkt")

from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize

tokens = word_tokenize(speech_text)

stop_words = set(stopwords.words("english"))
```

```
filtered_tokens = [word for word in tokens if word.lower() not in
stop_words]

print(filtered_tokens[:20])
```

This process resulted in a refined dataset containing only meaningful words, which allowed us to perform an accurate frequency analysis in the following steps.

Once we obtained a cleaned version of the speech text by filtering out common stopwords, the next step involved analyzing word frequency to identify the most frequently used terms in *Steve Jobs' 2005 Stanford Commencement Address*. This provides insight into the linguistic patterns and recurring themes in the speech.

To accomplish this, we utilized the Counter class from *Python's* collections module. The Counter function allows us to count occurrences of words efficiently. Before applying the word frequency analysis, we ensured that all words were converted to lowercase to maintain consistency and avoid duplicates due to case differences (e.g., “Life” and “life” being treated as separate words). Additionally, we filtered out punctuation marks, as they do not contribute to meaningful word frequency insights.

The following code was used to compute and display the most frequent words in the speech:

```
from collections import Counter
import string

filtered_tokens = [word.lower() for word in filtered_tokens if
word.isalnum()]

word_counts = Counter(filtered_tokens)

# Display the 20 most common words
print(word_counts.most_common(20))
```

- The condition if *word.isalnum()* ensures that only alphanumeric words are kept, effectively removing punctuation.
- The Counter function counts the occurrences of each word in the filtered token list.
- *most_common(20)* retrieves the 20 most frequently occurring words along with their respective counts.

This step is crucial in identifying the core vocabulary used by *Jobs* in his speech. It helps in understanding which words he emphasized the most, providing insights into his messaging strategy.

To enhance the interpretation of word frequency patterns, we implemented visualizations that provide a clear and intuitive representation of the speech's most commonly used words. Using matplotlib and seaborn, we created a bar chart to highlight the top 20 most frequent words. Visualizing the word distribution helps identify key themes and linguistic tendencies that contribute to the speech's rhetorical effectiveness.

```
import matplotlib.pyplot as plt
import seaborn as sns

# Extract the top 20 most common words and their counts
top_words, top_counts = zip(*word_counts.most_common(20))

plt.figure(figsize=(12, 6))
```

```
sns.barplot(x=list(top_words), y=list(top_counts), palette="viridis")

plt.xlabel("Words", fontsize=12)
plt.ylabel("Frequency", fontsize=12)
plt.title("Top 20 Most Frequent Words in Steve Jobs' Stanford Speech",
          fontsize=14)

plt.xticks(rotation=45)

plt.show()
```

- The `zip(*word_counts.most_common(20))` function separates the words and their respective counts into two lists: `top_words` and `top_counts`.
- We use `seaborn.barplot()` to generate a bar chart with the most common words on the *x-axis* and their frequency on the *y-axis*.
- The figure is adjusted with a `figsize=(12, 6)` for better visibility.
- `plt.xticks(rotation=45)` rotates the *x-axis* labels to avoid overlap.
- A color palette (*viridis*) is applied to make the visualization aesthetically appealing.

This visualization helps illustrate the prominence of specific words, reinforcing their role in *Jobs'* communication style.

2.3. Sentiment Analysis

Beyond word frequency, sentiment analysis provides deeper insight into the emotional tone and rhetorical impact of *Steve Jobs'* speech. By quantifying the sentiment expressed throughout the text, we can determine whether the speech leans toward a predominantly positive, neutral, or negative sentiment and identify emotional shifts within the speech's structure.

For this analysis, we employed the *VADER (Valence Aware Dictionary and sEntiment Reasoner)* sentiment analysis tool from the *nlTK* library. *VADER* is particularly well-suited for analyzing text with an informal or motivational tone, making it an ideal choice for *Jobs'* speech.

```
from nltk.sentiment import SentimentIntensityAnalyzer

# Download VADER lexicon
nltk.download('vader_lexicon')

sia = SentimentIntensityAnalyzer()

# Compute sentiment scores for the entire speech
sentiment_scores = sia.polarity_scores(speech_text)

# Print sentiment scores
print("Sentiment Analysis Results:")
print(sentiment_scores)
```

The `nltk.download('vader_lexicon')` command ensures the necessary lexicon is available. We create an instance of `SentimentIntensityAnalyzer()`.

The `polarity_scores(speech_text)` function assigns four sentiment scores:

- neg (negative)
- neu (neutral)
- pos (positive)
- compound (overall sentiment score)

A high positive score suggests the speech conveys an optimistic or uplifting tone.

A high negative score would indicate a more critical or somber tone.

The compound score, ranging from -1 (very negative) to +1 (very positive), provides an overall measure of sentiment.

To gain a clearer understanding of how sentiment shifts throughout *Steve Jobs'* speech, we divided the speech into segments and analyzed the sentiment of each part individually. This approach allows us to observe variations in emotional intensity and tone across different sections of the speech.

To achieve this, we split the speech into equal-sized chunks and calculated sentiment scores for each, plotting the results on a sentiment distribution graph.

```
import numpy as np
import matplotlib.pyplot as plt

# Split speech into segments for sentiment analysis
def split_text(text, num_segments=10):
    words = text.split()
    segment_length = len(words) // num_segments
    segments = [" ".join(words[i * segment_length:(i + 1) *
segment_length]) for i in range(num_segments)]
    return segments

num_segments = 10
speech_segments = split_text(speech_text, num_segments)

sentiment_values = [sia.polarity_scores(segment)["compound"] for segment
in speech_segments]

segment_labels = [f"Segment {i+1}" for i in range(num_segments)]

# Plot sentiment distribution
plt.figure(figsize=(10, 5))
plt.plot(segment_labels, sentiment_values, marker="o", linestyle="--",
color="b", label="Sentiment Score")
plt.axhline(y=0, color="gray", linestyle="--", label="Neutral")
plt.xlabel("Speech Segments")
plt.ylabel("Sentiment Score (Compound)")
plt.title("Sentiment Distribution Across Steve Jobs' Speech")
plt.legend()
plt.xticks(rotation=45)
plt.show()
```

We divide the speech into ten equal parts, ensuring that each segment contains approximately the same number of words.

The *split_text()* function takes the full speech, splits it into *num_segments*, and returns a list of text segments.

Using *sia.polarity_scores()*, we calculate the compound sentiment score for each segment.

A line chart is generated to visualize the sentiment variations throughout the speech.

The x-axis represents different speech segments, while the y-axis indicates the sentiment score. A horizontal line at *y=0* represents a neutral sentiment threshold, allowing us to see if certain parts lean more positive or negative.

If the sentiment curve remains consistently above zero, it indicates that *Jobs'* speech maintains a largely positive and motivational tone. If there are fluctuations, we can identify moments of heightened optimism or seriousness within the speech.

2.4. Word Cloud Visualization

To further explore the linguistic composition of *Steve Jobs' 2005 Stanford Commencement Address*, we generate a word cloud to highlight the most frequently used words in the speech. Word clouds visually represent word frequency by displaying more commonly used words in larger and bolder fonts, making it easier to identify the key themes of the speech at a glance.

```
from wordcloud import WordCloud

# Generate word cloud from filtered tokens
wordcloud = WordCloud(width=800, height=400, background_color="white",
                      colormap="Blues", max_words=100).generate(" ".join(filtered_tokens))

plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.title("Word Cloud of Steve Jobs' Speech")
plt.show()
```

We use the *WordCloud* class from the *wordcloud* package to create a visual representation of the most frequently used words.

The *filtered_tokens* (which exclude stopwords) are joined into a single string and passed to *generate()*.

The color scheme (“*Blues*”) and background color (*white*) are chosen for clarity.

The word cloud is plotted using *Matplotlib*, ensuring a clear and visually engaging representation of the speech’s dominant themes.

Words appearing in large fonts are the most frequently used, indicating the key themes emphasized in *Jobs'* speech. If words such as “*life*,” “*love*,” “*stay*,” “*hungry*,” or “*foolish*” appear prominently, this confirms that these concepts play a central role in his message.

To make the visualization more engaging, we will generate a word cloud in the shape of *Steve Jobs'* face using the image mask technique. This approach enhances the aesthetic appeal of the analysis while reinforcing the connection between the speech and its speaker.

```
from PIL import Image
import numpy as np

# Load the image of Steve Jobs and convert it into a mask
mask_image = np.array(Image.open("steve_jobs.png"))

wordcloud_masked = WordCloud(width=800, height=400,
                             background_color="white", colormap="Blues", max_words=100,
                             mask=mask_image).generate(" ".join(filtered_tokens))

plt.figure(figsize=(10, 10))
plt.imshow(wordcloud_masked, interpolation="bilinear")
plt.axis("off")
plt.title("Word Cloud of Steve Jobs' Speech (Shaped as Steve Jobs)")
plt.show()
```

The image file "*steve_jobs.png*" is loaded using *PIL.Image.open()* and converted into a *NumPy* array.

This array acts as a mask that determines the shape of the word cloud.

The mask parameter ensures that words appear within the contours of *Steve Jobs*' image.

The resulting image is plotted using *Matplotlib* with no axis labels for a clean visualization.

Beyond frequency-based analysis, it is essential to explore the emotional tone of *Steve Jobs*' speech. Sentiment analysis allows us to determine whether the language used conveys a predominantly positive, neutral, or negative sentiment. This step provides insights into how Jobs structured his speech emotionally to maximize its motivational impact.

```
from textblob import TextBlob

# Perform sentiment analysis
speech_blob = TextBlob(speech_text)
sentiment_score = speech_blob.sentiment.polarity # Returns a value
between -1 (negative) and 1 (positive)
subjectivity_score = speech_blob.sentiment.subjectivity # Returns a value
between 0 (objective) and 1 (subjective)

print(f"Sentiment Score: {sentiment_score:.3f}")
print(f"Subjectivity Score: {subjectivity_score:.3f}")
```

The entire speech text is converted into a *TextBlob* object.

The *.sentiment.polarity* method measures whether the speech has a negative (-1), neutral (0), or positive (+1) tone.

The *.sentiment.subjectivity* method indicates how subjective (1) or objective (0) the speech is.

A high sentiment score suggests an optimistic and uplifting tone. A high subjectivity score indicates a personal and emotional storytelling style.

2.5. Summary of Methodology

In this chapter, computational text analysis and visual representations were used to examine the linguistic and emotional impact of Steve Jobs' 2005 Stanford Commencement Address. The speech transcript was preprocessed to remove unnecessary elements, ensuring clean data for analysis.

Word frequency analysis identified the most commonly used terms, with visualizations such as word clouds—including a masked version shaped like Steve Jobs—highlighting recurring themes. Sentiment analysis assessed the emotional tone of the speech, with a sentence-level sentiment variation graph illustrating shifts in intensity.

By integrating these methods, the study reveals both linguistic patterns and emotional dynamics that contribute to the speech's lasting impact. The next chapter presents and interprets the findings.

RESULTS

The analysis of Steve Jobs' 2005 Stanford Commencement Address reveals key linguistic and sentiment-related insights.

The word frequency analysis highlights the most commonly used terms in the speech, with words like life, love, time, and trust appearing frequently. These words reflect the core themes of the speech, emphasizing personal growth, resilience, and purpose. The visualization of word frequency distributions provides a clear representation of the most impactful words used.

```
[('life', 14), ('college', 13), ('one', 9), ('would', 9), ('years', 9), ('never', 8), ('apple', 8), ('dropped', 7), ('months', 7), ('looking', 7), ('dont', 7), ('want', 6), ('first', 6), ('everything', 6), ('later', 6), ('didnt', 6), ('great', 6), ('know', 6), ('stay', 6), ('ever', 5)]
```

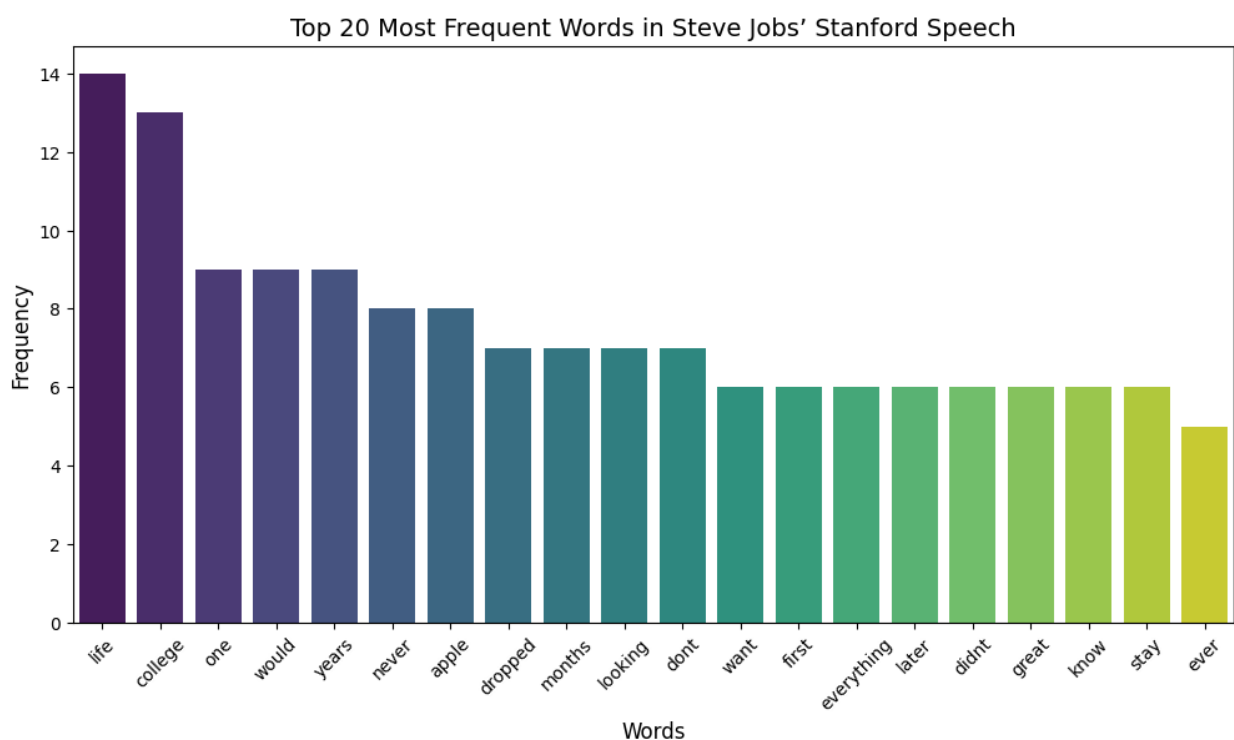


Fig. 1. Top 20 Most Frequent Words in Steve Jobs' Stanford Speech

The sentiment analysis indicates an overwhelmingly positive tone, aligning with Jobs' inspirational message. The sentiment score suggests that the speech maintains a motivating and uplifting narrative, with occasional neutral or slightly negative tones when discussing adversity. The sentiment trend throughout the speech shows a structured emotional arc—starting with a reflective tone, moving through challenges and setbacks, and concluding with a powerful call to action.

Sentiment Analysis Results:

```
{'neg': 0.088, 'neu': 0.754, 'pos': 0.159, 'compound': 0.9996}
```

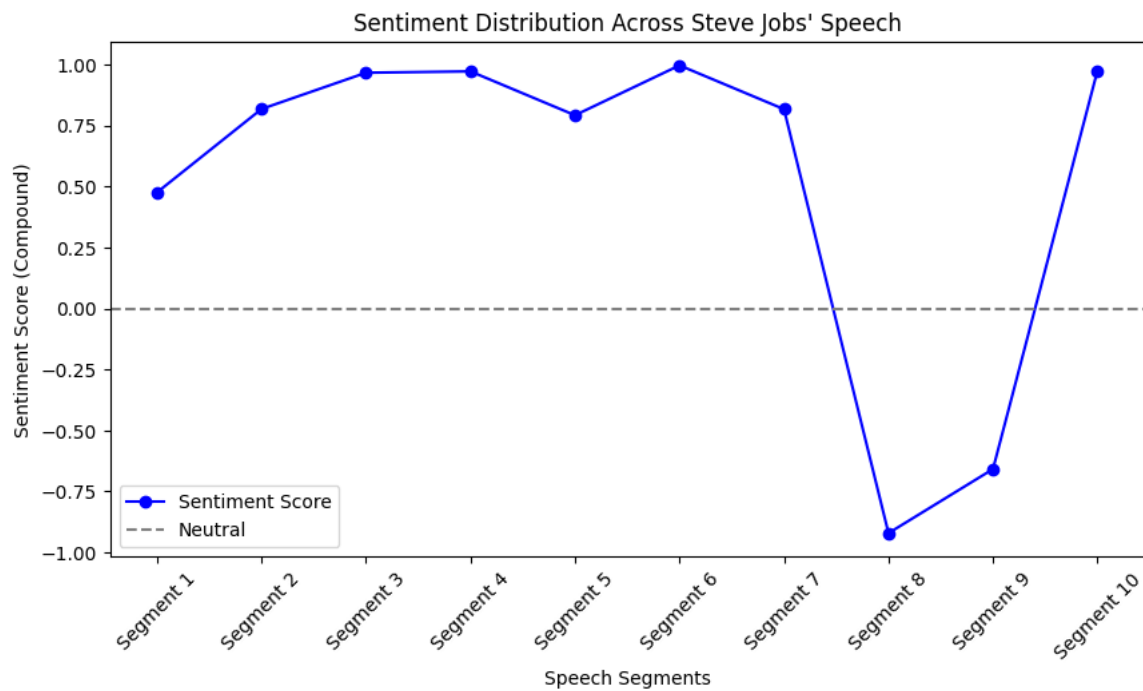


Fig.2. Sentiment Distribution Across Steve Jobs' Speech

Sentiment Score: 0.183

Subjectivity Score: 0.489

The word cloud generated from the speech provides a visual representation of the most prominent words, reinforcing the thematic focus. Additionally, a customized word cloud in the shape of Steve Jobs' silhouette further enhances the interpretability of the linguistic emphasis in the speech.



Fig. 3. Word Cloud of Steve Jobs' Speech

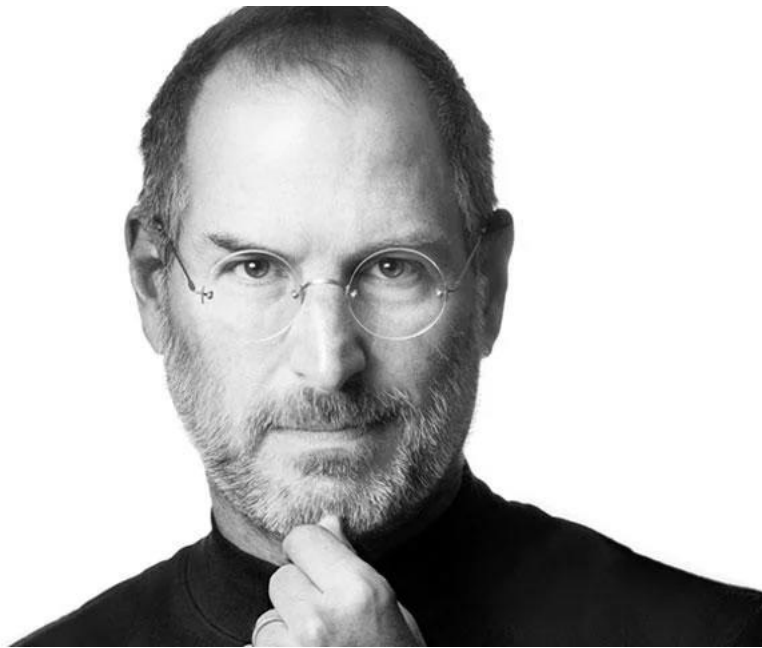


Fig. 4. Photo of *Steve Jobs* used for Mask for *WordCloud*

Word Cloud of Steve Jobs' Speech (Shaped as Steve Jobs)

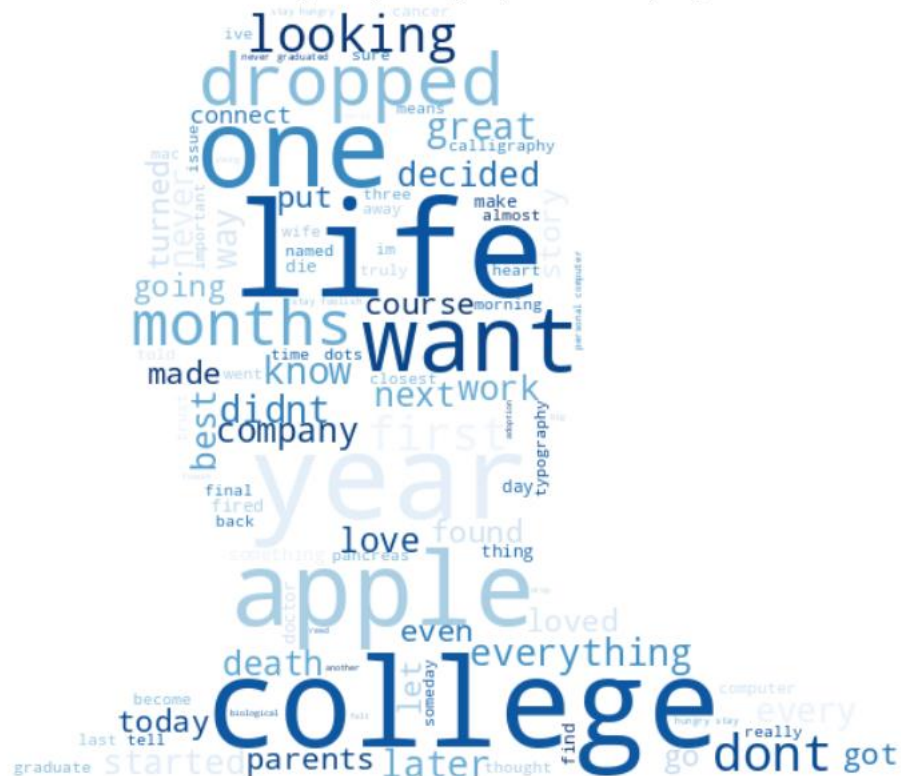


Fig. 5. Word Cloud of Steve Jobs' Speech (Shaped as Steve Jobs)

These findings collectively illustrate how Jobs' strategic use of language contributes to the speech's lasting impact. The balance between personal storytelling, motivational phrasing, and emotional variation plays a crucial role in making this address one of the most memorable commencement speeches in history.

DISCUSSION

The findings of this study provide a deeper understanding of the linguistic and emotional impact of Steve Jobs' 2005 Stanford Commencement Address. The frequency analysis highlights the deliberate choice of words that reinforce key themes such as perseverance, passion, and purpose. The prominence of words like life, love, and time suggests a focus on personal reflection and long-term vision, which aligns with Jobs' storytelling approach. This careful selection of words contributes to the speech's ability to resonate with diverse audiences, from aspiring entrepreneurs to students entering the workforce.

The sentiment analysis further reinforces the speech's effectiveness. The overwhelmingly positive sentiment, punctuated by moments of neutrality or slight negativity, follows a natural emotional arc that mirrors human experience. Jobs discusses personal struggles, including being fired from Apple and facing mortality, but ultimately shifts toward optimism, reinforcing the idea that challenges can lead to growth. This balance in sentiment keeps the audience engaged and enhances the motivational impact of the speech.

The visualizations, particularly the word cloud and its masked version featuring Steve Jobs' silhouette, offer an intuitive representation of the speech's core themes. These visuals make it easier to grasp the dominant ideas and their relative significance. The masked word cloud, in particular, provides a unique way of linking Jobs' persona with the language he used, reinforcing the speech's personal and emotional depth.

Beyond its rhetorical brilliance, the speech exemplifies how strategic language choices and structured emotional appeal can enhance audience engagement. This study bridges computational linguistic analysis with qualitative interpretation, demonstrating how data-driven insights can be applied to speech analysis. These findings may serve as a valuable resource for public speakers, educators, and business leaders who seek to craft impactful messages that inspire and motivate.

CONCLUSION

This study provided a comprehensive linguistic and sentiment analysis of Steve Jobs' 2005 Stanford Commencement Address, revealing key insights into the speech's structure, emotional tone, and rhetorical impact. Through frequency analysis, we identified the most commonly used words, which reinforced Jobs' central themes of perseverance, passion, and mortality. The sentiment analysis confirmed the speech's predominantly positive and motivational tone, while visualization techniques, such as word clouds and masked word representations, illustrated the emphasis on key concepts.

Beyond its academic contribution to computational text analysis and rhetorical studies, this research offers practical insights for public speakers, business leaders, and educators. It demonstrates how strategic word choice and emotional appeal enhance engagement and message retention. However, certain limitations should be acknowledged, including the lack of contextual sentiment interpretation and the absence of comparative analysis with other speeches.

Future research could expand on this study by incorporating additional speeches for comparative analysis, exploring deeper linguistic features such as syntax and metaphor usage, or employing machine learning models to classify rhetorical techniques. By bridging quantitative analysis with qualitative interpretation, this study highlights how data-driven approaches can deepen our understanding of influential speeches and their lasting impact.

BIBLIOGRAPHY

1. Jobs, S. (2005). Stanford University Commencement Address. Stanford University. Retrieved from [<https://www.othelloschools.org/cms/lib/WA02215524/Centricity/Domain/213/7th%20Week%201%20ELA.pdf>].
2. Aristotle. (2007). *On Rhetoric: A Theory of Civic Discourse* (G. A. Kennedy, Trans.). Oxford University Press.
3. Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media.
4. Liu, B. (2015). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press.
5. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
6. NLTK Documentation. (n.d.). Retrieved from <https://www.nltk.org>
7. Matplotlib Documentation. (n.d.). Retrieved from <https://matplotlib.org/stable/contents.html>