

INTERNATIONAL JOURNAL OF
BEHAVIOR STUDIES IN ORGANIZATIONS

Corpus Analysis: Global Leaders' Speech

Yasuo Nakatani 

Faculty of Economics, Hosei University, Tokyo, Japan

Received: 10 August 2024

Revised: 24 September 2024

Accepted: 7 October 2024

ABSTRACT

The importance of communication strategies for global CEOs has been recognized. It is believed that their effective messaging to stakeholders and society not only inspires employees but also impacts stock prices and business performance. However, previous research has primarily focused on qualitative studies, such as case studies and interviews. There have been few studies that collected large amounts of speech data from business leaders and analyzed it using reliable methods. This study employs corpus data analysis to examine the issues by analyzing a large amount of linguistic data from speeches of global business executives at the Oxford Union and TED Talks. Using the concordance software AntConc, a keyword analysis was conducted against representative large-scale English corpora to extract keywords frequently used by the global business leaders ($p < 0.05$). Similarly, a cluster analysis was performed on these characteristic words to investigate specifically how the communication strategies are constructed, how the audience is guided, and how persuasion is achieved. The results indicated that global business leaders effectively use significant communication strategies to achieve their goals.

Keywords:

Corpus Analysis, Global Leaders, Communication Strategies

How to cite this article:

Nakatani, Y. (2024). Corpus Analysis: Global Leaders' Speech. *International Journal of Behavior Studies in Organizations*, 12, 23-31. <https://doi.org/10.32038/jbso.2024.12.02>

Correspondence:
nakatani@hosei.ac.jp



Introduction

In modern business, the role of leaders in convincing and inspiring their followers, fostering empathy, and achieving goals has been emphasized (e.g., [Kotter, 2001](#)). To share and guide actions effectively, leaders often employ strategic Communication Strategies (CS) ([Conger, 1991](#); [Nakatani, 2010](#)). In business contexts, CS plays a critical role in persuading others, building empathy, and sharing action guidelines to accomplish objectives (e.g., [Johnson & Hackman, 2018](#)).

[Johnson and Hackman \(2018\)](#) highlight the importance of CS and argue for a more detailed analysis of top leaders' discourse. However, collecting spoken data for analysis is not always straightforward ([Nakatani, 2010](#)). As a result, many studies rely on qualitative methods such as observation and interviews, as seen in [Mayfield et al. \(2014\)](#), with fewer studies offering robust empirical validation (e.g., [Locker & Kaczmarek, 2013](#)). While these studies provide valuable insights into the examination of leaders' CS, a more detailed analysis of spoken data is essential for building reliability in the findings. Some research has addressed this by collecting business leaders' presentation scripts for corpus-based data analysis ([Almaghlouth, 2022](#)). However, in this field, vocabulary research dominates, and there has been limited use of corpus analysis for the study of effective CS, suggesting the need for further investigation ([Nakatani, 2016](#)).

In this study, we focus on these issues by collecting speech data from global business executives who have delivered talks at prominent forums such as The Oxford Union and TED Talks, creating a substantial Business Leaders' Speech Corpus (BLSC). We conducted a keyword analysis of these speeches, using the widely recognized F-BROWN Corpus as a reference corpus. Additionally, we extracted clusters of these keywords using a concordancer software to analyze how they were utilized as CS. The results indicate that global business leaders frequently and effectively employ distinctive CS to guide, persuade, and inspire action among their audiences.

Background

The Oxford Union (OU) and its leaders

The Oxford Union (OU) was established in 1825 as a debating society led by students of Oxford University. Its primary aim has been to foster free and open debate, encouraging students from various colleges and disciplines to refine their skills and prepare for future leadership roles. Over the years, the OU has attracted aspiring leaders not only from the United Kingdom but

from around the world (Gerald, 1970). One of the most notable early Presidents of the OU, William Gladstone, honed his debating abilities here before becoming a model figure of British parliamentary democracy. His legacy influenced subsequent OU Presidents like Edward Heath and Boris Johnson, both of whom went on to serve as Prime Ministers of the UK (Graham, 2005). A particularly prominent event hosted by the OU is the weekly formal debate, held every Thursday during the academic term. These debates feature not only students but also politicians, business leaders, and social activists who are at the forefront of their respective fields. Among the most famous debates in OU history is the 1933 “King and Country” debate, held in the prelude to World War II. The motion, “This House will under no circumstances fight for its King and Country,” was passed, leading to a lasting legend that this outcome may have influenced the aggressive ambitions of Hitler and Mussolini. Another noteworthy event occurred in 1964, when the OU invited Malcolm X, a radical Black leader, to debate. The debate was televised, and for many in the UK, it was their first exposure to the realities of racial issues in the world. This debate had a profound impact on white audiences and is seen as a precursor to later movements such as Black Lives Matter. Beyond debates, the OU has a long history of inviting world leaders to give talks and participate in Q&A sessions. Distinguished guests have included figures like Marie Theresa, the Dalai Lama, Michael Jackson, and former U.S. President Richard Nixon. In the business world, notable leaders such as Bernard Arnault of LVMH and James Quincey, CEO of Coca-Cola, have also addressed OU audiences (Hughes & Phillips, 2000). These global leaders engage with students, sharing their experiences and discussing contemporary challenges, thereby contributing to the development of future leaders. In light of its role in shaping future leaders, the OU serves as a rich source of data for this research. The actual recordings of guest speeches held at the OU have been transcribed into text and utilized as corpus data for analysis in this study.

TED Talk

Technology, Entertainment, and Design (TED) is an organization established in 1984 that focuses primarily on presentations delivered in English. Headquartered in New York, USA, TED hosts its annual TED Conference in Vancouver, Canada, where world leaders are invited to speak. In 2006, TED launched the TED Talk video distribution project, allowing prominent figures from politics, economics, and society to deliver speeches lasting approximately 18 minutes. These talks are broadcast globally, wielding significant influence (e.g., Nakatani 2016).

Research questions

In this study, we utilize speech data from TED Talks, where global leaders give presentations, as well as speech data from the Oxford Union (OU), and developed Business Leaders' Speech Corpus (BLSC). To ensure a robust dataset, we combined both sources for a comprehensive analysis. Our research question is: what kind of English-language business leaders use to attract, guide, persuade, inspire, and let the audience take action? To answer such questions, we set up the following hypotheses:

Hypothesis 1: Global business leaders use specific keywords to organize their speeches.

Hypothesis 2: They use effective communication strategies.

Method

The current study employs a corpus-based data analysis approach to test the two hypotheses, as detailed below.

Speech corpus analysis

The corpus data analysis involves the collection and analysis of large volumes of spoken or written text that have been converted into text data. Using tools such as concordancer, researchers could extract and statistically analyze lexical frequency and collocations from this data (Biber et al., 2002).

For this study, BLSC, which consisted of speech data from 99 leaders who spoke at the Oxford Union (OU) and TED Talks since 2018, were randomly selected in a 1:2 ratio, respectively. The selection of topics was also random. The larger number of TED Talk speakers justified the decision to set this ratio. To maximize the dataset, both OU debate and TED Talk speeches were treated as monologues delivered by leaders, and differences between the two formats were not considered in the analysis.

Analytical method

To enhance the reliability and validity of the extracted communication strategies (CS) used by business leaders, the following procedure was employed. The analysis involved a total of 231, 712 words from the BLSC data of 99 leaders.

Word list creation

Using the concordance software *AntConc* 4.0.10, word lists were generated from the corpus data of business leaders. A word list is a collection of words ordered by frequency of use within a given corpus. However, the most frequently used words in a specific corpus are not necessarily characteristic of that corpus. By comparing word frequencies with another corpus, words that appear relatively more often in the target corpus can be identified. In this study, the BLSC served as the Target Corpus, while the F- BROWN Corpus was used as the Reference Corpus.

Keyword analysis

The process of comparing word frequencies between the two corpora to identify characteristic vocabulary in the target corpus is known as keyword analysis. Using *AntConc* keyword function, a keyword analysis was conducted to identify the distinctive vocabulary of the business leaders. The statistical significance of keywords was determined through a Log-Likelihood test, with a rejection threshold of $p < 0.05$. This value is considered appropriate in corpus analysis research, as seen in [Nelson \(2006\)](#), and indicates that a word can be identified as a keyword with 99.5% confidence. The key indicator in this analysis is the "Keyness" value, and generally, a Keyness score of 15.13 or higher is considered statistically significant at $p < 0.05$ ([Nelson, 2006](#)).

Cluster and plot analysis

Cluster analysis was then performed on the identified keywords to extract characteristic expressions and Communication Strategies (CS). Cluster analysis identifies phrases that frequently occur with specific keywords, revealing stronger contextual associations. By expanding from individual words to phrases, the meaning and usage of the vocabulary become clearer. Furthermore, plot analysis of these phrases was conducted to pinpoint where specific CSs are employed within the structure of the English speeches. This analysis helps clarify how and in what context these distinctive expressions are used within particular paragraphs ([Nakatani, 2023](#)).

Results

Keywords

The analysis results reveal the top-ranking keywords in the business leader corpus from both the Oxford Union and TED Talk speeches. Table 1 presents the top 10 words that are most characteristic of presentations given by the business leaders. The ranking in Table 1 is based on the prominence of each keyword, with "we" being the most significant. The column "Freq-Tar" shows the frequency of the word in the target corpus (BLSC), where "we" appears 2,573 times. In contrast, "Freq-Ref" shows its frequency in the reference corpus (F-BROWN), where it appears 964 times.

Table 1.

Business leaders' keyword

Rank	Type	Freq_Tar	Freq_Ref	Likelihood
1	we	2573	964	3529.221
2	And	1987	560	3132.610
3	you	2997	2195	2483.617
4	that	4138	5257	1591.806
5	So	882	184	1560.436
6	this	1821	1356	1479.073
7	I	3384	4325	1283.408
8	it	2690	3365	1060.851
9	is	2902	4179	851.030
10	people	925	610	843.575

The "Likelihood" column represents the strength of the word as a keyword, with values above 15.13 typically considered statistically significant at $p < 0.05$. In the case of "we," the Likelihood value is 3529.221, indicating that it is a particularly strong keyword. The second-ranked keyword is "and."

From these results, it is evident that personal pronouns such as "we," "you," and "I" dominate the top 10 keywords in the BLSC. This suggests that business leaders effectively use personal pronouns to engage, persuade, and motivate their audiences.

Cluster analysis

Due to space limitations, this paper focuses on the cluster analysis of the most statistically significant keyword, "we." [Table 2](#) presents the top 10 results from the cluster analysis of "we" in the BLC. The most frequent cluster, ranked first, is "we don't do," which appears 84 times. The second most frequent cluster is "we need to," with 71 occurrences, followed by "we have to," ranked third.

Table 2.

We cluster

Rank	Cluster	Freq
1	we don' t	84
2	we need to	71
3	we have to	61
4	we' ve got	47
5	we' re going	44
6	we have a	32
7	we want to	31
8	and so we	29
8	I think we	29
10	we can do	28

Discussion

The importance of effective presentation skills for business leaders has been emphasized in numerous studies ([Mayfield et al., 2014](#)). However, to date, few have reliably identified the specific vocabulary used by these leaders, and even fewer have conducted a detailed examination of their communication strategies (CS) through large-scale corpus analysis. This study aimed to address this gap by analyzing the vocabulary and CS employed by global business leaders who spoke at the Oxford Union (OU) and TED Talks. By comparing the business leader corpus with the large-scale F-BROWN as the reference corpus, the study revealed that these leaders frequently used a significant number of keywords, supporting Hypothesis 1.

The most prominent keyword identified was "we," which was further analyzed through cluster analysis to uncover how "we" was used in various CS clusters. The findings showed that business leaders effectively used "we" in cluster expressions such as "we don't," "we need

to," and "we have to." As seen in the following examples, these expressions were employed strategically to guide, persuade, and motivate the audience to take action.

Example 1: "Indeed, if we don't do this, we're..."

This CS is used to urge the audience to take action.

Example 2: "What we need to start thinking about..."

Here, the speaker uses this expression to guide the audience towards their argument.

Example 3: "What we have to remember is this."

This cluster expression is employed to persuade the audience.

The results indicate that business leaders effectively utilize "we" cluster expressions as part of their CS to influence the audience. This finding also supports Hypothesis 2.

Through the large-scale corpus analysis of business leaders' speeches, this study successfully identified the specific expressions and CS they employed. However, due to space limitations, only the "we" cluster analysis was presented. Business leaders use many other keywords, which warrant further exploration. Future research could improve the reliability of findings by incorporating more speech data. Additionally, while this study focused on monologues, such as speeches, analyzing interactive data, such as debates, could help uncover other CS used by leaders in more dynamic settings.

References

- Almaghlouth, S. (2022). Deconstructing agency in the G20 leaders' declarations in the last decade: A corpus-assisted discourse study. *Helyon*, 8(12), 1-12. <https://doi.org/10.1016/j.heliyon.2022.e12439>
- Biber, D., Conrad, S., & Leech, G. (2002). *Student Grammar of Spoken and Written English*. Harlow: Pearson Educated Limited.
- Conger, J. A. (1991). Inspiring others: the language of leadership. *Academy of Management Perspectives*, 5(1), 31-45. <https://dx.doi.org/10.2307/4164992>
- Gerald, C. M. (1970). Old mortality at Oxford: Monsman. *Studies in philology*, 67(3), 359-389. <https://www.jstor.org/stable/4173687>
- Graham, F. (2005). *Playing at Politics*. Edinburg: Dunedin Academic Press.
- Hughes, D., & Phillips, B. (2000). *The Oxford Union Guide to Successful Public Speaking*. London: Virgin publishing.
- Johnson, C. E., & Hackman, M. Z. (2018). *Leadership: A Communication Perspective* (7th Ed.). IL: Waveland Press.

- Kotter, J. P. (2001). What leaders really do. *Harvard Business Review*, 68(3), 103-111.
- Locker, K. O., & Kaczmarek, S. (2013). *Business Communication: Building Critical Skills* (6th Ed.). NY: McGraw-Hill Education.
- Mayfield, J., Mayfield, M., & Sharbrough, W. C. (2014). Strategic vision and values in top leaders 'communications: Motivating language at a higher level. *International Journal of Business Communication*, 52(1), 97-121. <https://doi.org/10.1177/2329488414560282>
- Nakatani, Y. (2010). *Improving Oral Proficiency through Strategy Training – Focus on Language Testing, Learners' Corpus and Cognition*. Saarland, Germany: Lambert Academic Publishing.
- Nakatani, Y. (2016). Exploring business communication strategies based on CEFR. *International journal of languages, literature and linguistics*, 2-3, 86-89. <http://dx.doi.org/10.18178/IJLL.2016.2.3.72>"<http://dx.doi.org/10.18178/IJLL.2016.2.3.72>
- Nakatani, Y. (2023) Enhancing English debate instruction through scientific methods: utilizing corpus analysis with concordance software. In BaJpai, C.S., Paul, R., Rai, P. and Singh, A (Eds.) *Technology Integration in Higher Education: Opportunities and Challenges* (pp. 132-145). Delhi: NLUD Press
- Nelson, M. (2006). Semantic associations in business English: A corpus-based analysis. *English for specific purposes*, 25(2), 217-234. <https://dx.doi.org/10.1016/j.esp.2005.02.008>

Acknowledgement

I wish to express my gratitude to editors and reviewers for their insight, comments, and suggestions.

Funding Information

This research received no external funding.

Declaration of Conflict

The authors declare no conflict of interest.