Final Report: Coopetition in Israeli Academia

Ram Shiri, Amit Gitin, Lior Skudowitz, Nir Yemini

1. Introduction

The Israeli higher education landscape is shaped by a complex interplay between cooperation and competition. Universities in Israel often join forces in research initiatives, joint academic programs, and national infrastructure projects. At the same time, they compete fiercely over student enrollment, government funding and prestige in international rankings. This duality, is a strategic characteristic of the academic field in Israel.

The central question guiding this study is - How does Israeli universities communicate competition and cooperation, and how does that play out on their websites?

This project builds on Nir Yemini's foundational work analyzing Israeli academic institutions' strategic behavior. While his research focused on institutional policies and positioning in global academia, our work extends the analysis to the public discourse dimension — examining how universities communicate cooperation and competition through their own media output.

The research investigates how cooperation and competition shapes the strategic positioning of academic institutions. We focus solely on universities, as universities operate within a competitive framework emphasizing prestige, research output, and global visibility — unlike colleges, which often serve regional and vocational roles.

Our aim is to identify whether the language used in official university publications reflects cooperative or competitive strategies and nature.

Code & Data available at: **GitHub**

2. Data

We scraped over 3,000 news articles from the official websites of major Israeli universities, including: Ben-Gurion University, Tel Aviv University, the Technion, the Hebrew University, and Bar-Ilan University. Each article was annotated by the research team based on its tone (cooperative, competitive, or neutral), keywords, and structural metadata.

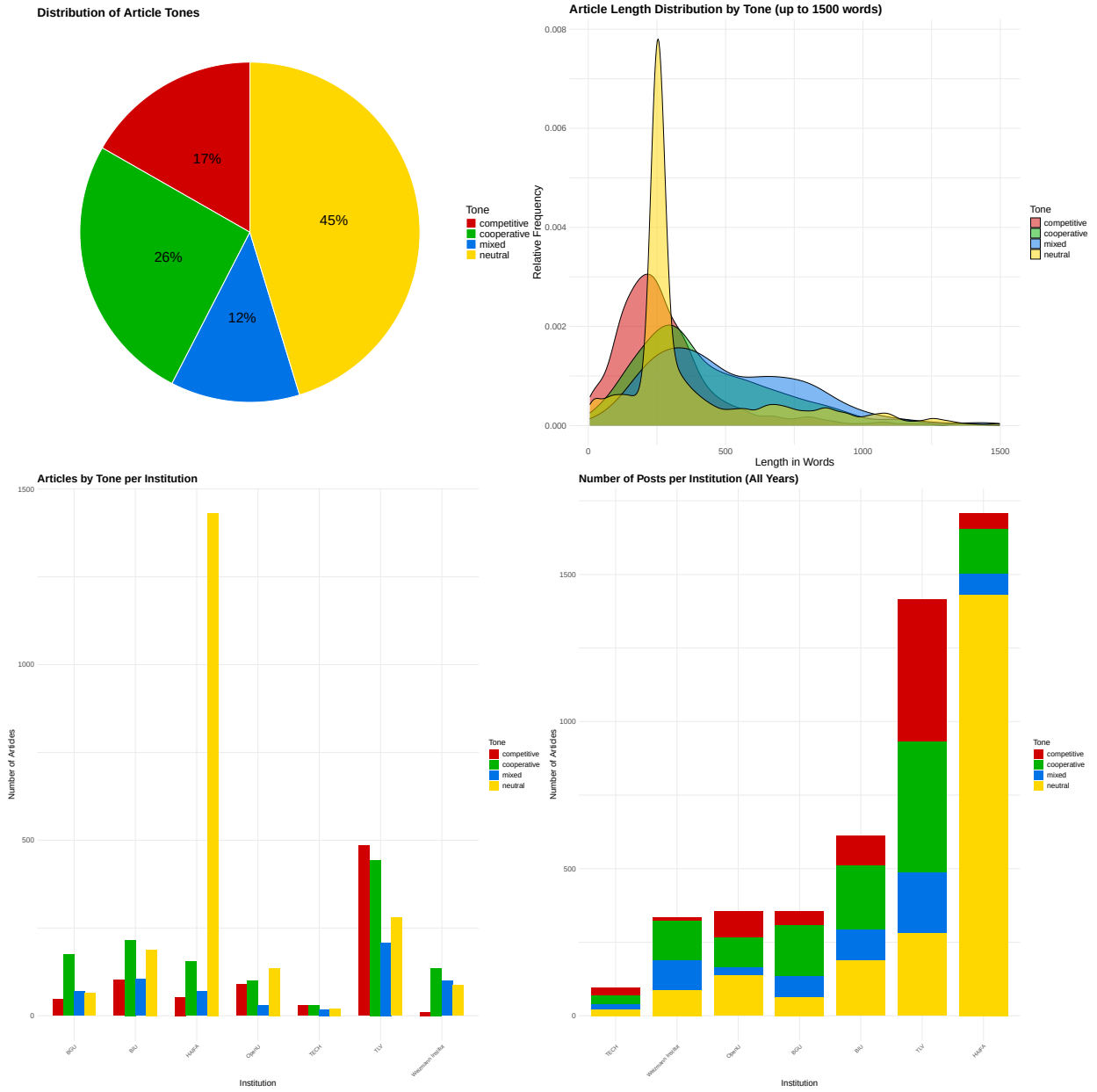
Dataset summary: Articles: 4,000+

```
Universities: 7 (Not including Ariel Raichman and Huji)

Features: tone, cooperation type, thematic tags, Shanghai ranking, article length, etc.

Storage: CSV file with 11 structured columns; full description in /data/README.md
```

All data has been saved as a CSV file in the /data folder, accompanied by a README.md file that describes each field. Text content has been cleaned, and categorical fields have been normalized. This structured dataset enables us to run visualizations and exploratory analyses.

3. Methods and Results

We analyzed how Israeli universities frame their messaging. Given the subjective, categorical nature of tone and lack of labeled data, we used exploratory visualizations rather than predictive modeling. All visualizations in this report were generated using R and ggplot2.



3.1 Distribution of Article Tones

Most articles use cooperative tone, followed by neutral, competitive, and a smaller mixed category.

3.2 Article Length by Tone

A plot that shows that neutral articles are typically shorter, while cooperative and mixed articles tend to be longer.

3.3 Articles by Tone per Institution

A grouped bar chart reveals that universities like BIU and BGU emphasize cooperative tone, while TLV uses more competitive and mixed framing.

3.4 Total Posts by Institution

A stacked bar chart shows total publication volume and tone makeup per institution. TLV and TECH publish the most, but tone proportions vary significantly.

4. Limitations and Future Work

Limitations

While this study provides meaningful insights into the strategic use of language in Israeli academic institutions, several limitations should be noted. First, the manual annotation of article tone and thematic tags introduces a level of subjectivity, which may affect consistency and accuracy. Second, there is a clear institutional imbalance: some universities publish significantly more articles than others, which may skew comparisons and overrepresent certain voices. Third, the Shanghai ranking was treated as a static binary indicator, although rankings may vary over time and do not necessarily reflect communication strategies. Lastly, the tagging process relied on keyword heuristics rather than advanced NLP methods, which limits the granularity and nuance of thematic analysis.

These limitations suggest caution in generalizing findings beyond the sample and point to opportunities for more automated, large-scale analysis in future research.

Future Work

Future work could improve the analysis in several directions. First, implementing automated tone classification using large language models (e.g., BERT or GPT-based sentiment classifiers) could reduce subjectivity and scale the annotation process. Second, applying time-series analysis would allow exploration of how institutional tone evolves over time in response to external events (e.g., political shifts, global rankings, academic reforms). Additionally, network analysis of cooperation mentions could reveal inter-institutional relationships and clustering patterns. Finally, integrating more detailed institutional metadata — such as faculty size, international partnerships, or funding levels — could strengthen the interpretability of tone-related trends.

Appendix

Variables in `ALL_articles.csv`

Variable	Description
`institution`	Publishing university name
`date`	Date of article publication
`title`	Article headline
`url`	URL to original article
`tone`	One of: cooperative, competitive, neutral, or mixed
`mentions_cooperation`	TRUE/FALSE – whether article mentions cooperation
`mentions_competition`	TRUE/FALSE – whether article mentions competition
`cooperation_type`	Type of cooperation (e.g., Domestic, International)
`thematic_tags`	Content tags (semicolon-separated)
`length_words`	Estimated article length in words

Source code

```
library(knitr)
library(tidyverse)
library(broom)
library(htmltools)
library(ggplot2)
library(patchwork)
library(lubridate)
library(forcats)

opts_chunk$set(echo=FALSE) # hide source code in the document


#
df <- read_csv("C:/Avoda/submission/ALL_articles.csv", show_col_types = FALSE) %>%
  mutate(
    tone = tolower(trimws(tone)),
    tone = case_when(
      tone %in% c("cooperative", "collaborative") ~ "cooperative",
      tone == "competitive" ~ "competitive",
      tone == "neutral"     ~ "neutral",
      tone == "mixed"       ~ "mixed",  #
      TRUE ~ NA_character_
    ),
    length_words = suppressWarnings(as.numeric(length_words)),
    Date = suppressWarnings(dmy(Date))
  ) %>%
  filter(!is.na(tone), !is.na(institution))


#      )     (
```

```r
tone_colors <- c(
  "cooperative" = "#00B200",  #
  "competitive" = "#D00000",  #
  "neutral"     = "#FFD700",  #
  "mixed"       = "#0073E6"   #
)

#
df_counts <- df %>%
  count(institution, tone) %>%
  complete(institution, tone, fill = list(n = 0))

#
institution_totals <- df_counts %>%
  group_by(institution) %>%
  summarise(total = sum(n), .groups = "drop")

df_counts_ordered <- df_counts %>%
  left_join(institution_totals, by = "institution") %>%
  mutate(institution = fct_reorder(institution, total))

#  Pie chart
pie_chart <- df %>%
  count(tone) %>%
  mutate(
    perc = n / sum(n),
    label = paste0(round(perc * 100), "%")
  ) %>%
  ggplot(aes(x = "", y = perc, fill = tone)) +
  geom_col(width = 1, color = "white") +
  coord_polar(theta = "y") +
  labs(title = "Distribution of Article Tones", fill = "Tone") +
  geom_text(aes(label = label), position = position_stack(vjust = 0.5), size = 8) +
  scale_fill_manual(values = tone_colors) +
  theme_void() +
  theme(
    plot.title = element_text(size = 20, face = "bold"),
    legend.title = element_text(size = 18),
    legend.text = element_text(size = 16)
  )

#  Density (with tone = mixed!)
length_density <- df %>%
  filter(length_words < 1500) %>%
  ggplot(aes(x = length_words, fill = tone)) +
  geom_density(alpha = 0.5) +
  labs(title = "Article Length Distribution by Tone (up to 1500 words)",
       x = "Length in Words", y = "Relative Frequency", fill = "Tone") +
  scale_fill_manual(values = tone_colors) +
  theme_minimal() +
```

```r
  theme(
    plot.title = element_text(size = 20, face = "bold"),
    axis.title = element_text(size = 18),
    axis.text = element_text(size = 14),
    legend.title = element_text(size = 16),
    legend.text = element_text(size = 14)
  )

#  Coopetition bar
#
coopetition_absolute <- df %>%
  filter(tone %in% c("cooperative", "competitive", "mixed", "neutral")) %>%
  count(institution, tone) %>%
  complete(institution, tone = c("cooperative", "competitive", "mixed", "neutral"), fill = list

coopetition_bar <- ggplot(coopetition_absolute,
                          aes(x = institution, y = n, fill = tone)) +
  geom_col(position = "dodge", width = 0.7) +
  labs(title = "Articles by Tone per Institution",
       x = "Institution", y = "Number of Articles", fill = "Tone") +
  scale_fill_manual(values = tone_colors) +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 18, face = "bold"),
    axis.text.x = element_text(angle = 45, hjust = 1),
    axis.text.y = element_text(size = 12),
    axis.title = element_text(size = 14),
    legend.text = element_text(size = 12),
    legend.title = element_text(size = 13)
  )

#  Stacked bar (All tones)
recent_posts_bar <- ggplot(df_counts_ordered, aes(x = institution, y = n, fill = tone)) +
  geom_col(position = "stack", width = 0.8) +
  labs(title = "Number of Posts per Institution (All Years)",
       x = "Institution", y = "Number of Articles", fill = "Tone") +
  scale_fill_manual(values = tone_colors) +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 18, face = "bold"),
    axis.text.x = element_text(angle = 45, hjust = 1),
    axis.text.y = element_text(size = 12),
    axis.title = element_text(size = 14),
    legend.text = element_text(size = 12),
    legend.title = element_text(size = 13)
  )

#  Combine
(pie_chart + length_density) /
  (coopetition_bar + recent_posts_bar) +
```

```r
plot_layout(heights = c(1, 1.3))
```