What can Reveal 1,018 Speeches of Fidel Castro?

Sergio Peignier[⋆], Patricia Zapata[◇]

[⋆]CMLA, ENS Paris-Saclay, CNRS UMR 8536, Université Paris-Saclay, France

[◇]Carrera de Lingüística e Idiomas, Universidad Mayor de San Andrés, La Paz, Bolivia

8 June 2018

Fidel Castro Discourse Analysis

↓

Understand his Political Success

Fidel Castro

- Major Cuban politician.
- Half-century government (1959-2008).
- Well known rhetoric skills.

29/09/1960 UN longest speech (4h39min).

Corpus

- Available online[a].
- **1,018** speeches.
- **+7,500,000** words.
- **+4,000,000** non-empty words.

[a] http://www.cuba.cu/gobierno/discursos/

Previous Works
Systematic non-lexicometric approaches

Work	Corpus size	Methodology
[Joyner 1964]	**3** discourses	Aristotelian Rhetoric Principles
[Nieto et al. 2002]	**1** conversation	Conversational Analysis
[Belisario 2016]	**3** discourses	Critical Discourse Analysis
[Reyes 2011]	6 months speeches ($\sim$ **dozen**)	Sociolinguistics / Critical Discourse Analysis

Complex and **time consuming task**

$\downarrow$

Use small (and possibly **non-representative**) **samples**

$\downarrow$

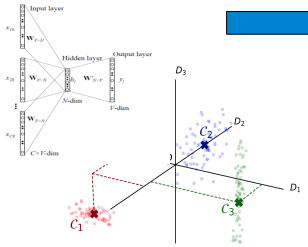
Possibly biased results

Key Points

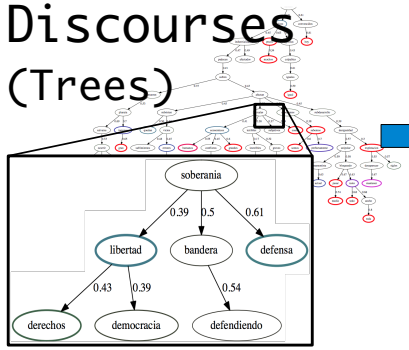
Corpus

1,018 speeches

Machine Learning Framework



Prototypic Discourses (Trees)



Charaudeau Linguistic Discourse Analysis

Machine Learning framework

↓

More representative systematic study.

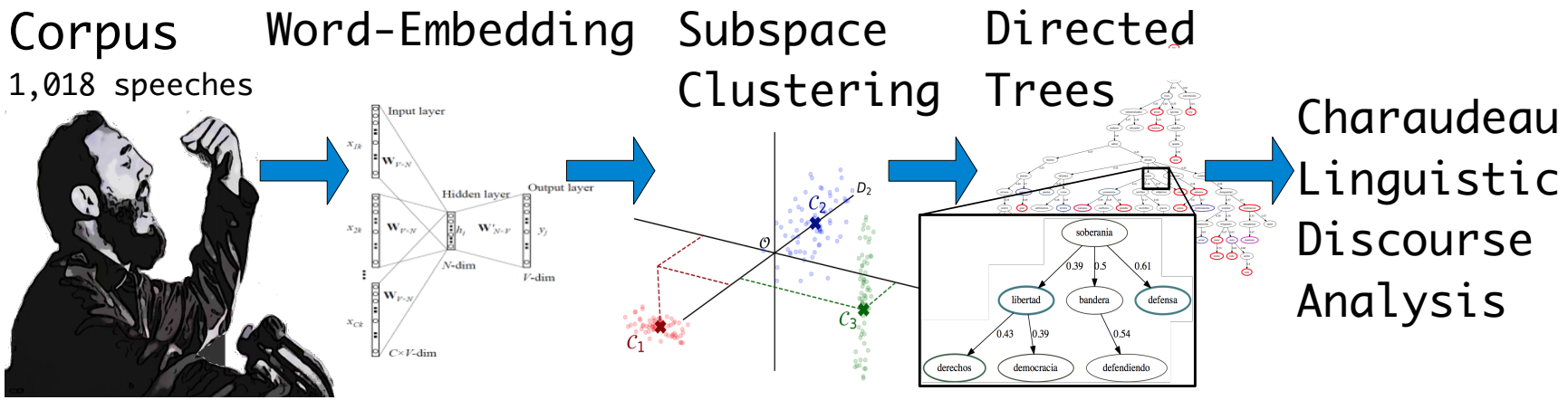
Outline

1. Machine Learning Framework
2. Discourse Analysis Methodology
3. Results
4. Conclusion and Perspectives

Outline

1. Machine Learning Framework
2. Discourse Analysis Methodology
3. Results
4. Conclusion and Perspectives

Overall Framework



Framework steps

- **Pre-processing**: data cleaning and empty words filtering.
- **Word-Embedding** $\rightarrow$ Word Vector Representations.
- **Subspace Clustering** $\rightarrow$ Contextually Linked Word Clusters. (Discursive Contexts)
- **Directed Trees** $\rightarrow$ Intelligible Structure (Prototypic Discourses).
- **Python 2.7** (re, nltk, gensim Word2Vec, SubCMedians)

Distributional Hypothesis: "Words that occur in similar contexts tend to have similar meanings" [Harris 1954]

Example (Fidel Castro)

$$\text{"} \ldots \overbrace{\textit{Dentro } \cancel{de} \textit{ } \cancel{la} \textit{ revolucion } \boxed{\textit{todo}} \textit{ contra } \cancel{la} \textit{ revolucion nada}}^{\textit{Context (width=2)}} \ldots \text{"}$$

Words $\rightarrow$ D-dimensional Vector Space

Word : $w = \boxed{\textit{todo}} \rightarrow x = (0.3, 0.4, -0.9, \ldots)$

Context :

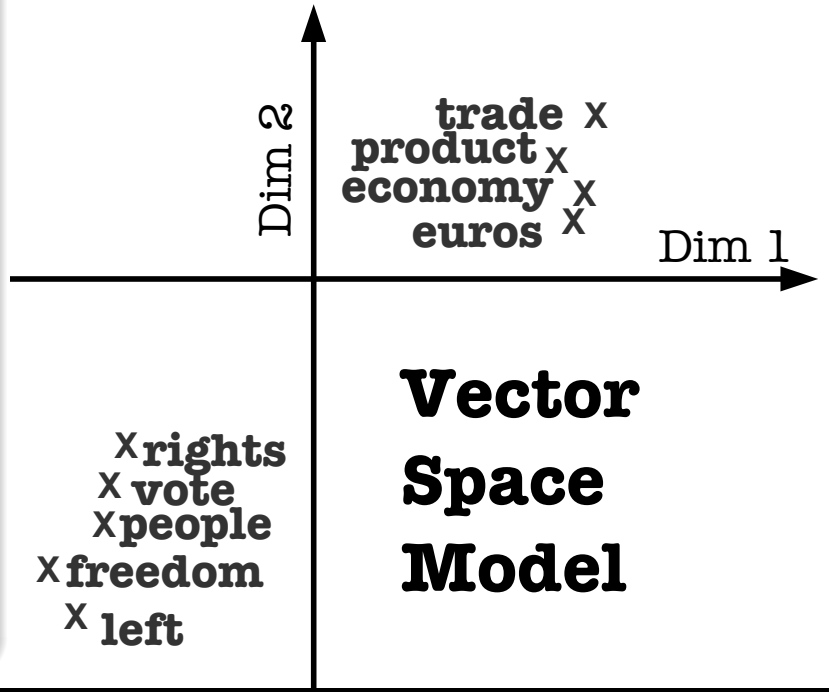
$c = \{\textit{dentro}, \textit{revolucion}, \ldots\} \rightarrow z = (0.1, -0.3, 0.2, \ldots)$

Probability that $\langle w, c \rangle$ **exists**

$$P(\textit{exists}|w, c) \approx p(x, z) = 1/(1 + e^{-x^\top z})$$

Objective function:

$$\overbrace{\log(P(\textit{exists}|w, c))}^{\textit{Maximize for existing } \langle w,c \rangle} + \overbrace{k \times \mathbb{E}(\log(P(\overline{\textit{exists}}|\tilde{w}, c)))}^{\textit{Maximize for } k \textit{ random words } \tilde{w}}$$

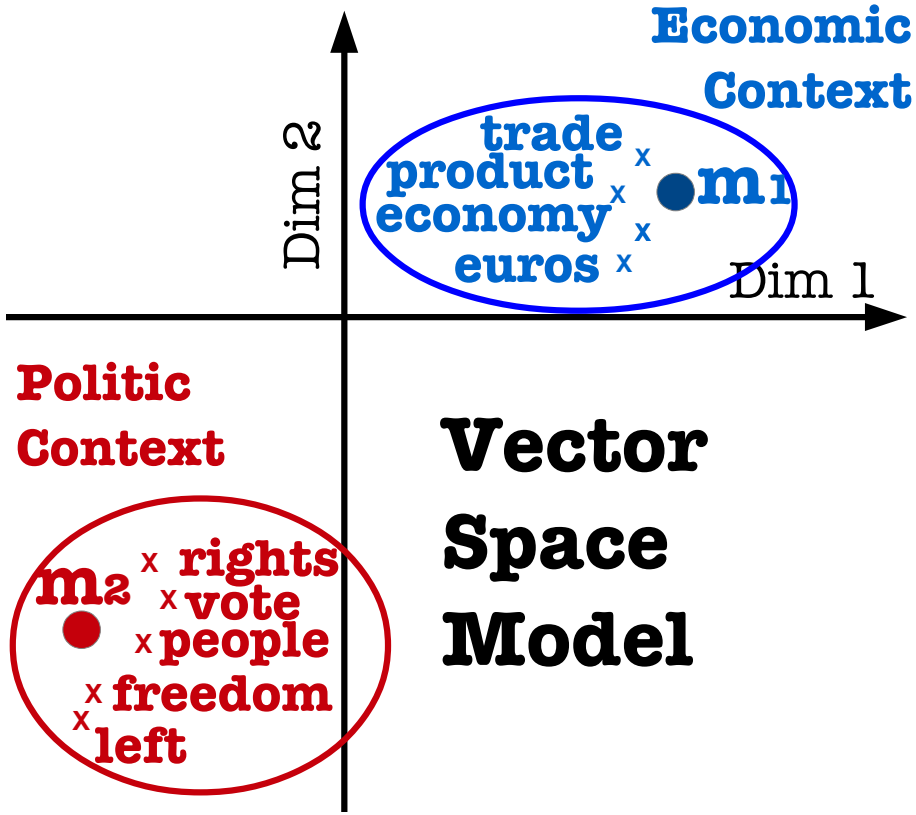


Subspace Clustering: Discursive Context Clusters

SubCMedians [Peignier et al. 2018]

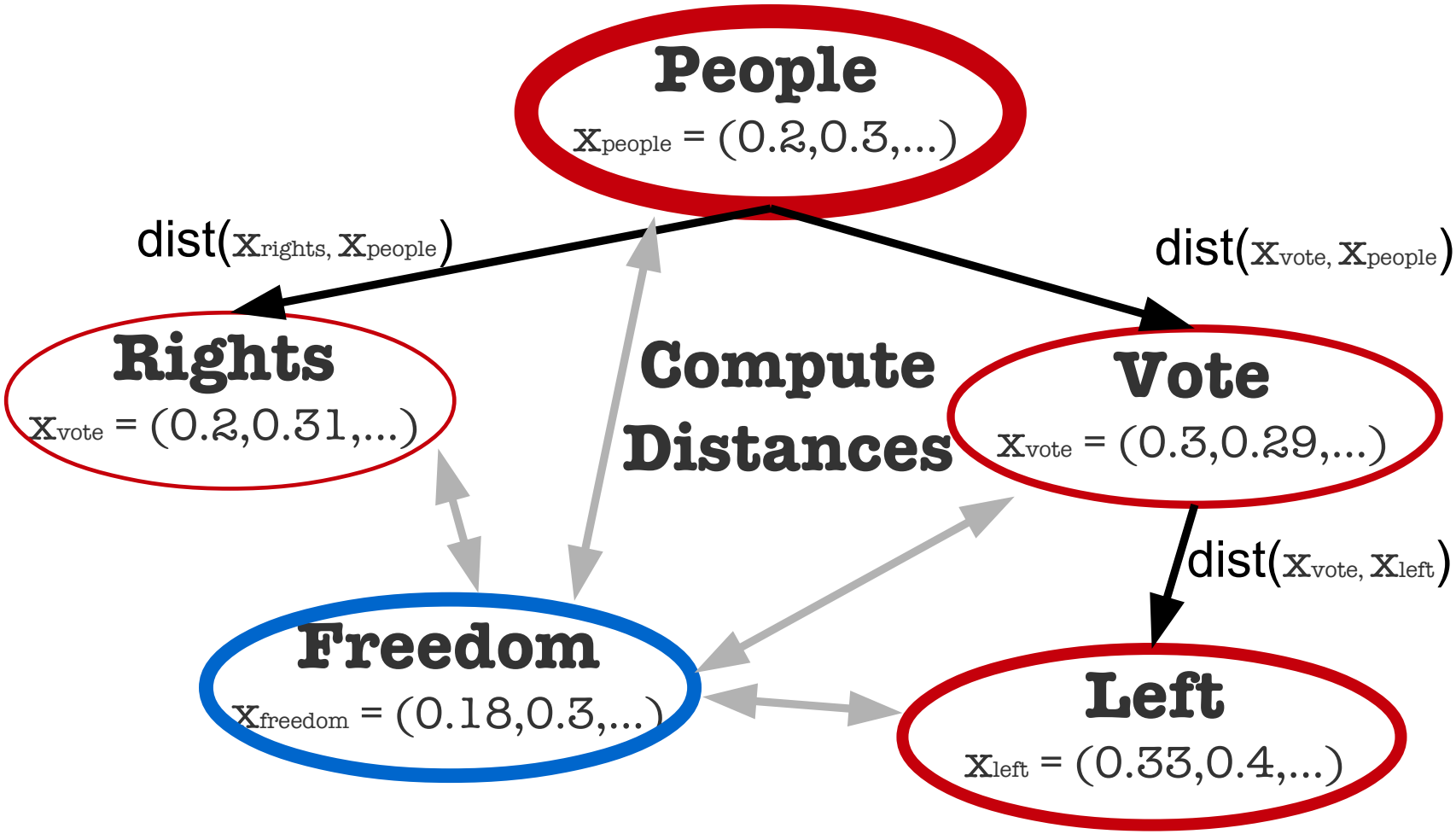
Find **clusters of words** in the **Vector Space Model**

Subspace Clustering is more adapted to **high-dimensional spaces** than clustering.



- **Word vectors** $\mathcal{X} = \{x_1, x_2, \dots\}$.
- **Subspace Centers** $\mathcal{H} = \{m_1, m_2, \dots\}$.
- **Minimise**: $SAE(\mathcal{S}, \mathcal{H}) = \sum_{s_i \in \mathcal{S}} min_{m_j \in \mathcal{H}}(distance(x_i, m_j))$.

Build Directed-Trees



- Most frequent word from cluster $\rightarrow$ root.
- Compute **distances** between **cluster words** and those in the **tree**.
- Include the closest word to the tree.

Outline

1 Machine Learning Framework

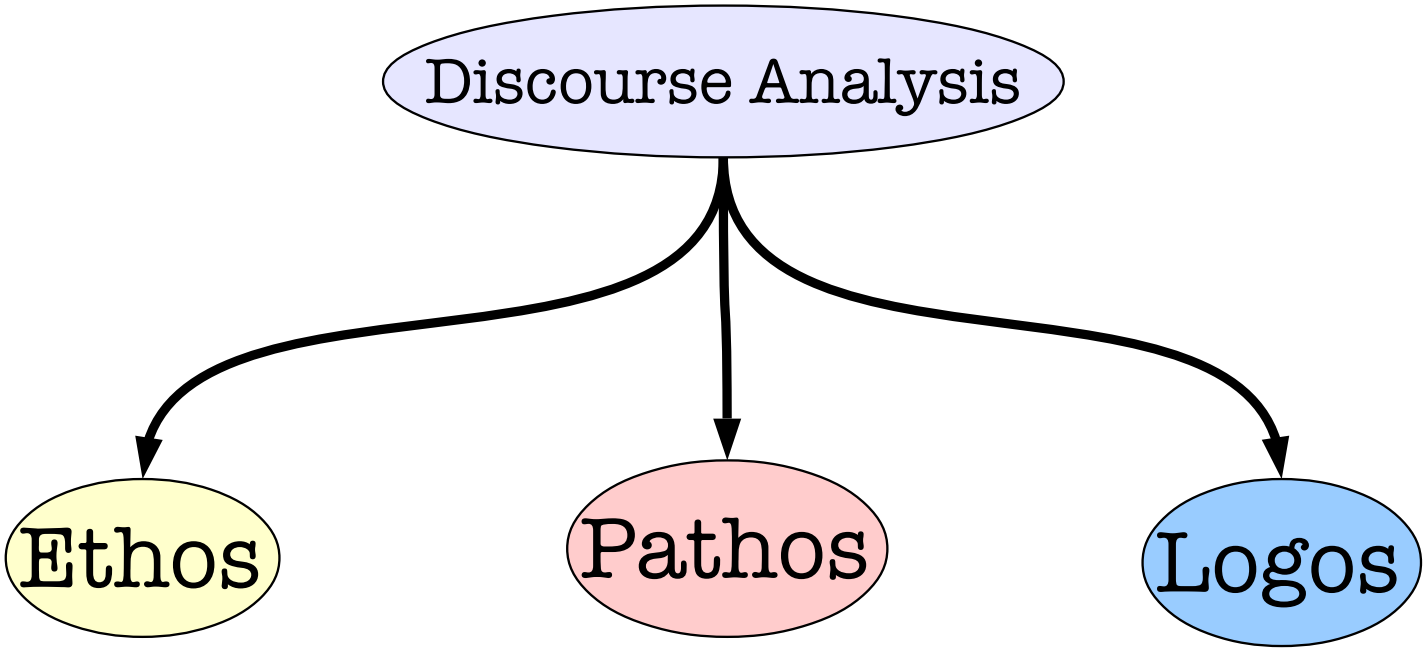
2 Discourse Analysis Methodology

3 Results

4 Conclusion and Perspectives

Charaudeau Discourse Analysis Methodology

- **Semio-pragmatic discourse analysis [Charaudeau 1995].**
- Based on the **Aristotelian classification of rhetoric art.**



Ethos: How the politician depicts himself

Discursive Identity

- **"Me"**
  Personal commitment.
- **"Us"**
  Identification with audience.
- **"Me-Spokesman"**
  Incarnates an ideology.

Speech Situation

- **Candidate.**
  e.g., Political party leader.
- **Ruler.**
  e.g., President / Minister

Embody Truthful Characters

- **Leader** (power).
- **Expert** (knowledge).
- **Witness** (veracity).
- **Committed** person.
- **Charismatic** person.

Pathos (1/2)

Persuade through **feelings** and **passions**.

Recruitment process

Lead the **audience** to **accept** the **speaker's project willingly**.

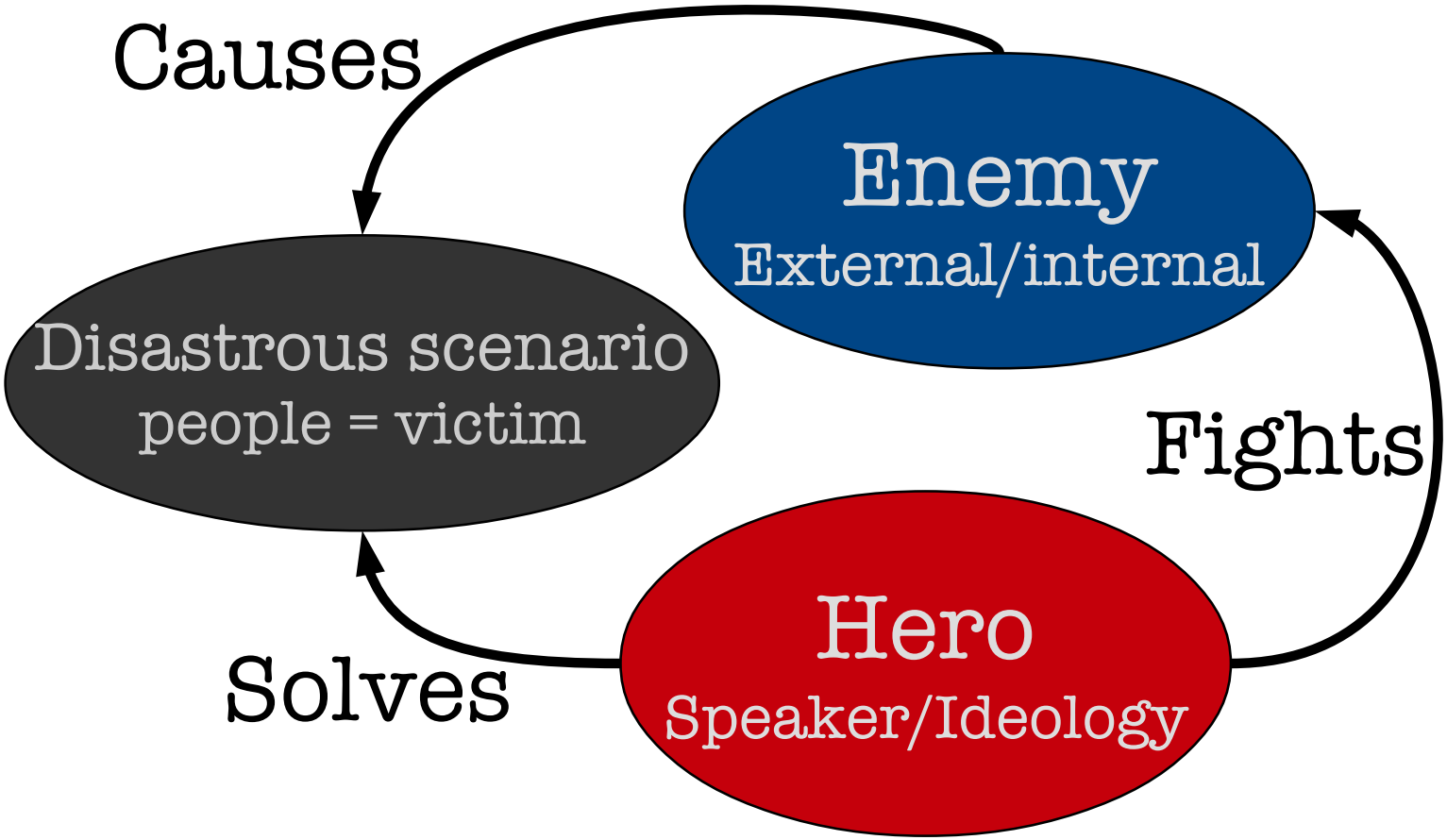
- **Moral Values**.
- **Social Welfare**.
- **Progress** (economy, technology).
- **Religious** belonging.
- **Ethnic** belonging.
- **Nationalism**.

Rhetoric of effects

Create **feelings** and **passions** in the **audience**.

- **Hope**
- **Pride**
- **Courage**
- **Fear**
- **Hatred**
- **Outrage**

Triadic Scenario



Logos

Convince through **logical reasoning** and **argumentation**.

Simplifying strategies

- **Singularization**
  Reduce the number of ideas described.
- **Essentialization**
  Express a complex idea with few words.
- **Formula = Singularization + Essentialization.**
- **Slogans.**

Argumentative strategies

- **Simplified causal arguments.**
- **Analogy with past events/characters.**
- **Detailed descriptions $\rightarrow$ Veracity.**

Outline

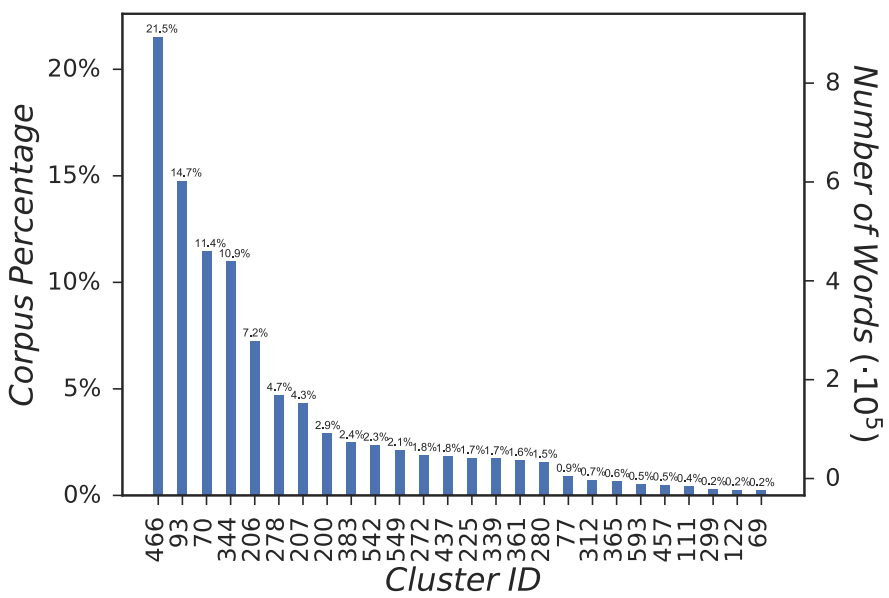
1 Machine Learning Framework

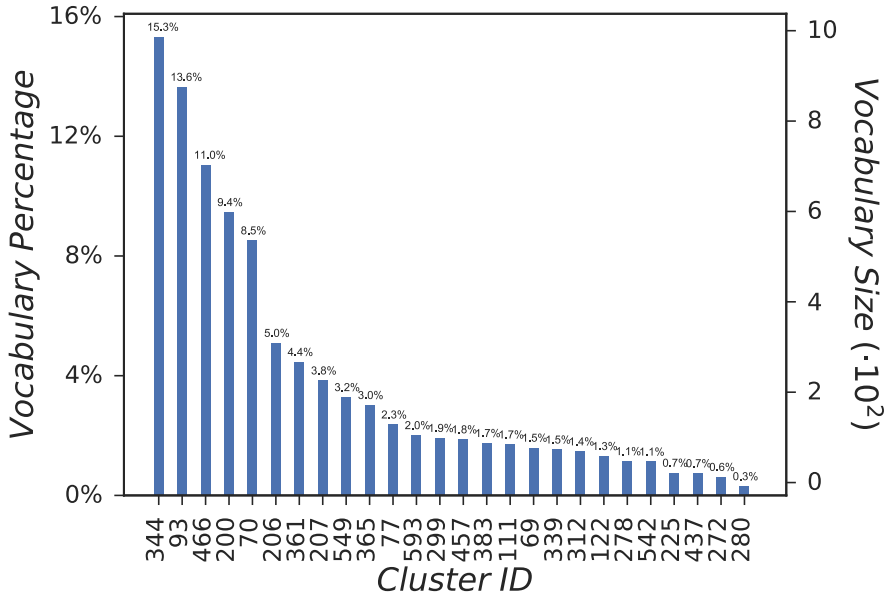
2 Discourse Analysis Methodology

3 Results

4 Conclusion and Perspectives

Cluster Model Structure





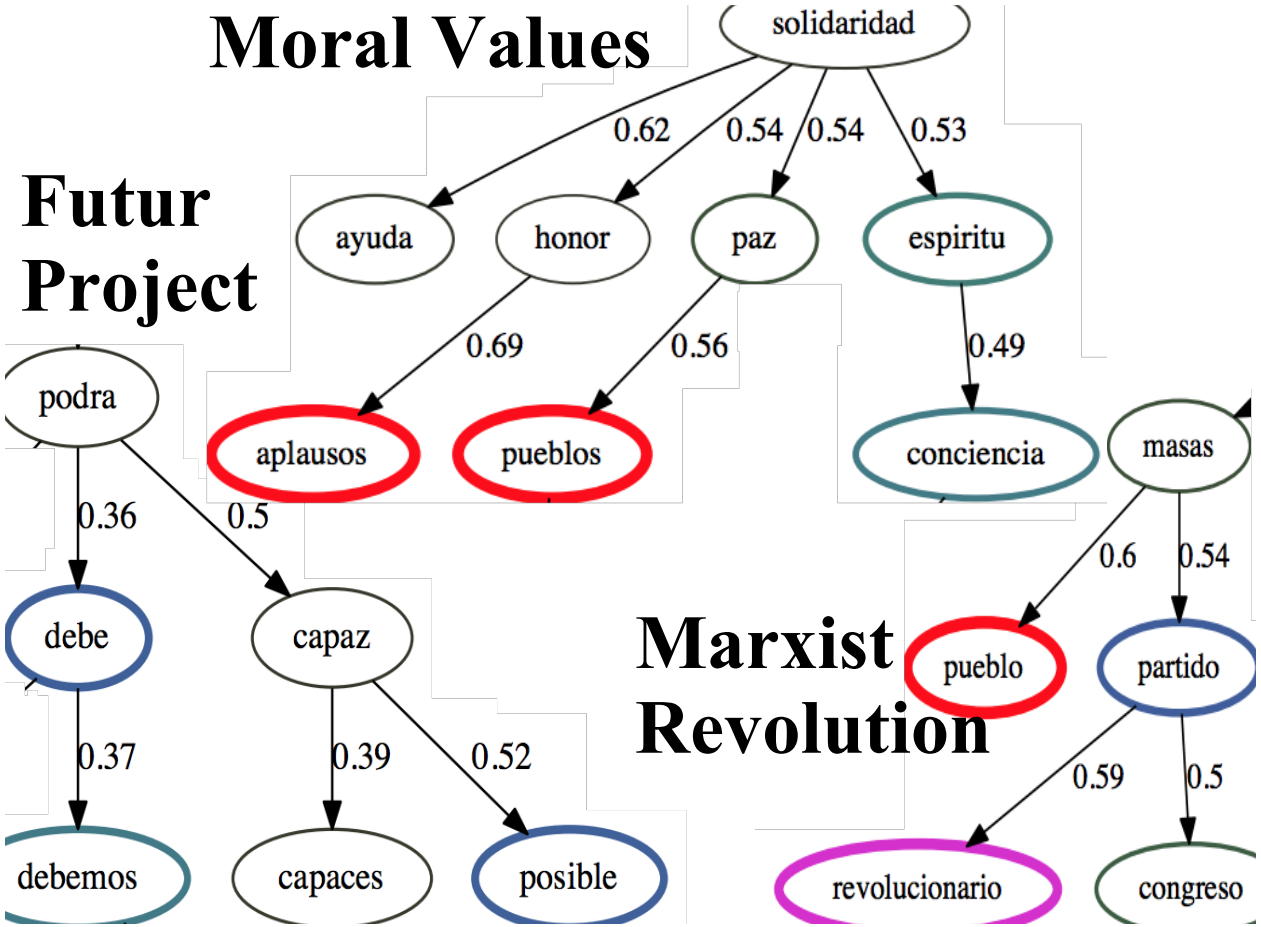
Four largest clusters: 466, 93, 70, 344

- $\sim$ 60% **Number of words.**
- $\sim$ 50% **Vocabulary.**
- $\sim$ Analysis of 595 discourses out of the 1,018.
- **Small clusters** $\rightarrow$ **Specific vocabulary** or **general words.**

Cluster 466
Political Marxist Project

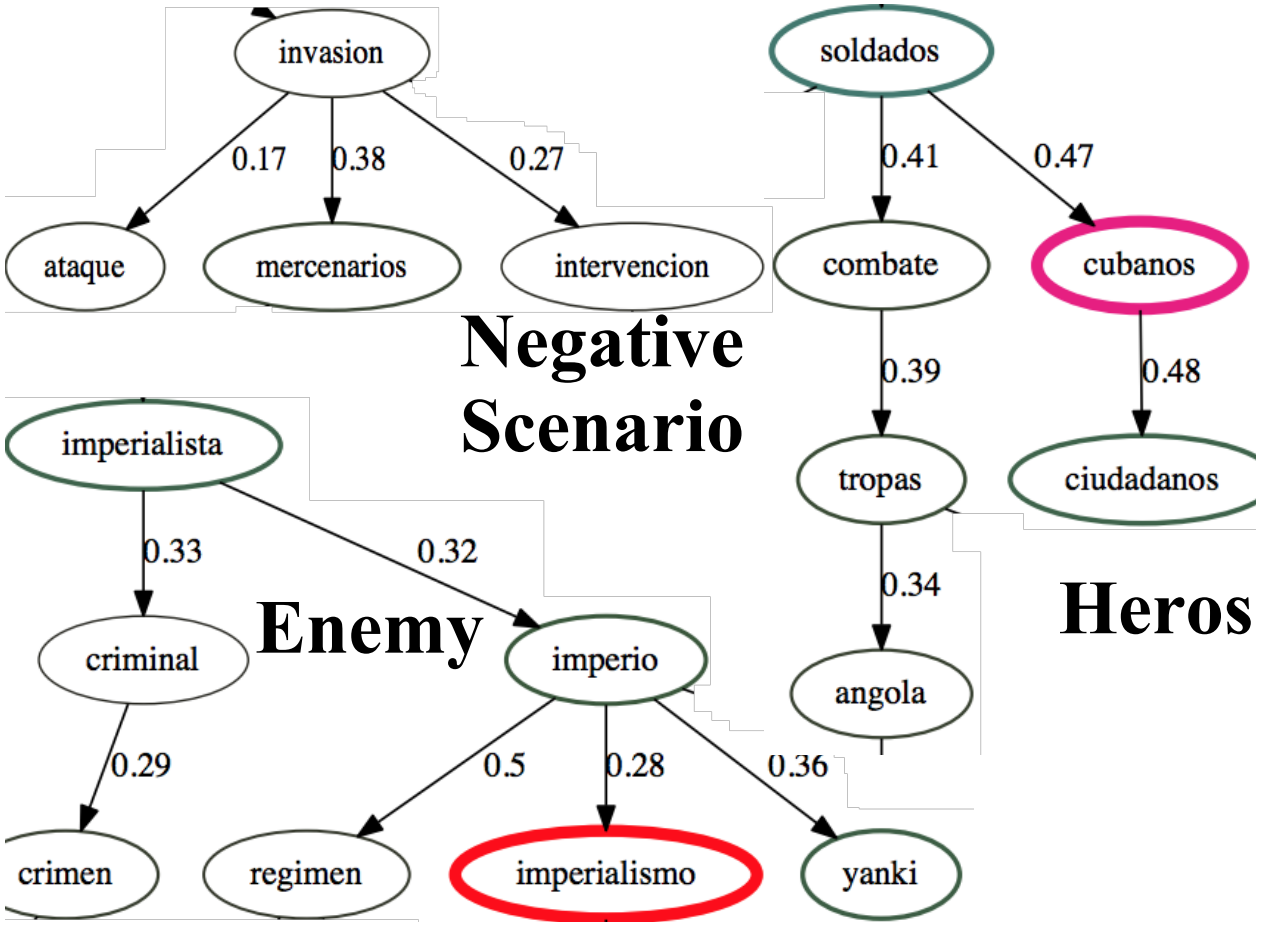
- **Futur Plan**
- **Marxist Revolution**
- **Defence** of the project

- **Identification** with **People**
- **Details** $\rightarrow$ **Veracity**



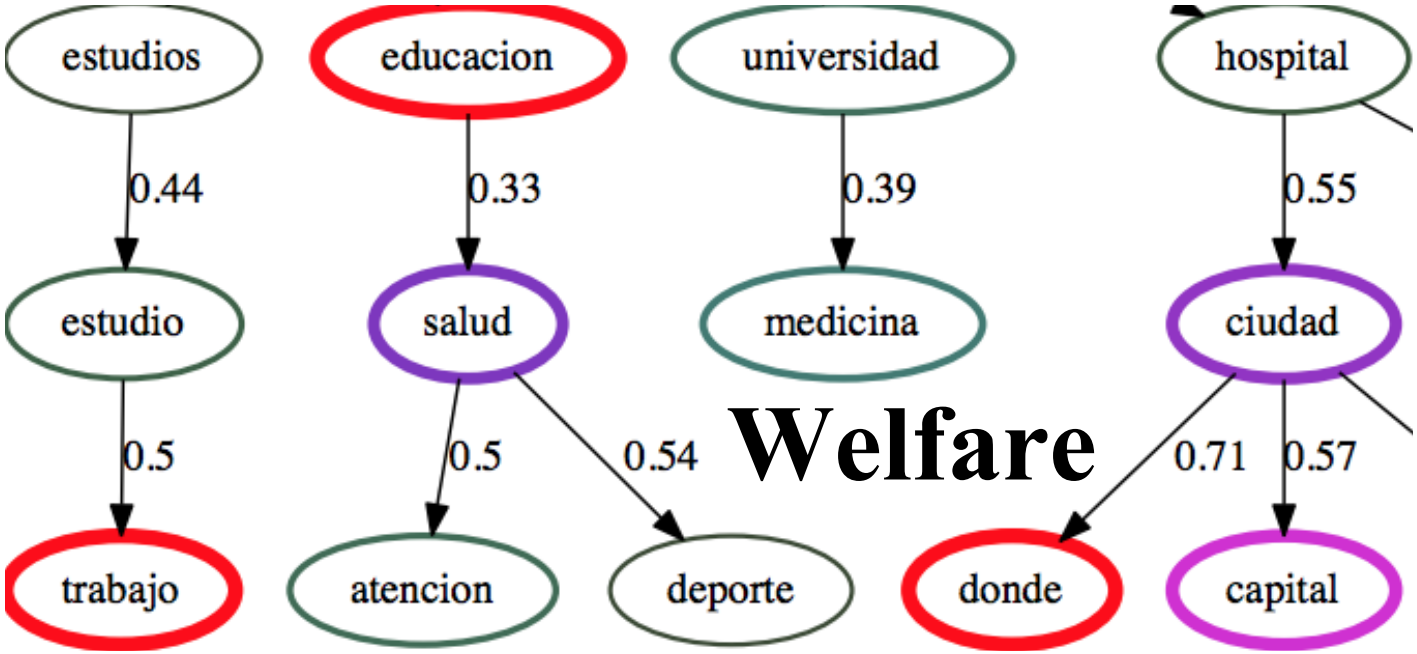
Cluster 344
Cold War Triadic Scenario

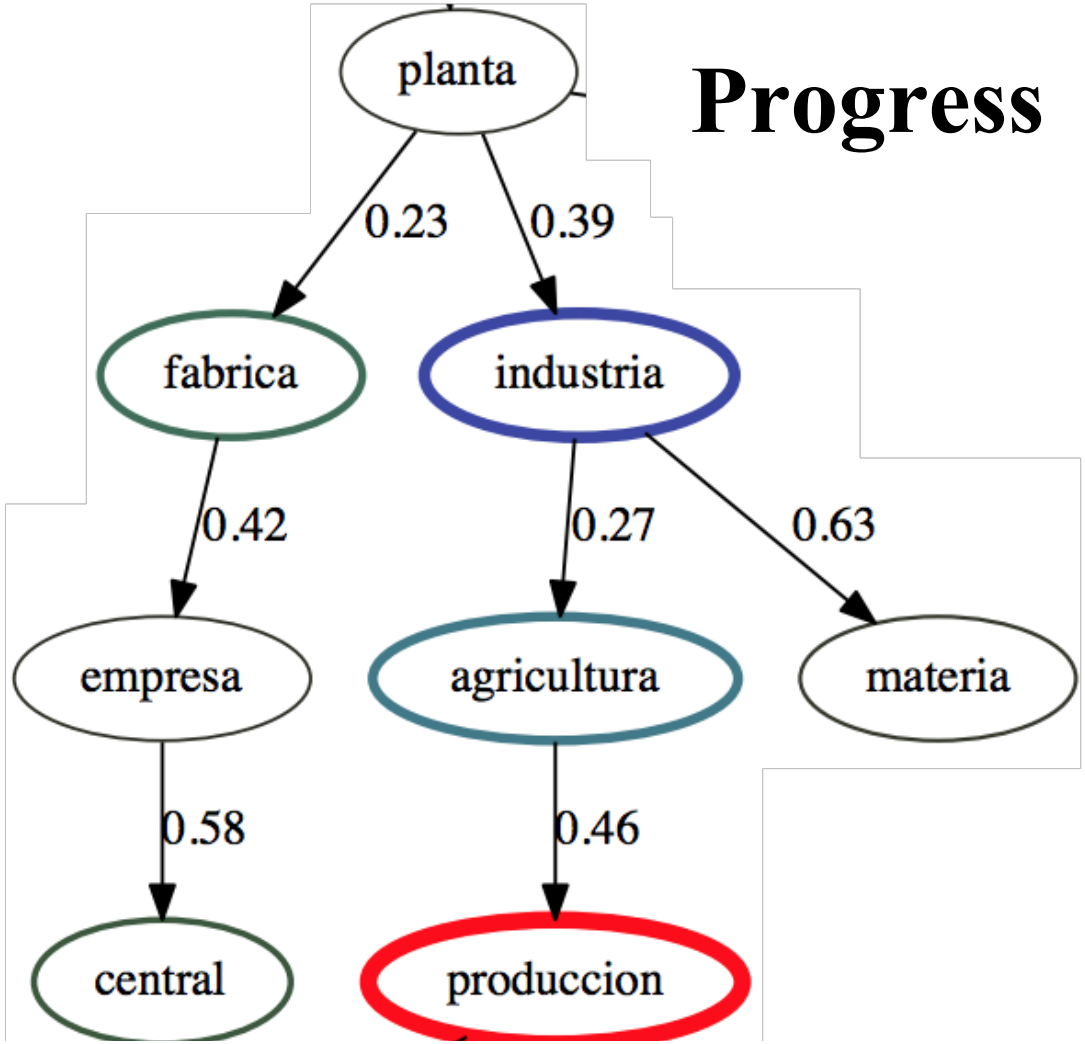
- **Enemies**
  Oligarchy, USA
- **Cold War conflicts**
- **Heroes**
  Communism, Cuban army, people
- **Feelings**
  Fear, Hatred, Outrage



Cluster 70
Social welfare, audience as beneficiary

- **Society as actor and beneficiary**
- **Social welfare**: Education, employment, health
- **Feelings** → Hope, Pride

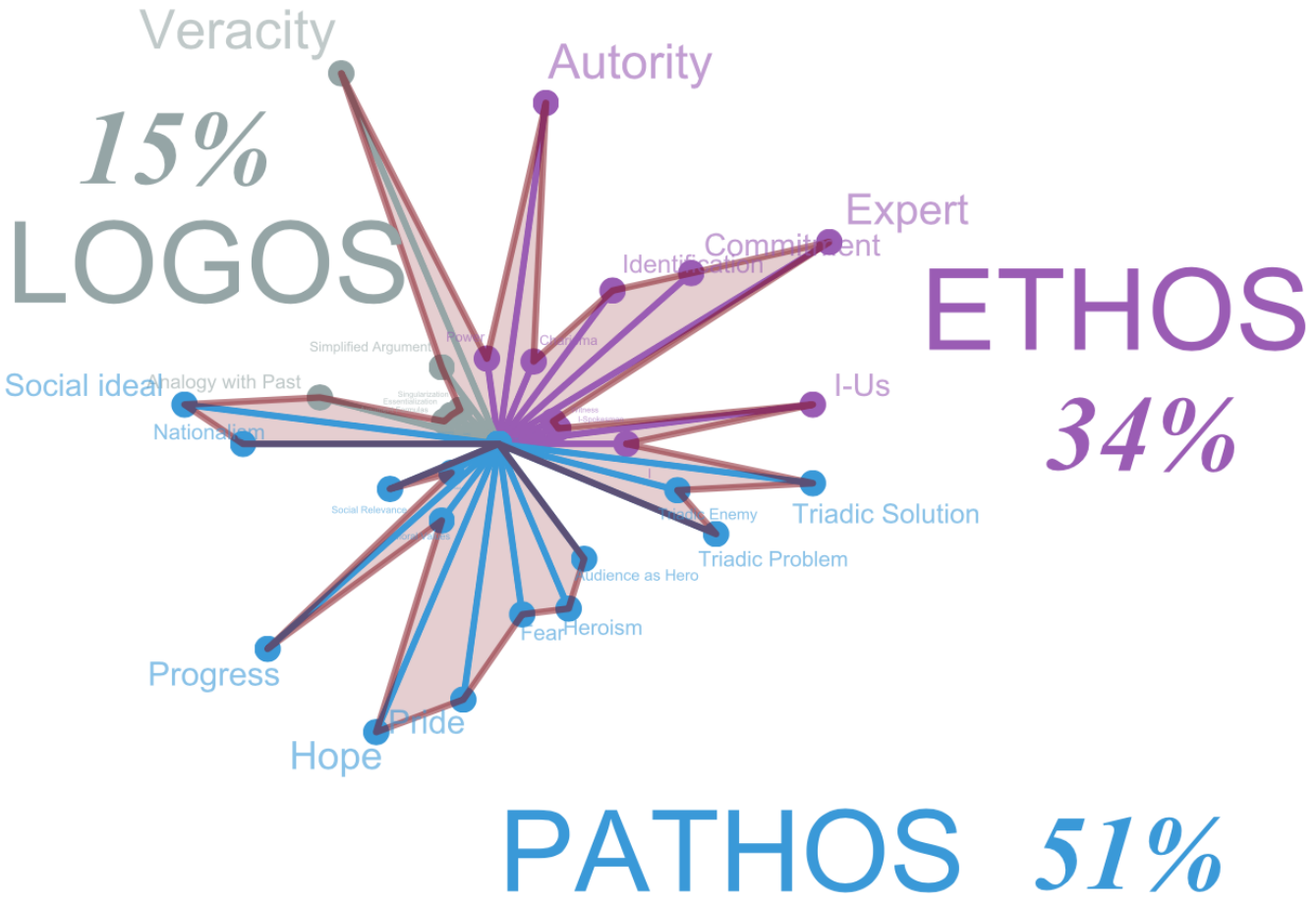




Progress

- **Economic progress**
  Infrastructures
- Needs/Obligations
- **Feelings**
  Hope, Pride
- **Ethos**
  Expert, Authority
- **Details** $\rightarrow$ **Veracity**

Castro Discursive Strategies



Pathos (51%)

Triadic scenario

Progress Welfare Nationalism

Hope Pride Fear Heroism

Ethos (34%)

Authority Expertise

Commitment Identification

Logos (15%)

Detailed descriptions →

Veracity

Outline

1 Machine Learning Framework

2 Discourse Analysis Methodology

3 Results

4 Conclusion and Perspectives

Conclusion and Perspectives

- **Broader** and **more representative** characterisation of Castro's speeches.
- Study relationship with **historical events**.
- Study **influence** on other politicians' discourses.
- Extension to **Automatic Discourse Analysis**.