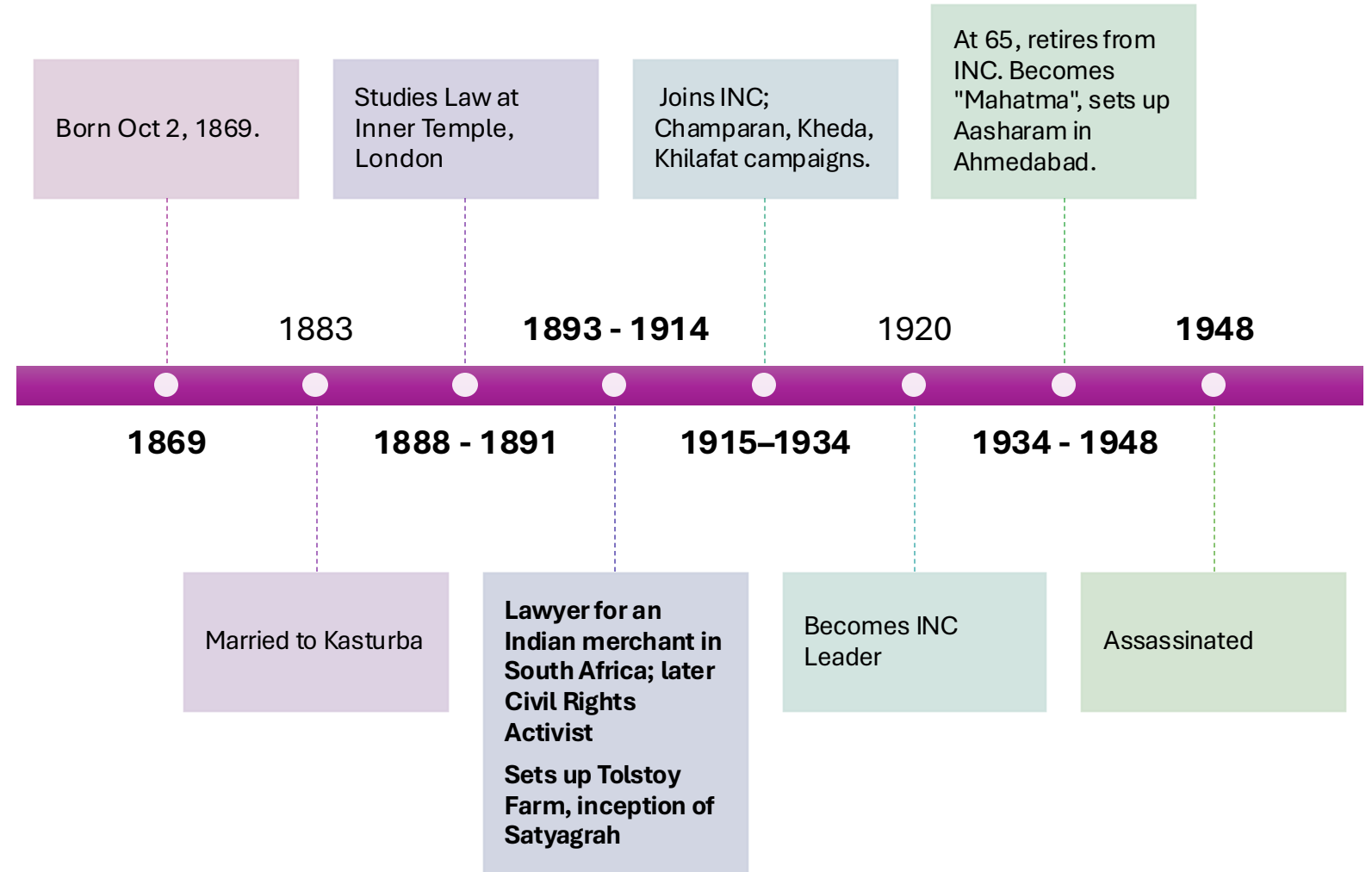# Thus spake Gandhi

Tracing Narrative Drift in Gandhi's Writings (1888-1948)

A word embedding–based exploration of conceptual change over time

Timeline

Born Oct 2, 1869.

Studies Law at Inner Temple, London

Joins INC; Champaran, Kheda, Khilafat campaigns.

At 65, retires from INC. Becomes "Mahatma", sets up Aasharam in Ahmedabad.

1883

1893 - 1914

1920

1948

1869

1888 - 1891

1915–1934

1934 - 1948

Married to Kasturba

Lawyer for an Indian merchant in South Africa; later Civil Rights Activist

Sets up Tolstoy Farm, inception of Satyagrah

Becomes INC Leader

Assassinated

Formative Years (South Africa) 1884-1915

Develops early methods of nonviolent protest and civil disobedience

Founds **Tolstoy Farm** and Phoenix Settlement as experimental communities

Participates as a volunteer in the **Boer War** and **Zulu Rebellion**

Coins and refines the concept of **Satyagraha**

Nationalist Leader and Mass Movements 1915-1934

Returns to India and joins the **Indian National Congress (INC)**

Leads landmark agitations:

Champaran Satyagraha (1917)

Kheda Satyagraha (1918)

Non-Cooperation Movement (1920–22)

Khilafat Movement (1920–24)

Civil Disobedience Movement (1930)

Salt March / Dandi March (1930)

Negotiates at **Round Table Conferences (1931, 1932)**

Signs the **Poona Pact (1932)** on Dalit representation

Gandhi - The Mahatma
1934-1948

Focuses on **social activism** beyond party politics

Establishes **Sevagram Ashram** (near Wardha)

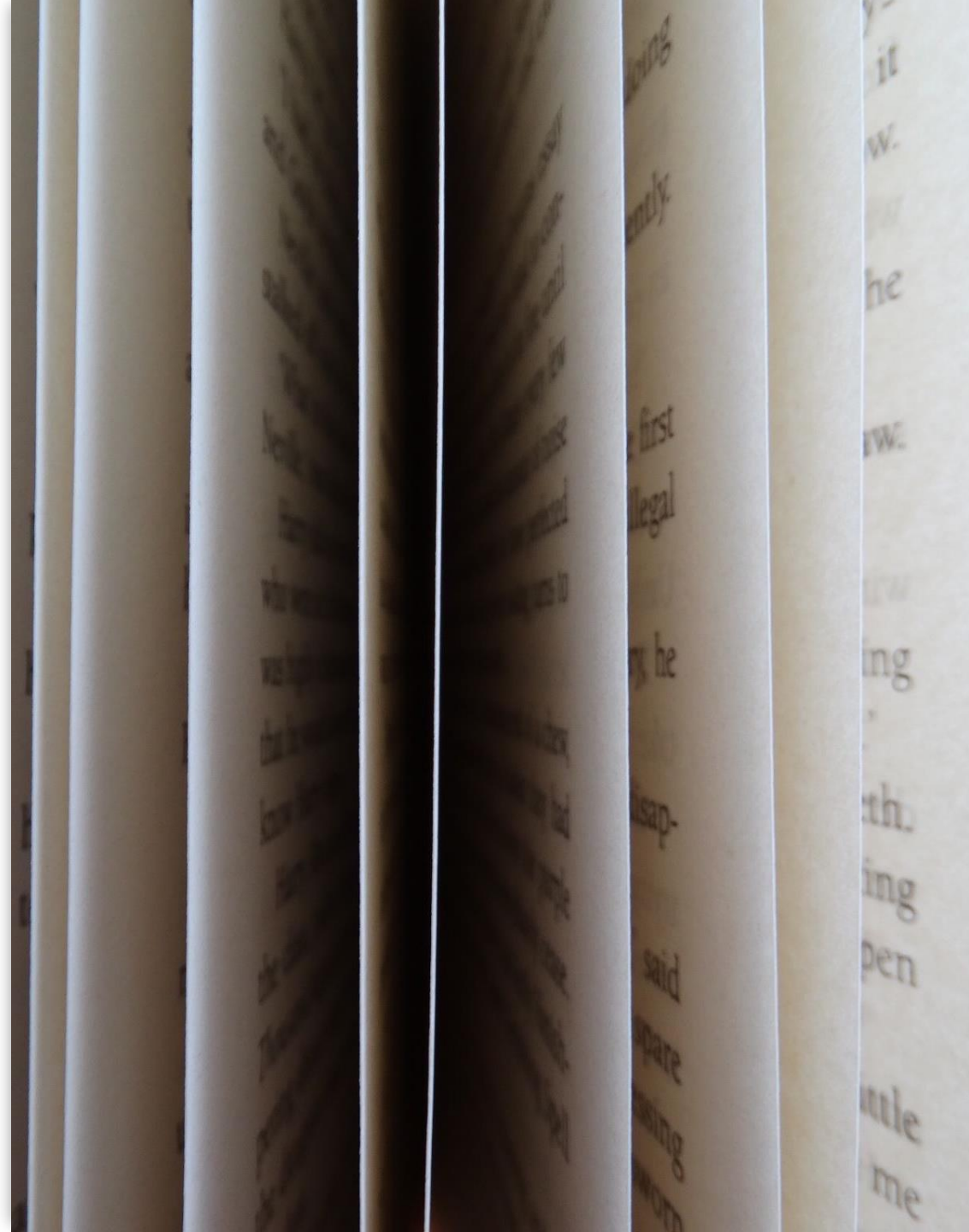Launches **Harijan** magazine to campaign against untouchability

Leads the **Quit India Movement** (1942)

Advocates for Hindu-Muslim unity, communal harmony, and the rights of the oppressed

Assassinated in January 1948.

Data

- Collected works of MK Gandhi publicly available [link]. Handwritten letters converted into text by OCR with minor defects.

- A rich compilation of **letters, speeches, articles, telegrams, and public addresses** authored by Gandhi, offering direct insight into his thought processes, political strategies, and social commentary during India's independence movement.

- 98 volumes, covering his entire adult life – (1884-1948)

Rationale and Method

- **Word2Vec models** provide dense vector representations of words based on their contextual usage. Comparing representation of keyword over time shows how usage evolves.

- Two choices for representing words as vectors: CBOW, Skip-gram

- **Skip-gram performs better on small to medium-sized corpora** (like specialized historical archives) because it learns rare word associations more effectively.

- **Skip-gram preserves local word order within the context window**, which is valuable for analysing rhetorical and ideological patterns where word positioning affects interpretation.

Skip Gram Representation – Input to word2vec

- Simulates a moving window over the text to collect pairs of (center word, surrounding context word) — the training data for a skip-gram model.

- **Example:**
  If tokens = `[At, the, stroke, of, midnight, India, is, a, free, nation]`

- context_window_size=3, target= `India`
  `→[(India, stroke), (India, of), (India, midnight), (India, is), (India, a), (India, free)]` etc.
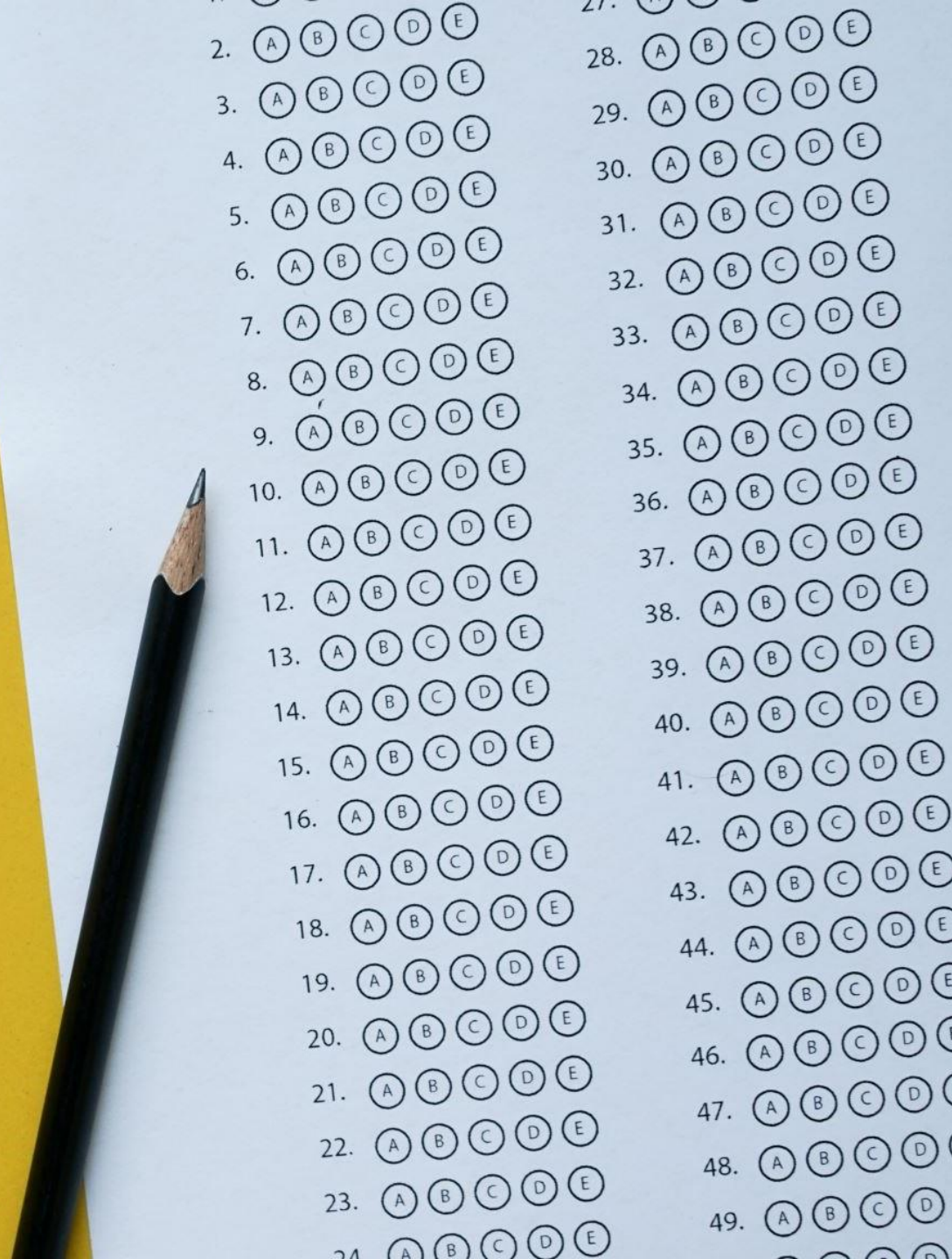
Steps - Preprocessing

- NLTK to remove stop words; lemmatize the tokens in the documents to their roots.

- Converting the documents so obtained into skip-gram representation.

- Training the word2vec network, on each phase.

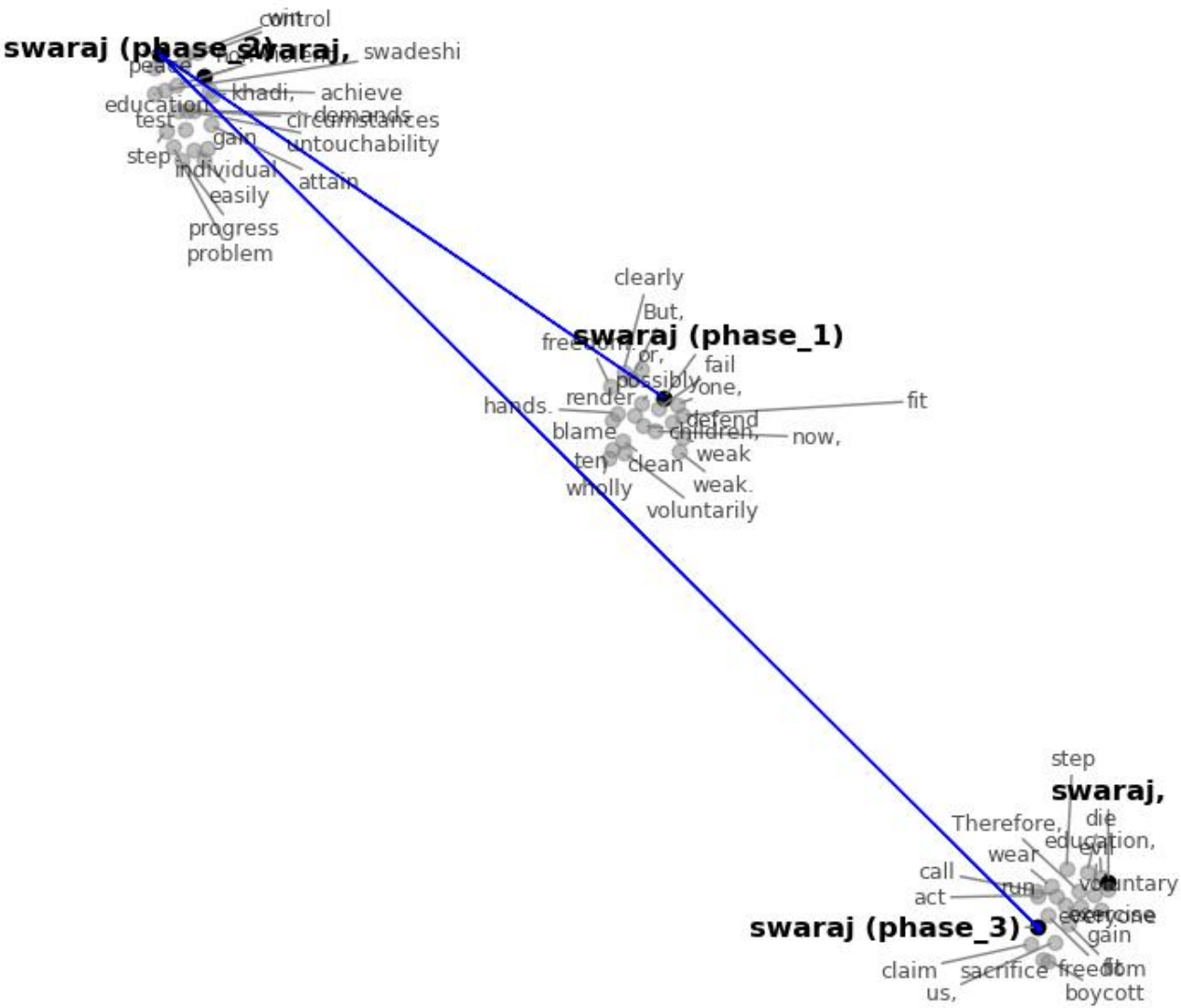- Saved the weights of the trained word2vec nets..

Rationale and Analysis

- After selecting a keyword, how to compare different embeddings?

- ~~Orthogonal Procrustes – aligns vector space, unstable results, poor interpretability.~~

- **Tune hyperparametes** (no. of training epochs, size of context window, shuffle the order of documents) till the embedding is stable.

- **Nearest neighbours** – compare the Jaccard score of sets of nearest neighbours in consecutive phases to get a measure of stability.
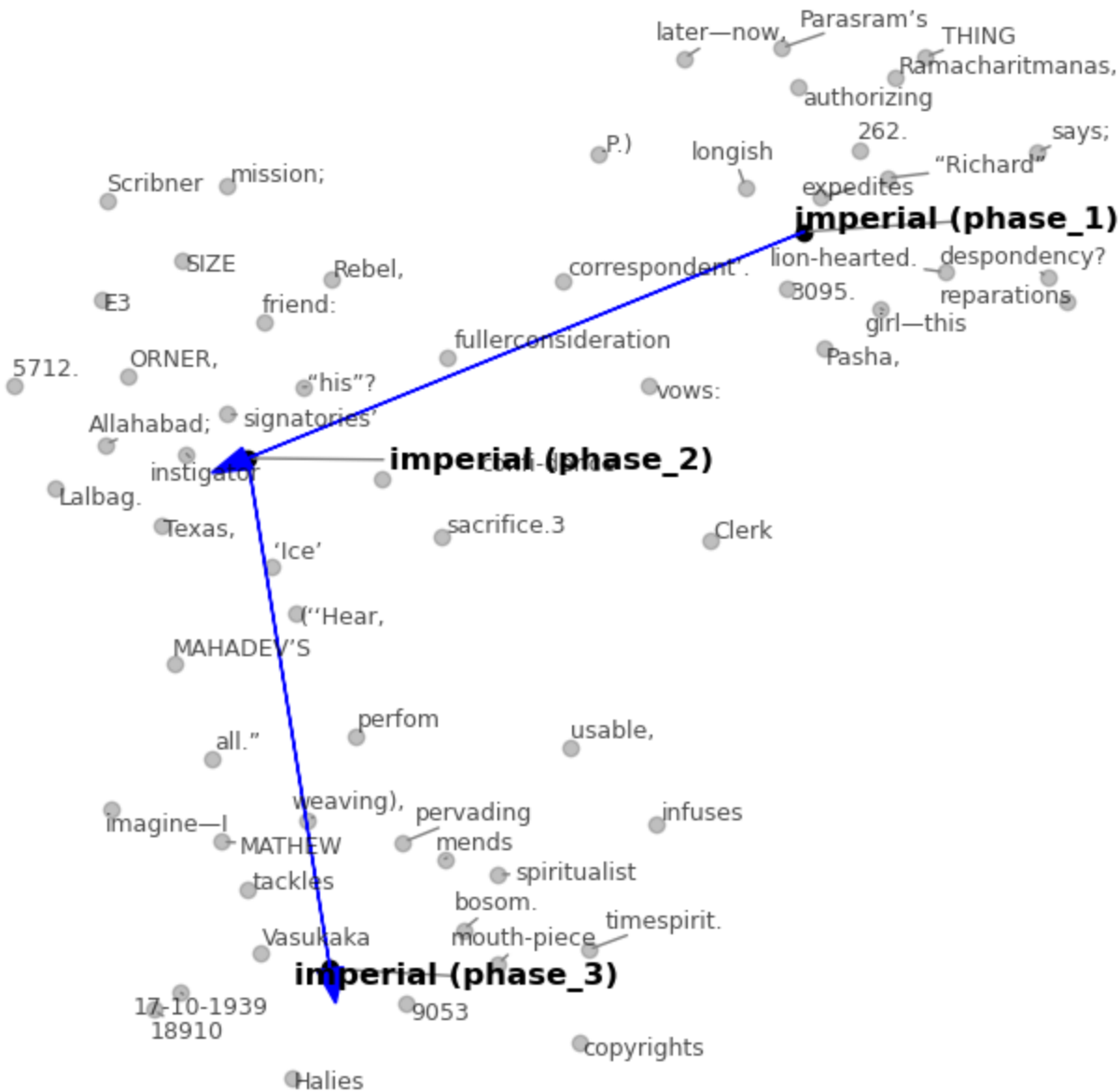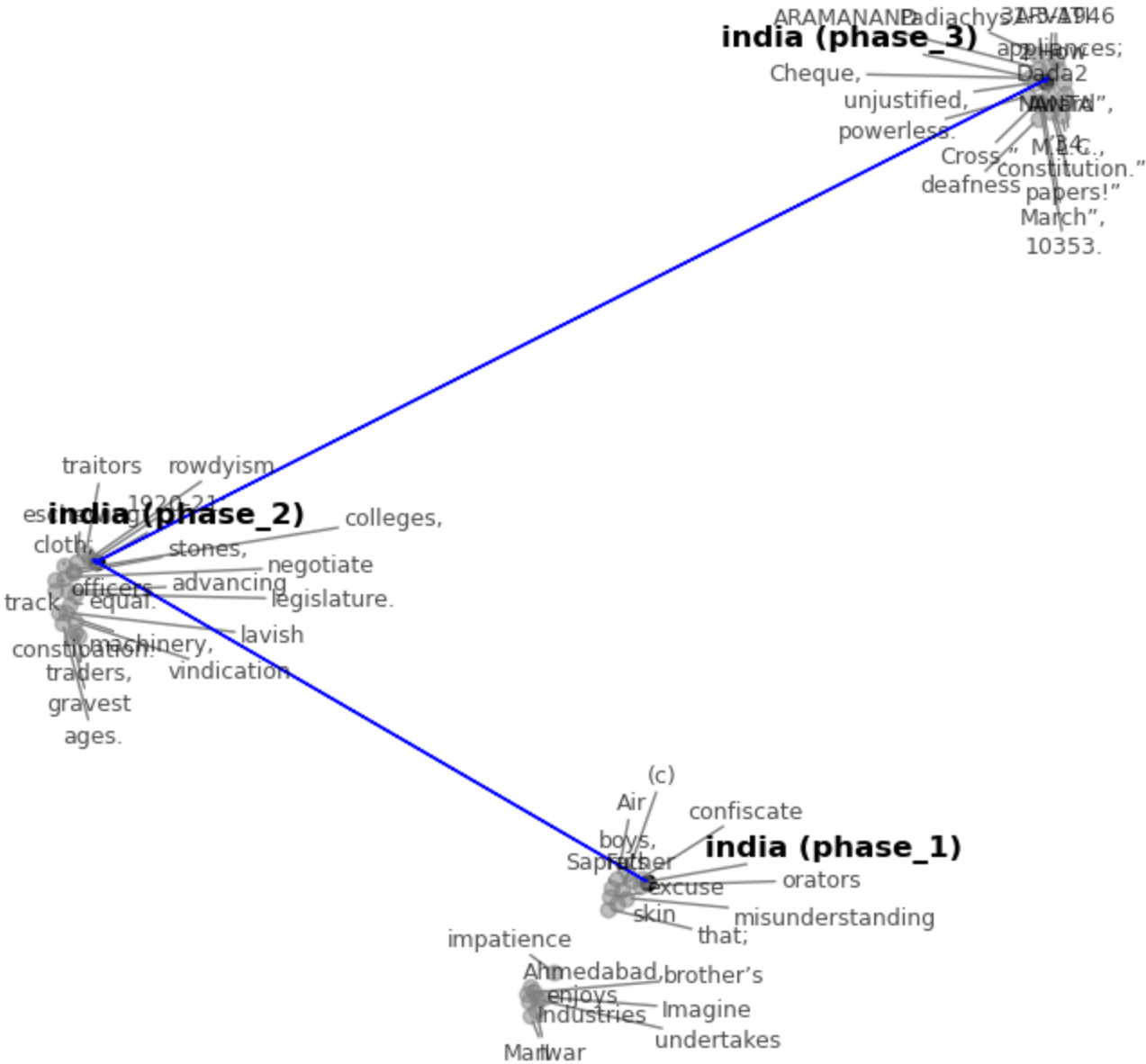
Semantic Drift Map for 'swaraj'

Swaraj
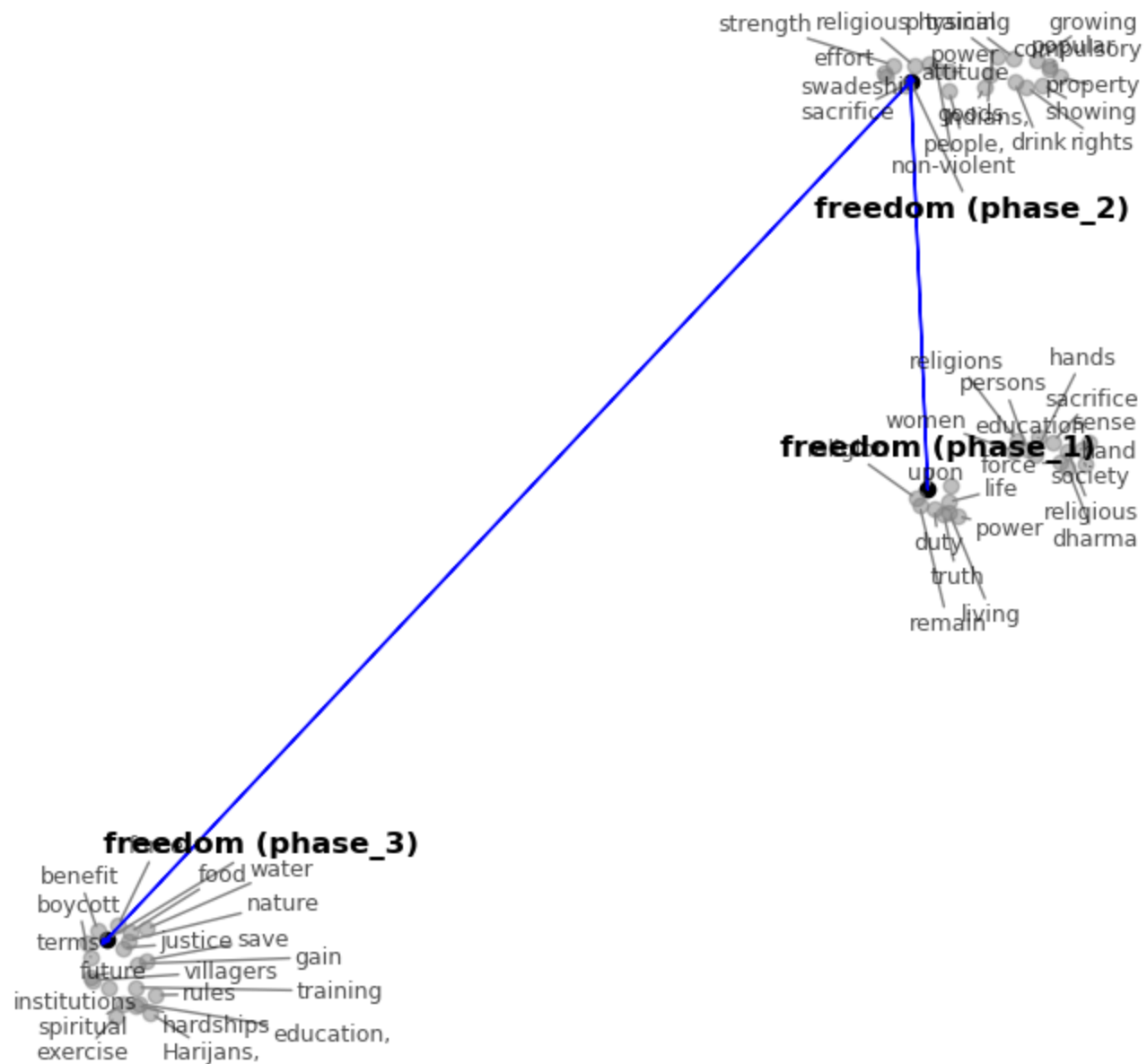
Semantic Drift Map for 'imperial'

Imperial

India

Semantic Drift Map for 'india'

Semantic Drift Map for 'freedom'
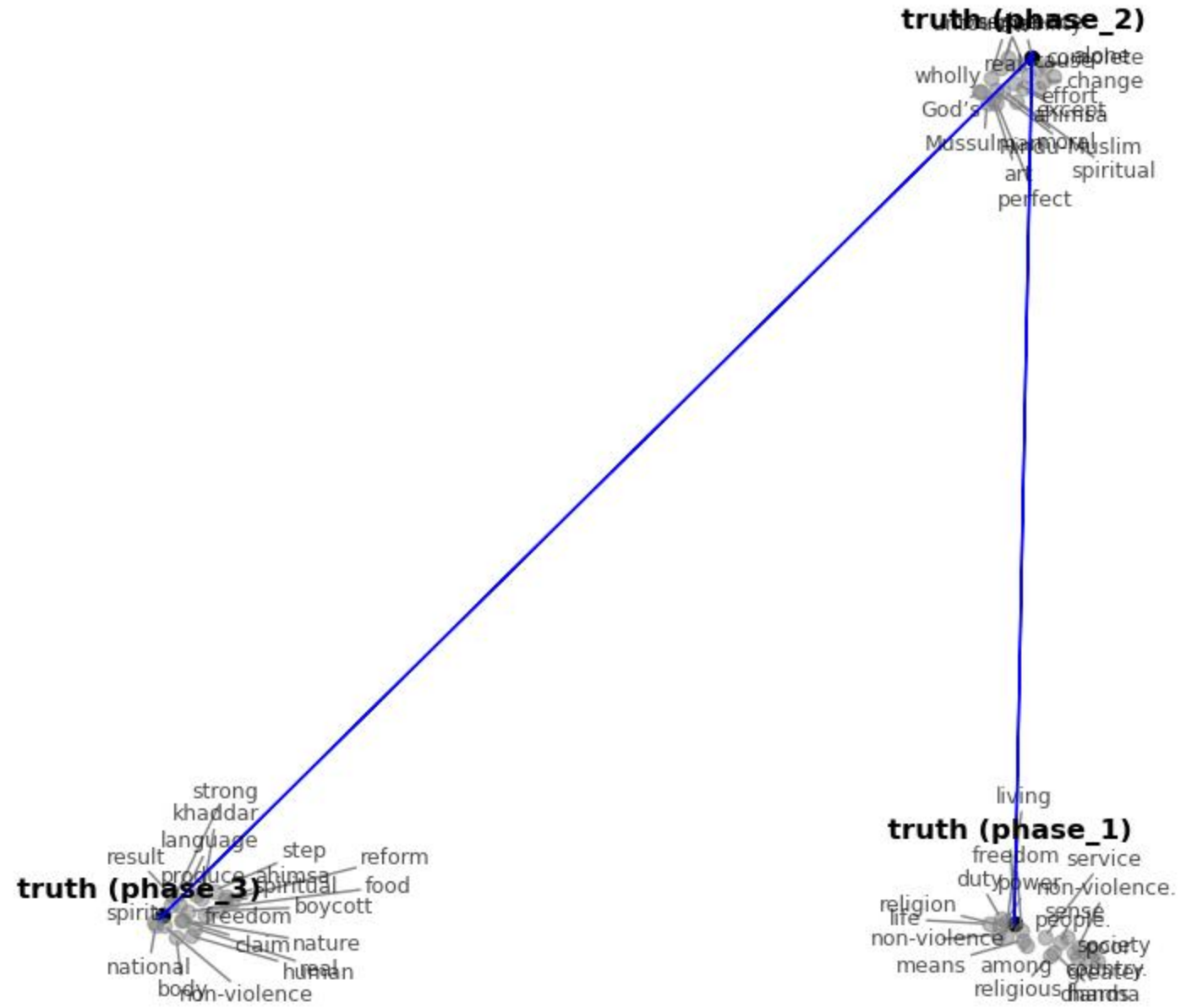
Freedom

Semantic Drift Map for 'truth'

Truth

Corruption