

# Constructing the *Nation*

*Identity and historical narratives in South Korean history textbooks, 1895–2016.*

Steven Denney · Aron van de Pol

*Leiden University*

---

RESEARCH QUESTION

How do the language and ideas of national identity in *state-sanctioned educational material* vary across regime types?

# Institutions and Narrative.

---

## *The curricular core.*

National history is the discourse *least likely* to change — where the state tells its most formal story of the nation.

## *Compressed history.*

1974–2002: the state was not a certifier but the *author*. One textbook, across Park, Chun, and the democratic transition.

---

COMPETING HYPOTHESES

# Regime type shapes textbook language.

---

EXPECTATION A

Authoritarian states compress the national vocabulary to a single official story.

EXPECTATION B

Democratic institutions dissolve that story into plural, contested meanings.

---

LITERATURE

Snyder (2000): democratization can *intensify* nationalism.

Jo (2025): elites may *entrench* rather than replace the inherited narrative.

*§ ii*

Corpus  
& periods.

# 67 textbooks, 1895–2016.

---

# 67

TEXTBOOKS

51 from NIKH digital archive · 16 OCR'd from existing library scans.

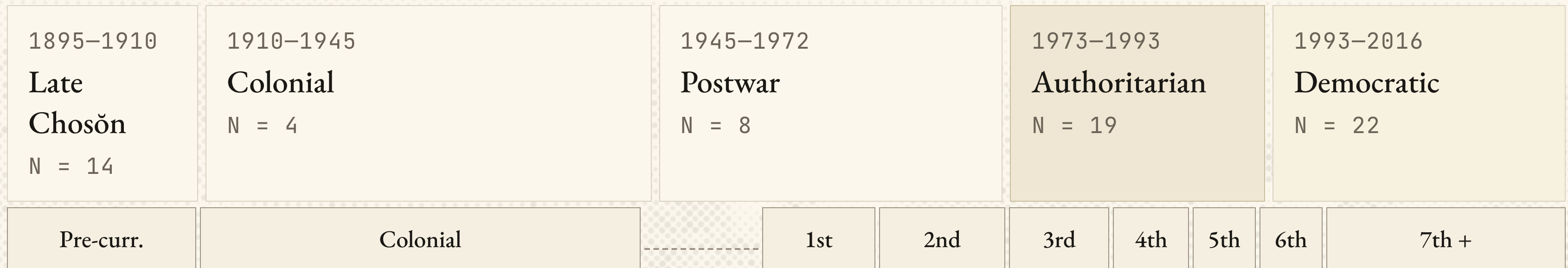
# 1895– 2016

COVERAGE

Kabo reforms through the 7th national curriculum.

Part of a cross-national corpus digitizing 1,000s of history textbooks.

# Five periods of textbook production.



*§ iii*

How we

*measure meaning.*

# LLM-assisted term discovery.

---

## 01 · EXTRACT

Three EXAONE models (2.4B / 7.8B / 32B) scan 4,000-char windows per textbook, returning identity-relevant nouns.

## 02 · SCORE

Candidates weighted 60 / 40 on semantic similarity and frequency; must appear  $\geq 3$  times.

## 03 · CONSENSUS

A term enters only if at least 2 of 3 models select it. *Twelve* terms survive.

---

VALIDATED AGAINST    BERTopic (recovers **12 / 12**) · Latent Dirichlet Allocation (recovers **7 / 12**)

# Three axes of semantic change.

---

Every sentence containing an identity term is embedded. For each term and each transition, we measure change along three complementary axes.

---

## LOGISTIC REGRESSION (LR)

### Discriminability

How easily a classifier can tell which period a sentence belongs to. Higher = more distinctly period-specific usage.

---

## CONTRASTIVE PCA (CPCA)

### Contextual spread

Whether a term's usage expands into new contexts or contracts into fewer. Positive = wider range of meanings.

---

## PROCRUSTES ANALYSIS (PA)

### Geometric relocation

How far a term moves in the embedding space after optimal alignment. Large = repositioned relative to other terms.

---

Applied to the two post-liberation transitions: Postwar → Authoritarian and Authoritarian → Democratic.

# Twelve terms that define the nation.

nation /  
ethnos      *minjok*      민족

state      *kukka*      국가

citizens      *kungmin*      국민

independence      *tongnip*      독립

unification      *t'ongil*      통일

liberation      *haebang*      해방

history      *yöksa*      역사

democracy      *minjujuüi*      민주주의

politics      *chöngch'i*      정치

society      *sahoe*      사회

territory      *yöngt'o*      영토

community      *kongdongch'e*      공동체

*§ iii*

What each word  
*is allowed to mean.*

## Nation-building vocabulary appears after 1945.

Before 1945, most of the twelve terms are absent or marginal in school history.

After 1945, the full set arrives together — a coordinated language of nation, state, politics, and citizenship.

National history installed a shared vocabulary for nation, state, and political community across all schools.

---

FIG. 1 · FOUR REPRESENTATIVE TERMS · PER 1,000 TOKENS

## Minjok peaks under authoritarianism.

---

*Mean frequency per 1,000 tokens across five periods.*

MEAN ABSOLUTE SHIFT MAGNITUDES · 12 TERMS × 3 AXES

## Variance dominates both transitions.

AXIS	POSTWAR → AUTH.	AUTH. → DEM.
Discriminative (LR)	0.46	0.41
<b>Variance (cPCA)</b>	<b>1.37</b>	<b>1.81</b>
Alignment (PA)	0.31	0.40

The dominant mode of change is not wholesale replacement of one vocabulary by another. It is expansion or contraction of the *range of contexts* in which terms appear. cPCA shifts are 3–5× larger than discriminative or alignment shifts.

# Selective appropriation, not uniform closure.

## EXPANDING — USABLE FOR LEGITIMATION

Independence 독립		+2.83
Nation 민족		+1.74
Citizens 국민		+1.59
History 역사		+1.49
Democracy 민주주의		+0.52
Unification 통일		+0.32

## COMPRESSING — HARDER TO CONTROL

Territory 영토		-0.18
State 국가		-0.85
Politics 정치		-0.89
Society 사회		-1.18
Liberation 해방		-1.98
Community 공동체		-2.89

*LR is uniformly negative across all terms (all become more period-specific). The split is on cPCA: widen what serves legitimation, compress what opens onto political or social analysis.*

---

WHAT THE SPLIT REVEALS

The regime *widened* the meanings of concepts useful for legitimation and *narrowed* the rest to a few repeated phrases.

SELECTIVE APPROPRIATION · NOT UNIFORM CLOSURE

---

EXPANSION

Ethnic belonging, colonial grievance, disciplined citizenship.

---

COMPRESSION

Society, politics, community -- pulled toward formulaic use.

---

FREEZING

Unification and democracy -- present but held in safe historical registers.

# Same word, wider range, new ideological load.

## POSTWAR

국사 교육은 우리 **민족**의 과거 생활을 정확하게 인식시켜 자아를 반성하고..  
"Recognizing our **nation's** past life, fostering self-reflection and patriotism..."

POSTWAR TEXTBOOK

## AUTHORITARIAN

일제의 식민지 지배는 우리 **민족**의 민족 운동을 탄압하고, 경제적 착취를 강행하  
며..  
"Colonial rule suppressed our **nation's** movement, enforced  
exploitation..."

1981 TEXTBOOK

**+1.74** cPCA — usage contexts widen

**0.63** PA — changing neighbouring words

# Keeps expanding — adds reconciliation to the *inherited frame*.

일제의 식민지 지배는 우리 **민족**의 민족 운동을 탄압하고, 경제적 착취를 강행하며...

"Colonial rule suppressed our **nation's** movement, enforced exploitation..."

1981 TEXTBOOK

7.4 남북 공동 성명에서 남북한은 자주 통일, 평화 통일, **민족** 대단결의 3대 원칙을 내세웠다.

"North and South Korea put forward three principles: self-reliant reunification, peaceful reunification, and great **national** unity."

2006 TEXTBOOK

**+1.22** cPCA — still expanding, less dramatically

# Frequent but frozen — confined to *anti-communist ritual*.

## HISTORICAL REFERENT

신라는 삼국을 통일하여 민족 문화 발전의 기틀을 마련하였다.

"Silla **unified** the Three Kingdoms and laid the foundation for national culture."

1979 TEXTBOOK

## ANTI-COMMUNIST FORMULA

민족적 숙원인 국토 통일을 위하여 공산주의와 대결할 수 있는 실력 배양에 전력을 집중한다.

"Concentrate all efforts on cultivating strength to confront communism for territorial **unification**."

1973 · YUSHIN ERA

Variance shift of only **+0.32** cPCA. The term was *retained* but *constrained* -- confined to Unified Silla and ritualized slogans, never substantive policy.

# Under democracy, unification enters *contemporary policy.*

AUTHORITARIAN · FROZEN

민족적 숙원인 국토 **통일**을 위하여 공산주의와 대결할 수 있는...

"...cultivating strength to confront communism for **unification**..."

1973 · YUSHIN ERA

DEMOCRATIC · POLICY DISCOURSE

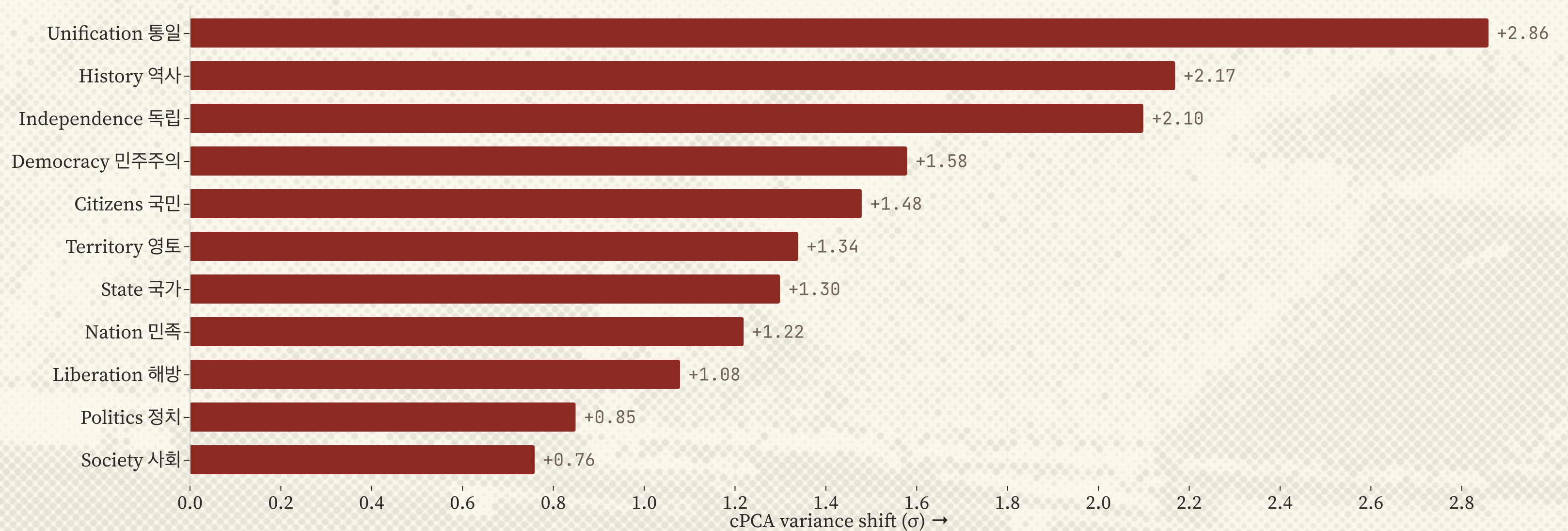
정상 회담 결과 **통일** 문제와 남북 관계의 기본 방향을 담은 6.15 남북 공동 선언이 발표되었다.

"The June 15 Joint Declaration... the basic direction for **unification** and inter-Korean relations."

2016 TEXTBOOK

The **+2.86** cPCA shift is the second-largest in the analysis. Unification had not been *absent* from authoritarian textbooks -- it had been *constrained*. Institutional change, not new referents, unlocked the range.

# All terms expand under democracy.



# Haebang reactivates its liberation sense.

양민으로서 억울하게 노비가 된 자들을 **해방** 시켜...

"**emancipating** those unjustly enslaved" — premodern royal policy.

1981 TEXTBOOK

독립 운동계는 민족 **해방** 을 부르짖는 민족주의 계열과 계급 **해방** 을 목표로 하는 사회주의 세력으로 양분되었다.

"...national **liberation** vs. **class liberation**" — political-movement framing.

2007 TEXTBOOK

+1.91 PA — largest shift in the analysis. Geometric relocation: dormant senses activate.

# Two terms never fully recover.

## SAHOE 사회 · SOCIETY

-1.18 → +0.76

cPCA · Post. → Auth.

cPCA · Auth. → Dem.

**Only 2/3 recovered.** The evaluative edge -- society as critical category -  
- does not fully return.

## HAEBANG 해방 · LIBERATION

-1.98 → +1.08

cPCA · Post. → Auth.

cPCA · Auth. → Dem.

**Only 1/2 recovered.** Three decades of democracy did not erase the  
narrowing.

*A form of banal constraint: a narrowing of conceptual tools that persists precisely because it is unremarkable.*

# Three phases of state-driven semantic change.

---

---

## *Phase I*

### Establish

Liberation installs a stable vocabulary. Every successor regime inherits it.

---

## *Phase II*

### Modulate

Selective appropriation: stretch what serves legitimation, flatten what doesn't.

---

## *Phase III*

### Pluralize

Democracy does not de-nationalize the textbook. It *de-solidifies* it: every term widens; the frame loosens.

*The core persists. What varies is how tightly the surrounding language is gathered into one frame.*

WHAT THIS PAPER OFFERS

# Contributions.

---

## METHOD

Replicable pipeline for corpus-driven concept discovery using LLMs, validated against BERTopic and LDA.

Multi-axis semantic design (LR, cPCA, PA) that separates different mechanisms of meaning change.

## SUBSTANTIVE & THEORETICAL

Regime type shapes the grand narrative states transmit through education.

Authoritarian schooling selectively appropriates — expanding some terms while compressing others.

Democratization does not per se replace the inherited narrative — it widens inside the frame authoritarianism built.

A national vocabulary can be *shared* across regimes.

The meanings those words *carry* can differ substantially with the institutions that produce them.