Youngchae's Voice: Audible Speech and Inner Thought in Mujeong

Injin Woo

Dept of Philosophy, SKKU
Introduction to Digital Humanities

30 October 2026

Youngchae's speech and thought

How do Youngchae's audible speech and represented thoughts vary across chapters 31–40?

Provisional claim

Counting audible utterances alone may not capture how the novel represents Youngchae's voice.

Literary check: Do close readings of chapters 31 and 35 support the pattern in the tagged text?

Dataset description

- **Text:** Yi Kwang-su's long novel *Mujeong*, chapters 31–40, transcribed from Wikisource.
- **Encoding:** a reviewed TEI file with ten chapter divisions, 127 paragraphs, and 26,599 characters of chapter text.
- **Derived dataset:** ten chapter-level plain-text files plus tables of tagged passages and chapter totals.

This is a separate encoding of the long novel, not KNoTE's 1910 short story *Mujeong*.

Dataset link: Download the TEI dataset ZIP from my website

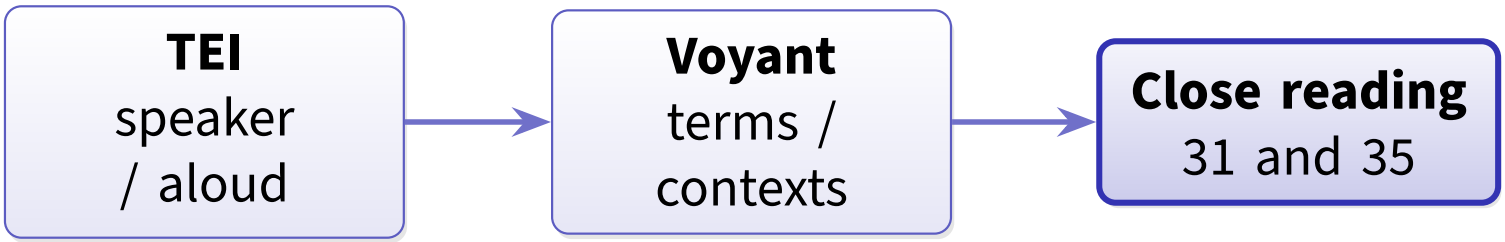
Dataset structure and content

Layer	Field	Use
TEI chapter	`div@n`	chapter number
TEI passage	`said@who`	attributed speaker
TEI passage	`said@aloud`	audible / unspoken
CSV row	chapter, paragraph, text	inspectable passage

Youngchae tags	Audible	Unspoken
Chapter 31	8	2
Chapter 35	0	15

Tools and method

1. **TEI:** extract passages with `who=#YC`; compare `aloud=true` and `aloud=false` by chapter.
2. **Voyant:** upload the ten chapter texts; use *Trends* for selected terms and *Contexts* to inspect their uses.
3. **Close reading:** compare the quantitative pattern with narration and diction in chapters 31 and 35.



Final project and expected outcomes

Final paper

1. Explain the dataset and tagging decisions.
2. Compare chapter-level speech and thought tags.
3. Closely read chapters 31 and 35, then reassess the claim.

Expected outcome — a hypothesis to test

Youngchae's voice may be present in represented thought even where audible speech is absent.

Datasheet for Datasets

- **Scope:** Why were chapters 31–40 selected?
- **Provenance:** Where did the source text come from, and who reviewed the TEI?
- **Annotation:** What does one `said` row represent, and how were ambiguous speakers handled?
- **Validation:** How was the TEI checked against the transcription?
- **Reuse:** Which files are needed to reproduce the chapter totals?

Datasheet questions (1–6)

1. What research question was this dataset created to address?
2. Why were chapters 31–40 selected rather than the entire novel?
3. What is the intended unit of analysis: chapter, paragraph, tagged passage, or speech event?
4. What is the source of the chapter texts?
5. Who created and reviewed the TEI encoding, and when?
6. How were the ten chapter-level plain-text files extracted from TEI?

Datasheet questions (7–12)

7. What does each row of `said_segments.csv` represent?
8. What criteria distinguish audible speech from unspoken thought?
9. How were ambiguous speakers handled?
10. Can one speech event produce more than one `said` tag?
11. How was the TEI text checked against the source transcription?
12. Which files and instructions are required to reproduce the tables and figures?