Use of Multi-Step Markov Chains in the Characterization of English Literary Works

Clement Leung

School of Science and Engineering

The Chinese University of Hong Kong, Shenzhen

Guangdong Provincial Key Laboratory of Future Networks of Intelligence

clementleung@cuhk.edu.cn

Chenjie Zeng

School of Humanities and Social Science

The Chinese University of Hong Kong, Shenzhen

chenjiezeng@link.cuhk.edu.cn

About the Presenters

Clement Leung

- FULL PROFESSORSHIPS at
  - University of London, UK; National University of Singapore; Chinese University of Hong Kong, Shenzhen, China; Hong Kong Baptist University; Victoria University, Australia
- Two US patents, five books and over 150 research articles
- Program Chair, Keynote Speaker, Panel Expert of major International Conferences
- Editorial Board of ten International Journals
- Listed in Who's Who in the World and Great Minds of the 21st Century
- Fellow of the British Computer Society, Fellow of the Royal Society of Arts, Chartered Engineer

Chenjie Zeng

- PhD Candidate
- Chinese University of Hong Kong, Shenzhen

Content

A Markov Chain Representation

Multi-Step Markov Model

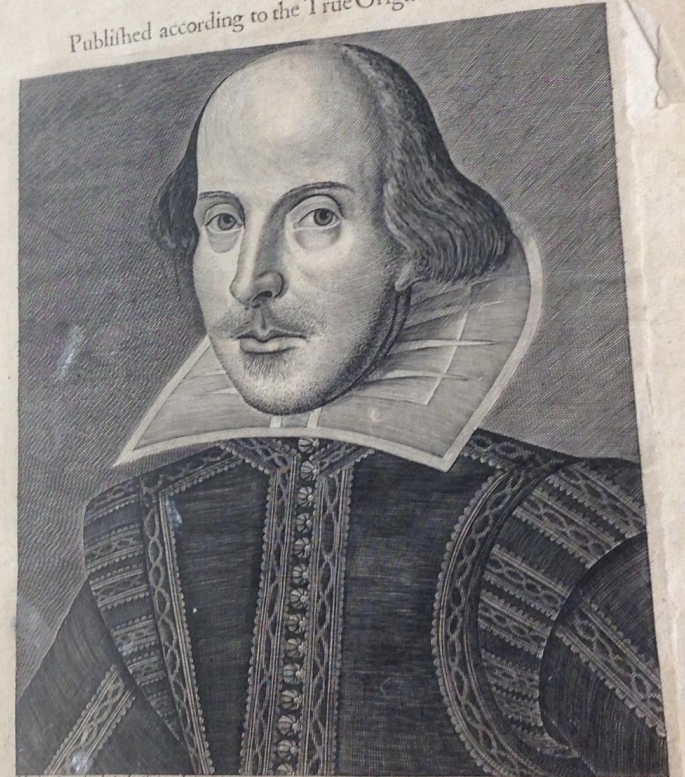**Linguistic Experiments**

Conclusion

A Markov Chain Representation

- Polarity and Emotions
- Iambic Pentameter

Polarity and Emotions

Polarity	Verse	Emotion
Positive	Tossing their heads in sprightly dance	Sprightly
	A poet could not but be gay	Gay
	In such a jocund company	Jocund
Negative	Which is the bliss of solitude	Bliss; Solitude
	I wandered lonely as a cloud	Lonely
	In vacant or in pensive mood	Vacant; Pensive

Iambic Pentameter
(like heartbeats)

- **stressed** sound: "/" (**ictus** syllable=4)
- Unstressed sound: "×" (non-ictus syllable=1)
- standard iambic pentameter:
    × / × / × / × / × /   (five pairs of iambs)
- Example:
    - 1   4 1   4 1   4   1   4 1   4
    - ×   /   ×   /   ×   /   ×   /   ×   /
    - To **be** or **not** to **be**, that **is** a **question.**

Iambic Pentameter

- u   / u /   u /   u /
- *They **flash** u**pon** that **in**ward **eye***
- u   / u   /   u / u /
- *Which **is** the **bliss** of **so**li**tude**;*
- u   /   u   /   u /   u   /
- *And **then** my **heart** with **plea**sure **fills**,*
- u   / u   / u / u /
- *And **dan**ces **with** the **da**ffo**dils**.*

a. "Daffodils" by William Wordsworth

MULTI-STEP MARKOV CHAIN CHARACTERIZATION WITH MEMORY

- Multi-Step Markov Model from Simple Markov Chain
- Linguistic Experiments

William Shakespeare's Henry IV, Part 1

Markov Model

- 1   4   1   4   1   4   1   3   1   4
- ×   /   ×   /   ×   /   ×   /   ×   /
- *His **bran**disht **Sword** did **blinde** men **with** his **beames**.*

$$\begin{pmatrix} 0 & 0 & 0.2 & 0.8 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

William Shakespeare's
Henry IV, Part 2

Markov Model

- *1   4   1   4   1   4   1   4   1   4*

- *× / × / × / × / × /*

- *His **spark**ling **eyes**, re**pleat** with **wrath**full **fire**.*

can be constructed simply as

$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

Convex Combination of Transition Matrices

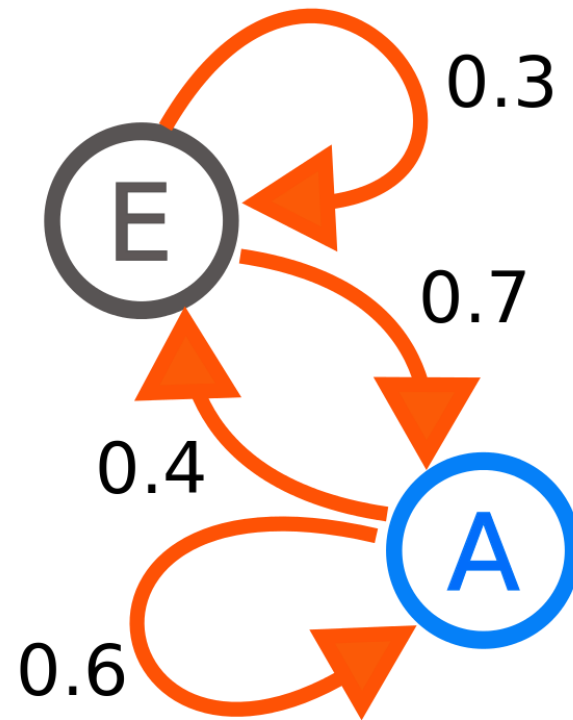$$T = r \begin{pmatrix} 0 & 0 & 0.2 & 0.8 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} + s \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$T' = \begin{pmatrix} 0 & 0 & 0.2r & 0.8r + s \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$U = \sum_{j=1}^{m} p_j T_j$$

Markov Chain is Memoryless

Simple Markov Chain Has No
Memory of Past States

Incorporate Memory by Forming the Cartesian Product of the Markov State Space

$$(a_1,b_1) \; (a_1,b_2) \; (a_1,b_3) \; (a_1,b_4) \; (a_1,b_5) \; \ldots$$

$$(a_2,b_1) \; (a_2,b_2) \; (a_2,b_3) \; (a_2,b_4) \; (a_2,b_5) \; \ldots$$

$$(a_3,b_1) \; (a_3,b_2) \; (a_3,b_3) \; (a_3,b_4) \; (a_3,b_5) \; \ldots$$

$$(a_4,b_1) \; (a_4,b_2) \; (a_4,b_3) \; (a_4,b_4) \; (a_4,b_5) \; \ldots$$

$$(a_5,b_1) \; (a_5,b_2) \; (a_5,b_3) \; (a_5,b_4) \; (a_5,b_5) \; \ldots$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
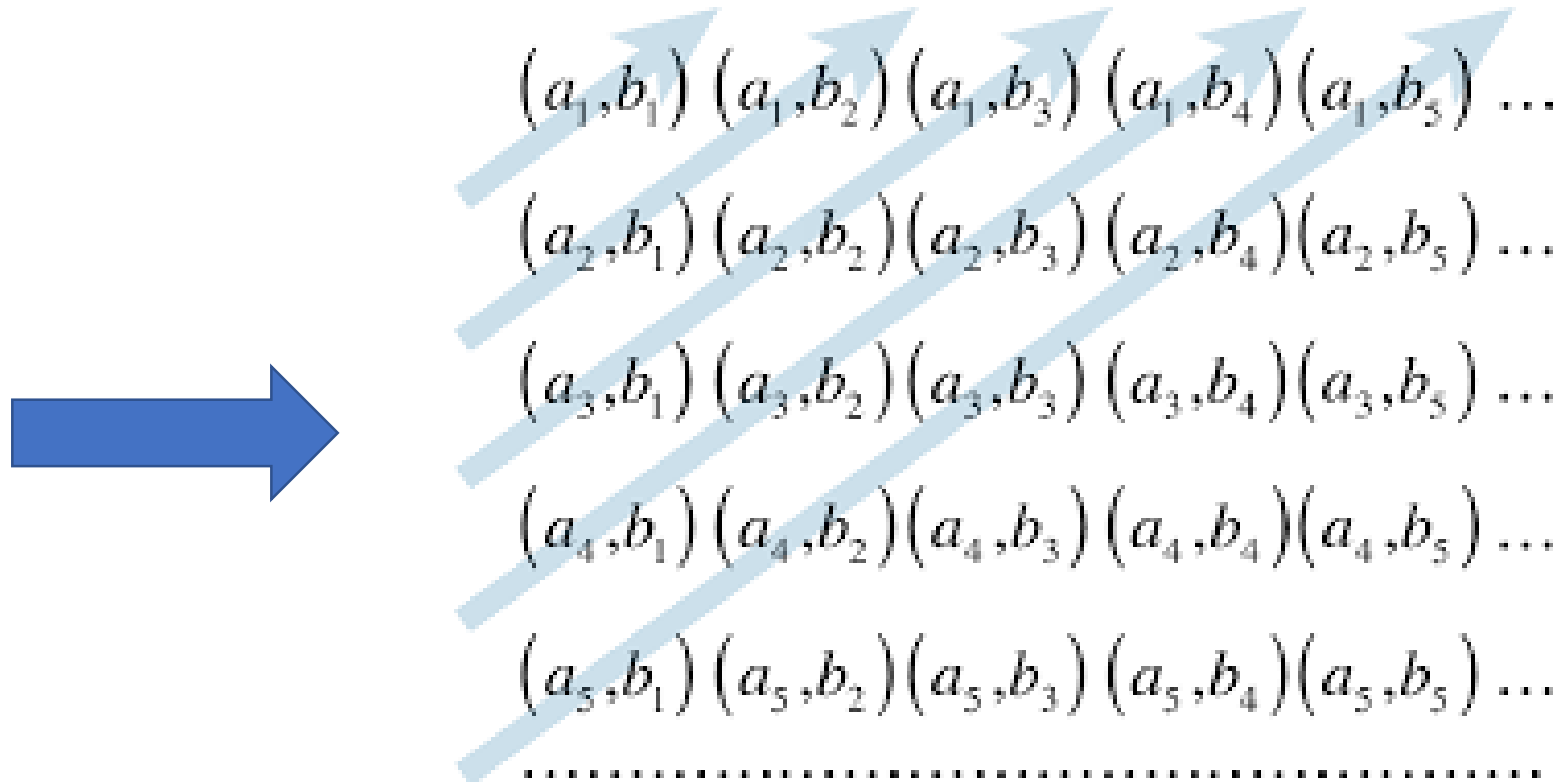
Cartesian Product of the Underling State Space: $\Omega = S \times S$

Multi-Step Markov Model

1   4   1   4   1   4   1   4   1   2
✕ / ✕ /   ✕ /   ✕ / ✕ /
*His **cush**es on his **thi**ghs, **gal**lantly **arm**'d,*

1   4   1   4   1   4   1   4   1   3
✕ / ✕ /   ✕ /   ✕ / ✕ /
*Rise **from** the **groun**d like **feat**hered **Mer**cury,*

1   4   1   3   1   4   1   4   1   4
✕ / ✕ /   ✕ /   ✕ / ✕ /
*And **Vaul**ted **with** such **ease** on**to** his **seat***

$$\Omega = S \times S = \{\ (1, 1), (1, 2), (1, 3), (1, 4),$$
$$(2, 1), (2, 2), (2, 3), (2, 4),$$
$$(3, 1), (3 ,2), (3, 3), (3, 4),$$
$$(4, 1), (4, 2), (4, 3), (4, 4)\ \}$$

Space Reduction Eliminates Impossible Transitions

1   4   1   4   1   4   1   4   1   2
✕ / ✕ / ✕ / ✕ / ✕ /
*His **cush**es on his **thi**ghs, **gall**antly **arm**'d,*

1   4   1   4   1   4   1   4   1   3
✕ / ✕ / ✕ / ✕ / ✕ /
*Rise **from** the **groun**d like **feat**hered **Mer**cury,*

1   4   1   3   1   4   1   4   1   4
✕ / ✕ / ✕ / ✕ / ✕ /
*And **Vaul**ted **with** such **ease** on**to** his **seat***

$$\Omega = S \times S = \{ \cancel{(1, 1)}, (1, 2), (1, 3), (1, 4),$$
$$(2, 1), \cancel{(2, 2)}, \cancel{(2, 3)}, \cancel{(2, 4)},$$
$$(3, 1), \cancel{(3, 2)}, \cancel{(3, 3)}, \cancel{(3, 4)},$$
$$(4, 1), \cancel{(4, 2)}, \cancel{(4, 3)}, \cancel{(4, 4)} \}.$$

Sparse Matrix Approximations Simplify Computation and Comparison

$$H' = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0.09 & 0.18 & 0.73 & 0 & 0 & 0 \end{pmatrix} \doteqdot \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Linguistic Experiment on Exclamatory Lines

Table 1– Exclamatory Lines

| References	Frequency	Total lines	Percentage
Act One (Total)	36	596	33%
Scene I	11	177	6%
Scene II	6	150	4%
Scene III	11	90	12%
Scene IV	6	111	5%
Scene V	2	37	5%
Scene VI	0	31	0%
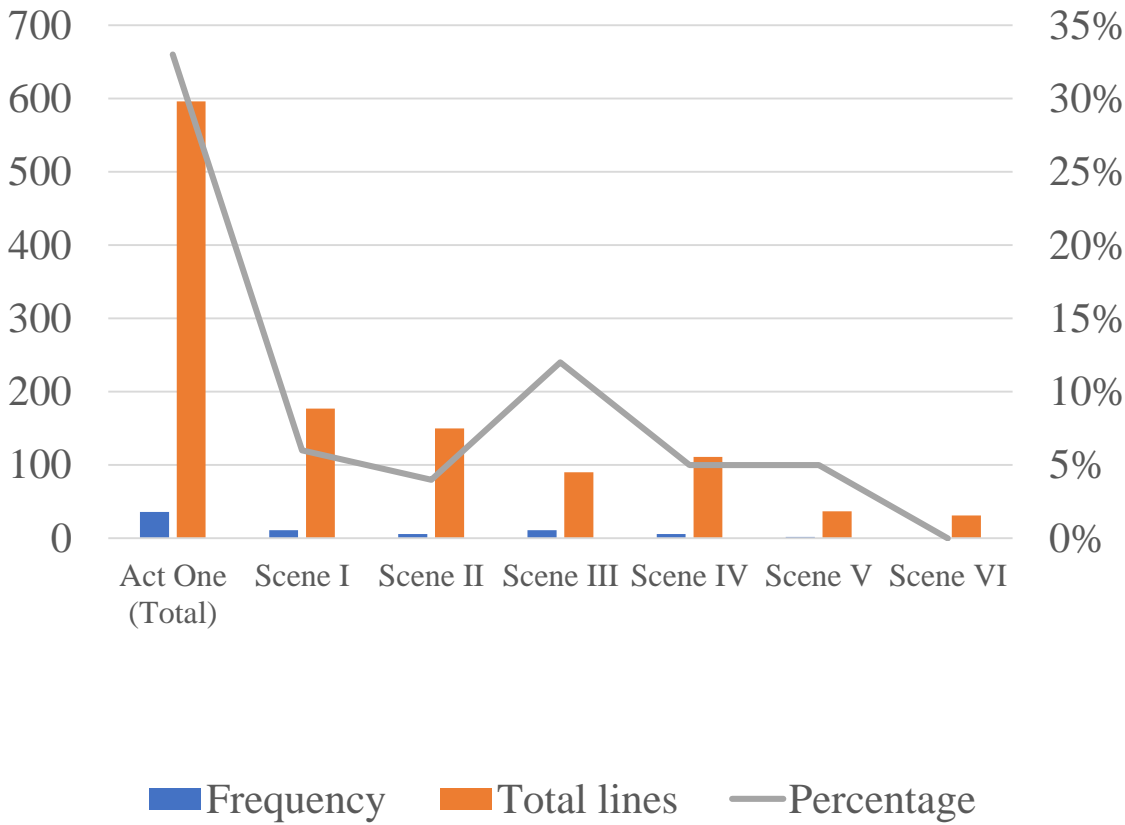
*(Source: Act one, Part one, **King Henry the Sixth**)*

Linguistic Experiment on Exclamatory Lines

References	Frequency	Total lines	Percentage
Act One (Total)	36	596	33%
Scene I	11	177	6%
Scene II	6	150	4%
Scene III	11	90	12%
Scene IV	6	111	5%
Scene V	2	37	5%
Scene VI	0	31	0%
Act Two (Total)	26	556	26%
Scene I	11	82	13%
Scene II	6	134	4%
Scene III	3	82	4%
Scene IV	3	134	2%
Scene V	3	124	2%
Act Three (Total)	39	474	30%
Scene I	15	201	7%
Scene II	16	137	12%
Scene III	6	91	7%
Scene IV	2	45	4%
Act Four (Total)	47	557	62%
Scene I	12	194	6%
Scene II	6	56	11%
Scene III	7	53	13%
Scene IV	3	46	7%
Scene V	3	55	5%
Scene VI	5	57	9%
Scene VII	11	96	11%
Act Five (Total)	39	561	38%
Scene I	3	62	5%
Scene II	3	21	14%
Scene III	15	195	8%
Scene IV	16	175	9%
Scene V	2	108	2%

Table – Exclamatory Lines
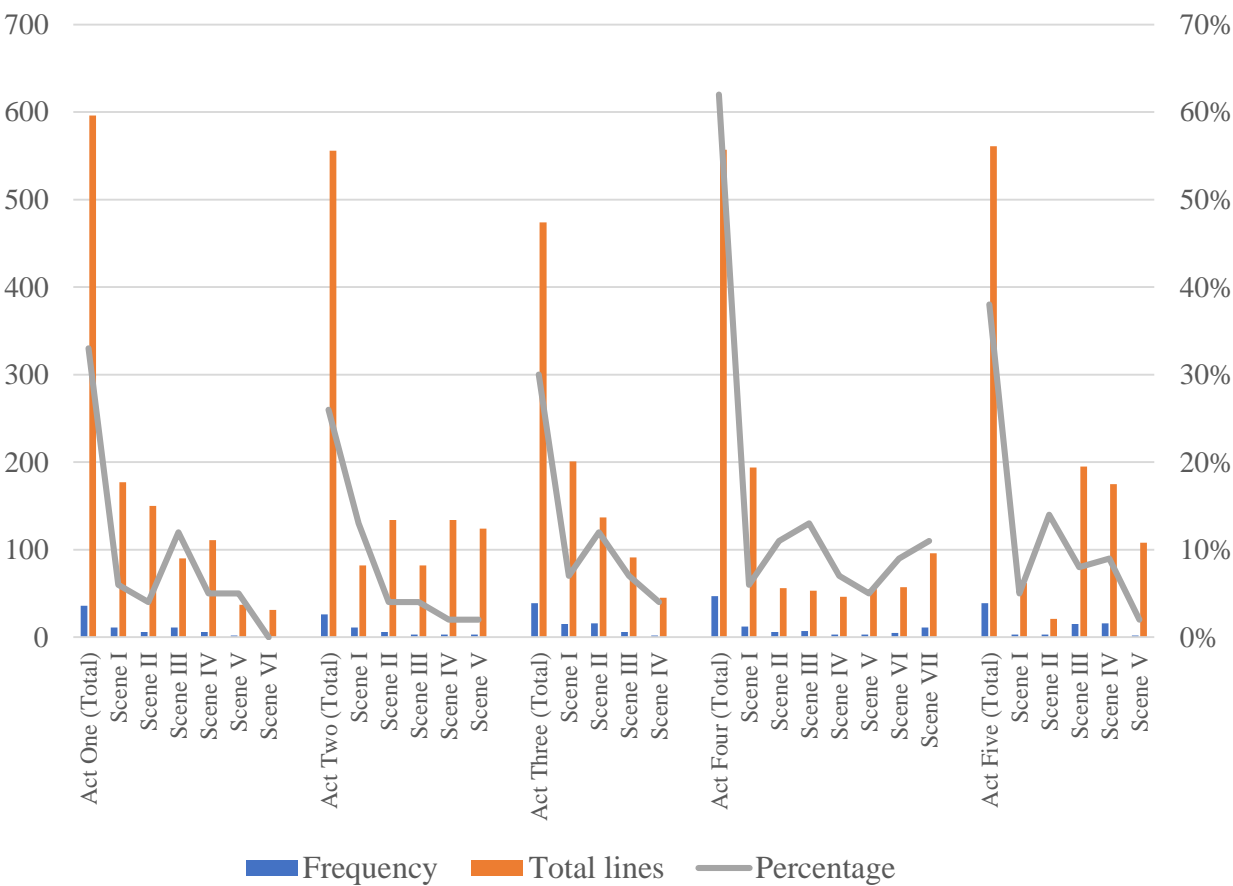
(Source: Part one, King Henry the Sixth)

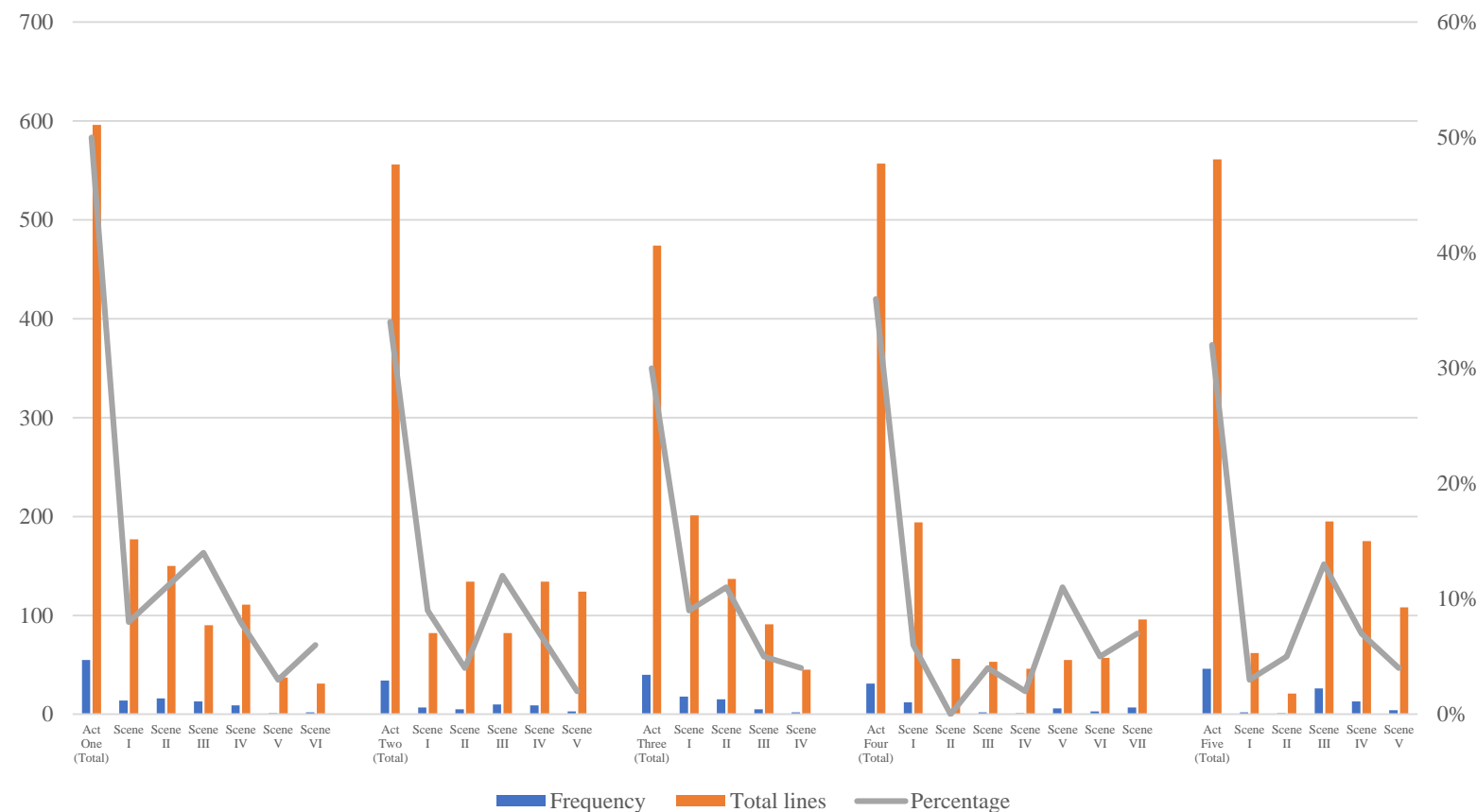# Linguistic Experiment on Interrogative Lines

Figure 2. Interrogative lines in Henry VI, Part I

Conclusion

- Simple Markov Chains provide unique representation of literary passages
- Multi-Step Transition Matrices enable versatile and detailed passage and style characterizations
- Sparse Transition Matrices from reduced state space greatly smplify computation