Things Hidden Behind
The Numbers of İnce Memeds

Bilkent Library Lunchtime Lecture
March 5, 2015

Fazlı Can
Computer Engineering Department
Bilkent University
“Change of Writing Style with Time” (2001)

Jon M. Patton
Miami University, OH

Süha Oğuzertem
Bilgi Üniversitesi
İnce Memed (1969)
İnce Memed... (1969)
İnce Memed... (1954)
İnce Memed... (Cumhuriyet, 1954)
İnce Memed... Digital! (2001)

Engin Demir
THK Üniversitesi
“İnce Memed – Köy ve Kasaba Arasında”
Sami Arpa (2014)
“İnce Memedlerin Sayılarında Gizlenenler” Yaşar Kemal Onur Günü Bilgi Üniversitesi (2014)
Yaşar Kemal [1923 – 2015]
What I Talk About

• Text Mining (TM)
• Change of Style with Time
• İnce Memeds
• İnce Memeds vs. 20th Turkish Century Literature
• Change in Turkish Language
• Conclusions & TM in Literary Scholarship
Text Mining

• Find hidden patterns: Unconscious/Conscious

• Find the pattern in text that is under scrutiny
  – Author
  – Time
  – Genre
  – Gender
  – Sentiment
  – …
Music Mining

- Find the pattern
- Find/Guess:
  - Composer
  - Time
  - Genre
  - Gender
  - Sentiment
  - ...

Şükrü Tunar

Antonio Vivaldi

Zeki Müren
A Literary Hoax – Süha Oğuzertem, 2003
Gizli Kalmış Bir İstanbul Masalı
An İstanbul Fairy Tale that Remained Secret

Nurten Ay

Cem Behar

1991
Nurten Ay – Cem Behar: Are they similar?

Machine Learning for Authorship Attribution: A Book is in Search for an Author

Cem Behar: Cem Behar

Nurten Ay: Cem Behar
“A Literary Hoax” 2007

Gizli Kalmış Bir İstanbul Masalı

Ali Teoman
Is text mining always successful?

*La Vie Devant Soi – The Life Before Us*
How Do They Work?

• Training Process: Teach style

• Text in suspect: Check its style

• Is it similar to the style taught?

• Daily life example: Look at real estate listings and guess
Text Mining Methods

• Supervised
  – Statistical Methods
  – Machine Learning

• Unsupervised
  – Clustering
How to Define Text?

• Divide text into blocks containing the same number of words:
Define Text with Objective Style Markers

• Frequency of function words
  – bu, ki, ve ...

• Frequency of TOKEN (words in text) lengths
  – “Memed yavaş yavaş kendine geliyor.”
    3 words of length 5

  –
    2 words of length 7

• Frequency of TYPE (words in dictionary) lengths
  – “Memed yavaş kendine geliyor”
    2 unique words of length 5
    2 unique words of length 7
Style Markers...

- Frequency of sentence lengths
- Frequency of lexical categories of words
  - noun, adjective, pronoun, adverb, ...
- Frequency of syllable lengths
  - “Memed yavaş yavaş kendine geliyor.”
    3 words with 2 syllables
    2 words with 3 syllables
Text Mining

Binboğalar

Fırat Suyu

bu ki ve ...
[ 3 5 12 ...]

[ .... ]

[ .... ]
Change of Style with Time

Çetin Altan ve Yaşar Kemal
Distinction of Old & New

Yaşar Kemal: Distinction of Old & New
Binboğalar... 1971/いただいてる Suyu... 1998

Function Words

Old
New
## Old or New?

*Binboğalar... – Fırat Suyu ...*

<table>
<thead>
<tr>
<th>Success Rate</th>
<th>New</th>
<th>Old</th>
</tr>
</thead>
<tbody>
<tr>
<td>81.25</td>
<td>(13)</td>
<td>(12)</td>
</tr>
<tr>
<td>75.00</td>
<td>(12)</td>
<td>(13)</td>
</tr>
<tr>
<td>93.75</td>
<td>(15)</td>
<td>(16)</td>
</tr>
</tbody>
</table>

- **Token Lengths**:
  - New: 13
  - Old: 12

- **Type Lengths**: New: 12, Old: 13

- **Functional Words**:
  - New: 15
  - Old: 16

*Each Novel is 16 Blocks*

$\frac{16}{16} = 100\%$ Success Rate
Change in Word Usage
Old (*Binboğalar*) ➔ New (*Fırat Suyu...*)

- Average Token Length: 5.92 ➔ 6.03 letters
- Average Type Length: 6.96 ➔ 7.09 letters
- Length difference between old and new is statistically significant (not by chance)
İnce Memed Tetralogy


- Publication years: 1955 - 1987
- Writing years: 1947 - 1986
- Novel Events: 1924-1938
- Writing Style
  - Fethi Naci: The same style in all
  - Konur Ertop: Changed his style
  - Süha Oğuzertem: Changed his style
**İnce Memed Tetralogy**

<table>
<thead>
<tr>
<th>Novel, Date of Publication</th>
<th>No. of Tokens (N)</th>
<th>No. of Types (V)</th>
<th>Avg. Token Length</th>
<th>Avg. Type Length</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>İnce Memed</em> [1], 1955</td>
<td>86,457</td>
<td>17,110</td>
<td>5.80</td>
<td>8.01</td>
</tr>
<tr>
<td><em>İnce Memed</em> 2, 1969</td>
<td>107,348</td>
<td>21,146</td>
<td>5.85</td>
<td>8.24</td>
</tr>
<tr>
<td><em>İnce Memed</em> 3, 1983</td>
<td>156,876</td>
<td>26,805</td>
<td>5.81</td>
<td>8.42</td>
</tr>
<tr>
<td><em>İnce Memed</em> 4, 1987 (*)</td>
<td>164,474</td>
<td>28,350</td>
<td>5.91</td>
<td>8.48</td>
</tr>
<tr>
<td><em>İnce Memed</em> 1-4, 1955-1987</td>
<td>515,155</td>
<td>55,394</td>
<td>5.85</td>
<td>8.82</td>
</tr>
</tbody>
</table>
İnce Memed Tetralogy

• Analysis with six style markers
  – Sentence length
  – Most frequent (function) words
  – Syllable count
  – Word types (noun, ...)
  – Token length
  – Type length
Correct Classification Rates

[Bar chart showing success rates for different classification methods across four volumes (Vol 1 to Vol 4). The methods include Sentence Length, Most Frequent Words, Syllable Counts, Word Types, Token Lengths, and Type Lengths.]
Sentence Length Changes

Percentage of Sentences vs. Words Per Sentence

- Vol1
- Vol2
- Vol3
- Vol4

Words per Sentence

Percentage of Sentences
Sentence Lengths

Sentence Length

İnce Memed 1    ...    İnce Memed 4
New Words: Longer

<table>
<thead>
<tr>
<th>Volume</th>
<th>Token Length</th>
<th>Type Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>1</td>
<td>5.80</td>
<td>0.081</td>
</tr>
<tr>
<td>2</td>
<td>5.85</td>
<td>0.113</td>
</tr>
<tr>
<td>3</td>
<td>5.81</td>
<td>0.143</td>
</tr>
<tr>
<td>4</td>
<td>5.91</td>
<td>0.161</td>
</tr>
</tbody>
</table>
20th Century Turkish Novel

20th Century Turkish Literature
40 Novels: 10 Novels/Quarter Century
Token Length: Longer with Time

Average Word in Text (Year) = 5.9741 + 0.00368 \times (\text{Year} - 1900)

- 1900: 5.9741 letters
- 2000: 6.3421 letters
- Increase in 100 Years: 0.368 Letters
Type Length: Longer with Time

Avg. Word Length in Dictionary (Year) =
8.0127 + 0.00609 *(Year - 1900)

- 1900: 8.0127 letters
- 2000: 8.6217 letters
- Increase in 100 Years: 0.609 Letters
Language Change Quantification: Using Time-Separated Parallel Translations

Study Based on Time-Separated Translations

Average Stem Length (No. of Letters) Change From Old to New
Average Suffix Length (No. of Letters)
Change From Old to New
Old vs. New

• Words: Longer
• Stems: Shorter
• Suffixes: Longer
• Meanings per stem: More
• Vocabulary richness: Lower
Fahim Bey ve Biz, Fahim Bey and Us
Abdülhak Şinasi Hisar, 1941
(Varlık 1966 ed., published by Yaşar Nabi)
"New generations are unable to get the taste of authors who died just yesterday [...] A huge drama with an unknown finishing time for Turkish authors." Yaşar Nabi
Conclusion

• Text mining: Works well as expected.

• Turkish words: Significantly longer.

• Literary works: A social picture.
  – İnce Memeds – words become longer: A reflection of change in Turkish.
Conclusion ...

• Objective style studies can be used to support/oppose thesis of literary scholars.
  – Support: İnce Memed
  – Oppose: Being different from different/: Gizli Kalmış Bir İstanbul Masalı
Conclusion ...

• Can be used to hypothesize new claims.
Other Questions

• *İnce Memed* Sentiment Analysis: Always the same flow?

• Text to Audio Sentiment Reflection: Is it possible?

• Ali Teoman - Aimed different style in each work: Successful?

• *Kara Kitap*: Did Galip become Celâl with time?

• “Success with Style”: Is it possible to predict commercial success of Turkish literary works?
Thank You