

Things Hidden Behind The Numbers of İnce Memeds

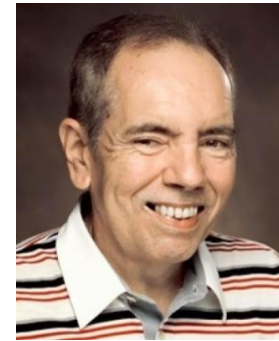
Bilkent Library Lunchtime Lecture
March 5, 2015

Fazlı Can

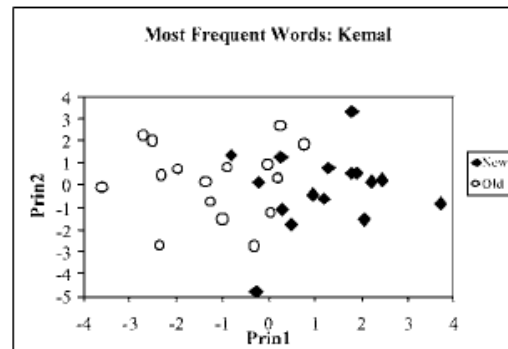
**Computer Engineering Department
Bilkent University**



“Change of Writing Style with Time” (2001)

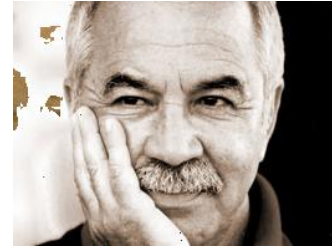
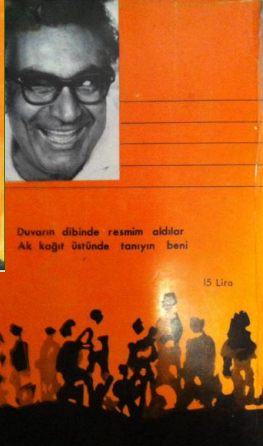
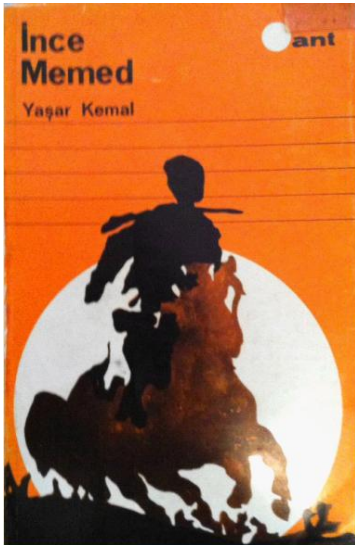


Jon M. Patton
Miami University, OH



Süha Oğuzertem
Bilgi Üniversitesi

Ince Memed (1969)

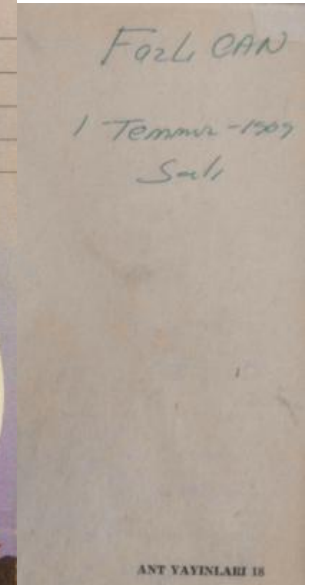
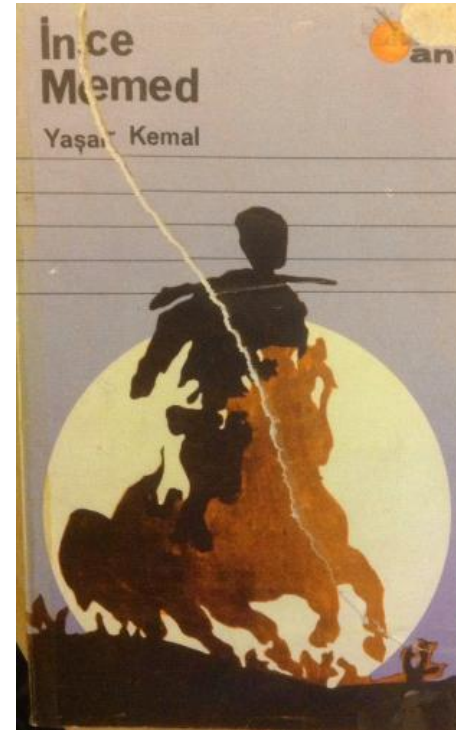
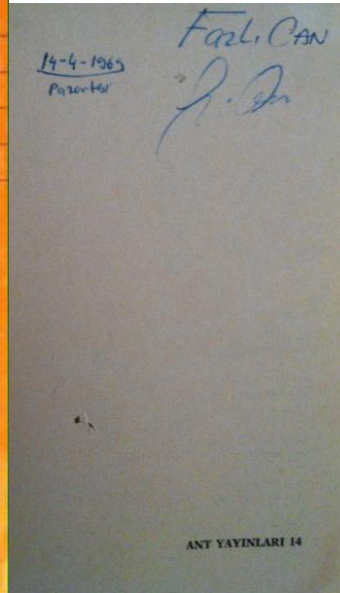
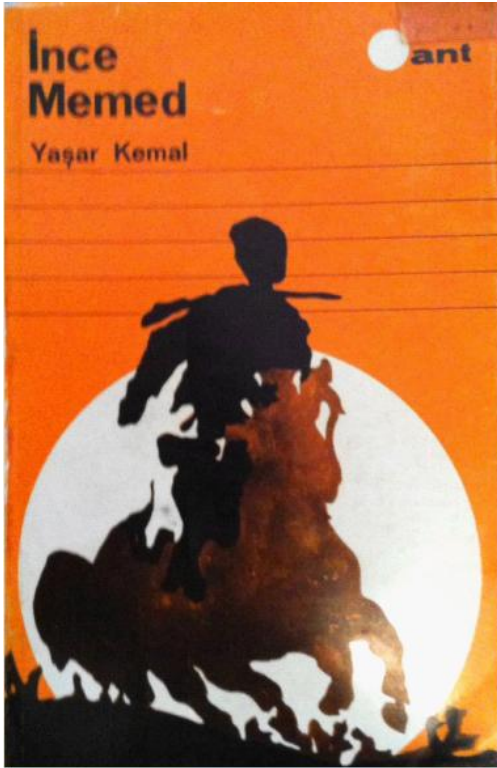


Erdal Öz

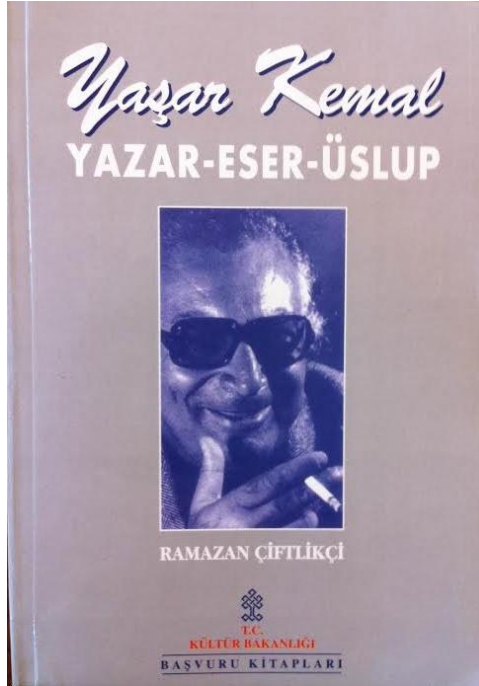


**Büyük Sinema
Kızılay Ankara**

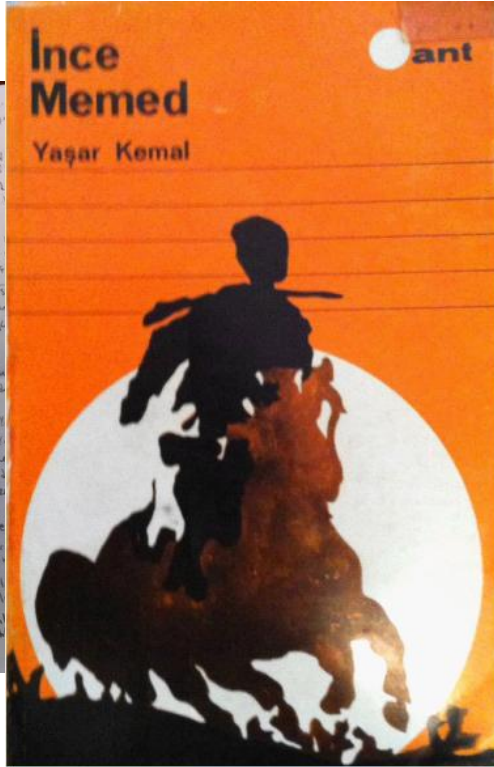
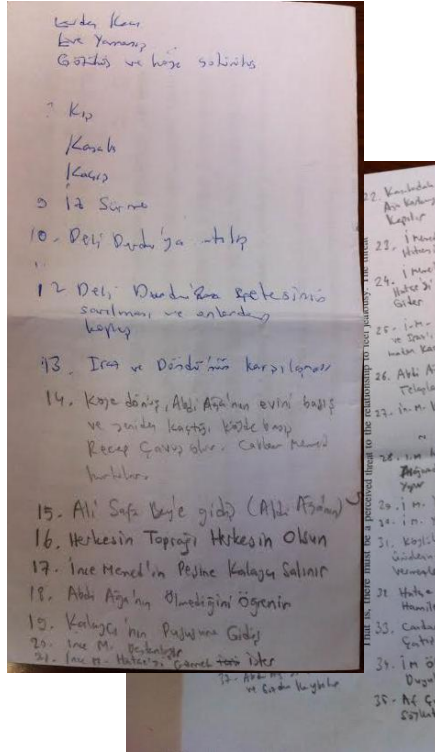
İnce Memed... (1969)



İnce Memed... (1954)



İnce Memed... Digital! (2001)



Engin Demir
THK Üniversitesi

“İnce Memed – Köy ve Kasaba Arasında”

Sami Arpa (2014)



Sami Arpa
EPFL

“İnce Memedlerin *Sayılarında Gizlenenler*” Yaşar Kemal Onur Günü Bilgi Üniversitesi (2014)



Yaşar Kemal Fahri Doktora Töreni ve Sempozyum

güncelleme: 04.11.2014

Tarih: 12 Kasım 2014, Çarşamba

Saat: 10.00-20.00

Yer: santralistanbul Kampüsü, ÇSM-403/404

İstanbul Bilgi Üniversitesi olarak Yaşar Kemal'le onurlanacağımız, fahri doktora unvanı tevcih edeceğimiz bu güne tüm BİLGİ'liler davetlidir.

PROGRAM:

10.00-18.00 Sempozyum: Binbir Kültürün Elçisi Yaşar Kemal

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



Zeki Müren



Antonio Vivaldi

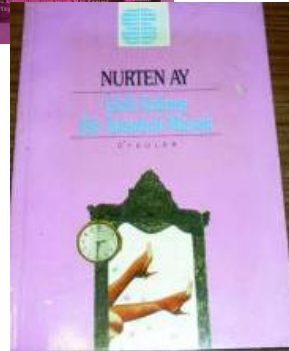
A Literary Hoax – Süha Oğuzertem, 2003

Gizli Kalmış Bir İstanbul Masalı

An İstanbul Fairy Tale that Remained Secret



Nurten Ay



1991



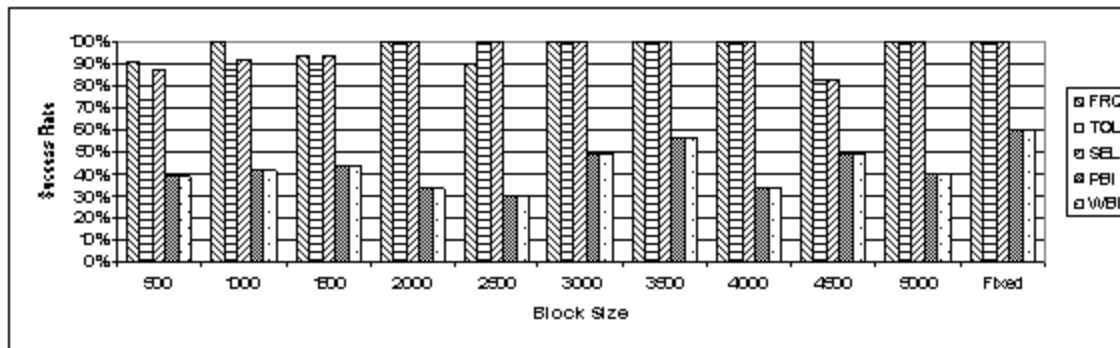
Cem Behar

Nurten Ay – Cem Behar: Are they similar?

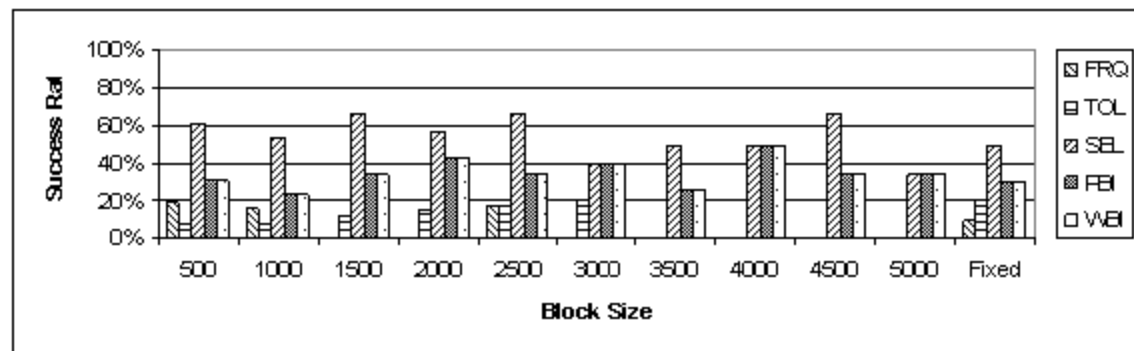
Machine Learning for Authorship Attribution: A Book is in Search for an Author

F. Can, Y. Dedeoğlu, M. Koç, M. Tekkalmaz, S. Oğuzertem; 2006 (unpublished).

Cem Behar:Cem Behar



Nurten Ay:Cem Behar



“A Literary Hoax” 2007

Gizli Kalmış Bir İstanbul Masalı

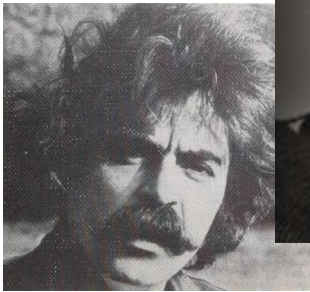


Ali Teoman

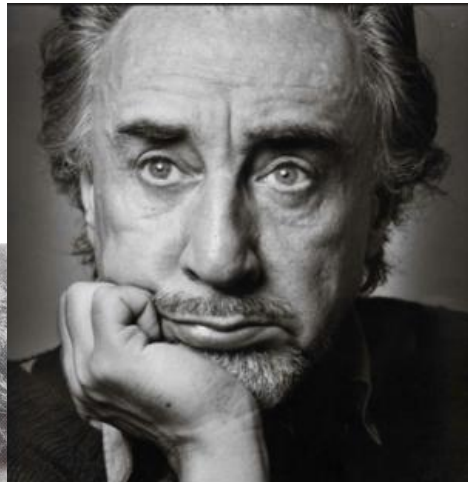


Is text mining always successful?

La Vie Devant Soi – The Life Before Us

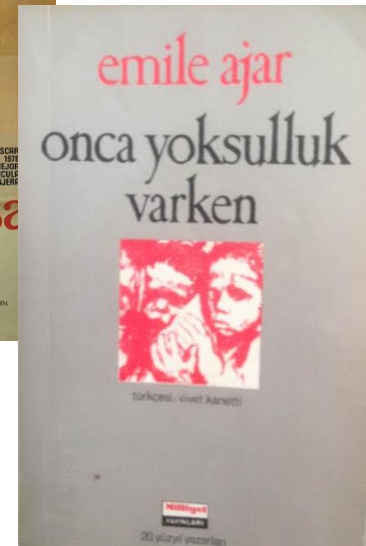


Émile Ajar



Romain Gary

2 Goncourt Awards



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:

Anavarza ovasının güneyinden Ceyhan ırmağı geçer. Irmak Hemite dağından Anavarza kayalıklarına kadar öyle büyük kıvrıntılar yapmadan döz iner. Bazı yerlerde sular toprağı derinden oymuştur. Altı oyulmuş toprak zaman zaman büyük gümbürtülerle suyun üstüne çöker. Bazı yerlerde kılıçla kesilmişcesine suya inen dik yarlar, çöküntülerden dolayı diş diş olur, ırmağın kıyılarında küçük, kumlu koyucuklar yapar. Bazı yerlerde de ırmak ovaya çakıltaşlarını sererek, geniş, yayılır. Buralarda sığ suyun aydınlık dibinden binlerce iri sazın balığı ışığa batmış, ardı ardına tırkenmiş, oradan oraya dalgalanarak kayar. Bir de ırmağın kıyılarında küçük sazlıklar vardır. Sazlıklarda çok iri yeşil kurbağalar, bulut
rengi, uzun boyumlu balıkçılar dolaşırlar. İlgün ağaçları, hayıtlar, söğütler, zincarlar, böğürtlenler yer yer kıyıları örерler. Sarıca arılar, kırmızı eşek arıları, boncuklu arılar, mavi yoz anlar peteklerini sıcakta büyütürler. Arı vizalıtısından bükük uğıldar. Ve sert bir böcek sırtına benzeyen uzun gagalarıyla mavi, yanar döner pırıltılı gurruk kuşları gelir kılıç gibi kesilmiş yarlara ince, uzun deliklerini günlerce oyar, yuvalarını ta toprağın dibine, derinine yaparlar. Hemite dağından Anavarza kalesine kadar, suyun bu kıyısında, yani Anavarza kalesi gecesinde, Hemite, Orhaniye, Selimiye, Endel, Kesikelli köyleri vardır. Irmak zaman zaman yatak değıştirdiğinden bazan bu köylerden uzaklaşır, ta ötelere gider, bazan da evlerin içlerine kadar girer. Ulu sellerin köylerin yanına alıp götürdüğü zamanlar da olur. Ceyhan Anavarzanın yanında geniş bir büvet yapar. Anavarza kalesinin dibi büyük bir göl gibi olur, suları durmadan döner. Büvetin suları yüzlerce yerde küçük küçük, birer hortum gibi çukurlaşp kabarak, bağdöndürücü bir hızla ince ince köpüklenerek durmadan döner dururlar. Büvete düşmüş bir dal parçası, bir yaprak, hiçbir yere akıp gitmeden, büvetten büvete geçerek suyun yüzünde dönerler. Ve sulara kelebek düştüğünde, suların üstünde yüzbinlerce kelebek uçtuğunda, bir insandan daha iri yayın balıkları geniş bıyıklı ağızlarını açarak suyun yüzüne fırlarlar. Bir kelebek yığınını kapıp sulara geri dönerler. Bu sıralar su sarı köpükten olur.
Anavarza kalesinin kayalıkları kuzeyden güneye uzanmış bir gemiye benzer. Üstündeki eskimiş, dökülmüş örenleri, yıkıntılarıyla. Anavarza gemisi her zaman durgun bir denizde hiç sallanmadan ağır ağır ilerler. Yüce, sarp, mor Anavarza kayalıklarına çıkıp yönümüzü gündoğuya dönersek, ilk olarak başı dumanlı Hemite dağını görürüz. Eđer tanyerleri ışımışsa, dumanlar yücelere doğru ağımışsa, Hemite dağının tepesindeki top ağaçları, ermiş türbesini de görürüz. Hemite dağından sonra tepeler yassılır. Bozkuyu köyünün yöreleri çıplaktır. Boz topraklardır.
Uzaklardan apak gözükür. Çiğcik köyünün oralarda toprak değışir, kararır, ormanın yeşilli başlar, ortalık çiçeklenir. Çiğcıkta tarlaların ortasında, bir tarla Bizans mozayığı vardır. Yabanıl, büyük bir bahçe gibi ovanın ortasında çiçeklenir. Kuzeyde Kadirli, Kadirli'de Sülemiş tepesi... Sülemiş tepesinde tek tük mersin çalıları, ağır kokulu. Sülemiş tepesi çıplaktır ama hep yeşildir. Dibinden Savrun çayı kaynayarak ovaya iner. Anavarza ovasının kuzey batısı Kozan yolu... Sumbas çayı Kozan yolunu ikiye biçip Anavarza kayalıklarının önüne gelir. Sumbasın öte gecesı, Anavarza kalesinin batıdan yüzü. Hacılar kövü... Onun ötesinde de Aslanlı kövü... Aslanlı

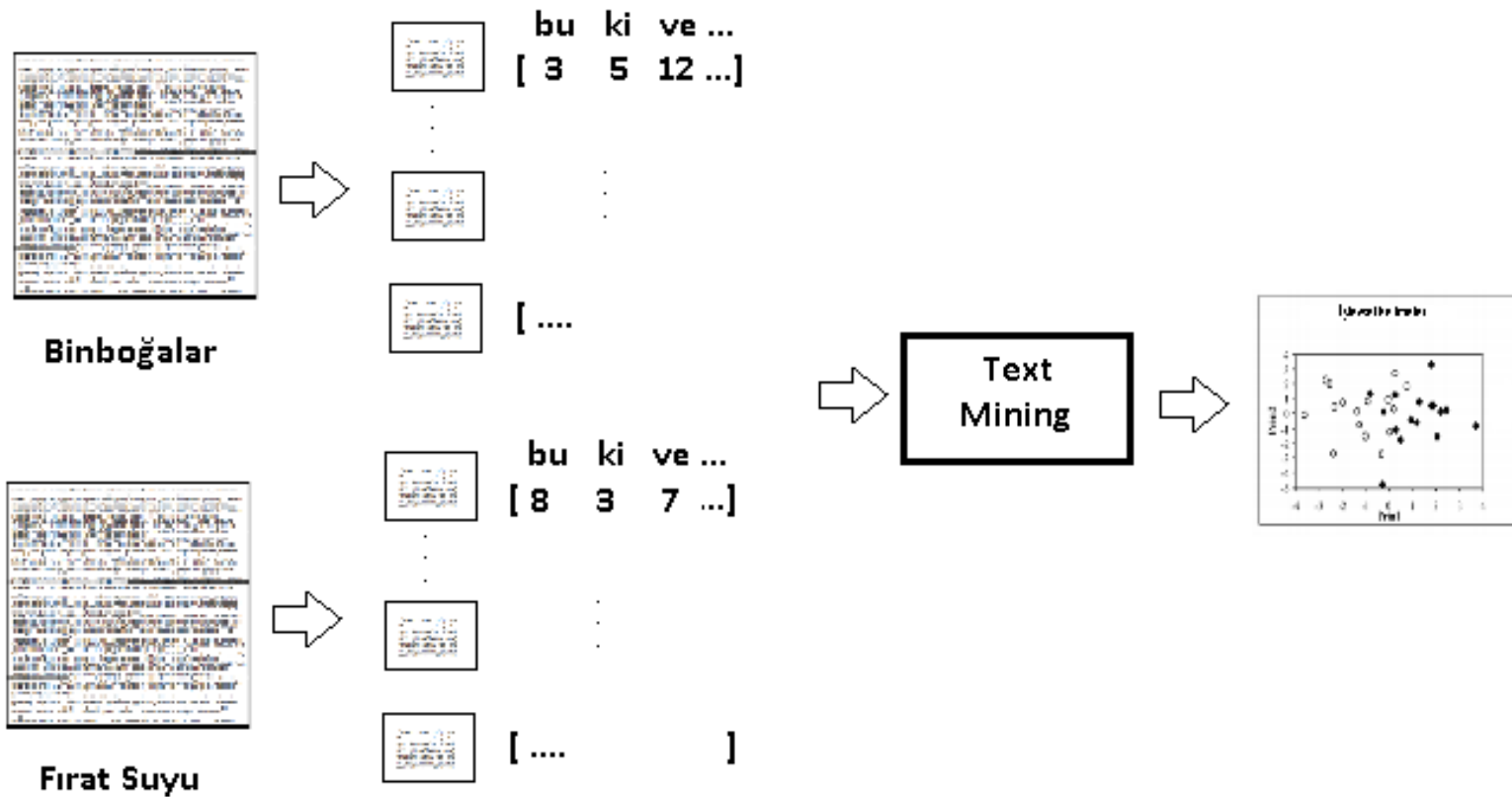
Define Text with Objective Style Markers

- 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



Change of Style with Time

Yaşar Kemal: Binboğalar Efsanesi/Fırat Suyu Kan Akıyor Baksana 1971/1998
Çetin Altan: Columns 1960-1969/2000



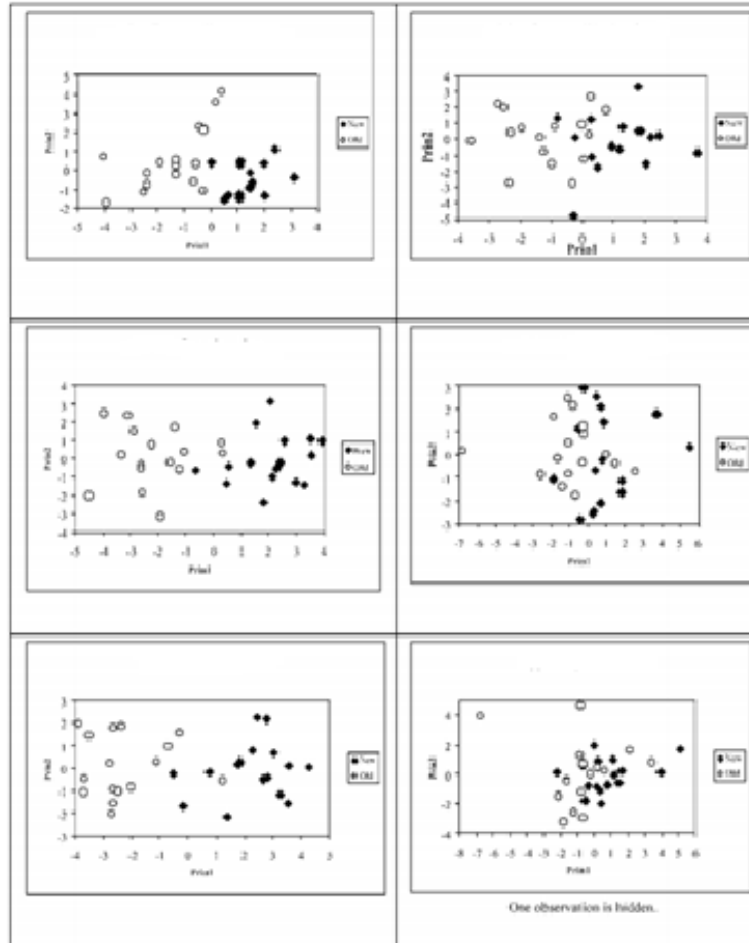
Çetin Altan ve Yaşar Kemal

Distinction of Old & New

Can, F., Patton J.M. Change of Writing Style with Time. *Computers and the Humanities*, 2004.

Çetin Altan

Yaşar Kemal

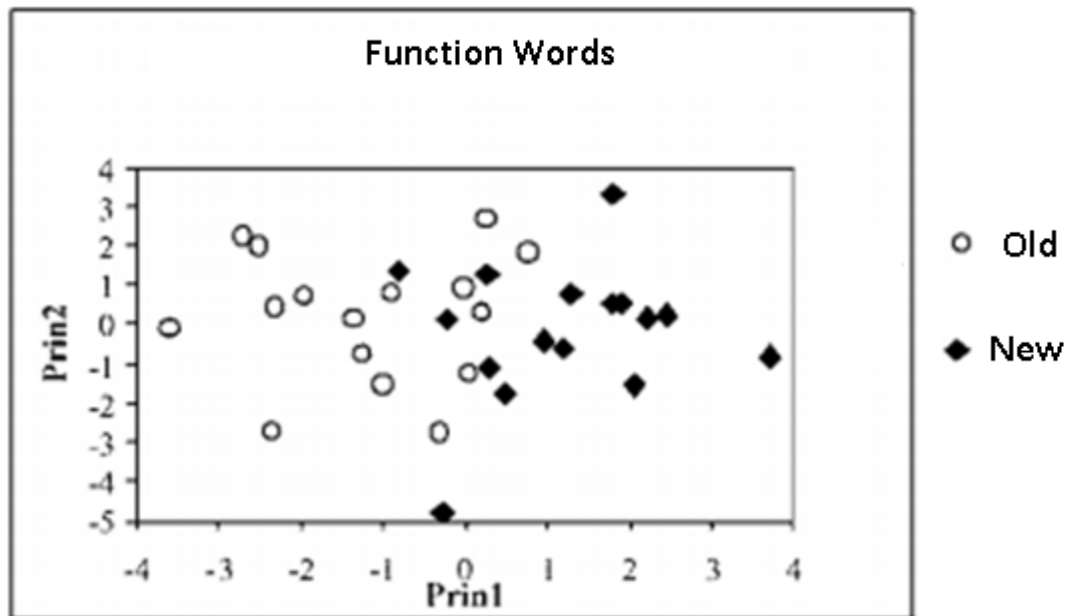


Function Words

Token Lengths

Type Lengths

Yaşar Kemal: Distinction of Old & New *Binboğalar...* 1971/*Fırat Suyu...* 1998



Old or New?

Binboğalar... – Fırat Suyu ...

Success Rate

New	Old		
81.25	75.00		
(13)	(12)	←	Token Lengths
75.00	81.25		
(12)	(13)	←	Type Lengths
93.75	100.00		
(15)	(16)	←	Functional Words



Each Novel is 16 Blocks
 $16/16 = 100\%$ 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

Can, F., Patton J.M A Stylometric Analysis of Yaşar Kemal's *İnce Memed* Tetralogy. *Computers and the Humanities*, 2004 and A Detailed Stylometric Investigation of the *İnce Memed* Tetralogy. Tech. Report, Miami Univ. 2004.

- 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

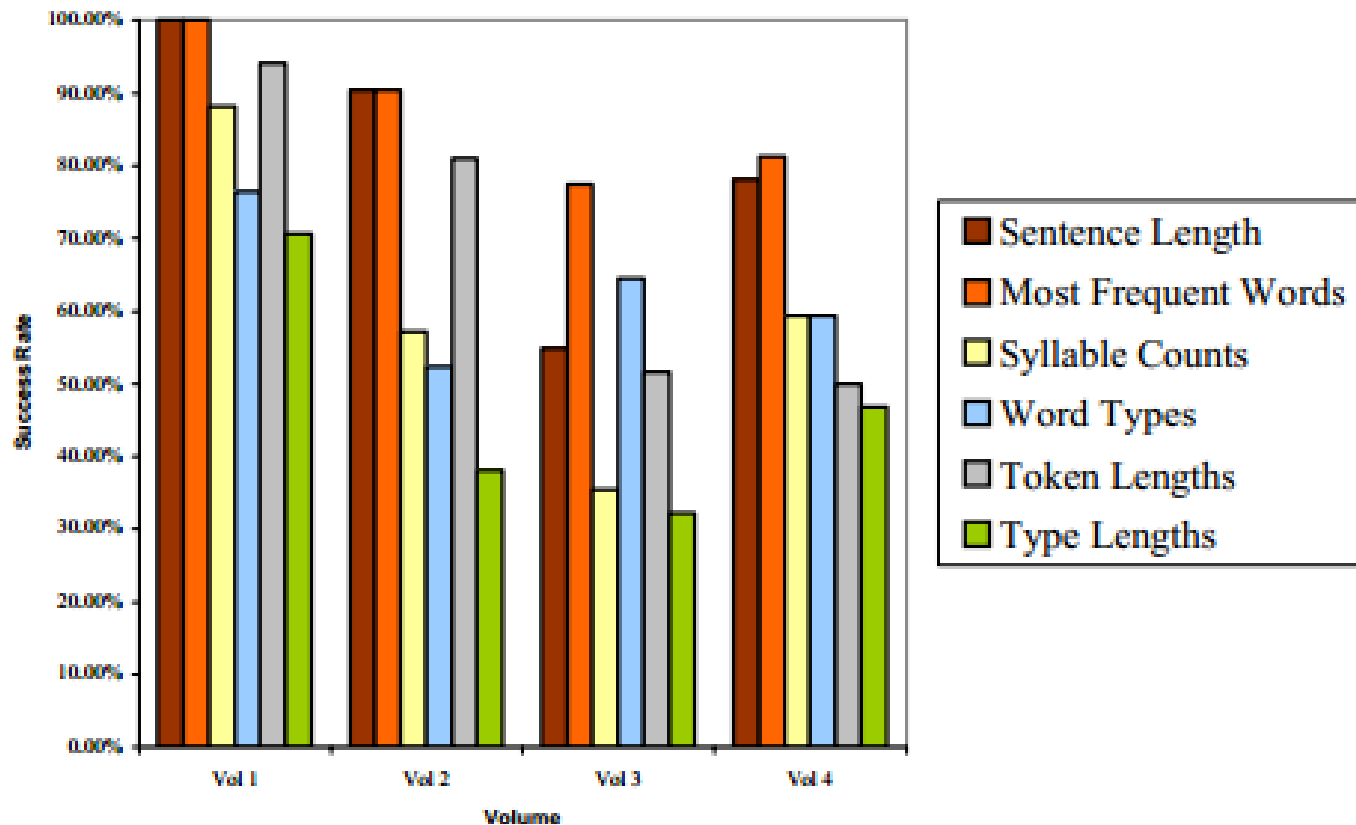
Ince Memed Tetralogy

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

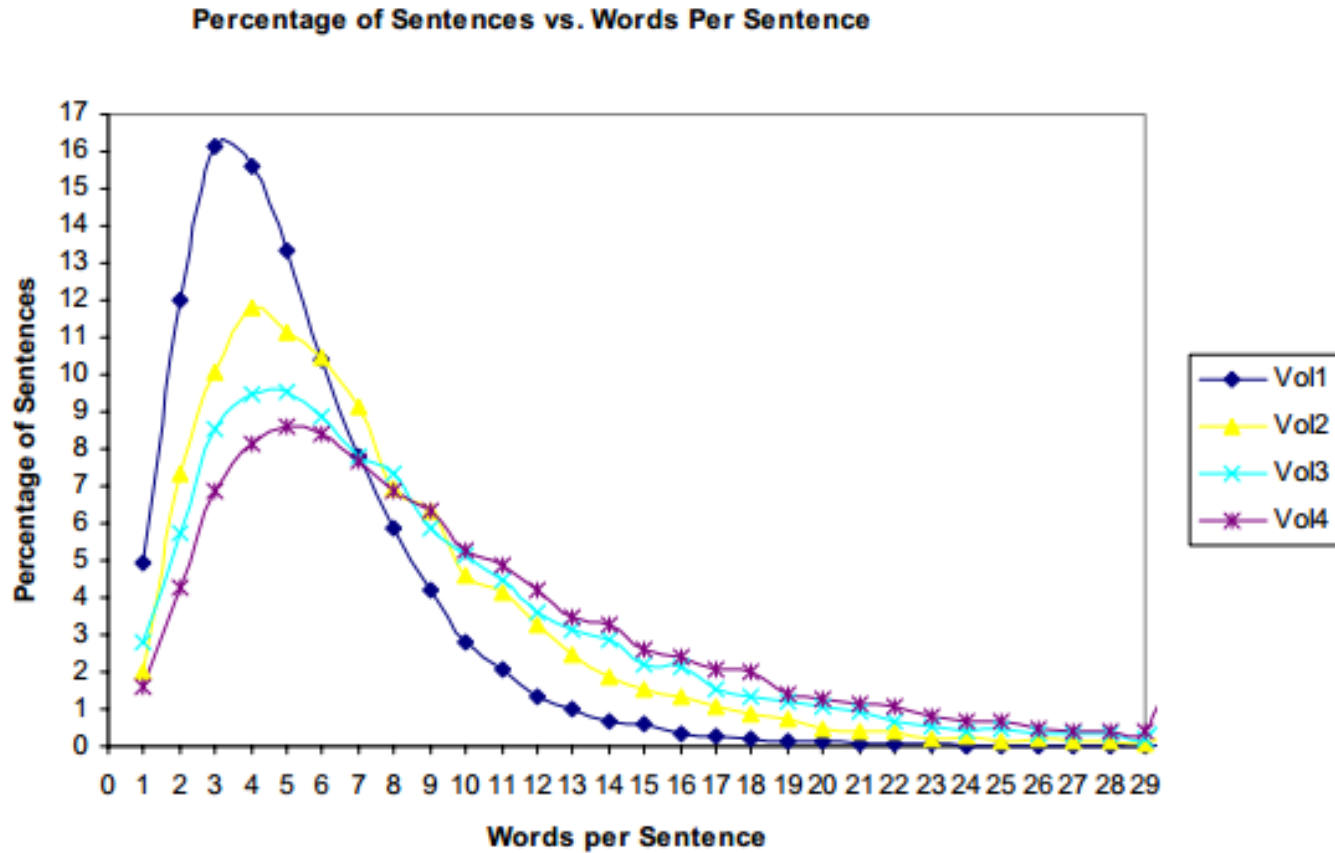
Ince Memed Tetralogy

- Analysis with six style markers
 - Sentence length
 - Most frequent (function) words
 - Syllable count
 - Word types (noun, ...)
 - Token length
 - Type length

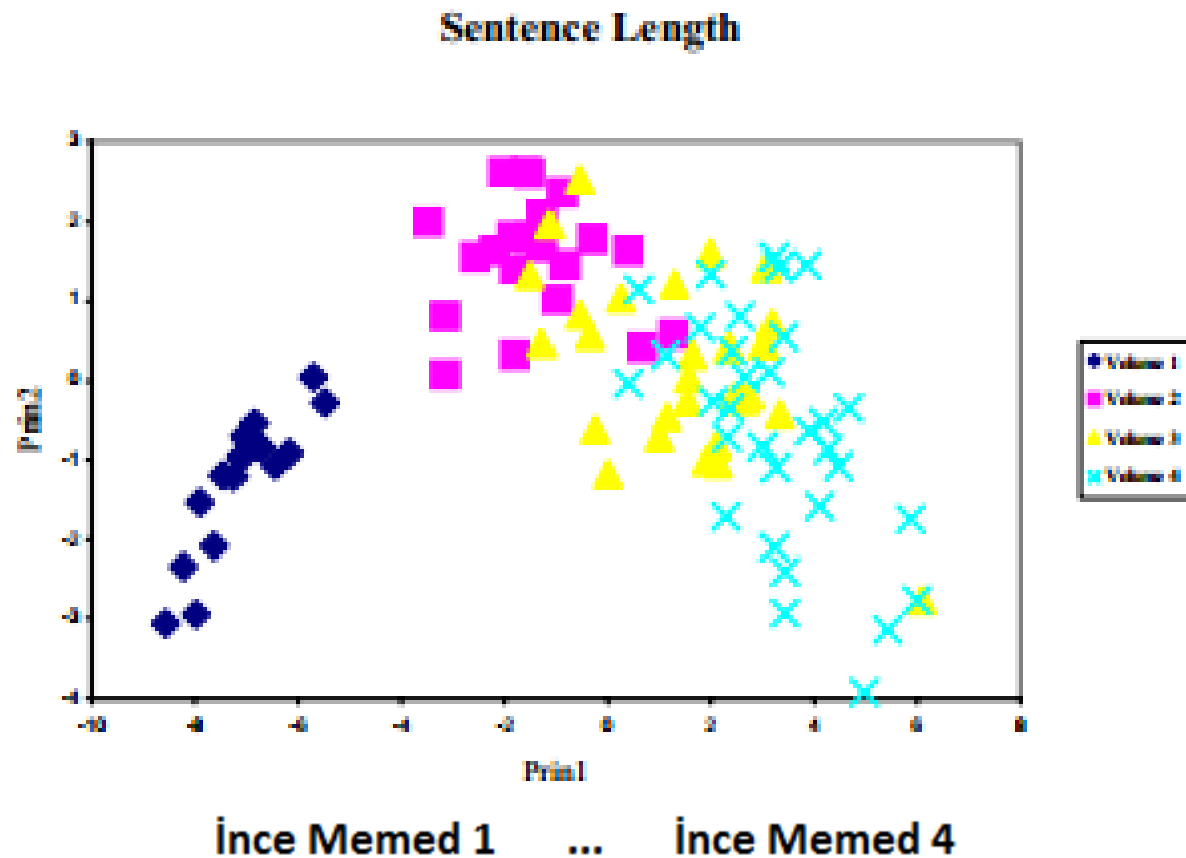
Correct Classification Rates



Sentence Length Changes



Sentence Lengths



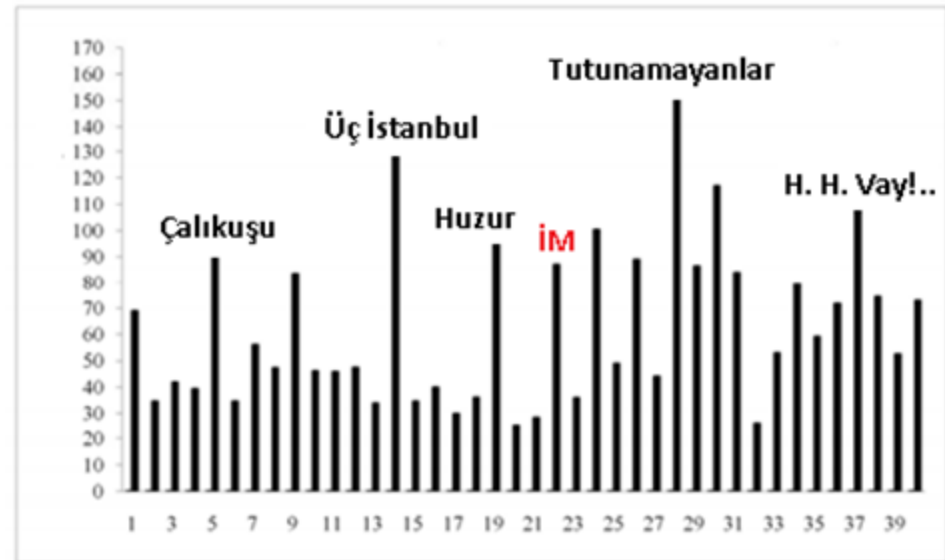
New Words: Longer

Volume	Token Length		Type Length	
	Mean	Std. Dev.	Mean	Std. Dev.
1	5.80	.081	7.03	.082
2	5.85	.113	7.18	.116
3	5.81	.143	7.20	.128
4	5.91	.161	7.25	.111

20th Century Turkish Novel

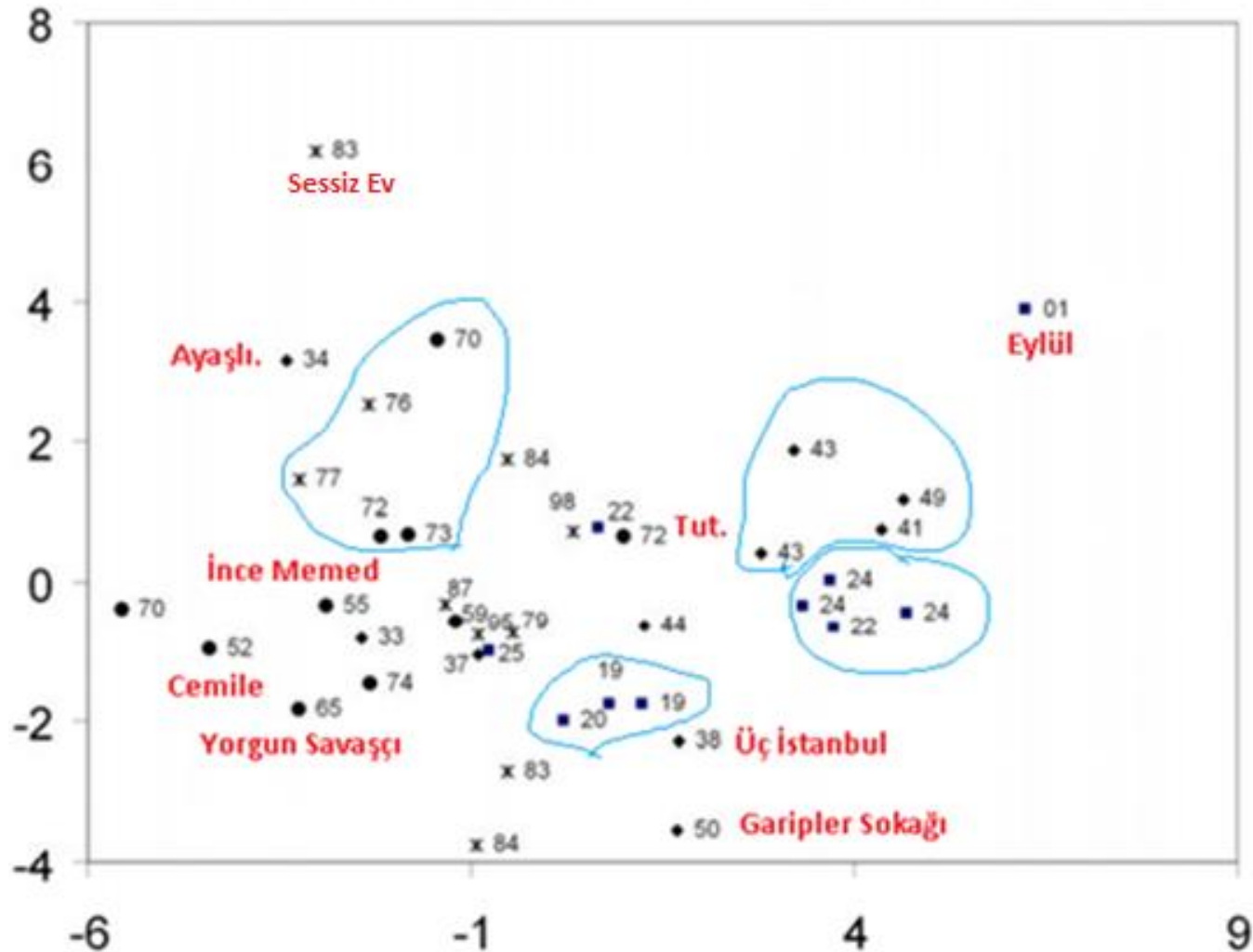
Can, F., Patton, J. M. Change of word characteristics in 20th century Turkish literature: A statistical analysis. *Journal of Quantitative Linguistics*, 2010.

1	Mehmet Rauf, <i>Eylül</i> , 1901
2	Hüseyin Rahmi Gürpınar, <i>Toraman</i> , 1919
3	Ömer Seyfettin, <i>Efruz Bey</i> , 1919
4	Refik Halit Karay, <i>İstanbul'un Bir Yüzü</i> , 1920
5	Reşat Nuri Güntekin, <i>Çalıkuşu</i> , 1922
6	Yakup Kadri Karaosmanoğlu, <i>Nur Baba</i> , 1922
7	Halide Edip Adıvar, <i>Kalb Ağrısı</i> , 1924
8	Salâhattin Enis, <i>Zaniyeler</i> , 1924
9	Halit Ziya Uşaklıgil, <i>Kırık Hayatlar</i> , 1924
10	Peyami Safa, <i>Sözde Kızlar</i> , 1925
11	Mahmut Yesari, <i>Tipi Dindi!</i> , 1933
12	M. Şevket Esendal, <i>Ayaşlı İle Kiracıları</i> , 1934
13	Cahit Uçuk, <i>Dikenli Çit</i> , 1937
14	Mithat Cemal Kuntay, <i>Üç İstanbul</i> , 1938
15	Abdülhak Şinasi Hisar, <i>Fahim Bey ve Biz</i> , 1941
16	Sabahattin Ali, <i>Kürk Mantolu Madonna</i> , 1943
17	Kemal Bilbaşar, <i>Denizin Çağırışı</i> , 1943
18	Sait Faik, <i>Medarı Maişet Motoru</i> , 1944
19	Ahmet Hamdi Tanpınar, <i>Huzur</i> , 1949
20	Oktay Akbal, <i>Garipler Sokakı</i> , 1950
21	Orhan Kemal, <i>Cemile</i> , 1952
22	Yaşar Kemal, <i>İnce Memed</i> [1], 1955
23	Yusuf Atılgan, <i>Aylak Adam</i> , 1959
24	Kemal Tahir, <i>Yorgun Savaşçı</i> , 1965
25	Tarık Buğra, <i>İbiş'in Rüyası</i> , 1970
26	Fakir Baykurt, <i>Turpan</i> , 1970
27	Çetin Altan, <i>Büyük Gözaltı</i> , 1972
28	Oğuz Atay, <i>Tutunamayanlar</i> , 1972
29	Adalet Ağaoğlu, <i>Ölmeye Yatmak</i> , 1973
30	Füruzan, <i>Kırkyedililer</i> , 1974
31	Pınar Kür, <i>Yarın Yarın</i> , 1976
32	Ferid Edgü, <i>O: Hakkari'de Bir Mevsim</i> , 1977
33	Selim İleri, <i>Ölüm İlişkileri</i> , 1979
34	Orhan Pamuk, <i>Sessiz Ev</i> , 1983
35	Latife Tekin, <i>Sevgili Arsız Ölüm</i> , 1983
36	Mehmet Eroğlu, <i>İssızlığın Ortasında</i> , 1984
37	Attilâ İlhan, <i>Hacı Hanım Vay!...</i> , 1984
38	Kaan Arslanoğlu, <i>Devrimciler</i> , 1987
39	Nedim Gürsel, <i>Boğazkesen: Fatih'in Romanı</i> , 1995
40	Ahmet Altan, <i>Kılıç Yarısı Gibi</i> , 1998



20th Century Turkish Literature

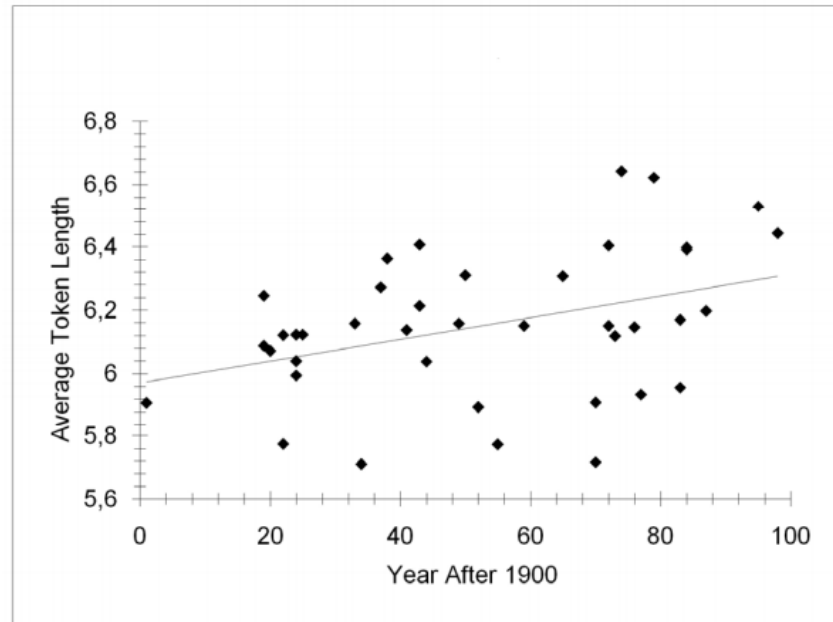
40 Novels: 10 Novels/Quarter Century



Token Length: Longer with Time

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

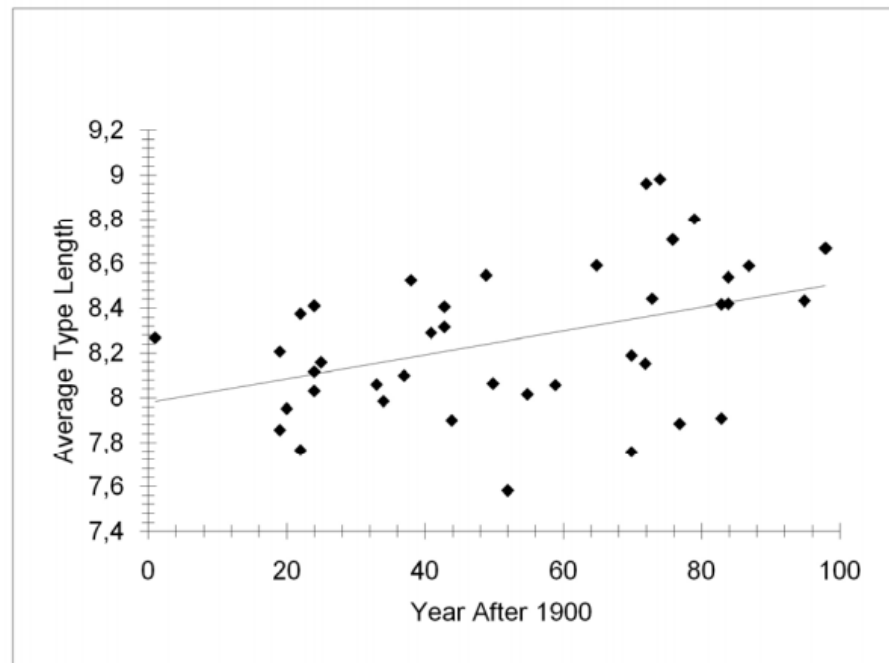
- 1900: 5.9741 letters
- 2000: 6.3421 letters
- Increase in 100 Years:
0.368 Letters



Type Length: Longer with Time

$$\text{Avg. Word Length in Dictionary (Year)} = 8.0127 + 0.00609 * (\text{Year} - 1900)$$

- 1900: 8.0127 letters
- 2000: 8.6217 letters
- Increase in 100 Years:
0.609 Letters



Language Change Quantification: Using Time-Separated Parallel Translations

Altintas, K., Can, F., Patton, J. M., "Language change quantification using time-separated parallel translations." *Literary & Linguistic Computing*. Vol. 22, No. 4 (November 2007), pp. 375-393.



1957 - 1999

1954 - 1999



1944 - 1999



1944 - 1999

1946 - 1999

1943 - 1999



1947 - 2002

Study Based on Time-Separated Translations

Altintas, K., Can, F., Patton, J. M., "Language change quantification using time-separated parallel translations." *Literary & Linguistic Computing*. Vol. 22, No. 4 (November 2007), pp. 375-393.

Alphonse Daudet

Değirmenimden Mektuplar (Letters from my Windmill)

DM-1944: Sabri Esat Sivayuşgil, Milli Eğitim, Ankara, 1989 ¹.

DM-1990: Rabia Ergüven, İnkilap Kitabevi, İstanbul, 1990.

Fyodor Dostoyevsky

Beyaz Geceler (White Nights)

BG-1957: Nihal Yalaza Taluy, Varlık Yayınları, İstanbul, 1957.

BG-1999: Mehmet Özgül, Cumhuriyet Gazetesi, İstanbul, 1999.

Uysal Kız (The Gentle Maide)

UK-1954: D. Sorakin, S. Aytekin, Maarif, Ankara, 1954.

UK-1999: Mehmet Özgül, Cumhuriyet Gazetesi, İstanbul, 1999.

Michel de Montaigne

Denemeler (Essays) ²

D-1947: Sabahattin Eyüboğlu, Milli Eğitim, Ankara, 1947.

D-2002: Celal Öner, Oda Yayınları., İstanbul, 2002.

William Shakespeare

Hamlet

H-1944: Orhan Burian, Maarif, Ankara, 1944.

H-1999: Bülent Bozkurt, Remzi Kitapevi, İstanbul, 1999.

Macbeth

M-1946: Orhan Burian, Milli Eğitim, Ankara, 1946.

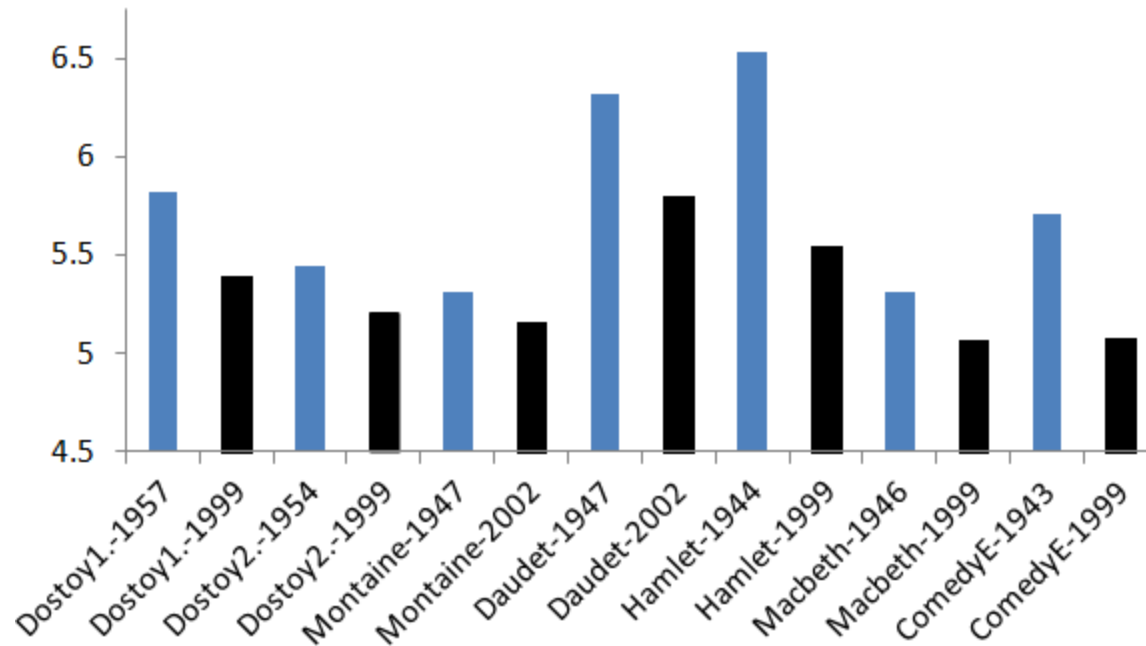
M-1999: Orhan Burian (Edited by Publisher), Cumhuriyet Gazetesi, İstanbul, 1999.

Yanlışlıklar Komedyası (Comedy of Errors)

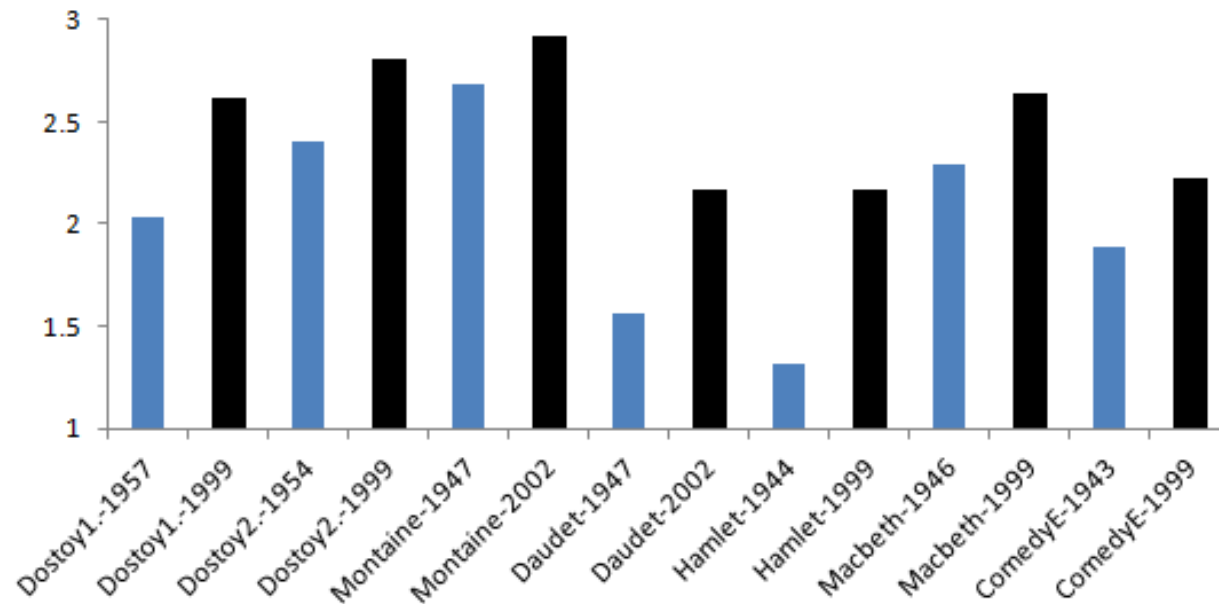
YK-1943: Avni Givda, Maarif, Ankara, 1943.

YK-1999: Bülent Bozkurt, Remzi Kitapevi, İstanbul, 1999.

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

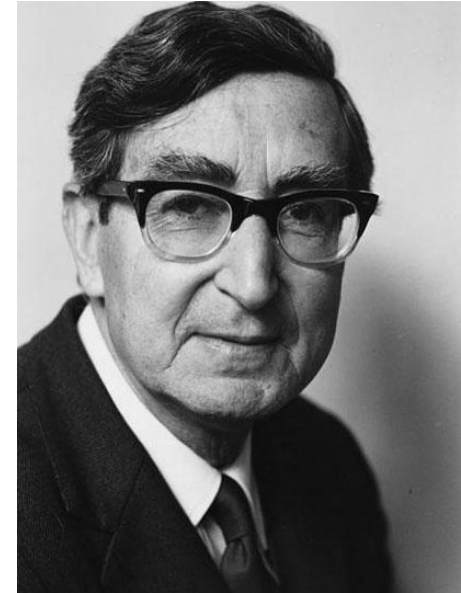
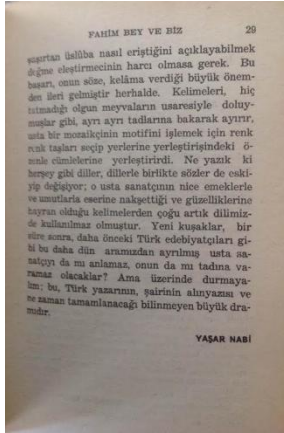
- Words: Longer
- Stems: Shorter
- Suffixes: Longer
- Meanings per stem: More
- Vocabulary richness: Lower

Fahim Bey ve Biz, Fahim Bey and Us **Abdülhak Şinasi Hisar, 1941** **(Varlık 1966 ed., published by Yaşar Nabi)**



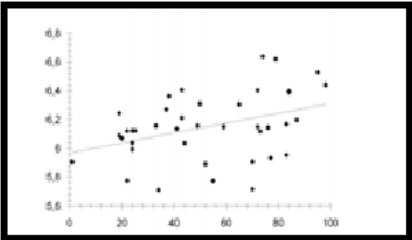
***Fahim Bey ve Biz* (Varlık ed. 1966)**

Foreword, Yaşar Nabi , p. 29



**"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**

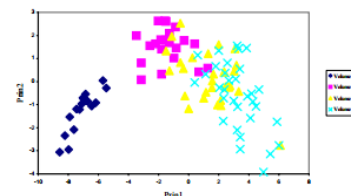
Conclusion

- Text mining: Works well as expected.
- Turkish words: Significantly longer.

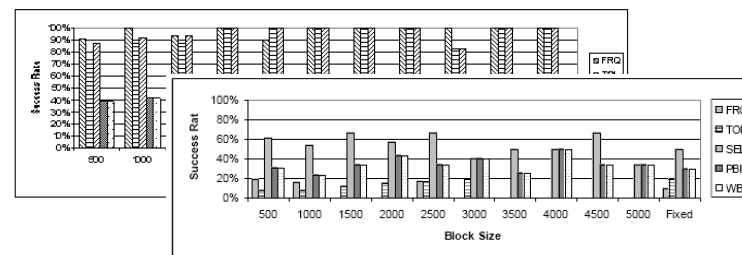
A scatter plot with a positive linear regression line. The x-axis ranges from 0 to 100 with major ticks every 20 units. The y-axis ranges from 0.0 to 0.6 with major ticks every 0.2 units. There are approximately 25 data points represented by small black squares. The regression line starts at approximately (0, 0.05) and ends at approximately (100, 0.35).
- 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*
S. Oğuzertem “1: romantic, 2: realist, 3/4: postmodern.”



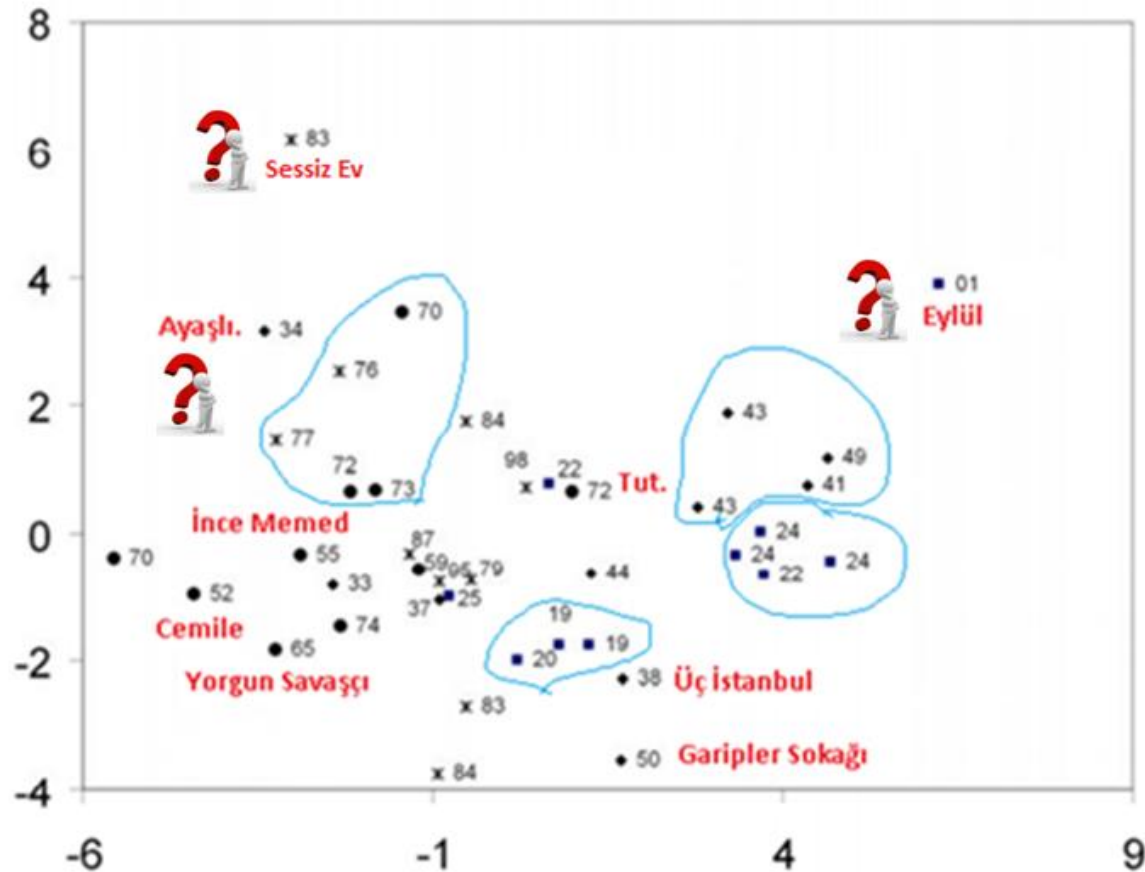
- Oppose: Being different from different/: *Gizli Kalmış Bir İstanbul Masalı*



(a) Bayesian

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