



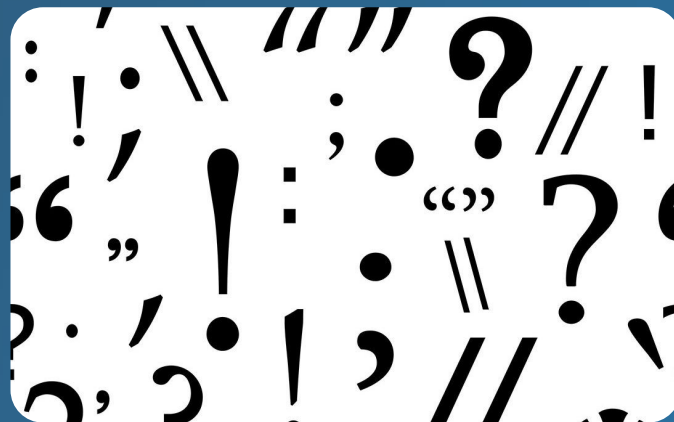
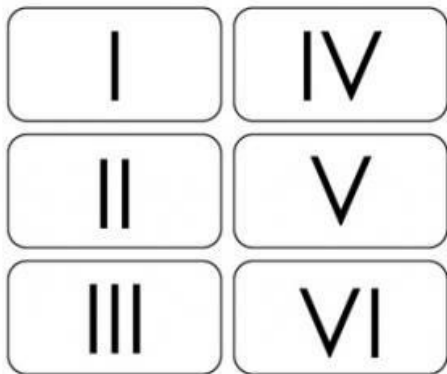
# Outline

- Introduction
- Preprocessing The Data
- Collocations Extraction
- Comparison Of Methods
- Summary

# Introduction

- Aim of the project
- Describing the dataset
- The produced results

# Preprocessing The Data



# Collocations Extraction Methods

1 Collocations By Frequency

2 Mutual Information

3 Hypothesis Testing : T-test

## Collocation By Frequency

Donem - 20

C(w1;w2)	collocation	tag pattern
15135	"('B�y�k', 'Millet')"	AN
11883	"('Millet', 'Meclisi')"	NN
10906	"('soru', '�nergesi')"	NN
9785	"('sıralarından', 'alkışlar')"	NN
7968	"('Meclisi', 'Başkanlığına')"	NN
7759	"('yazılı', 'soru')"	AN
6166	"('arz', 'ederim')"	NN
6150	"('Devlet', 'Bakanı')"	NN
5794	"('Kabul', 'edenler')"	NN
5334	"('yazılı', 'cevabı')"	AN

## Collocation By Frequency

Donem - 21

C(w1;w2)	collocation	tag pattern
18189	"('sıralarından', 'alkışlar')"	NN
17993	"('Büyük', 'Millet')"	AN
16761	"('soru', 'önergesi')"	NN
13162	"('Millet', 'Meclisi')"	NN
9978	"('yazılı', 'soru')"	AN
9381	"('BAŞKAN', 'Teşekkür')"	AN
9217	"('önergesi', 'Başkanlığa')"	NN
9111	"('değerli', 'milletvekilleri')"	AN
8976	"('Türkiye', 'de')"	AN
8805	"('Kabul', 'edenler')"	NN

## Collocation By Frequency

Donem - 22

C(w1;w2)	collocation	tag pattern
90229	"('le', 'ri')"	NN
26526	"('Baş', 'kan')"	NN
25384	"('me', 'si')"	NN
25252	"('BAŞ', 'KAN')"	NN
24955	"('soru', 'önergesi')"	NN
21252	"('so', 'ru')"	NN
19452	"('yazılı', 'soru')"	AN
19363	"('ko', 'nu')"	NN
19116	"('Kabul', 'edenler')"	NN
18576	"('öner', 'ge')"	NN



## Collocation By Mutual Information

Donem - 23

$I(w_1;w_2)$	$C(w_1)$	$C(w_2)$	$C(w_1;w_2)$	collocation
14.05182413349557	1178	1202	1178	"('SUHA', 'OKAY')"
13.808325367680274	1423	1423	1423	"('SIRRI', 'SAKIK')"
13.803227605298126	1427	1377	1376	"('KÂTÎP', 'ÜYELER')"
13.697143864515985	1537	1529	1529	"('BEKÎR', 'BOZDAĞ')"
13.658904371971014	1507	1572	1501	"('Mehdi', 'Eker')"
13.541201235215599	1401	1666	1363	"('NEVZAT', 'KORKMAZ')"
13.515575918036705	1681	1571	1515	"('SUAT', 'KILIÇ')"
13.514463241289054	1743	1178	1177	"('HAKKI', 'SUHA')"
13.512796558994163	1708	1361	1331	"('Ad', 'yaman')"
13.453973342110512	1498	1450	1194	"('Erkan', 'Akçay')"

## Collocation By Mutual Information

Donem - 24

$I(w1;w2)$	$C(w1)$	$C(w2)$	$C(w1;w2)$	collocation
13.826911353605638	1565	1589	1564	"('TABİİ', 'KAYNAKLAR')"
13.761212952888028	1587	1662	1585	"('SATAŞMALARA', 'İLİŞKİN')"
13.749916870930475	1571	1548	1450	"('Lütfi', 'Baydar')"
13.734437313554585	1639	1657	1602	"('Mülkiye', 'Birtane')"
13.723622299585225	1707	1664	1663	"('BEKİR', 'BOZDAĞ')"
13.653132918217107	1789	1576	1572	"('Kadir', 'Gökmen')"
13.5898199164518	1737	1696	1572	"('Veysel', 'Eroğlu')"
13.564337007371533	1532	1885	1514	"('Ruhsar', 'Demirel')"
13.563135520062849	1909	1897	1897	"('VELİ', 'AĞBABA')"
13.517706926766607	1844	1969	1843	"('Nedret', 'Akova')"

## Collocation By Mutual Information

Donem - 25

I(w1;w2)	C(w1)	C(w2)	C(w1;w2)	collocation
13.66212592790625	8	8	8	"('Zeynep', 'Altıok')"
13.492200926463937	9	9	9	"('Tacettin', 'Bayır')"
13.492200926463937	9	9	9	"('Dengir', 'Mir')"
13.340197833018886	10	10	10	"('ÖZDOĞAN', 'ENC')"
13.340197833018886	10	10	10	"('yazılmıştır', 'İÇİNDEKİLER')"
13.340197833018886	10	10	10	"('belirtilmiş', 'alıntı')"
13.340197833018886	10	10	10	"('Tarihinden', 'İtibaren')"
13.340197833018886	10	10	10	"('Maddesi', 'Uyarınca')"
13.340197833018886	10	10	10	"('GÖKÇEN', 'ÖZDOĞAN')"
13.340197833018886	10	10	10	"('BAŞKANLARININ', 'KONUŞMALARI')"

## Collocation By Hypothesis Testing : T - test

Donem - 26

t	C(w1)	C(w2)	C(w1;w2)	collocation
203.17708536748637	164122	53687	42428	"('Sayın', 'Başkan')"
176.19650568589998	35656	32089	31195	"('sıralarından', 'alkışlar')"
145.87541517692785	21409	21526	21340	"('AK', 'PARTİ')"
137.167515305873	32335	22460	18910	"('Teşekkür', 'ederim')"
135.13933431893773	38144	19499	18360	"('ilişkin', 'açıklaması')"
133.51450149519994	18834	18596	17872	"('Büyük', 'Millet')"
132.70648015924584	72688	18834	17790	"('Türkiye', 'Büyük')"
131.99920106300746	92864	164122	19370	"('BAŞKAN', 'Sayın')"
126.30829865352594	37019	17585	16039	"('Kabul', 'edenler')"
122.63328688995487	92864	32335	15430	"('BAŞKAN', 'Teşekkür')"

## Collocation By Hypothesis Testing : T - test

Donem - 27

t	C(w1)	C(w2)	C(w1;w2)	collocation
181.7100913956161	114123	44886	33804	"('Sayın', 'Başkan')"
175.92705641358512	35669	32160	31127	"('sıralarından', 'alkışlar')"
146.44625704675255	35210	22404	21568	"('ilişkin', 'açıklaması')"
146.19729168311767	21523	34192	21487	"('AK', 'PARTİ')"
130.4369755361739	17075	17173	17059	"('Mikrofon', 'otomatik')"
130.40879111573335	17173	17208	17052	"('otomatik', 'cihaz')"
130.28685017095134	31534	17362	17059	"('tarafından', 'kapatıldı')"
130.2283017124624	17208	31534	17043	"('cihaz', 'tarafından')"
122.82295497028802	34192	35669	15273	"('PARTİ', 'sıralarından')"
116.1243610697605	70584	114123	14701	"('BAŞKAN', 'Sayın')"

# Comparison Of Methods

- Collocation By Frequency
- Collocation By Mutual Information
- Collocation By T- test





## All Corpus ( All Donems ) - Collocation By Frequency

C(w1;w2)	collocation	tag pattern
31127	"('sıralarından', 'alkışlar')"	NN
21568	"('ilişkin', 'açıklaması')"	NN
17052	"('otomatik', 'cihaz')"	AN
17043	"('cihaz', 'tarafından')"	NN
13058	"('Büyük', 'Millet')"	AN
12071	"('CHP', 'sıralarından')"	NN
11810	"('Kanun', 'Teklifi')"	NN
11438	"('değerli', 'milletvekilleri')"	AN
11418	"('Teşekkür', 'ederim')"	NN
10473	"('Merkezî', 'Yönetim')"	AN

## All Corpus ( All Donems ) - Mutual Information

I(w1;w2)	C(w1)	C(w2)	C(w1;w2)	collocation
17.47851285606217	661	663	661	"('MÜSAVAT', 'DERVİŞOĞLU')"
17.43141773832687	685	684	684	"('KEMALBAY', 'PEKGÖZEGÜ')"
17.4188360157865	691	685	685	"('SERPİL', 'KEMALBAY')"
16.78504657166309	799	1044	778	"('Müsavat', 'Dervişoğlu')"
16.77889744304015	821	922	703	"('Metanet', 'Çulhaoğlu')"
16.661553357294743	765	1168	765	"('RUŞTU', 'TİRYAKİ')"
16.535096384428538	1275	661	661	"('DURSUN', 'MÜSAVAT')"
16.453703283226492	1349	722	722	"('YAVUZ', 'AĞIRALIOĞLU')"
15.660041245442947	2330	2214	2206	"('SARUHAN', 'OLUÇ')"
15.52379163293885	1465	2342	1335	"('Saruhan', 'oluç')"



## All Corpus ( All Donems ) - Hypothesis Testing : T - test

t	C(w1)	C(w2)	C(w1;w2)	collocation
174.91134475405363	193554	167438	31127	"('sıralarından', 'alkışlar')"
172.40712491797947	936663	272166	33804	"('Sayın', 'Başkan')"
145.89676977194608	125433	97304	21487	"('AK', 'PARTİ')"
145.73256584070825	351193	57109	21568	"('ilişkin', 'açıklaması')"
130.4053869332727	56258	57547	17059	"('Mikrofon', 'otomatik')"
130.37538924022556	57547	57125	17052	"('otomatik', 'cihaz')"
130.03143772992468	172060	53186	17059	"('tarafından', 'kapatıldı')"
129.92702396234773	57125	172060	17043	"('cihaz', 'tarafından')"
122.32534258536487	97304	193554	15273	"('PARTİ', 'sıralarından')"
112.84139755960638	138366	143014	13058	"('Büyük', 'Millet')"

# Summary

- The results found
- What's next?

Thank you  
for listening ! :)

