

**GTU Department of Computer Engineering  
CSE 484/654 Natural Language Processing  
Fall 2021 - Homework 3 Report**

**Akif KARTAL  
171044098**

## 1) Problem Definition

The problem is to correct Turkish sentences written using English letters only by using Turkish language model.

## 2) Solution

The homework was finished as expected in homework pdf file.

### 2.1) Finding $2^N$ possible sentences.

In order to find possible sentences following algorithm has been written.

```
self.__english = ["i", "o", "u", "s", "c", "g"]
self.__turkish = {"i": "ı", "o": "ö", "u": "ü",
                  "s": "ş", "c": "ç", "g": "ğ"}
```

Here we have English letters and corresponded Turkish letters by using this list and dictionary we created following algorithm.

```
def ____findPossibleSentences(self, sentence):
    occurrenceDict = dict()
    resultSet = set()
    resultSet.add(sentence)
    size = 0
    for ele in self.__english:
        occurrenceDict[ele] = [m.start() for m in re.finditer(ele, sentence)]
        size = size + len(occurrenceDict[ele])

    for key1, value1 in occurrenceDict.items():
        if len(value1) > 0:
            for index in value1:
                tempList = []
                for sent in resultSet:
                    liste = list(sent)
                    liste[index] = self.__turkish[key1]
                    tempList.append(''.join(liste))
                for item in tempList:
                    resultSet.add(item)

    return resultSet
```

In this algorithm, first for loop marks all the letters that may change on a Turkish alphabet version and save them into a dict. In second loop by using these letters and indexes of them we are finding all possible sentences.

### 2.1.1) Test Results of this algorithm

```
D:\github\pythons\venv\Scripts\python.exe D:/GTU/4.yıl-
Test Sentence: beklenen olumlu tepkiyi alamadi
Number of Possible Sentences: 64
-----
Test Sentence: guzel vakit geciyor oneririm
Number of Possible Sentences: 1024
-----
Test Sentence: konforlu ve temiz bir yer
Number of Possible Sentences: 32
-----
Test Sentence: sicak ve eglenceli bir ortam
Number of Possible Sentences: 256
-----
Test Sentence: bu yurt okula cok uzak
Number of Possible Sentences: 128
-----
```

As you can we found the possible sentences successfully.

### 2.2) Changes on my HW2 Language Model

In my hw2 there are missing point I changed them to get better results.

- At line 157;  
Change;  
allSentences = tokenize.sent\_tokenize(self.\_\_corpusTestFile.read())  
to;  
allSentences = self.\_\_corpusTestFile.readlines()
- At line 103 remove res.append(' ')
- Change syllable file

After changes probability results

```
1 bilgisayar işletim sistemlerinin en temel parçası
2 20.013196985394455
3 bununla birlikte lenin
4 23058.56097751246
5 bilgisayar işletim şıştemlerinin en temel parçası
6 1.7322789540788036e+31
7 bününla birlıkte lenın
8 42779810012.187836
```

As you can see correct sentences has less probability which is wrong result.

### 2.3) Performance of the language model with NGrams

In following test, **first 6 sentence** comes from train dataset(text). Also, we will see that **3 grams** gives best result.

### 2.3.1) 1 Grams

Test Sentence	Sentence With the Highest Probability	Result
bilgisayar isletim sistemlerinin	bilgişayar isletim sistemlerinin	F
ticari linux dagitimlariinin gelirlerini	tıçarı linux dağıtımlarının gelirlerini	F
orta duzeyde zengin bir ailede	örta düzeyde zengin bır ailede	F
yeni adıyla turkiye komunist partisi	yeni adıyla turkiye komunist partışı	F
ilde bircok etkinlik duzenlenmektedir	ılde birçok etkinlik düzenlenmektedir	F
cok sayida koyun yetistirilir	çok şayida koyun yetistirılır	F
beklenen olumlu tepkiyi alamadi	beklenen ölümlü tepkiyi alamadı	F
guzel vakit geciyor oneririm	guzel vakit geçiyor öneririm	F
konforlu ve temiz bir yer	konforlü ve temiz bır yer	F
sicak ve eglenceli bir ortam	şıcak ve eğlençeli bır ortam	F
bu yurt okula cok uzak	bü yurt öküla çok uzak	F
temizlik berbat	çamasir deterjanını bile vermiyorlar	F
idare eder tavsiye ederim	ıdare eder tavsiye ederim	F
fiyatlar cok pahali	fiyatlar çok pahalı	F
oteller burdan ucuzdur	öteller bürdan ucuzdür	F
yurtda mutfak var is gorur	yurtda mutfak var is görür	F
bence parasi olanlar kacirmasin	bençe paraşı olanlar kaçırmaşın	F
odalarda calisma masasi yok	ödalarda çalışma maşaşı yok	F
koskoca yurtda kutuphane yok	köskaç yurtda kütüphane yok	F
nice dostluklariin kurulduđu yer	nıçe dostlukların kürüldüğü yer	F
cadirda kalirim daha iyi	çadirda kalırım daha iyi	F
resimlere inanmayin bes para etmez	resimlere inanmayın beş para etmez	F
cok bisey beklemeyin	çok bisey beklemeyin	F
bu zamanda zor bulunacak bir yurt	bü zamanda zor bülünacak bır yurt	F
ogrenci dostu bir yurt	öğrençi döştü bır yurt	F
calisanlariin davranislari cok iyi	çalışanların davranışları çok iyi	F
mac yayini yok bilginiz olsun	maç yayını yok bilginiz olsun	F
guzel gunlerim olsun diyenlere onerilir	guzel gunlerim olsun diyenlere önerilir	F
hayatinizde en az bi kere burda kalmalisiniz	hayatınızda en az bi kere bürda kalmalışınız	F
universiteye en yakin yer burasi sanirim	univerşıteye en yakin yer büraşı şanırim	F
kayit yaparken gezdik teras super	kayıt yaparken gezdik teraş şuper	F
koskoca yurtda utu yok	köskaç yurtda utü yok	F
okula yakin olmasi disinda iyi bir tarafi yok	öküla yakin olmaşı dışında iyi bır tarafı yok	F
gencliginize yazik etmeyin buralarda	gençliğınıza yazık etmeyin büralarda	F
sozlesme imzalatiyorlar bulasmayin derim ben	sözleşme imzalatiyorlar bulaşmayın derim ben	F
paraniza yazik olur burda	paranıza yazık olur bürda	F
bu yurtda ogrenci kalitesi gercekten cok dusuk	bü yurtda öğrençi kaliteşi gercekten çok düşük	F
tek kelimeyle guzel bir yer etrafida yesillik	tek kelimeyle guzel bır yer etrafida yesillik	F

### 2.3.2) 2 Grams

Test Sentence	Sentence With the Highest Probability	Result
bilgisayar isletim sistemlerinin	bilgişayar işletim sistemlerinin	F
ticari linux dagitimlerinin gelirlerini	tiçarı linux dağıtımlarının gelirlerini	F
orta duzeyde zengin bir ailede	orta duzeyde zengin bir ailede	F
yeni adıyla turkiye komunist partisi	yeni adıyla turkiye kömünist partisi	F
ilde birçok etkinlik duzenlenmektedir	ilde birçok etkinlik duzenlenmektedir	F
cok sayida koyun yetistirilir	çok şayida koyun yetistirilir	F
beklenen olumlu tepkiyi alamadi	beklenen olumlu tepkiyi alamadı	F
guzel vakit geciyor oneririm	güzel vakit geçiyor oneririm	F
konforlu ve temiz bir yer	konforlu ve temiz bir yer	P
sicak ve eglenceli bir ortam	sıcak ve eğlençeli bir ortam	F
bu yurt okula cok uzak	bu yurt okula çok uzak	P
temizlik berbat	çamaşır deterjanını bile vermiyorlar	F
idare eder tavsiye ederim	ıdare eder tavsiye ederim	F
fiyatlar cok pahali	fiyatlar çok pahalı	P
oteller burdan ucuzdur	oteller burdan ucuzdur	P
yurtda mutfak var is gorur	yurtda mutfak var iş görür	P
bence parasi olanlar kacirmasin	bence paraşı olanlar kaçırmasin	F
odalarda calisma masasi yok	ödalarda calisma maşası yok	F
koskoca yurtda kutuphane yok	koskoca yurtda kütüphane yok	P
nice dostluklarin kuruldugu yer	nice dostluklarin kuruldugu yer	F
cadirda kalirim daha iyi	cadırda kalirim daha iyi	F
resimlere inanmayin bes para etmez	reşimlere inanmayin beş para etmez	F
cok bisey beklemeyin	çok bişey beklemeyin	P
bu zamanda zor bulunacak bir yurt	bu zamanda zor bulunaçak bir yurt	F
ogrenci dostu bir yurt	öğrençi dostu bir yurt	F
calisanlarin davranislari cok iyi	çalışanlarin davranislari çok iyi	F
mac yayini yok bilginiz olsun	maç yayini yok bilginiz olsun	F
guzel gunlerim olsun diyenlere önerilir	güzel günlerim olsun diyenlere önerilir	F
hayatinizda en az bi kere burda kalmalisiniz	hayatinizda en az bi kere burda kalmalışiniz	F
universiteye en yakin yer burasi sanirim	üniverşıteye en yakın yer burası şanirim	F
kayit yaparken gezdik teras super	kayıt yaparken gezdik teras super	F
koskoca yurtda utu yok	koskoca yurtda ütu yok	F
okula yakin olmasi disinda iyi bir tarafi yok	öküla yakın olmaşı disinda iyi bir tarafi yok	F
gencliginize yazik etmeyin buralarda	gençliğinize yazik etmeyin buralarda	F
sozlesme imzalatiyorlar bulasmayin derim ben	sözleşme imzalatiyorlar bulaşmayın derim ben	F
paraniza yazik olur burda	paraniza yazik olur burda	F
bu yurtda ogrenci kalitesi gercekten cok dusuk	bü yurtda öğrenci kalitesi gercekten çok düşük	F
tek kelimeyle guzel bir yer etrafida yesillik	tek kelimeyle güzel bir yer etrafida yeşillik	P

### 2.3.3) 3 Grams

Test Sentence	Sentence With the Highest Probability	Result
bilgisayar isletim sistemlerinin	bilgisayar işletim sistemlerinin	F
ticari linux dagitimlarinin gelirlerini	ticari linux dağıtımlarının gelirlerini	P
orta duzeyde zengin bir ailede	orta düzeyde zengin bir ailede	P
yeni adiyla turkiye komunist partisi	yeni adıyla türkiye komünist partisi	P
ilde bircok etkinlik duzenlenmektedir	ilde birçok etkinlik düzenlenmektedir	P
cok sayida koyun yetistirilir	çok sayıda koyun yetiştirilir	P
beklenen olumlu tepkiyi alamadi	beklenen olumlu tepkiyi alamadı	F
guzel vakit geciyor oneririm	güzel vakit geçiyor öneririm	F
konforlu ve temiz bir yer	konforlu ve temiz bir yer	P
sicak ve eglenceli bir ortam	sıcak ve eğlenceli bir ortam	P
bu yurt okula cok uzak	bu yurt okula çok uzak	P
temizlik berbat	çamaşır deterjanini bile vermiyorlar	F
idare eder tavsiye ederim	idare eder tavsiye ederim	P
fiyatlar cok pahali	fiyatlar çok pahali	F
oteller burdan ucuzdur	oteller burdan ucuzdur	P
yurtda mutfak var is gorur	yürtda mutfak var iş görür	F
bence parasi olanlar kacirmasin	bençe parası olanlar kaçırmasın	F
odalarda calisma masasi yok	odalarda çalışma masası yok	P
koskoca yurtda kutuphane yok	köşkoça yürtda kütüphane yok	F
nice dostluklarin kuruldugu yer	nice dostlukların kurulduğu yer	P
cadirda kalirim daha iyi	çadırdı kalırım daha iyi	F
resimlere inanmayin bes para etmez	reşimlere inanmayın beş para etmez	F
cok bisey beklemeyin	çok bişey beklemeyin	P
bu zamanda zor bulunacak bir yurt	bu zamanda zor bulunacak bir yurt	P
ogrenci dostu bir yurt	öğrenci dostu bir yurt	F
calisanlarin davranislari cok iyi	çalışanların davranışları çok iyi	P
mac yayini yok bilginiz olsun	mac yayını yok bilginiz olsun	F
guzel gunlerim olsun diyenlere onerilir	güzel günlerim olsun diyenlere önerilir	P
hayatinizda en az bi kere burda kalmalisiniz	hayatınızda en az bi kere burda kalmalısınız	F
universiteye en yakin yer burasi sanirim	üniversiteye en yakın yer burası sanırım	F
kayit yaparken gezdik teras super	kayıt yaparken gezdik teras şüper	F
koskoca yurtda utu yok	köşkoça yürtda utü yok	F
okula yakin olmasi disinda iyi bir tarafi yok	okula yakın olması dışında iyi bir tarafı yök	F
gencliginize yazik etmeyin buralarda	gençliğinize yazık etmeyin buralarda	P
sozlesme imzalatiyorlar bulasmayin derim ben	sözleşme imzalatıyorlar bulaşmayın derim ben	P
paraniza yazik olur burda	paranıza yazık olur bürda	F
bu yurtda ogrenci kalitesi gercekten cok dusuk	bu yürtda öğrenci kalitesi gerçekten çok düşük	P
tek kelimeyle guzel bir yer etrafida yesillik	tek kelimeyle güzel bir yer etrafıda yeşillik	P

### 2.3.4) 4 Grams

Test Sentence	Sentence With the Highest Probability	Result
bilgisayar isletim sistemlerinin	bilgisayar işletim sistemlerinin	P
ticari linux dagitimlariinin gelirlerini	ticari linux dağıtımlarının gelirlerini	P
orta duzeyde zengin bir ailede	orta düzeyde zengin bir ailede	P
yeni adıyla turkiye komunist partisi	yeni adıyla türkiye komünist partisi	P
ilde bircok etkinlik duzenlenmektedir	ilde birçok etkinlik düzenlenmektedir	P
cok sayida koyun yetistirilir	çok sayıda koyun yetiştirilir	P
beklenen olumlu tepkiyi alamadi	beklenen olumlu tepkiyi alamadı	F
guzel vakit geciyor oneririm	guzel vakit geçiyör öneririm	F
konforlu ve temiz bir yer	konforlü ve temiz bir yer	F
sicak ve eglenceli bir ortam	sicak ve eğlenceli bir ortam	F
bu yurt okula cok uzak	bu yurt öküla cök uzak	F
temizlik berbat	camaşır deterjanını bile vermiyörlar	F
idare eder tavsiye ederim	idare eder tavsiye ederim	P
fiyatlar cok pahali	fiyatlar çok pahali	F
oteller burdan ucuzdur	öteller burdan ücüzdür	F
yurtda mutfak var is gorur	yurtda mütfağ var iş görür	F
bence parasi olanlar kacirmasin	bence paraşı olanlar kaçırmaşın	F
odalarda calisma masasi yok	odalarda çalışma masası yök	F
koskoca yurtda kutuphane yok	köşkoça yurtda kütüphane yok	F
nice dostluklarin kurulduđu yer	nıçe döstluklarin kurulduđu yer	F
cadirda kalirim daha iyi	çadırdı kalırım daha iyi	P
resimlere inanmayin bes para etmez	resimlere inanmayın beş para etmez	P
cok bisey beklemeyin	cök bişey beklemeyin	F
bu zamanda zor bulunacak bir yurt	bu zamanda zör bulunacak bır yürt	F
ogrenci dostu bir yurt	öğrenci doştü bır yürt	F
calisanlarin davranislari cok iyi	çalışanların davranışları cök iyi	F
mac yayini yok bilginiz olsun	mac yayını yok bilginiz olşün	F
guzel gunlerim olsun diyenlere onerilir	güzel günlerim ölsün diyenlere önerilir	F
hayatinizda en az bi kere burda kalmalisiniz	hayatinızda en az bı kere burda kalmalısınız	F
universiteye en yakin yer burasi sanirim	universiteye en yakın yer burası sanırım	F
kayit yaparken gezdik teras super	kayıt yaparken gezdık teraş şüper	F
koskoca yurtda utu yok	kosköça yürtda utu yök	F
okula yakin olmasi disinda iyi bir tarafi yok	okula yakın olması dışında iyi bir tarafı yök	F
gencliginize yazik etmeyin buralarda	gençliğinize yazik etmeyin buralarda	F
sozlesme imzalatiyorlar bulasmayin derim ben	sözleşme imzalatiyörlar bülasmayin derım ben	F
paraniza yazik olur burda	paraniza yazik olur bürda	F
bu yurtda ogrenci kalitesi gercekten cok dusuk	bu yürtda öğrenci kalitesi gerçekten cok düşük	F
tek kelimeyle guzel bir yer etrafida yesillik	tek kelimeyle güzel bir yer etrafida yeşillik	F

### 2.3.5) 5 Grams

Test Sentence	Sentence With the Highest Probability	Result
bilgisayar isletim sistemlerinin	bilgisayar işletim sistemlerinin	P
ticari linux dagitimlarinin gelirlerini	ticari linux dağıtımlarının gelirlerini	P
orta duzeyde zengin bir ailede	orta düzeyde zengin bir ailede	P
yeni adıyla turkiye komunist partisi	yeni adıyla türkiye komünist partisi	P
ilde birçok etkinlik duzenlenmektedir	ilde birçok etkinlik düzenlenmektedir	P
cok sayida koyun yetistirilir	çok sayıda koyun yetiştirilir	P
beklenen olumlu tepkiyi alamadi	beklenen olumlu tepkiyi alamadı	F
guzel vakit geciyor oneririm	guzel vakit geciyor öneririm	F
konforlu ve temiz bir yer	konforlü ve temiz bir yer	F
sicak ve eglenceli bir ortam	sıcak ve eğlenceli bir ortam	F
bu yurt okula cok uzak	bu yurt okula çok uzak	F
temizlik berbat	camaşır deterjanını bile vermiyörlar	F
idare eder tavsiye ederim	idare eder tavsiye ederim	P
fiyatlar cok pahali	fiyatlar çok pahali	F
oteller burdan ucuzdur	oteller burdan ucüzdür	F
yurtda mutfak var is gorur	yurtda mutfak var ısı görür	F
bence parasi olanlar kacirmasin	bençe parası olanlar kacırmaşın	F
odalarda calisma masasi yok	odalarda çalışma masası yok	F
koskoca yurtda kutuphane yok	köşkoça yurtda kütüphane yok	F
nice dostluklarin kuruldugu yer	nıçe doştlüklerin kurulduğı yer	F
cadirda kalirim daha iyi	cadirda kalırım daha iyi	F
resimlere inanmayin bes para etmez	reşimlere inanmayın beş para etmez	F
cok bisey beklemeyin	çok bisey beklemeyın	F
bu zamanda zor bulunacak bir yurt	bu zamanda zor bulunaçak bir yürt	F
ogrenci dostu bir yurt	ogrençi döştü bir yürt	F
calisanlarin davranislari cok iyi	çalışanların davranışlari çok iyi	F
mac yayini yok bilginiz olsun	mac yayını yok bilginiz ölsün	F
guzel gunlerim olsun diyenlere onerilir	güzel günlerım ölsün diyenlere önerilir	F
hayatinizda en az bi kere burda kalmalisiniz	hayatinızda en az bı kere bürda kalmalışınız	F
universiteye en yakin yer burasi sanirim	universiteye en yakin yer bürasi sanırım	F
kayit yaparken gezdik teras super	kayıt yaparken ğezdik teraş şuper	F
koskoca yurtda utu yok	kösköça yürtda ütü yok	F
okula yakin olmasi disinda iyi bir tarafi yok	okula yakın ölması dışında iyi bır tarafi yok	F
gencliginize yazik etmeyin buralarda	ğencliginize yazık etmeyin büralarda	F
sozlesme imzalatiyorlar bulasmayin derim ben	sözleşme imzalatiyörlar bülasmayın derım ben	F
paraniza yazik olur burda	paraniza yazık olur burda	F
bu yurtda ogrenci kalitesi gercekten cok dusuk	bü yurtda öğrenci kaliteşi gercekten çok düşük	F
tek kelimeyle guzel bir yer etrafida yesillik	tek kelimeyle güzel bir yer etrafida yeşillik	F

### How to run

Just install needed python libraries and run 171044098.py file. Before run change test sentence in test\_sentences.txt file. After running is finished check output files to see chosen sentences.