

An Overview of Resources Available for Turkish Natural Language Processing Applications

Tunga GÜNGÖR
Computer Engineering, Boğaziçi University
TurcLing 2016 Keynote Speaker

An Overview of Resources Available for Turkish Natural Language Processing Applications

Tunga Güngör
Boğaziçi University
Computer Engineering

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

1

Turkish WordNet

- ✓ Syntactic and structural quality was measured
- ✓ Coverage was measured using two sources
 - ✓ George Orwell's 1984 novel & a general-purpose corpus
 - ✓ Coverage of words was found to be 87%

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

4

Outline

- ✓ Turkish WordNet
- ✓ METU Turkish Corpus
- ✓ Turkish Treebank
- ✓ Boun Resources
- ✓ Kemik Group (YTU)
- ✓ ITU Turkish NLP Pipeline
- ✓ Turkish Labeled Text Corpus
- ✓ TS Corpus

- ✓ Other Resources

- ✓ Conclusions
- ✓ Paper List

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

2

Turkish WordNet

- ✓ Basic statistics (March 2004)

	Frequency
Synsets	11,628
Synset members	16,095
Average synset size	1.38
Unlexicalized concepts	987
Definitions	4,913

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

5

Turkish WordNet

- ✓ Part of BalkaNet project
 - ✓ WordNets in Central and Eastern European languages
- ✓ Linked to Princeton WordNet and Euro WordNet

- ✓ Methodology
 - ✓ Base concepts of Euro WordNet translated
 - ✓ Synonyms, antonyms and hyponyms extracted automatically from monolingual Turkish dictionary
 - ✓ Base concept set enlarged

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

3

Turkish WordNet

- ✓ Parts of speech

	Frequency	Percentage
Nouns	8,691	74.7 %
Verbs	2,556	22.0 %
Adjectives	381	3.3 %

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

6

Turkish WordNet

✓ Semantic relations

	Frequency
Hypemym	11,251
Holo_member	946
Holo_part	1,423
Holo_portion	176
Causes	100
Be_in_state	577
Near-antonym	1,400
Subevent	127
Also_see	270
Verb_group	896
Category_domain	384
Total	17,550

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

7

Turkish WordNet

```
<SYNSET><ID>ENG20-02687362-
n</ID><POS>n</POS><SYNONYM><LITERAL>bant<SENSE>1</SENSE></LITERAL><LITERAL>şerit<SENSE>2</SENSE></LITERAL><SYNONYM><LR>ENG20-04172147-n</TYPE>hypemym</TYPE></LR><BCS>2</BCS></SYNSET>
```

```
<SYNSET><ID>ENG20-02728698-
n</ID><POS>n</POS><SYNONYM><LITERAL>bant<SENSE>2</SENSE></LITERAL></SYNONYM><LR>ENG20-13089369-
n</TYPE>hypemym</TYPE></LR><DEF>Düz, ensiz, yassı bağ<DEF><STAMP>ozemc
2003/05/23</STAMP><BCS>3</BCS></SYNSET>
```

```
<SYNSET><ID>ENG20-02873277-
n</ID><SYNONYM><LITERAL>kaset<SENSE>1</SENSE></LITERAL><LITERAL>bant<SENSE>3</SENSE></LITERAL></SYNONYM><DEF>a cassette that contains magnetic tape</DEF><POS>n</POS></LR>ENG20-03569739-
n</TYPE>hypemym</TYPE></LR><LR>ENG20-02872759-n</TYPE>holo_part</TYPE></LR><STAMP>orhanb
2004/06/25</STAMP></SYNSET>
```

```
<SYNSET><ID>ENG20-03569739-
n</ID><POS>n</POS><SYNONYM><LITERAL>bant<SENSE>4</SENSE></LITERAL></SYNONYM><LR>ENG20-03604997-
n</TYPE>hypemym</TYPE></LR><STAMP>ozemc 2003/05/23</STAMP><BCS>3</BCS></SYNSET>
```

```
<SYNSET><ID>ENG20-05878892-
n</ID><POS>n</POS><SYNONYM><LITERAL>bant<SENSE>5</SENSE></LITERAL></SYNONYM><LR>ENG20-07888704-
n</TYPE>hypemym</TYPE></LR><STAMP>ozemc 2003/05/23</STAMP><BCS>3</BCS></SYNSET>
```

```
<SYNSET><ID>ENG20-05879002-
n</ID><POS>n</POS><SYNONYM><LITERAL>bant<SENSE>6</SENSE></LITERAL><LITERAL>frekans
band<SENSE>1</SENSE></LITERAL></SYNONYM><LR>ENG20-05878892-
n</TYPE>hypemym</TYPE></LR><LR>ENG20-05894694-
n</TYPE>category_domain</TYPE></LR><BCS>2</BCS></SYNSET>
```

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

10

Turkish WordNet

```
<SYNSET>
<ID>ENG20-02768681-n</ID>
<POS>n</POS>
<SYNONYM><LITERAL>kitap<SENSE>1</SENSE></LITERAL></SYNONYM>
<LR>ENG20-03856368-n</TYPE>hypemym</TYPE></LR>
<DEF>Ciltli veya ciltsiz olarak bir araya getirilmiş, basılı veya yazılı kağıt yaprakların
bütünü.</DEF>
<BCS>1</BCS>
</SYNSET>
```

```
ENG20-03856368-n :
<LITERAL>ürün<SENSE>3</SENSE></LITERAL>
```

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

8

METU Turkish Corpus

- ✓ Post-1990 written Turkish sampled from various genres
- ✓ 2M words
- ✓ 291 sources
- ✓ 520 samples
- ✓ Balances among authors and genres
- ✓ 14 genres
- ✓ Annotated wrt TEI (Text Encoding Initiative) guidelines and XCES (XML-based Corpus Encoding Standard)

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

11

Turkish WordNet

```
<SYNSET>
<ID>ENG20-09970518-n</ID>
<POS>n</POS>
<SYNONYM><LITERAL>öğrenci<SENSE>2</SENSE></LITERAL>
<LITERAL>talebe<SENSE>1</SENSE></LITERAL></SYNONYM>
<LR>ENG20-09415396-n</TYPE>hypemym</TYPE></LR>
<LR>ENG20-13057333-n</TYPE>holo_member</TYPE></LR>
<DEF>Öğrenim görmek amacıyla herhangi bir öğretim kurumunda okuyan kimse</DEF>
<BCS>1</BCS>
</SYNSET>
```

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

9

METU Turkish Corpus

```
<oesHeader>
<fileDesc>
<title>
<h>file>00017113</h>
<title>
<extent>
<wordCount>2008</wordCount>
<byteCount>17929</byteCount>
<extent>
<sourceDesc>
<bibliStruct>
<analytic>
<h>file>Anadolu Da lar n n "BİRİNCİ" Prof. Dr. Turhan BAYTOP</h>
<h>author>Nalan MAHSEREC</h>
<analytic>
<imprint>
<publisher>Bilim ve Ötopya</publisher>
<pubDate>Mart 2000</pubDate>
<pubPlace>istanbul</pubPlace>
<imprint>
<idno>1301 - 6717</idno>
<bibliStruct>
<sourceDesc>
```

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

12

METU Turkish Corpus

```

<textClass>
<catRef>Makale</catRef>
<textClass>
<profileDesc>
<revisionDesc>
<change>
<changeDate>12.10.2000</changeDate>
<respname>Sedef</respname>
<h.item>The header part was changed.</h.item>
<change>
<revisionDesc>
<catHeader>
<text>
<body>
<p>Oktay biraz önce, <q>Hadî biz de Sitem'in yan na gidelim, <q> demi ş. Sitem'in, kuca nda Tomurcuk Bayle Y lanî
nçirlerden yana gitti ini o da görmü şü çüknü. Ben omuz silmekle yetindim, Oktay da üstelemedi. Sitem ikimizin
yüzüne kar da görünmez kap lar kapam t. Benim de elinden kay p gidivermeden korkan Oktay beni
<h>oyalamak</h> için geçen yaz Giray A abeysyle Kiraz! Yaylası yapt klar bir gezintiyi anlatmaya ba lad. <p><p>O
gün ve sonras nda olanlar elbet sana da anlattım lard r, Dalya. Gene de o karga a, o a k rî k, o panik, o kafa
karma as ya anmadan bilinemez. <p>
</body>
</text>

```

Turkish Treebank

```

<Set sentences="1">
<S No="1">
<W IX="1" LEM="" MORPH="" IG="[1,"iyi+Adj"]" REL="[2,1,(SENTENCE)]">iyi</W>
<W IX="2" LEM="" MORPH="" IG="[1,"+Punc"]" REL="[7,1,(COORDINATION)]">,</W>
<W IX="3" LEM="" MORPH="" IG="[1,"git+Verb+Pos+Imp+A2pl"]" REL="[7,1,(MODIFIER)]">gidin</W>
<W IX="4" LEM="" MORPH="" IG="[1,"de+Con"]" REL="[3,1,(RELATIVIZER)]">de</W>
<W IX="5" LEM="" MORPH="" IG="[1,"iskele+Noun+A3sg+Pnon+Acc"]" REL="[7,1,(OBJECT)]">iskeleyi</W>
<W IX="6" LEM="" MORPH="" IG="[1,"bir+Def"]" REL="[7,1,(DETERMINER)]">bir</W>
<W IX="7" LEM="" MORPH="" IG="[1,"yika+Verb+Pos+Imp+A2pl"]" REL="[8,1,(SENTENCE)]">yıkayın</W>
<W IX="8" LEM="" MORPH="" IG="[1,"+Punc"]" REL="[10,1,(COORDINATION)]">,</W>
<W IX="9" LEM="" MORPH="" IG="[1,"sonra+Adv"]" REL="[10,1,(MODIFIER)]">sonra</W>
<W IX="10" LEM="" MORPH="" IG="[1,"dinlen+Verb+Pos+Aor+A2pl"]" REL="[11,1,(SENTENCE)]">dinlenirsiniz</W>
<W IX="11" LEM="" MORPH="" IG="[1,"+Punc"]" REL="[12,1,(OBJECT)]">,</W>
<W IX="12" LEM="" MORPH="" IG="[1,"dedi+Verb+Pos+Past+A3sg"]" REL="[13,1,(SENTENCE)]">dedi</W>
<W IX="13" LEM="" MORPH="" IG="[1,"+Punc"]" REL="[1,1,)>,</W>
</S>
</Set>

```

Turkish Treebank

- ✓ Subcorpus of METU Turkish Corpus
- ✓ 10,000 sentences with morphological and syntactic annotations
- ✓ Syntactic relations:

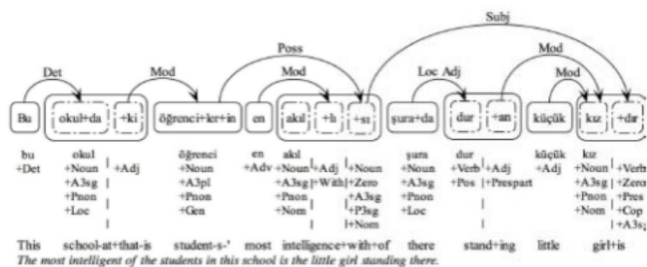
Sentence	Subject
Object	Modifier
Modifier	Determiner
Focus-Particle	Question-Particle
Vocative	Classifier
Dative Adjunct	Ablative Adjunct
Locative Adjunct	Instrum. Adjunct
Coordination	Relativizer
Possessive	

Turkish Treebank

- ✓ Annotation tool

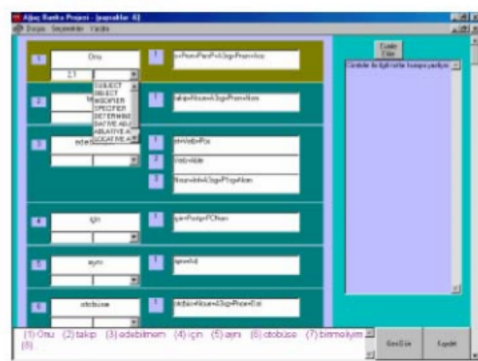


Turkish Treebank



Turkish Treebank

- ✓ Annotation tool



Boun Resources

- ✓ Developed at Boğaziçi University
- ✓ Resources:
 - ✓ Stochastic morphological parser
 - ✓ Morphological disambiguator
 - ✓ Web corpus

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

19

Boun Resources

- ✓ Lexicon and morphotactics
 - ✓ Based on PC-KIMMO implementation (K. Oflazer)
 - ✓ Morphological features from Xerox implementation
- ✓ 55K root word lexicon
- ✓ 33 two-level rules hand-compiled to FSTs
 - ✓ Improved morphotactics
- ✓ 11,000 words/sec

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

22

Boun Resources

- ✓ Morphological parser
 - ✓ Lexicon: Stores the root words and some morphological features
 - ✓ Morphotactics: (Morphosyntax) is the word grammar for the ordering of morphemes
 - ✓ Morphophonemics: Phonological rules describing morphological alternations
- ✓ Stochastized using the morphological disambiguator and web corpus

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

20

Boun Resources


- ✓ Lexicon and morphotactics
 - ✓ An example output for the word *alın*:
 - ✓ *alın* [Verb]+[Pos]+[Imp]+[A2sg]
 - ✓ *alın* [Noun]+[A3sg]+[Pnon]
 - ✓ *alın* [Noun]+[A3sg]+[Pnon]-[Verb+Pres+A3sg]
 - ✓ *al* [Noun]+[A3sg]+Hn[P2sg]
 - ✓ *al* [Noun]+[A3sg]+Hn[P2sg]-[Verb+Pres+A3sg]
 - ✓ *al* [Noun]+[A3sg]+[Pnon]+NHn[Gen]
 - ✓ *al* [Noun]+[A3sg]+[Pnon]+NHn[Gen]-[Verb+Pres+A3sg]
 - ✓ *al* [Verb]+[Pos]+[Imp]+YHn[A2p]
 - ✓ *al* [Verb]-Hn[Verb+Pass]+[Pos]+[Imp]+[A2sg]
 - ✓ *al* [Adj]-[Noun]+[A3sg]+Hn[P2sg]

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

23

Boun Resources

- ✓ Lexicon and morphotactics
 - ✓ Implemented as a single finite-state transducer
 - ✓ An example for Turkish nominal inflection:

- ✓ Output side is input to two-level rules transducer
- ✓ Special symbols are used for phonological alternations
- ✓ Compose lexicon/morphotactics transducer with the morphophonemics transducer

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

21

Boun Resources

- ✓ Morphological disambiguator
 - ✓ Based on averaged perceptron algorithm
 - ✓ *alın*[Verb]+[Pos]+[Imp]+[A2sg]
 - ✓ *alın*[Noun]+[A3sg]+[Pnon]
 - ✓ *alın*[Noun]+[A3sg]+[Pnon]-[Verb+Pres+A3sg]
 - ✓ *al*[Noun]+[A3sg]+Hn[P2sg]
 - ✓ *al*[Noun]+[A3sg]+Hn[P2sg]-[Verb+Pres+A3sg]
 - ✓ *al*[Noun]+[A3sg]+[Pnon]+NHn[Gen]
 - ✓ *al*[Noun]+[A3sg]+[Pnon]+NHn[Gen]-[Verb+Pres+A3sg]
 - ✓ *al*[Verb]+[Pos]+[Imp]+YHn[A2p]
 - ✓ *al*[Verb]-Hn[Verb+Pass]+[Pos]+[Imp]+[A2sg]
 - ✓ *al*[Adj]-[Noun]+[A3sg]+Hn[P2sg]
- ✓ 97.8% success rate

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

24

Boun Resources

✓ Morphological disambiguator

Gloss	Feature
Morphological parse trigram	(1) $t_{i-2} t_{i-1} t_i$
Morphological parse bigram	(2) $t_{i-2} t_i$ & (3) $t_{i-1} t_i$
Morphological parse unigram	(4) t_i
Morpheme tag with previous tag	(5) $t_{i-1} m_i$
Morpheme tag with second to previous tag	(6) $t_{i-2} m_i$
Root trigram	(7) $r_{i-2} r_{i-1} r_i$
Root bigram	(8) $r_{i-2} r_i$ & (9) $r_{i-1} r_i$
Root unigram	(10) r_i
Morpheme tag trigram	(11) $m_{i-2} m_{i-1} m_i$
Morpheme tag bigram	(12) $m_{i-2} m_i$ & (13) $m_{i-1} m_i$
Morpheme tag unigram	(14) m_i
Individual morpheme tags	(15) m_j for $j = 1, \dots, n_i$
Individual morpheme tags with position	(16) $j m_j$ for $j = 1, \dots, n_i$
Number of morpheme tags	(17) n_i

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

25

Boun Resources

Web Corpus

Coverage statistics

Vocabulary size	Coverage %
1K	50.5
5K	67.0
10K	73.0
20K	81.5
50K	89.0
100K	93.2
200K	94.7
400K	96.2
1M	98.5
3M	99.2
4M	100.0

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

28

Boun Resources

Web Corpus

- ✓ Formed of two subcorpora:
 - ✓ NewsCor: Milliyet, Ntvmsnbc, Radikal
 - ✓ GenCor: General sampling of web pages
- ✓ About 500M tokens
- ✓ Cleaned using morphological parser and heuristics
- ✓ Converted to XCES XML format
 - ✓ `<p><s><q>Cthulhunun Çağırısı</q> ve ardından <q>Deliliğin Dağlarında</q> adlı eserleri Türkçeye çevrilen Howard Phillips Lovecraft korku ve gerilim ustası bir yazar.</s></p>`

April 4, 2016

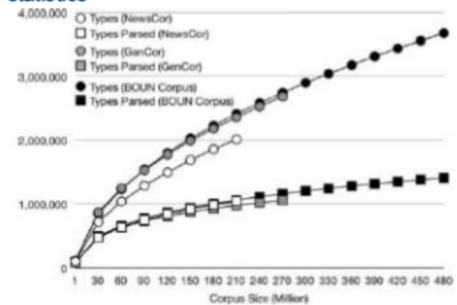
Turkic Computational Linguistics - TurCLing
2016

26

Boun Resources

Web Corpus

Type statistics



April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

29

Boun Resources

Web Corpus

Corpus	Words	Tokens	Types	Tokens parsed (%)	Types parsed (%)
Milliyet	59M	68M	1.1M	96.7	63.5
Ntvmsnbc	75M	86M	1.2M	96.4	55.8
Radikal	50M	58M	1.0M	97.0	65.7
NewsCor	184M	212M	2.2M	96.7	52.2
GenCor	239M	279M	3.0M	94.6	39.5
BounCor	423M	491M	4.1M	95.5	38.4

April 4, 2016

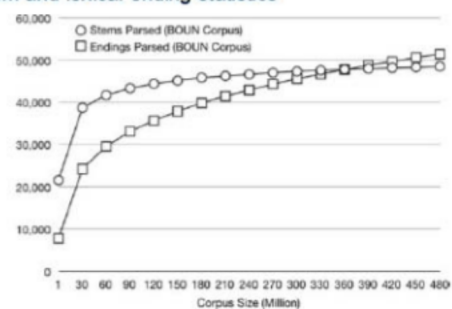
Turkic Computational Linguistics - TurCLing
2016

27

Boun Resources

Web Corpus

Stem and lexical ending statistics



April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

30

Kemik Group (YTU)

✓ Datasets

Dataset Name	Explanation	Download	Size
2000 Parsed Sentence	2000 Parsed Turkish Sentences	2000cumle_oge.rar	0.1 MB
20 million Turkish Tweets	20 million Turkish Tweets	20milyonTweet.rar	640 MB
1200 words	Most used Words in Turkish Children Literature	1200kelime.rar	0.1 MB
2 million Turkish Tweets	2 million Turkish Tweets	2milyonTweet.rar	91 MB
42 K news	42K Turkish news texts in 13 classes	42bin_haber.rar	42 MB
3000 Tweets	3000 Turkish Tweets for sentiment analysis (3 classes)	3000tweets.rar	1.6 MB
2500 Columns	2500 Columns from 50 different Columnists	2500koseyazisi.rar	5 MB
1500 Columns	1500 Columns from 30 different Columnists, in raw and arff formats.	30Columnistb.zip	7 MB

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

31

Kemik Group (YTU)

Dataset: 2000 parsed Turkish sentences

✓ Manually tagged with syntactic labels

Label	Explanation
Zaman	Zaman Tümlacı (Time Complement)
DT	Dolaylı Tümlaç (Indirect Object)
İle	Biriktelik (With)
Y	Yüklem (Verb)
Özne	Özne (Subject)
Nasıl	Nasıl (How)
Bsiz	Belirtisiz Nesne (Indefinite Object)
Bili	Belirtili Nesne (Definite Object)
O	Noktalama İşareti (Punctuation)
İçin	Sebeup (Cause)
Kadar	Kadar (Until)

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

34

Kemik Group (YTU)

✓ Datasets

Dataset Name	Explanation	Download	Size
157 blogs	157 Blogs for sentiment analysis, in raw and arff formats.	ruh_hali.zip	0.7 MB
105 Reviews	105 positive, negative and neutral movie reviews, in raw and arff formats.	film_yorumlari.zip	0.2 MB
69 Authors	910 articles from 69 different Turkish authors, in raw and arff formats.	69yazar.zip	3.2 MB
1150 News	1150 Turkish news texts in 5 classes, in raw and arff formats.	1150haber.rar	3.7 MB
630 Articles	630 articles from 18 different Turkish authors, in raw and arff formats.	630koseyazisi.zip	3.5 MB
270 Articles	270 articles from 18 different Turkish authors, in raw and arff formats.	270koseyazisi.zip	1.6 MB
140 Poems	140 poems from 7 different Turkish poets, in raw and arff formats.	140sir.zip	0.4 MB
90 Articles	90 articles from 9 different Turkish authors, in raw and arff formats.	90koseyazisi.zip	0.7 MB

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

32

Kemik Group (YTU)

Dataset: 2000 parsed Turkish sentences

✓ Example sentence:

Akşam	Zaman
çay	DT
bahçesinde	DT
arkadaşıyla	İle
buluşacak	Y
.	O

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

35

Kemik Group (YTU)

✓ Datasets

Dataset Name	Explanation	Download	Size
75 News	75 Turkish news texts in 5 classes, in raw and arff formats.	75haber.zip	0.4 MB
Word Meanings	The meanings of Turkish 10535 words.	Kelime_anlamlari.zip	0.3 MB
Term Groups	2750 Turkish terms in 27 term groups.	Terim_gruplari.zip	0.1 MB
Semantic Roles	5849 Turkish words, word types, suffixes and semantic roles in arff format.	semanikRol.zip	0.1 MB
28567 Turkish Web Pages	28567 Turkish web pages in 758 hierarchical classes under 13 main category.	tk_original.zip	173 MB
Turkish syllable statistics	Turkish syllable statistics extracted from 24 million Turkish words	hece_ist.zip	0.2 MB
Turkish Semantic Relation Database	127203 semantic relations automatically derived from TDK and Wik dictionaries.	TDK_viki.rar	0.8 MB

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

33

Kemik Group (YTU)

Dataset: 2000 parsed Turkish sentences

✓ Example sentence (with morphological labels):

Akşam	000...000	Zaman
çay	010...000	DT
bahçesinde	010...000	DT
arkadaşıyla	010...000	İle
buluşacak	100...000	Y
.	000...000	O

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

36

Kemik Group (YTU)

Dataset: 2000 parsed Turkish sentences

✓ 113 morphological labels

FIL_KOKISIM_KOKOZEL_KOKFIIIL_DONUSUM_ENBAGLAC_KOKISIM_DONUSUM_LEISIM_BIRLIKTELİK_LE
 FIIIL_EDILGENSESULI_NISIM_OOGUL_LERISIM_KISI_ONLAR_LERISIM_BULUNMA_LIISIM_KALMA_DEISM_I
 USIKULI_SELISIM_TAMLAMA_IISIM_SAHİPLİK_Ö_IJAMIR_KOKFIIIL_GENİSZAMAN_IRFIIIL_YETERSİZLİK_EFII
 L_İSTEK_EISIM_YONELME_EISIM_SAHİPLİK_SEN_INFIIIL_SUREKLİLİK_EREKFIIIL_SIMDIRKIZAMAN_IYORFIIIL
 L_KISI_SİZFIIIL_BERABERLİK_İSFIIIL_GECMİSZAMAN_DIFIIIL_KISI_BENİMEK_HİKAYE_DIISIM_YOKLULUK_SIZI
 SİM_SAHİPLİK_ONLAR_LERIFIIIL_ETTIRGEN_TIRFIIIL_MASTAR_MEKFIIIL_GECMİSZAMAN_MISFIIIL_DONUS
 UM_MISISIM_BELİRTME_IFIIIL_MASTAR_CESİM_TARAINDAN_CEFIIIL_ZAMAN_INCEISIM_BULUNMA_LIKI
 SİM_DURUM_LIKFIIIL_DONUSUM_MEFIIIL_OLUMSUZLUK_MEFIIIL_DONUSUM_İSİMEK_RVAYET_MISISIM_K
 İSİ_BİZ_İZFIIIL_EMİR_SİZ_İNİSİM_TAMLAMA_INFIIIL_EDILGEN_ILSORU_KOKİSİM_BULUNMA_KİSİM_SAHİP
 LİK_BEN_İMİSİM_ILGI_CIFIIIL_TANIMLAMA_ICFIIIL_KISI_SENFIIIL_BELİRTME_DIKİSİM_CIKMA_DENZAMAN
 _KOKSAYI_KOKFIIIL_DONUSUM_ECEKFIIL_ÖLECEKZAMAN_ECEKEDAT_KOKFIIIL_ÖLDÜRĞAN_TİSİM_SAH
 İPLİK_BİZ_İMİZMEK_ZAMAN_KENİMEK_KOKMEK_SART_SEFIIIL_DONUSUM_MİSİM_TANIMLAMA_DIRFIIIL
 _KISI_ONLARSAYI_KESİR_DÉFIIIL_SART_SEFIIIL_DONUSUM_MEZİSİM_KISI_BEN_İMİSİM_KISI_SEN_SINFIIIL
 _L_DONUSUM_ILİSİM_GIBI_CEFIIIL_YETENEK_EBİLFIIL_BERİ_ELİSİM_DONUSUM_LESFIIIL_EMİR_O_SINFIIIL
 _L_DONUSUM_INTIFIIIL_İMSİ_IPFIIIL_DONUSUM_ESİSİM_SAHİPLİK_SİZ_İNİZFIIIL_ZORUNLULUK_MELUFIIIL
 DONUSUM_İKSAYI_SIRA_İNCİZAMAN_BELİRTME_KİSİM_KÜÇÜLTME_CIKISIM_ANDIRMA_SIFIIIL_DEVAMLİ
 LİK_DIKCEFIIIL_EMİR_ONLAR_SINLERFIIIL_OLUMSUZLUK_DENFIIIL_KISI_SİZİSİM_KISI_SİZ_SİNİZZAMİR_S
 AHİPLİK_İMİSİM_KÜÇÜLTME_CEGİZSAYI_ÜLESTİRME_ERFIIIL_EMİR_SÜRESİM_İNİZFIIIL_ETTIRGEN_TEK
 RAR_TFIIIL_OLUMSUZLUK_SİZİNFIIL_DONUSUM_ESİYEFIIIL_SÜRERLİK_EGELYANKI_KOKFIIIL_SÜRERLİK
 EDÜRYANKI_DONUSUM_TİFIIIL_GIBI_CESİNEFIIIL_TEZLİK_IVERİSİM_ANDIRMA_İMSİSAYI_TOPLULUK_İZÜ
 NLEM_KOKYANKI_DONUSUM_DASAYI_KOSE_GEN

Kemik Group (YTU)

✓ Dataset: Turkish Syllable Statistics

- ✓ Collected from 24M words
- ✓ 2500 syllables obtained
- ✓ Noise reduction using a threshold

Kemik Group (YTU)

Dataset: 2M Turkish tweets

- ✓ In the form: date – user – tweet
- ✓ Collected during July-August 2012 period

Dataset: 2500 Columns

- ✓ 50 columnists
- ✓ 50 texts for each
- ✓ Collected during 2012-2013 period

Dataset: Word meanings

- ✓ An online dictionary containing 10.5K words and their meanings

Kemik Group (YTU)

✓ Software and Tools

Software	Explanation	Download	Size
Extended Zemberek	Adding new words to Zemberek Dictionary	ExtendedZemberek.zip	845KB
Text2arff v3.1	A program that automatically extracts various features of Turkish texts and saves in arff format	text2arffV3.zip	3 MB
Author Recognition	Author Recognition is done using Artificial Neural Networks. Also, for use in Wek ARFF format files is created.	Author Recognition SetUp	5.98 MB
Document Similarity	İki doküman arasındaki benzerliğin Jacard, Cosine, Dice ve Tanimoto benzerlikleri kullanılarak ölçülmektedir.	Document Similarity (updating)	2.46 MB
Haberci	News Summarization and Archiving Program	haberci.zip	1.5 MB
Text2arffEng V1.0	A program that automatically extracts various features of English texts and saves in arff format.	text2arffEng.zip	4 MB
Text2arff v2.0	A program that automatically extracts various features of Turkish texts and saves in arff format.	text2arff.zip	2.9 MB

Kemik Group (YTU)

Dataset: Term Groups

- ✓ 27 disciplines
- ✓ 2750 terms grouped wrt discipline

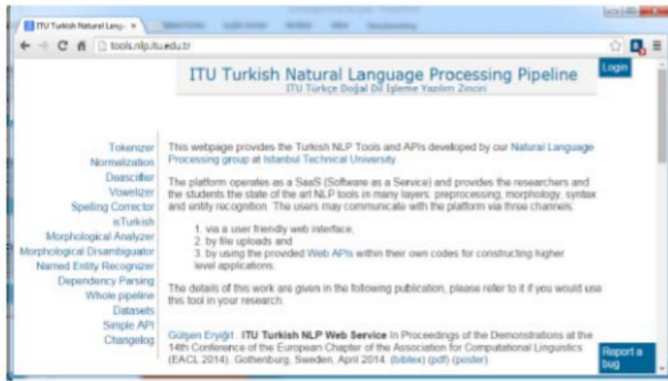
Class	Frequency	Class	Frequency	Class	Frequency
anatomi	100	matematik	202	din	156
gramer	45	astronomi	76	denizcilik	35
zooloji	132	fizik	182	cografya	58
ticaret	68	yiyecek	123	botanik	88
tarih	140	evesyasi	78	biyoloji	73
spor	126	hukuk	84	bilgisim	30
sosyoloji	65	felsefe	154	askerlik	127
psikoloji	81	ekonomi	62		
muzik	107	tip	157		
kimya	132	edebiyat	69		

Kemik Group (YTU)

✓ Software and Tools

Software	Explanation	Download	Size
Text Summarization	A program that automatically summarizes Turkish texts.	metin_ozetleme.rar	4 MB
Text Similarity	A program that calculates Turkish text similarities to each other. It works on txt, doc and pdf file formats.	doküman_benzerligi.zip	4.8 MB
Noun Phrase Extraction from Html documents	A program that automatically extracts the noun phrases in Turkish html web collections.	isim_tamlama.zip	4.8 MB
N-Gram and Syllable Statistics	A program that calculates N-gram and syllable statistics either from Turkish dictionary of spelling or any given Turkish text.	ngram_hoce_istatistik.zip	303KB
Focused Web Spider	A Customizable Focused Web Spider Specialized for Turkish Language	fws.zip	2.96 MB
Abstract Feature Extractor (AFE)	A program that extracts features from an input dataset of text documents using abstract feature extraction (AFE).	yfu.zip	24 KB
Author Detection	An Author Detection program which utilizes Artificial Immune System and Anti-Colony Optimization Methods	author_detection.zip	7.07 MB

ITU Turkish NLP Pipeline



ITU Turkish NLP Pipeline

Preprocessing

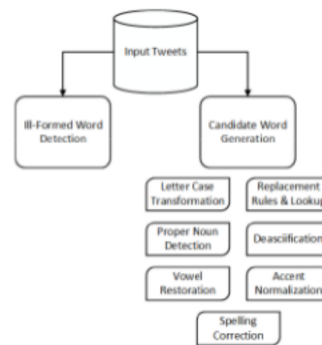
- ✓ A cascaded approach for text normalization
- ✓ Two stages:
 - ✓ Ill-formed word detection
 - ✓ Using morphological analyzer and abbreviation list
 - ✓ Candidate word generation
 - ✓ Rule-based and machine learning
 - ✓ Seven normalization layers

ITU Turkish NLP Pipeline

- ✓ Preprocessing
 - ✓ Tokenizer
 - ✓ Deascifier (diacritic restorer)
 - ✓ Vowelizer
 - ✓ Spelling corrector
 - ✓ Normalizer
- ✓ Morphological processing
 - ✓ Morphological analyzer
 - ✓ Morphological disambiguator
- ✓ Multiword expression handling
 - ✓ Named entity recognizer
- ✓ Syntactic processing
 - ✓ Dependency parsing

ITU Turkish NLP Pipeline

✓ Different layers of normalization



ITU Turkish NLP Pipeline

- ✓ User interface via three channels:
 - ✓ Web interface
 - ✓ Uploading a file
 - ✓ Using Web APIs

ITU Turkish NLP Pipeline

Preprocessing

- ✓ Ill-formed word detection: 86%
- ✓ Candidate generation: 71%

ITU Turkish NLP Pipeline

Preprocessing

- ✓ Vowel and diacritic restoration
- ✓ Two stage model:
 - ✓ Discriminative sequence classifier
 - ✓ Uses CRF to produce the most probable words
 - ✓ Views as a character level sequence labeling task
 - ✓ Language validator
 - ✓ Uses morphological analyzer to select the most probable valid word among the n-best sequences

April 4, 2016

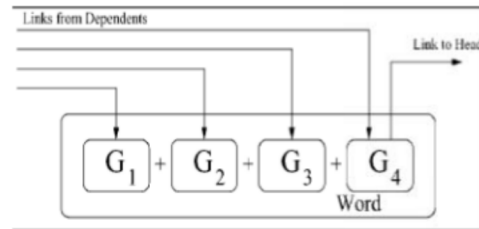
Turkic Computational Linguistics - TurCLing
2016

49

ITU Turkish NLP Pipeline

Dependency parser

- ✓ State-of-the-art syntactic parser for Turkish
- ✓ Dependencies based on inflectional groups rather than words



April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

52

ITU Turkish NLP Pipeline

Preprocessing

- ✓ Spell correction
- ✓ Finding words in a dictionary most similar to the misspelled word
- ✓ A probabilistic model is trained
- ✓ A set of rules for transformation are defined

April 4, 2016

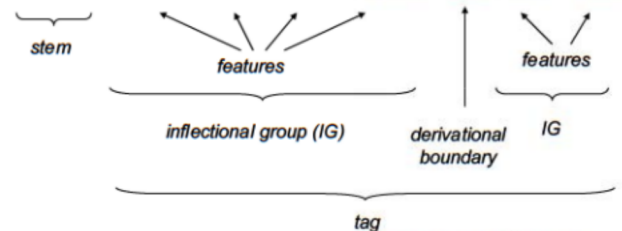
Turkic Computational Linguistics - TurCLing
2016

50

ITU Turkish NLP Pipeline

Dependency parser

- ✓ **masa** + Noun + A3sg + Pnon + Nom ^DB + Adj + With



- ✓ IGs can independently act as heads or modifiers in syntactic dependencies

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

53

ITU Turkish NLP Pipeline

Named entity recognizer

- ✓ Entity types:

person	date	money
location	time	percentage
organization		

- ✓ Uses CRF as the probabilistic model
- ✓ Raw tags & IOB tags
- ✓ Uses morphological processor and gazetteers
- ✓ Morphological, lexical, and gazetteer features

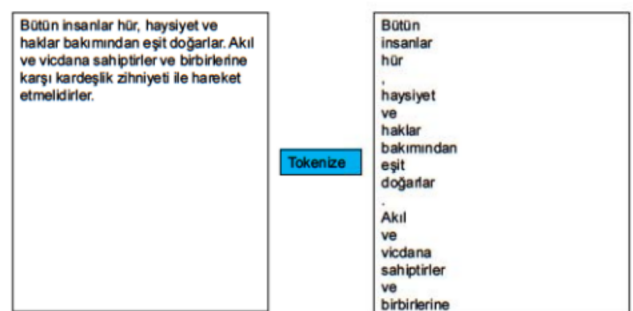
April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

51

ITU Turkish NLP Pipeline

Preprocessing – Tokenizer



April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

54

ITU Turkish NLP Pipeline

Preprocessing – Morphological Disambiguator

<S> <S>+BSTag
 Bütün bütün+Adj bütün+Noun+A3sg+Pnon+Nom Bütün+Noun+Prop+A3sg+Pnon+Nom
 insanlar insan+Noun+A3pl+Pnon+Nom
 insan+Noun+A3sg+Pnon+Nom+DB+Verb+Zero+Pres+A3pl
 hür hür+Adj hür+Noun+A3sg+Pnon+Nom
 . .+Punc
 haysiyet haysiyet+Noun+A3sg+Pnon+Nom
 ve ve+Conj

Disambiguate

<S> <S>+BSTag
 Bütün bütün+Adj
 insanlar insan+Noun+A3pl+Pnon+Nom
 hür hür+Adj
 . .+Punc
 haysiyet haysiyet+Noun+A3sg+Pnon+Nom
 ve ve+Conj
 haklar hak+Noun+A3pl+Pnon+Nom

ITU Turkish NLP Pipeline



ITU Turkish NLP Pipeline

Preprocessing – Named Entity Recognizer

<DOC> <DOC>+BDTag
 <S> <S>+BSTag
 Başkan başkan+Noun+A3sg+Pnon+Nom
 İbrahim İbrahim+Noun+Prop+A3sg+Pnon+N
 om
 Turhan Turhan+Noun+Prop+A3sg+Pnon+No
 m
 . .+Punc
 İzmirde İzmir+Noun+Prop+A3sg+Pnon+Loc
 saat saat+Noun+A3sg+Pnon+Nom
 14:00'te 14:00'te+?
 düzenlediği düzenle+Verb+Pos+DB+Adj+PastPar
 t+P3sg

Send to Ner

<DOC> <DOC>+BDTag
 O
 <S> <S>+BSTag O
 Başkan başkan+Noun+A3sg+Pnon+Nom O
 İbrahim İbrahim+Noun+Prop+A3sg+Pnon+N
 om B-PERSON
 Turhan Turhan+Noun+Prop+A3sg+Pnon+No
 m I-PERSON
 . .+Punc O
 İzmirde İzmir+Noun+Prop+A3sg+Pnon+Loc
 saat saat+Noun+A3sg+Pnon+Nom
 B-LOCATION
 14:00'te 14:00'te+?
 B-TIME

Turkish Labeled Text Corpus

- ✓ Collected from Turkish academic papers
- ✓ Titles, abstracts, and keywords
- ✓ 35 disciplines
- ✓ Topic classification wrt ULAKBİM *Ulusal Veri Tabanları* (National Data Sources) classification
- ✓ 5 main data sources:
 - ✓ Social and humanities (16 classes)
 - ✓ Medical (4 classes)
 - ✓ Life sciences (6 classes)
 - ✓ Engineering and natural sciences (8 classes)
 - ✓ Law (1 class)

ITU Turkish NLP Pipeline

Preprocessing – Dependency Parser

1	Bütün	bütün	Adj	Adj	-
2	insanlar	insan	Noun	Noun	A3pl Pnon Nom
3	hür	hür	Adj	Adj	-
4	.	.	Punc	Punc	-
5	haysiyet	haysiyet	Noun	Noun	A3sg Pnon Nom
6	ve	ve	Conj	Conj	-
7	haklar	hak	Noun	Noun	A3pl Pnon Nom
8	bakımından	bakım	Noun	Noun	A3sg P3sg Abl

Disambiguate

1	Bütün	bütün	Adj	Adj	-	2	MODIFIER
2	insanlar	insan	Noun	Noun	A3pl Pnon Nom	3	SUBJECT
3	hür	hür	Adj	Adj	-	10	COORDINATION
4	.	.	Punc	Punc	-	3	PUNCTUATION
5	haysiyet	haysiyet	Noun	Noun	A3sg Pnon Nom	7	COORDINATION

Turkish Labeled Text Corpus

- ✓ 200 documents/subject
- ✓ Each document is in XML form

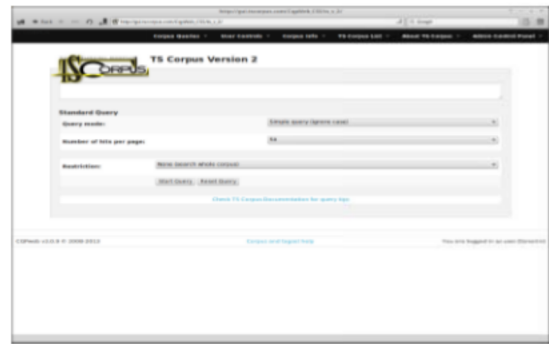
```
<makale>
<Etiket>Sınıf Etiket1</Etiket>
<Başlık>Makale Başlığı<Başlık>
<Özet>Bu özet, Türkçe Etiketli Metin
Derlemi formatına örnek olması için
olusturulmuştur. Derlemedeki her bir
belge bu xml formatında
saklanmaktadır. Bütün alanlardaki
veriler birer örnektir.</Özet>
<Anahtar>Metin Derlemi, xml örneği,
Türkçe</Anahtar>
<Kaynak>Metin Derlemi Konferansı 2014
Bildiriler Kitabı</Kaynak>
</makale>
```

TS Corpus

- ✓ A project that aims at building Turkish corpora, NLP tools, and linguistic datasets.
- ✓ Started in 2011, first release published on March 2012

TS Corpus – Main corpus

- ✓ Corpus can be queried online



TS Corpus

- ✓ General purpose corpora
 - ✓ TS Corpus
 - ✓ Wikipedia
 - ✓ Tweets
 - ✓ Idioms and proverbs
- ✓ Specialized corpora
 - ✓ Abstracts
 - ✓ Newspaper corpus
 - ✓ Gezi corpus
 - ✓ Constitutions

TS Corpus – Main corpus

- ✓ Corpus can be queried online
 - ✓ Simple query mod
 - ✓ CQP syntax

P-attributes in this corpus:	
word	Main word-token attribute
Lemma	Lemma(using tr_morph)
Morph	Morph(using tr_morph)
PosTag	PosTag(using tr_morph)

S-attributes in this corpus:	
<text>	Text
<text_id>	Text ID

TS Corpus – Main corpus

- ✓ Based on Boun Web Corpus
- ✓ 491M tokens and 4.9M word types
- ✓ Words with POS tags and morphological annotations
- ✓ Based on CWB (IMS Open Corpus Workbench) CQP (Corpus Query Processor) structure
 - ✓ Queries can be written using regular expressions or CQP commands

TS Corpus – Main corpus

- ✓ Corpus can be queried online
 - ✓ Simple query mod
 - ✓ CQP syntax



TS Gezi Corpus

- ✓ Collected from Turkish and foreign press, during Taksim Gezi Park protests
- ✓ 9 sources (BBC, Cumhuriyet, France 24, Hürriyet, NY Times, Spiegel, Telegraph, Washington Post, Zaman)
- ✓ 2,968 articles
- ✓ 1.1M tokens
- ✓ Specialized corpus
- ✓ Covers a period of 1 month
- ✓ Can be queried wrt source, language, genre, period

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

79

TS Idioms and Proverbs Corpus

No	Filename	Solution 1 to 11	Page 1 / 1
1	ts_proverb_and_idioms_corpus	0.00	avrat dümdüz girmek
2	ts_proverb_and_idioms_corpus	0.00	baba günü
3	ts_proverb_and_idioms_corpus	0.00	besler hurmayla eloğlu kaşşar yamayla
4	ts_proverb_and_idioms_corpus	0.00	gezer kız gezer , bu çeyizi kim dözer
5	ts_proverb_and_idioms_corpus	0.00	gibi yer olmaz bağdat gibi diyar olmaz
6	ts_proverb_and_idioms_corpus	0.00	hakko tanı hakko
7	ts_proverb_and_idioms_corpus	0.00	kızna tahtı kuter kız bahı kocadan arar
8	ts_proverb_and_idioms_corpus	0.00	kızna tahtı kumuş baht kuramamış
9	ts_proverb_and_idioms_corpus	0.00	ylan edü yılan , kan çiçek edü gerçek
10	ts_proverb_and_idioms_corpus	0.00	yürektan yana
11	ts_proverb_and_idioms_corpus	teyaz	0.00 yansır

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

82

TS Constitution Corpus

- ✓ Collection of three constitutions (1924, 1961, 1982)
- ✓ 32M tokens
- ✓ 5.3K word types
- ✓ Grouped wrt constitution and article number
- ✓ Can be queried wrt constitution year

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

80

Other Resources

- ✓ Zemberek
<https://github.com/ahmetaa/zemberek-nlp>
- ✓ TRmorph and other tools
<https://github.com/coltekin/>
- ✓ Morphological disambiguator (D. Yuret)
<http://www.denizyuret.com/2006/06/learning-morphological-disambiguation.html>

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

83

TS Idioms and Proverbs Corpus

- ✓ 10K idioms and proverbs
- ✓ 27.3K tokens
- ✓ Query returns the full idiom or proverb

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

81

Conclusions

- ✓ Currently available NLP resources for Turkish language
 - ✓ Semantic relations knowledge base
 - ✓ Treebank
 - ✓ Morphological analysis and disambiguation
 - ✓ Named entity recognition
 - ✓ Text normalizer
 - ✓ Syntactic analysis
 - ✓ Various corpora

April 4, 2016

Turkic Computational Linguistics - TurCLing
2016

84

Paper List

- Hajim Sak, Tunga Göngr, and Murat Saracilar: "Resources for Turkish Morphological Processing". Language Resources and Evaluation, Vol. 45, No. 2, pp. 249-261, 2011.
- Hajim Sak, Murat Saracilar, and Tunga Göngr: "On-the-fly Lattice Rescoring for Real-time Automatic Speech Recognition". INTERSPEECH 2010, pp. 2450-2453, 2010.
- Hajim Sak, Murat Saracilar, and Tunga Göngr: "Morphology-based and Sub-word Language Modeling for Turkish Speech Recognition". Acoustics Speech and Signal Processing (ICASSP), pp. 5403-5406, 2010.
- Hajim Sak, Murat Saracilar, and Tunga Göngr: "Integrating Morphology into Automatic Speech Recognition". IEEE Workshop on Automatic Speech Recognition & Understanding (ASRU), 2009, pp. 334-339.
- Ebu Ansoy, Doğan Can, Siddika Parlak, Hajim Sak, and Murat Saracilar: "Turkish Broadcast News Transcription and Retrieval". IEEE Transactions on Audio, Speech & Language Processing, Vol. 17, No. 5, pp. 874-883, 2009.
- Hajim Sak, Tunga Göngr, and Murat Saracilar: "Turkish Language Resource: Morphological Parser, Morphological Disambiguator and Web Corpus". In GoTAL 2008, volume 5221 of LNCS, 2008, pages 417-427. Springer.
- E. Ansoy, M. Kurnia, M. Saracilar, T. Hirsimäki, J. Pylkönen, T. Alama and H. Sak. 2008. "Statistical Language Modeling for Automatic Speech Recognition of Agglutinative Languages". Speech Recognition: Techniques, Technology and Applications, I-Tech Education and Publishing, Vienna, Austria.
- Tuncay Akauçur, Siddika Parlak, Hajim Sak, Murat Saracilar. "Turçe Haber Programları İçin Di Modelleme Yakaçınlarını Karşılaması". IEEE 19. Sinyal İşleme, İletişim ve Uygulama Konferansı (SIU), Didim, Türkiye, 2008.
- Ebu Ansoy, Hajim Sak, and Murat Saracilar. "Language modeling for automatic Turkish broadcast news transcription". In Proceedings of Interspeech 2007 - Europe, pp. 2391-2394, 2007. [PDF](#) [Bibtex](#)
- Hajim Sak, Tunga Göngr, and Murat Saracilar. "Morphological disambiguation of Turkish text with perceptron algorithm". In ICiLING 2007, volume LNCS 4394, pages 107-118, 2007. [PDF](#) [Bibtex](#)
- Hajim Sak, Tunga Göngr, and Yapar Saifan. "A corpus-based concatenative speech synthesis system for Turkish". Turkish Journal of Electrical Engineering and Computer Sciences, 14(2):209-223, 2006. [PDF](#) [Bibtex](#) [Gizli](#) [TTS outloud](#)
- Hajim Sak, Tunga Göngr, and Yapar Saifan. "Generation of synthetic speech from Turkish text". In 13th European Signal Processing Conference (EUSIPCO 2005), 2005. [PDF](#) [Bibtex](#)
- Zeki Göngr, Kemal Oflaz. "Parating Turkish using the lexical functional grammar formalism". 1994.
- Göhan Tur, Dilek Hakkıoğlu, Kemal Oflaz. "A statistical information extraction system for Turkish". 2003.
- Deniz Yuret, Ferhan T. "Learning morphological disambiguation rules for Turkish". 2009.
- Oflaz, K., Tur, G. "Morphological disambiguation by voting constraints". 1997.

Turkic Computational Linguistics - TurCLing

April 4, 2016

2016

85

Paper List

- Altın, B., Güz, M.C., Dri, B. "A Semantic Kernel for Text Classification Based on Iterative Higher-Order Relations Between Words and Documents", 13th International Conference on Artificial Intelligence and Soft Computing, ICASIS 2014, Zakopane, Poland, 2014. [PDF](#)
- Altın, B., Güz, M.C., Dri, B. "A Novel Higher-Order Semantic Kernel for Text Classification", 10th International Conference on Electronic Computer and Computation, ICECCO 2013, Ankara, Turkey, 2013. [PDF](#)
- Dural, B., Dri, B. "Türçe Arama Motoru Kümelenme Sonucu Çıkarılması", IEEE 21st Sinyal İşleme ve İletişim Uygulama Kurultayı, SIU 2013, Kibris, 2013. [PDF](#) (in Turkish)
- Çetin, M., Amasyalı, M.F., "Eğitilmiş ve Genelleştirilmiş Terim Ağlarında Yöntemlerle Duygu Analizi", IEEE 21st Sinyal İşleme ve İletişim Uygulama Kurultayı, SIU 2013, KKTCC, 2013. [PDF](#) (in Turkish)
- Brinç, G., Dri, B., Sönmec, C. "Abstract Feature Extraction for Text Classification", Turkish Journal of Electrical Engineering and Computer Sciences, Vol.20, No.5, pp. 1037-1059, 2012. [PDF](#)
- Amasyalı, M.F., Balo, S., Yurt, E.M., Kılıç, E. "Türçe Metinlerin Sınıflandırılmasında Metin Temsil Yöntemlerinin Performans Karşılaştırılması", EMO Bilimsel Dergi, Cilt 2, Sayı 4, Sayı 95-104, 2012. [PDF](#) (in Turkish)
- Çiğiz, M. Ö., Dri, B. "Mikroblog Kullanıcılarının Kategorizasyonu", IEEE 20. Sinyal İşleme ve İletişim Uygulama Kurultayı, SIU 2012, Fethiye, Türkiye, 2012. [PDF](#) (in Turkish)
- Brinç, G. "Sesli Bilgiyi Kullanarak Boyut İndirgeme Yöntemlerinin Metin Sınıflandırılmasında Etkilerinin Karşılaştırılması", IEEE 20. Sinyal İşleme ve İletişim Uygulama Kurultayı, SIU 2012, Fethiye, Türkiye, 2012. [PDF](#) (in Turkish)
- Türkmen, H., Dri, B., Brinç, G., Doğan, R. "Konuşma Dil Kullanılarak Demografik Bilgilerin Sınıflandırılması", IEEE 19. Sinyal İşleme ve İletişim Uygulama Kurultayı, SIU 2011, Antalya, Türkiye, 2011. [PDF](#) (in Turkish)
- Doğan, S., Dri, B. "Türçe Dokümanları İçin N-Gram Tabanlı Yeni Bir Sınıflandırma: Yazı, Tür ve Cinsiyet", Türkiye Bilimsel Vakti Bilgiye Bilimleri ve Mühendisliği Dergisi, Vol.3, s.11-20, 2010. [PDF](#) (in Turkish)
- Amasyalı, M.F., Davletov, F., Torayev, A., Çiftçi, Ü. "Türçe Metinleri İçin Öznelik Çıkarım Yazarı", IEEE 19. Sinyal İşleme ve İletişim Uygulama Kurultayı, SIU 2010, Diyarbakır, Türkiye, 2010. [PDF](#) (in Turkish)
- Brinç, G., Dri, B. "Impact of a New Attribute Extraction Algorithm on Web Page Classification", 5th International Conference on Data Mining, Las Vegas, Nevada, ASD, 2009. [PDF](#)
- Brinç, G., Dri, B., Sönmec, A.C.S. "A New Method For Attribute Extraction with Application on Text Classification", ICSCW-2009 Fifth International Conference on Soft Computing, Computing with Words and Perceptions in Application, Decision and Control, Gazimagosa, K.K.T.C., 2009. [PDF](#)
- Yıldız, H.K., Uta, N., Geçtavi, M., Dri, B., Amasyalı, M.F. "Metin Sınıflandırılmasında Yeni Bir Öznelik Çıkarım Metodu", 15. Sinyal İşleme ve Uygulama Kurultayı, SIU 2007, Eskişehir, Türkiye, 2007. [PDF](#) (in Turkish)

Turkic Computational Linguistics - TurCLing

April 4, 2016

2016

88

Paper List

- Levent, V.E., Dri, B. "Türçe Dokümanlarında Yayıncı Sınırları Açıkla ve Yazar Tanıma", Akademik Bilgi 2014, Mersin Üniverstesi, Mersin, 2014. [PDF](#) (in Turkish)
- Yaslı, M., Dri, B. "Author Recognition by Abstract Feature Extraction", IEEE 20th Signal Processing and Communications Applications Conference, SIU 2012, Fethiye, Turkey, 2008. [PDF](#) (in Turkish)
- Z.Kaban, B.Dri, "Genre and Author Detection in Turkish Texts Using Artificial Immune Recognition Systems", IEEE 18th Signal Processing and Communications Applications Conference, SIU, Aydın, Turkey, 2008. [PDF](#) (in Turkish)
- Türkoğlu, F., Dri, B., Amasyalı, M.F., "Author Attribution of Turkish Texts by Feature Mining", Third International Conference on Intelligent Computing, ICIC 2007, Guangzhou, China, 2007, LNCS Volume 468/12007 Springer. [PDF](#)
- Amasyalı, M.F., Dri, B. "Automatic Turkish Text Categorization in Terms of Author, Genre and Gender", 11th International Conference on Applications of Natural Language to Information Systems-NLDB2006, Austria, LNCS Volume 3999, Springer, 2006. [PDF](#)
- Türkoğlu, F., Dri, B., Amasyalı, M.F., "Yarık Öznelik Yöntemleri ile Türçe Dokümanları Yazarlarını Belirleme", 15th Turkish Symposium on Artificial Intelligence and Neural Network, TAINN-2006, Muğla, Türkiye, 2006. [PDF](#) (in Turkish)
- Amasyalı, M.F., Dri, B. "Identifying the Poets of the Anonymous Poem", International Twelfth Turkish Symposium on Artificial Intelligence and Neural Networks, TAINN, Çanakkale, Türkiye, 2003. [PDF](#) (in Turkish)
- Dri, B., Amasyalı, M.F. "Automatic Authorship Attribution in Turkish Language", 13th International Conference on Artificial Neural Network and 10th International Conference on Neural Information Processing, ICANN/ICNIP 2003, Istanbul, Turkey, 2003. [PDF](#)
- Safin, G., Amasyalı, M.F., "Derivative Information Extraction from Large Text Collections", "Genel Metin Koleksiyonlarından İlerit Bilgi Çıkarım", EMO Bilimsel Dergi, Cilt 4, Sayı 7, 13-20, 2014. [PDF](#) (in Turkish)
- Özcan, S., Amasyalı, M.F., "Turkish Commonsense Database (CSDb) and CSOYUN (A Game with a Purpose)", Sigma Journal of Engineering and Natural Sciences, Year: 2014, Volume: 32, Issue: 2, (June 2014), pp.116-127. [PDF](#) (in Turkish)
- Aygül, M., Karaslıoğlu, G., Amasyalı, M.F., "Kogulmuş Rasgele Alanlar ile Basit Cümlelerin Öğelerine Ayrılması", Sigma Mühendislik ve Fen Bilimsel Dergi, Cilt no: 32 Sayı: 19 Sayı: 23-30, 2014. [PDF](#) (in Turkish)
- Yener, Y., Kaplan, S.S., Amasyalı, M.F., "Nonrobot: Hayat Bilgisi Toplama Amacı Oynan", Akademik Bilgi, 2014, Mersin, Türkiye. [PDF](#) (in Turkish)
- Grinç, C., Dri, B. "Sesli Rasgele Alanlar ile Türçe Wikipea Sayfalarının Semantik İlişkilerinin Çıkarılması", IEEE 22. Sinyal İşleme ve İletişim Uygulama Kurultayı, SIU 2014, Trabzon, (23-25 Nisan), 2014. [PDF](#) (in Turkish)
- Yıldız, T., Dri, B., Yıldırım, S., "Analysis of Lexico-syntactic Patterns for Meronym Extraction from a Turkish Corpus", 6th Language & Technology Conference, Human Language Technologies as a Challenge for Computer Science and Linguistics, LTC 2013, Poznan, Poland. [PDF](#)

Turkic Computational Linguistics - TurCLing

April 4, 2016

2016

86

Paper List

- Şeker, Ş.E., Dri, B. "Davranışsal Türçe Metin Sınıflandırma Tasarımı ve Kodlanması", Akademi Sempozyumu Yenilikler ve Uygulamaları, ASYU, İstanbul, Turkey, 2006. [PDF](#) (in Turkish)
- Amasyalı, M.F., Yıldırım, T., "Otomatik Haber Metinleri Sınıflandırma", 13.Sinyal İşleme ve Uygulama Kurultayı, SIU 2004, Kuşadası, Türkiye, 2004. [PDF](#) (in Turkish)
- Ağan, C., Dri, B. "Aritmetik Bir Problemi Analiz ve Çözme: APAC", Akademi Sistemleri ve Yenilikler Uygulamaları, ASYU 2012, Trabzon, Türkiye, 2012. [PDF](#) (in Turkish)
- Topçu, S., Şen, C., Amasyalı, M.F., "Türçe Söhbet Robotu", Akademi Sistemleri ve Uygulamaları, ASYU 2012, Trabzon, Türkiye, 2012. [PDF](#) (in Turkish)
- Amasyalı, M.F., Dri, B. "Bir Soru Cevaplama Sistemi: BayBilim", Türkiye Bilimsel Vakti Bilgiye Bilimleri ve Mühendisliği Dergisi, Sayı:37-51, 2009. [PDF](#) (in Turkish)
- Yüksel, Ç., Çabuk, H., Moan, Z., Amasyalı, M.F., Dri, B. "Metin Analizi ve Sorgulama (MAVİB)", 11. Sinyal İşleme ve Uygulama Kurultayı, SIU 2003, İstanbul, Türkiye, 2003. [PDF](#) (in Turkish)
- Yüksel, Ç., Çabuk, H., Moan, Z., Amasyalı, M.F., Dri, B. "Text Analyzing and Interrogating", Xth International Twelfth Turkish Symposium on Artificial Intelligence and Neural Networks, TAINN, Çanakkale, Türkiye, 2003. [PDF](#) (in Turkish)
- Göran, A., Arslan, S.N., Kılıç, E., Dri, B. "Metin Özetleme İçin Çözümlü Seçim Metotları", IEEE 22.Sinyal İşleme ve İletişim Uygulama Kurultayı, SIU 2014, Trabzon, 2014. [PDF](#) (in Turkish)
- Sami, M., Dri, B. "HTML Dokümanların Otomatik Özetlenmesi", Akademi Sistemleri ve Uygulamaları, ASYU, Kayseri, 2010. [PDF](#) (in Turkish)
- Uzundere, E., Değde, E., Dri, B., Amasyalı, M.F., "Türçe Haber Metinleri İçin Otomatik Özetleme", Akademi Sistemleri ve Yenilikler ve Uygulamaları Sempozyumu, ASYU 2008, Isparta, Türkiye, 2008. [PDF](#) (in Turkish)
- Gökkuş, H., Dri, B. "Bilgi Bilime Dayalı Doküman Sınıflandırma", 18. Sinyal İşleme ve Uygulama Kurultayı, SIU 2010, Diyarbakır, Türkiye, 2010. [PDF](#) (in Turkish)
- Gökkuş, H., Dri, B. "Bilgi Bilime Dayalı Doküman Sınıflandırma", DED Mühendislik Bilimleri Dergisi, Cilt 12, Sayı 3, s. 77-85, 2010. [PDF](#) (in Turkish)
- Dri, B., Karagül, M.Y., "Design of a Genetic Algorithm for Constructing the Mathematical Model of Frequency Distribution of Words in Turkish Documents", The 18th International Symposium on Computer and Information Sciences, ISICIS, Antalya, Turkey, 2001
- Dri, B., Karagül, M.Y., "Türk Dilinin Bilgi Bilim Yapısından Yararlanarak Türçe Metinlerin Farklı İlgilere Ayrılması ve Kodlanması", 9. Sinyal İşleme ve Uygulama Kurultayı, SIU-2001, Gazı Magosa, Kıbrıs, 2001. [PDF](#) (in Turkish)

Turkic Computational Linguistics - TurCLing

April 4, 2016

2016

89

Paper List

- Yıldız, T., Yıldırım, S., Dri, B. "An Integrated Approach to Automatic Synonym Detection in Turkish Corpus", 9th International Conference on Natural Language Processing, PoNLP, Springer LNLP proceedings, Warsaw, Poland, (17-19 September), 2014. [PDF](#)
- Yıldız, T., Dri, B., Yıldırım, S., "Analysis of Lexico-syntactic Patterns for Meronym Extraction from a Turkish Corpus", 6th Language & Technology Conference, Human Language Technologies as a Challenge for Computer Science and Linguistics, LTC 2013, Poznan, Poland. [PDF](#)
- Yıldız, T., Yıldırım, S., Dri, B. "Extraction of Part-Whole Relations from Turkish Corpus", Computational Linguistics and Intelligent Text Processing, CICLing, Greece, 2013. [PDF](#)
- Özkoş, C., Amasyalı, M.F., "Türçe Öğelerinden Hayat Bilgisi Çıkarım", Türkiye Bilimsel Vakti, Bilgiye Bilimleri ve Mühendisliği Dergisi, Sayı:09, Aralık 2012. [PDF](#) (in Turkish)
- Temizer, M., Dri, B., "Automatic Subject-Object-Verb Relation Extraction", International Symposium on Innovations in Intelligent Systems and Applications, INISTA, Trabzon, 2012. [PDF](#)
- Amasyalı, M.F., Yener, Y., Kaplan, S.S., "Kelime Yönelimleri", IEEE 20th Signal Processing and Communication Applications Conference, SIU 2012, Antalya, Türkiye. [PDF](#) (in Turkish)
- Amasyalı, M.F., Çetin, M., Akbulut, C., "A Comparison about Semantic Text Representations for Text Classification", Akademi Sistemleri ve Uygulamaları, ASYU 2012, Trabzon, Türkiye. [PDF](#)
- Yazar, E., Amasyalı, M.F., "Automatic Extraction of Semantic Relationships Using Turkish Dictionary Definitions", EMO Bilimsel Dergi, Vol. 1, No. 1, pp. 1-13, 2011. [PDF](#) (in Turkish)
- Özdemir, C., Amasyalı, M.F., "Automatic Sentence Generation Using Common Sense Databases", InETr 2010, İstanbul, Türkiye. [PDF](#) (in Turkish)
- Amasyalı, M.F., İnk, B., Ersen, M.Z., "Construction of Turkish Commonsense Database", Akademik Bilgi, AB- 2010, Muğla, Türkiye, 2010. [PDF](#) (in Turkish)
- Amasyalı, M.F., Beken, A., "Türçe Kelimelerin Anlamlar Benzerliklerinin Ölçülmesi ve Metin Sınıflandırılmasında Kullanılması", IEEE 17th Signal Processing and Communication Applications Conference, SIU 2009, Antalya, Türkiye. [PDF](#) (in Turkish)
- Amasyalı, M.F., "Arana Motorları Kullanarak Bulunan Anlamlar Benzerlik Ölçütüne Dayalı Kelime Sınıflandırılması", IEEE 14th Signal Processing and Communication Applications Conference, SIU 2006, Antalya, Türkiye. [PDF](#) (in Turkish)
- Amasyalı, M.F., "Türçe WordNet'in Otomatik Olarak Oluşturulması", IEEE 13th Signal Processing and Communication Applications Conference, SIU 2005, Kayseri, Türkiye, 2005. [PDF](#) (in Turkish)
- Altın, B., Güz, M.C., Dri, B., "A Simple Semantic Kernel Approach for SVM using Higher-Order Paths", International Symposium on Innovations in Intelligent Systems and Applications, INISTA, Italy, Bari, Albergo, 2014. [PDF](#)

Turkic Computational Linguistics - TurCLing

April 4, 2016

2016

87

Paper List

- Dri, B. "A Text Compression System Based on the Morphology of Turkish Language", 15th Int. Symposium on Computer and Information Sciences, ISICIS, İstanbul, Turkey, 2000
- Özger, Z.B., Dri, B., "Sınıflandırma Tabanlı Türçe Soru Algılama", Akademi Sistemleri ve Uygulamaları, ASYU, İzmir, 2014. [PDF](#) (in Turkish)
- Meriç, M., Dri, B., "Twitter Üzerinde Duygu Analizi", IEEE 22. Sinyal İşleme ve İletişim Uygulama Kurultayı, SIU 2014, Trabzon, 2014. [PDF](#) (in Turkish)
- Çiğiz, M.Ö., Dri, B., "Content Mining of Microblogs", State of the Art Applications of Social Network Analysis, Lecture Notes in Social Networks, p.359-369, Fazıl Can, Tansel Özyer, Faruk Polat (Eds.), Springer, Print ISBN978-3-319-05911-2, 2014. [PDF](#)
- Özger, Z.B., Dri, B., Grinç, C., "Question Identification on Turkish Tweets", International Symposium on Innovations in Intelligent Systems and Applications, INISTA, Italy, Bari, Albergo, 2014. [PDF](#)
- Çetin, M., Amasyalı, M.F., "Active Learning for Turkish Sentiment Analysis", International Symposium on Innovations in Intelligent Systems and Applications, INISTA, Bulgaria, 2013. [PDF](#)
- Kayahan, D., Serçin, A., Dri, B., "Twitter ile TV Program Rejtinglerinin Belirlenmesi", IEEE 21. Sinyal İşleme ve İletişim Uygulama Kurultayı, SIU 2013, Kibris, 2013. [PDF](#) (in Turkish)
- Çengiz, C., Dri, B., "Türçe Tweetlerin Soru İfadelerini Bulmak", IEEE 21. Sinyal İşleme ve İletişim Uygulama Kurultayı, SIU 2013, Kibris, 2013. [PDF](#) (in Turkish)
- Çiğiz, M. Ö., Dri, B., "Content Mining of Microblogs", International Conference on Advances in Social Network Analysis and Mining, FOSINT, İstanbul, Turkey, 2013. [PDF](#)
- Çiğiz, M. Ö., Dri, B., "Mikroblog Kullanıcılarının Sınıflandırılması", IEEE 20. Sinyal İşleme ve İletişim Uygulama Kurultayı, SIU 2012, Fethiye, Türkiye, 2012. [PDF](#) (in Turkish)
- Parlak, S., Dri, B., "Duygu Tanımlama Farklı Veri Setleri Arasında Bir Çalışma", Akademi Sistemleri ve Uygulamaları, ASYU, İzmir, 2014. [PDF](#) (in Turkish)
- Parlak, S., Dri, B., "İnan Sözlüğü Duygu Çıkarımı", IEEE 21. Sinyal İşleme ve İletişim Uygulama Kurultayı, SIU 2013, Kibris, 2013. [PDF](#) (in Turkish)
- Türkmen, H.J., Karagül, M.E., "Dyphonic Speech Reconstruction: Construction of a Novel System for Effective and Efficient Communication", IEEE Engineering in Medicine and Biology Magazine, Volume 29, pp. 135-138, 2010. [PDF](#)
- Türkmen, H.J., Karagül, M.E., "Dyphonic Speech Reconstruction: Construction of a Novel System for Effective and Efficient Communication", IEEE Engineering in Medicine and Biology Magazine, Volume 29, pp. 135-138, 2010. [PDF](#)
- Türkmen, H.J., Karagül, M.E., "Dyphonic Speech Reconstruction: Construction of a Novel System for Effective and Efficient Communication", IEEE Engineering in Medicine and Biology Magazine, Volume 29, pp. 135-138, 2010. [PDF](#)
- Türkmen, H.J., Karagül, M.E., "Dyphonic Speech Reconstruction: Construction of a Novel System for Effective and Efficient Communication", IEEE Engineering in Medicine and Biology Magazine, Volume 29, pp. 135-138, 2010. [PDF](#)
- Türkmen, H.J., Karagül, M.E., "Dyphonic Speech Reconstruction: Construction of a Novel System for Effective and Efficient Communication", IEEE Engineering in Medicine and Biology Magazine, Volume 29, pp. 135-138, 2010. [PDF](#)

Turkic Computational Linguistics - TurCLing

April 4, 2016

2016

90

Paper List

- Ozger, B.Z., Diri, B., "Türkiye Dokümanları İçin Kurallı Tabanlı Varkit İsmi Tanıma", Türkiye Bilgiyi Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 2012 [PDF](#) (in Turkish)
- Ozger, B., Diri, B., "Türkiye Metinlerinde Şartlı Rastgele Alanlar Varkit İsmi Tanıma", IEEE 19. Signal Processing and Communication Applications Conference, SIU 2011, Antalya, Turkey, 2011 [PDF](#) (in Turkish)
- Dalkılıç, F.E., Geçici, S., Diri, B., "Türkiye Kurallı Tabanlı Varkit İsmi Tanıma", IEEE 18. Sinyal İşleme ve Uygulama Kurultayı, Diyarbakır, Turkey, 2010 [PDF](#) (in Turkish)
- Yıldız, E., Tantiş, A.C., Diri, B., "The Effect of Parallel Corpus Quality vs Size in English-to-Turkish SMT", 4th International Conference on Computer Science, Engineering and Applications, ICCSEA, Chennai, India, 2014 [PDF](#)
- Katmaç, H., Diri, B., "moBraille", 19. Türkiye'de İnternet, İnter-4/14, İzmir, 2014 [PDF](#) (in Turkish)
- Yıldız Emay, Tantiş A.C., Diri, B., "İstatistiksel Bilgisayarlı Çeviride Paralel Derlemin Büyüme Ölçütü ve Kalitesinin Etkileri", Akülü Sistemlerde Yenilikler ve Uygulamaları, ASYU, İzmir, 2014 [PDF](#) (in Turkish)
- Özdemir, C., Diri, B., "Web-Ear, an Information Retrieval System that uses Reported Speech Expressions in Turkish", Int. Symposium on Innovations in Intelligent Systems and Applications, INISTA, Istanbul, Turkey, 2011 [PDF](#)
- Şenar, S.E., Diri, B., "Türkiye Metinleri İçin Olay Sınıflama", Türkiye Bilgiyi Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, Vol.4, s.63-74, 2011 [PDF](#) (in Turkish)
- Savaşan, S., Diri, B., "Türkiye İçeriklerinden Otomatik Etik Etiket Buluntu Oluşturma", Yıldız Teknik Üniversitesi, Sigma Dergisi, Cilt 29, Sayı 2, s. 196-199, 2011 [PDF](#) (in Turkish)
- Briçok, G., Dağdas, S., Diri, B., "A Focused Web Spider Specialized for Turkish Language", Int. Symposium on Innovations in Intelligent Systems and Applications, INISTA, Kayseri, Turkey, 2010 [PDF](#)
- Şenar, S.E., Diri, B., "TimeML and Turkish Temporal Logic", International Conference on Artificial Intelligence, ICAIT'10, USA, 2010 [PDF](#)
- Şenar, S.E., Diri, B., "Reichenbach'e Alınan Zaman Tanımları ve TimeML", Akademik Bilgiyi, Muğla, 2010 [PDF](#) (in Turkish)
- Dağcı, F., Ayaz, R., Diri, B., "Web Browser Word Based Translator", Int. Informatica Congress, BR, TEK, Eskişehir, Turkey, 2005 [PDF](#)
- İlgiz, S., Amasyalı, M.F., Taş, E., "Bir Türkçe Metindeki Kelimelerin Çinko İçerisinde Durumlarına Bağlı olarak Etkilerine Ayrılması", Bilgiyi Zihne/101, İstanbul, 2001
- Kemal Oflaz, and [Mural Sarıca](#) (Editors), *Studies in Turkish Language Processing*, in preparation for Springer Verlag – Theory and Applications of Natural Language Processing Book Series.
- Rahim Derikhanjan, [Yusuf Saçan](#), [Barın Yankıoğlu](#), and Kemal Oflaz, *SentiTurkNet: A Turkish Polarity Lexicon for Sentiment Analysis*, Language Resources and Evaluation (to appear)

Turkish Computational Linguistics - TurCLing

April 4, 2016

2016

91

Paper List

- [İsmail Duray El-Kahout](#) and Kemal Oflaz, *Exploiting Morphology and Local Word Reordering in English to Turkish Phrase-based Statistical Machine Translation*, IEEE Transactions on Audio, Speech and Language Processing, 2010
- [Göken Eryiğit](#), [Joakim Nivre](#), and Kemal Oflaz, *Dependency Parsing of Turkish*, Computational Linguistics, Vol.34 No.3, September 2008.
- [Göken Eryiğit](#), 2013. *Bilgi Bilimsel Çözümlere*, Türkiye Bilgiyi Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 633-55. [PDF](#) (bibtex)
- [Joakim Nivre](#), [John Hall](#), [Jens Nilsson](#), [Almas Chaney](#), [Göken Eryiğit](#), [Sandra Kübler](#), [Stefelav Marinov](#), and [Erwin Marsi](#), *Maltparser: A language-independent system for data-driven dependency parsing*, *Natural Language Engineering Journal*, 13 no. 2, pp. 95-35, 2007. [PDF](#) (bibtex)
- [Göken Eryiğit](#), [Joakim Nivre](#), and [Kemal Oflaz](#), *The incremental use of morphological information and lexicization in data-driven dependency parsing*, *Computer Processing of Oriental Languages, Beyond the Orient: The Research Challenges Ahead*, [Springer LNLI](#), 428, s. pages 499-507, 2008. [PDF](#) (bibtex)
- [Göken Eryiğit](#), [Cüneyd Tantiş](#) and [Egri Adalı](#), 2006. *Yarınca e-posta ilemlerinde, harar destek makinesini, naive bayes ve belgeli tabanlı öğrenme yöntemlerinin karşılaştırılması*, Türkiye Bilgiyi Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 1:27-36. [PDF](#) (bibtex)
- [Alper Tokgöz](#), [Göken Eryiğit](#), *Transition-based Dependency DAG Parsing Using Dynamic Oracles*, ACL-UCLNP SRW 2015, Beijing, China, July 2015.
- [Tuğba Pamay](#), [Umur Sübübaçak](#), [Dilare Torunoğlu-Selamet](#), [Göken Eryiğit](#), *The Annotation Process of the ITU Web Treebank*, The 9th Linguistic Annotation Workshop at NAACL, Colorado, USA, May 2015.
- [Göken Eryiğit](#), [Kübra Adalı](#), [Dilare Torunoğlu-Selamet](#), [Umur Sübübaçak](#), [Tuğba Pamay](#), *Annotation and Extraction of Multiword Expressions in Turkish Treebanks*, The 11th Workshop on Multword Expressions at NAACL, Colorado, USA, May 2015.
- [Ezgi Yıldırım](#), [Fatih Samet Çetin](#), [Göken Eryiğit](#), *Tanel Tanel: The Impact of NLP on Turkish Sentiment Analysis*, in *Proceedings of the TURKLANG14 International Conference on Turkish Language Processing*, Istanbul, 06-07 November 2014. [PDF](#) (bibtex)
- [Gökhan Çelikkaya](#), [Göken Eryiğit](#), *A Mobile Assistant for Turkish*, The Impact of NLP on Turkish Sentiment Analysis, in *Proceedings of the TURKLANG14 International Conference on Turkish Language Processing*, Istanbul, 06-07 November 2014. [PDF](#) (bibtex)
- [Göken Eryiğit](#), *ITU Turkish NLP Web Interface*, accepted for publication in *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014)*, Gothenburg, Sweden, April 2014. [PDF](#) (bibtex)
- [Dilare Torunoğlu](#) and [Göken Eryiğit](#), *A Cascaded Approach for Social Media Text Normalization of Turkish*, accepted for publication in *9th Workshop on Language Analysis for Social Media (LASM) at EACL*, Gothenburg, Sweden, April 2014. [PDF](#) (bibtex)

Turkish Computational Linguistics - TurCLing

April 4, 2016

2016

94

Paper List

- [Wajdi Zaghouani](#), [Nizar Habash](#), [Osama Obeid](#), [Behrang Mohit](#) and [Kemal Oflaz](#), *Building an Arabic Machine Translation Post-Edited Corpus: Guidelines and Annotation*, in *Proceedings LREC 2016*, Slovenia, May 2016
- [Wajdi Zaghouani](#), [Houda Bouamor](#), [Abdelati Hawari](#), [Mona Diab](#), [Osama Obeid](#), [Mahmoud Ghoneim](#), [Sawsan Alghatani](#), and [Kemal Oflaz](#), *Guidelines and Framework for a Large Scale Arabic Dialectized Corpus*, in *Proceedings LREC 2016*, Slovenia, May 2016
- [Houda Bouamor](#), [Hasan Sajad](#), [Nadir Durani](#), [Kemal Oflaz](#), *CMUQIG@QALB-2015 Shared Task: Combining Character Level MT and Error-Driven Finite-State Recognition for Arabic Spelling Correction* in *Proceedings of the Workshop on Arabic NLP*, Beijing China, August 2015.
- [Houda Bouamor](#), [Wajdi Zaghouani](#), [Osama Obeid](#), [Mona Diab](#) and [Kemal Oflaz](#), *A Pilot Study on Arabic Multi-Genre Corpus Dialectization*, in *Proceedings of the Workshop on Arabic NLP*, Beijing China, August 2015.
- [Wajdi Zaghouani](#), [Nizar Habash](#), [Houda Bouamor](#), [Ala Rozovsky](#), [Behrang Mohit](#) and [Kemal Oflaz](#), *Correction Annotation for Non-Native Arabic Texts: Guidelines and Corpus*, (to appear) in *Proceedings of 9th Linguistic Annotation Workshop*, Denver Colorado, June 2015
- [Serena Jelliss](#), [Weston Feeley](#), [Houda Bouamor](#), [Alon Levi](#), [Nizar Habash](#) and [Kemal Oflaz](#), *Domain and Dialect Adaptation for Machine Translation into Egyptian Arabic*, in *Proceedings of Workshop on Arabic Natural Language Processing at EMNLP 2014*, October 2014, Doha, Qatar.
- [Serena Jelliss](#), [Houda Bouamor](#), [Wajdi Zaghouani](#) and [Kemal Oflaz](#), *CMUQIG@QALB-2014 An SMT-based System for Automatic Arabic Error Correction*, in *Proceedings of Workshop on Arabic Nat. Language Processing at EMNLP 2014*, October 2014, Doha, Qatar.
- [Houda Bouamor](#), [Hanan Abuhikmatobekir](#), [Behrang Mohit](#) and [Kemal Oflaz](#), *A Human Judgement Corpus and a Metric for Arabic MT Evaluation*, in *Proceedings of EMNLP 2014*, October 2014, Doha, Qatar.
- [Kemal Oflaz](#), *Turkish and Its Challenges for Natural Language Processing*, Invited paper, Language Resources and Evaluation, April 2014
- [Wajdi Zaghouani](#), [Behrang Mohit](#), [Nizar Habash](#), [Osama Obeid](#), [Nadir Tomah](#) and [Kemal Oflaz](#), *Large-scale Arabic Error Annotation: Guidelines and Framework*, in *Proceedings of LREC 2014*, 9th International Conference on Language Resources and Evaluation, Iceland, May 2014.
- [Houda Bouamor](#), [Nizar Habash](#), and [Kemal Oflaz](#), *A Multidialectal Parallel Corpus of Arabic*, in *Proceedings of LREC 2014*, 9th International Conference on Language Resources and Evaluation, Iceland, May 2014.
- [Ahmed Salama](#), [Houda Bouamor](#), [Behrang Mohit](#) and [Kemal Oflaz](#), *YouTube Arabic Comment Corpus*, in *Proceedings of LREC 2014*, 9th International Conference on Language Resources and Evaluation, Iceland, May 2014.

Turkish Computational Linguistics - TurCLing

April 4, 2016

2016

92

Paper List

- [Kübra Adalı](#) and [Göken Eryiğit](#), *Vowel and Dialect Restoration for Social Media Texts*, accepted for publication in *9th Workshop on Language Analysis for Social Media (LASM) at EACL*, Gothenburg, Sweden, April 2014. [PDF](#) (bibtex)
- [Muhammed Şahin](#), [Umur Sübübaçak](#), and [Göken Eryiğit](#), *Redefinition of Turkish Morphology Using Flag Dialects*, in *Proceedings of the Tenth Symposium on Natural Language Processing (SNLP2013)*, Phuket, Thailand, October 2013. [PDF](#) (bibtex)
- [Umur Sübübaçak](#), [Göken Eryiğit](#), *Representation of Morphosyntactic Units and Coordination Structures in the Turkish Dependency Treebank*, in *Proceedings of the SPML 2013 14th Workshop on Statistical Parsing of Morphologically Rich Languages*, EMNLP 2013, Seattle, USA, 19-21 October 2013. [PDF](#) (bibtex)
- [Gökhan Çelikkaya](#), [Dilare Torunoğlu](#), [Göken Eryiğit](#), *Named Entity Recognition on Real Data: A Preliminary Investigation for Turkish*, in *Proceedings of the 7th International Conference on Application of Information and Communication Technologies, AICT2013*, Baku, Azerbaijan, 9-11 October 2013. [PDF](#) (bibtex)
- [Göken Eryiğit](#), [Fatih Samet Çetin](#), [Meltem Yanık](#), [Tanel Tanel](#), [İlyas Çiçekli](#), *TURKSENT: A Sentiment Annotation Tool for Social Media*, in *Proceedings of the 7th Linguistic Annotation Workshop & Interoperability with Discourse*, ACL 2013, Sofia, Bulgaria, 4-9 August 2013. [PDF](#) (bibtex)
- [Gökhan Alkan Şenar](#), [Göken Eryiğit](#), *Initial explorations on using CRFs for Turkish Named Entity Recognition*, in *Proceedings of the 24th International Conference on Computational Linguistics, COLING 2012*, Mumbai, India, 8-15 December 2012. [PDF](#) (bibtex)
- [Raahin Ehsani](#), [Muazfer Ege Alper](#), [Göken Eryiğit](#), [Egri Adalı](#), 2012. *Disambiguating Main POS tags for Turkish*, in *Proceedings of the 24th Conference on Computational Linguistics and Speech Processing (ROUING 2012)*, Chung-U, Taiwan, 21-22 September 2012. [PDF](#) (bibtex)
- [Göken Eryiğit](#), *The Impact of Automatic Morphological Analysis & Disambiguation on Dependency Parsing of Turkish*, in *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, LREC 2012, Istanbul, May 2012. [PDF](#) (bibtex)
- [Melmet Tahir Çelikkaya](#), [Süleyman Acar](#), [Göken Eryiğit](#), *Word alignment for English-Turkish language pair*, in *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, LREC 2012, Istanbul, May 2012. [PDF](#) (bibtex)
- [Göken Eryiğit](#), [Tuğba Pamay](#), [Ozan Arkan Can](#), *Multiword Expressions in Statistical Dependency Parsing*, in *Proceedings of the Workshop on Statistical Parsing of Morphologically Rich Languages SPML at IJWPT*, (12th International Conference on Parsing Technology), Dublin, October 2011. [PDF](#) (bibtex)
- [Göken Eryiğit](#), *ITU Treebank Annotation Tool*, in *Proceedings of the Linguistic Annotation Workshop at ACL 2007*, Prague, 24-30 June 2007. [PDF](#) (bibtex)
- [John Hall](#), [Jens Nilsson](#), [Joakim Nivre](#), [Göken Eryiğit](#), [Bertalanfogai](#), [Mattiias Nilsson](#), and [Markus Sears](#), 2007. *Single mail or blended? a study in multilingual parser optimization*, in *Proceedings of the 11th Conference on Computational Natural Language Learning (CoNLL-X)*, Prague, 24-30 June. [PDF](#) (bibtex)

Turkish Computational Linguistics - TurCLing

April 4, 2016

2016

95

Paper List

- [Mahmoud Azab](#), [Ahmed Salama](#), [Kemal Oflaz](#), [Hitaki Shima](#), [Jun Aoki](#) and [Taro Matsuura](#), *An English Reading Tool as an NLP Approach*, in *Proceedings of JACL 2013* – Demo Session, Nagoya, Japan, October 2013
- [Mahmoud Azab](#), [Ahmed Salama](#), [Kemal Oflaz](#), [Hitaki Shima](#), [Jun Aoki](#) and [Taro Matsuura](#), *An NLP-based Reading Tool for Aiding Non-native English Readers*, in *Proceedings of SINALP-2013*, Hissar, Bulgaria, September 2013
- [Osama Obeid](#), [Wajdi Zaghouani](#), [Behrang Mohit](#), [Nizar Habash](#), [Kemal Oflaz](#) and [Nadir Tomah](#), *A Web-based Annotation Framework for Large-Scale Text Correction*, in *Proceedings of JCNLP-2013* – Demo Session, Nagoya, Japan, October 2013
- [Houda Bouamor](#), [Behrang Mohit](#) and [Kemal Oflaz](#), *SumMT: A Framework of Summarization and MT*, in *Proceedings of JCNLP-2013*, Nagoya, Japan, October 2013. [Best Paper Award](#)
- [Elif Eyoğar](#), [Duygu Özge](#) and [Kemal Oflaz](#), *Multi-Info HMMs for Word Alignment*, in *Proceedings of ACL 2013 English Workshop on Statistical Machine Translation*, Sofia, Bulgaria, August 2013
- [Ahmed Salama](#), [Kemal Oflaz](#) and [Susan Hagan](#), *Typing for Improved Readability using Lexical and Syntactic Information*, in *Proceedings of ACL 2013*, Sofia, Bulgaria, August 2013
- [Elif Eyoğar](#), [Duygu Özge](#) and [Kemal Oflaz](#), *Simultaneous Word-Morpheme Alignment for Statistical Machine Translation*, in *Proceedings of NAACL 2013*, Atlanta, Georgia, USA, June 2013
- [Nathan Schneider](#), [Behrang Mohit](#), [Chris Dyer](#), [Kemal Oflaz](#) and [Noah A. Smith](#), *Supersense Tagging for Arabic: the MT-in-the-Middle Approach*, in *Proceedings of NAACL 2013*, Atlanta, Georgia, USA, June 2013
- [Mahmoud Mahmoud Azab](#), [Houda Bouamor](#), [Behrang Mohit](#), and [Kemal Oflaz](#), *Building North vs. South London: Learning When to Transfer from Arabic*, in *Proceedings of NAACL 2013*, Atlanta, Georgia, USA, June 2013
- [Emad Mohamed](#), [Behrang Mohit](#) and [Kemal Oflaz](#), *Generating Colloquial Arabic from Standard Arabic*, in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, Jeju, South Korea, July 2012.
- [Nathan Schneider](#), [Behrang Mohit](#), [Kemal Oflaz](#) and [Noah A. Smith](#), *Coarse Lexical Semantic Annotation with Supersenses: An Arabic Case Study*, in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, Jeju, South Korea, July 2012.
- [Emad Mohamed](#), [Behrang Mohit](#) and [Kemal Oflaz](#), *Annotation and Learning Morphological Segmentation of Egyptian Colloquial Arabic*, in *Proceedings of LREC-2012*, Istanbul, Turkey, May 2012
- [Behrang Mohit](#), [Nathan Schneider](#), [Rishav Bhowmik](#), [Kemal Oflaz](#) and [Noah A. Smith](#), *Recall-Oriented Learning of Named Entities in Arabic Wikipedia*, in *Proceedings of EACL-2012*, Avignon France, April 2012.
- [Revan Yantzeri](#) and [Kemal Oflaz](#), *Syntax-to-Morphology Mapping in Factored Phrase-Based Statistical Machine Translation from English to Turkish*, in *Proceedings of ACL 2010*, Uppsala, Sweden, July 2010 (Recently mentioned as one of the interesting papers at the conference, in a recent [Natural Language Processing Blog entry](#))

Turkish Computational Linguistics - TurCLing

April 4, 2016

2016

93

Paper List

- [Göken Eryiğit](#), [Egri Adalı](#) and [Kemal Oflaz](#), 2006. *Türkiye cümlelerin kurallı tabanlı bağliktli analizi*, in *Proceedings of the 15th Turkish Symposium on Artificial Intelligence and Neural Networks*, pages 17-24, Muğla, 21-24 June. [PDF](#) (bibtex)
- [Joakim Nivre](#), [John Hall](#), [Jens Nilsson](#), [Göken Eryiğit](#), and [Stefelav Marinov](#), 2006. *Labeled pseudo-projective dependency parsing with support vector machines*, in *Proceedings of the 10th Conference on Computational Natural Language Learning (CoNLL-X)*, pages 221-225, New York, NY, 8-9 June. [PDF](#) (bibtex)
- [Göken Eryiğit](#) and [Kemal Oflaz](#), 2006. *Statistical dependency parsing of Turkish*, in *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 89-96, Trento, 3-7 April. [PDF](#) (bibtex)
- [Cüneyd Tantiş](#) and [Göken Eryiğit](#), 2006. *Applied Soft Computing Technologies: The Challenge of Complexity*, chapter *Performance Analysis of Naive Bayes Classification, Support Vector Machines and Neural Networks for Spam Categorization*, pages 496-504, Springer Advances in Soft Computing. [PDF](#) (bibtex)
- [Göken Eryiğit](#) and [Egri Adalı](#), 2004. *An affix stripping morphological analyzer for Turkish*, in *Proceedings of the International Conference on Artificial Intelligence and Applications*, pages 299-304, Innsbruck, 16-18 February. [PDF](#) (bibtex)
- [Cüneyd Tantiş](#) and [Göken Eryiğit](#), 2004. *Probabilistic Turkish word root generation*, in *Proceedings of the 3rd Asia Pacific International Symposium on Information Technology*, pages 443-447, Istanbul, 13-14 January. [PDF](#) (bibtex)
- [Göken Eryiğit](#), [Cüneyd Tantiş](#), [Egri Adalı](#), and [Yasar Erenler](#), 2003. *Sentelek Türkçe sözlük kökleri örneği*, in *Proceedings of the 12th Turkish Symposium on Artificial Intelligence and Neural Networks*, pages 133-136, Çanakkale, 2-4 July. [PDF](#) (bibtex)
- [Göken Eryiğit](#) and [Egri Adalı](#), 2002. *Sözlüklerde köke ulama yöntemi*, in *Proceedings of the 19th TDB Bilgiyi Kurultayı*, pages 165-169, Istanbul, 3-6 September. [PDF](#) (bibtex)
- [Pembis, F.C.](#) and [Güngör, T.](#), *A Task-based Learning Approach for Document Structure Analysis and Its Application to Web Search*, *Natural Language Engineering*, Vol.21(4), 2015, p.569-605.
- [Talek, H.](#) and [Güngör, T.](#), *A High Performance Context-based Classification Approach for Language Identification*, *Pattern Recognition Letters*, Vol.33(16), 2012, p.2077-2084.
- [Sak, H.](#), [Sarçar, M.](#) and [Güngör, T.](#), *Morphological and Discriminative Language Models for Turkish Automatic Speech Recognition*, *IEEE Transactions on Audio, Speech and Language Processing*, Vol.20(8), 2012, p.2341-2351.
- [Çogur, L.](#) and [Güngör, T.](#), *Optimization of Dependency and Punning Usage in Text Classification*, *Pattern Analysis and Applications*, Vol.18, 2012, p.48-58.
- [Sak, H.](#), [Güngör, T.](#) and [Sarçar, M.](#), *Resources for Turkish Morphological Processing*, *Journal of Language Resources and Evaluation*, Vol.45, 2011, p.246-261.

Turkish Computational Linguistics - TurCLing

April 4, 2016

2016

96

Paper List

- Pembe, F.C. and Gönğör, T., *Structure-Preserving and Query-Based Document Summarization for Web Searching*, Online Information Review, Vol.30(4), 2006, p.696-719.
- Merat, H.M., Sankur, B., Özsoy, S., Gönğör, T. and Sevinç, E., *Natural Language Watermarking via Morphosyntactic Alterations*, Computer Speech and Language, Vol.23(1), 2009, p.107-125.
- Çelik, A. and Gönğör, T., *Time-Efficient Spam E-mail Filtering Using N-gram Models*, Pattern Recognition Letters, Vol.29(1), 2008, p.19-33.
- Pembe, F.C. and Gönğör, T., *An Evaluation of Structure-Preserving and Query-Based Summaries in Web Search Tasks*, Computer Engineering, Vol.1(2), 2007, p.62-66.
- Gönğör, O. and Gönğör, T., *Towards an Bilgiyazın İnteraktif Sözlük Kulanarak Kavramlar Arasındaki İlişkilerin Belirlenmesi*, Bilgiyazın Bilim ve Mühendisliği, TÜBİTAK, 2007.
- Altınyurt, L., Ohan, Z. and Gönğör, T., *Towards Combining Rule-Based and Statistical Part of Speech Tagging in Agglutinative Languages*, Computer Engineering, Vol.1(1), 2007, p.66-69.
- Sak, H., Gönğör, T. and Salkan, Y., *A Corpus-Based Computational Speech Synthesis System for Turkish*, Turkish Journal of Electrical Engineering and Computer Sciences, Vol.14(2), p.209-223, 2008. [A *Lexical of sentences* synthesized by the system]
- Özgür, L., Gönğör, T. and Gürgeç, F., *Adaptive Anti-Spam Filtering for Agglutinative Languages: A Special Case for Turkish*, Pattern Recognition Letters, Vol.25(16), 2004, p.1819-1831.
- Karaböklü, S., Deniz, C., Keleş, M., Özsoy, S. and Gönğör, T., *Türk İleri Dil (TİD) ve Türkçe Uyarılarına İngilizce Bilgi Bilim Kongresi*, Celal Bayar University, September 2015, Manisa, (accepted).
- Özgür, L. and Gönğör, T., *Two-Stage Feature Selection for Text Classification*, 30th International Symposium on Computer and Information Sciences (ISCIS 2015), September 2015, London, (accepted).
- Deniz, C., Çelik, K., Kurbay, E., Aydın, Y., Özgür, A., Karal, G. and Gönğör, T., *Question Analysis for a Closed Domain Question Answering System*, 16th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2015), Ed. A.Gelbukh, 2015, Cairo, Egypt – LNCS (Lecture Notes in Computer Science), Vol.9042, 2015, p.469-482, Springer-Verlag, Berlin Heidelberg.
- Beğen, E., Kaya, M., Keleş, S., Karadağ, B., Gönğör, T., Day, M. and Çiftçi, H., *Sözlük Medya Ortamında Türkçe Dil Özelliklerine Dair Bir Araştırma*, 2015, Akademik Bilgi Konferansı, Anadolu University, February 2015, Eskişehir.
- Bal, O., Gülnoza, C., Karadağ, A. and Gönğör, T., *Özellikle Türkçe ve İngilizce Dil Üzerine Yöneltilmiş Bir Bilgi Çıkarma*, 31. Ulusal Bilgi Kurultayı, November 2014, p.125-130, Ankara.
- Çelik, A., Bağcı, V.H., Turgut, U.O. and Gönğör, T., *Bankacılık Alanında Doğal Dil İşleme Detekli Davranış Gözlemi*, 8. Ulusal Yazılım Mühendisliği Sempozyumu (UYMS 2014), METU, September 2014, p.477-483, Güzelyurt, K.K.T.C.
- Özdik, S., Sankur, B., Gönğör, T., Yılmaz, M.B., Kırdoğru, B., Ağca, O., İşler, M., Uşu, C. and Ahal, M., *Türkçe Dilbilgi Metin Çerçevesi*, 22th IEEE Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SIU-14), April 2014, Trabzon, Turkey.

April 4, 2016 Turkic Computational Linguistics - TurCling 97

Paper List

- Tezi, S., Gönğör, A.M. and Gönğör, T., *Combining Turkish-English Bilingual Corpus and Developing an Algorithm for Sentence Alignment*, Third International Bulgarian-Turkish Conference on Computer Science, October 2008, Istanbul, p.291-296.
- Altınyurt, L., Ohan, Z. and Gönğör, T., *A Composite Approach for Part of Speech Tagging in Turkish*, Third International Bulgarian-Turkish Conference on Computer Science, October 2008, Istanbul, p.19-24.
- Sak, H., Gönğör, T. and Salkan, Y., *Generational Synthetic Speech from Turkish Text*, 13th European Signal Processing Conference (EUSIPCO 2005), September 2005, Antalya.
- Özgür, L., Gönğör, T. and Gürgeç, F., *Spam Mail Detection Using Artificial Neural Network and Bayesian Filter*, Fifth International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2004), Eds. Z.R.Yong, R.Everson and H.Yin, August 2004, Beier – LNCS (Lecture Notes in Computer Science), Vol.3177, 2004, p.505-510, Springer-Verlag, Berlin Heidelberg.
- Özgür, L., Gönğör, T. and Gürgeç, F., *Yapay Zeka ile Spam Filtresinin İki Sorun Mesajının Tespiti*, AKIS Sistemlerde Yenilikler ve Uygulamaları Sempozyumu (ASYUNISTA), Eds. T.Yıldırım, S.Altınkay, B.Diri, S.N.Engin, L.Ozyılmaz, M.F.Arıyavuz, M.Avalı,Ü.Bakırcı, B.Bolat and B.Kapanoğlu, June 2004, Istanbul, p.219-222.
- Gönğör, T., *Lexical and Morphological Statistics for Turkish*, International Twelfth Turkish Symposium on Artificial Intelligence and Neural Networks (TANN 2003), Vol.E-5, July 2003, Çanakkale, p.409-412.
- Senkal, M., Gönğör, T. and Karadağ, O., *An Approach for Machine Translation Between Turkish and Spanish*, International Twelfth Turkish Symposium on Artificial Intelligence and Neural Networks (TANN2003), Vol.E-5, July 2003, Çanakkale, p.471-473.
- Özgür, L., Gönğör, T. and Gürgeç, F., *Adaptive Turkish-Spanish Filtering*, International Twelfth Turkish Symposium on Artificial Intelligence and Neural Networks (TANN 2003), Vol.E-5, July 2003, Çanakkale, p.453-455.
- Karadağ, O., Gönğör, T. and Kıpman, E., *Bir Uygulama Alanında Türkçe Metin Arayış ve İnceleme Sistemleri*, Akademik Bilim Konferansı, Cukurova Üniversitesi, February 2003, Adana.
- Gönğör, T. and Kuru, S., *Representational Turkish Morphology in ATN*, International Second Turkish Symposium on Artificial Intelligence and Artificial Neural Networks (TANN 1993), 1993, Istanbul, p.94-104.
- Kuru, S. and Gönğör, T., *Towards a Two Level Grammar for the Turkish Noun Phrase*, International Second Turkish Symposium on Artificial Intelligence and Artificial Neural Networks (TANN 1993), 1993, Istanbul.
- Akin, H.L., Kuru, S., Gönğör, T., Hamzaoğlu, I. and Ahal, D., *A Spelling Checker and Corrector for Turkish*, International Second Turkish Symposium on Artificial Intelligence and Artificial Neural Networks (TANN 1993), 1993, Istanbul, p.113-120.
- Kuru, S. and Gönğör, T., *Knowledge Transformation: Generating Causal Relations from Mathematical Models*, Fourth International Symposium on Computational Information Sciences (ISCIS 1989), 1989, Izmir.

April 4, 2016 Turkic Computational Linguistics - TurCling 100

Paper List

- Deniz, C., Çelik, K., Özgür, A., Gönğör, T., Kurbay, E., Aydın, Y. and Karal, G., *Türkçe Sözcük Çeşitliliği Sistemlerinde Kurul Tabanlı Dil Çıkarma*, 22th IEEE Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SIU-14), April 2014, Trabzon, Turkey.
- Aydın, C.R., Erkan, A., Gönğör, T. and Takcı, H., *Sözlük Kulanarak Türkçe için Kavram Madenciliği Metotları Geliştirme*, Akademik Bilgi Konferansı, Mersin University, February 2014, Mersin.
- Aydın, C.R., Erkan, A., Gönğör, T. and Takcı, H., *Dictionary-based Concept Mining: An Application for Turkish*, International Conference on Foundations of Computer Science and Technology (FST 2014), January 2014, Zurich.
- Aydın, C.R., Erkan, A., Gönğör, T. and Takcı, H., *Sözlük Tabanlı Kavram Madenciliği*, Türkçe için Uygulama, 30. Ulusal Bilgi Kurultayı, November 2013, Ankara.
- Gönğör, T., *A Machine Learning Approach for Displaying Query Results in Search Engines*, 15th International Conference on Computer Analysis of Images and Patterns (CAIP 2013), Ed. R.Wilson et al., August 2013, York, England – LNCS (Lecture Notes in Computer Science), Vol.8047, 2013, p.180-187, Springer-Verlag, Berlin Heidelberg.
- Gençor, A.E., Gönğör, T., Güner, A. and Özsoy, A.S., *Input-Evaluation: A New Mechanism for Collecting Data Using Games with a Purpose*, 17th IEEE Symposium on Computers and Communications (ISCC-12), July 2012, Cappadocia, Turkey.
- Gençor, A.E., Gönğör, T., Güner, A. and Özsoy, A.S., *Grid-Değerlendirmesi Arayış Çıkarılabilir Yeni Sistemler için Yeni Bir Makarizma*, 20th IEEE Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SIU-12), April 2012, Fethiye, Turkey.
- Güner, A., Gençor, A.E., Gönğör, T. and Özsoy, A.S., *Dil Çıkarma ve Türkçe Dil Gözlemi*, 26. Ulusal Dilbilim Kurultayı, May 2012, Isparta, Turkey.
- Bekler, M. and Gönğör, T., *Using Genetic Algorithms with Lexical Chains for Automatic Text Summarization*, 4th International Conference on Agents and Artificial Intelligence (ICAART 2012) - Special Session on Machine Learning (SML), February 2012, p.595-600, Vilamoura, Portugal.
- Sak, H., Sarıgür, M. and Gönğör, T., *Discriminative Ranking of ASR Hypotheses with Morphological and N-best-list Features*, IEEE Automatic Speech Recognition and Understanding Workshop (ASRU 2011).
- Uzun-Per, M., Taşcı, H. and Gönğör, T., *Developing a Concept Extraction System for Turkish*, International Conference on Artificial Intelligence (ICAI-11), July 2011, p.855-860, Las Vegas.
- Sak, H., Sarıgür, M. and Gönğör, T., *On-the-fly Latent Recording for Real-time Automatic Speech Recognition*, Interspeech, September 2010, Makuhari, Japan.
- Gönğör, O. and Gönğör, T., *Morphological Annotation of a Corpus with a Collaborative Multi-layer Game*, 11th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2010), Ed. A.Gelbukh, March 2010, Iasi, Romania – LNCS (Lecture Notes in Computer Science), Vol.6008, 2010, p.74-85, Springer-Verlag, Berlin Heidelberg.
- Sak, H., Sarıgür, M. and Gönğör, T., *Morphology-based and Sub-word Language Modeling for Turkish Speech Recognition*, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2010), March 2010, Dallas, Texas.
- Pembe, F.C. and Gönğör, T., *A Tree Learning Approach to Web Document Sectional Hierarchy Extraction*, 2nd International Conference on Agents and Artificial Intelligence (ICAART 2010), January 2010, Valencia.

April 4, 2016 Turkic Computational Linguistics - TurCling 98

Thank you...

Questions?

For the presentation, please send an e-mail to gungor@boun.edu.tr

April 4, 2016 Turkic Computational Linguistics - TurCling 101

Paper List

- Sak, H., Sarıgür, M. and Gönğör, T., *Integrating Morphology into Automatic Speech Recognition*, IEEE Automatic Speech Recognition and Understanding Workshop (ASRU 2009), December 2009, Merano, Italy.
- Taşcı, H. and Gönğör, T., *Link-based Keyword Selection in Text Categorization*, 24th International Symposium on Computer and Information Sciences (ISCIS 2009) September 2009, Northern Cyprus.
- Sak, H., Gönğör, T. and Sarıgür, M., *A Stochastic Finite-State Morphological Parser for Turkish*, ACL-UCLNLP 2009, August 2009, p.273-276, Singapore, ACL.
- Özgür, L. and Gönğör, T., *Analysis of Stemming Alternatives and Dependency Pattern Support in Text Classification*, 10th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2009), Ed. A.Gelbukh, March 2009, Mexico City – Research in Computing Science, Vol.41, 2009, p.195-206, IJN.
- Taşcı, H. and Gönğör, T., *An Evaluation of Existing and New Feature Selection Metrics in Text Categorization*, 23rd International Symposium on Computer and Information Sciences (ISCIS 2008), October 2008, Istanbul.
- Pembe, F.C. and Gönğör, T., *Towards a New Summarization Approach for Search Engine Results: An Application for Turkish*, 23rd International Symposium on Computer and Information Sciences (ISCIS 2008), October 2008, Istanbul.
- Sak, H., Gönğör, T. and Sarıgür, M., *Turkish Language Resources, Morphological Parser, Morphological Disambiguator and Web Corpus*, 6th International Conference on Natural Language Processing (ICNL 2008), Ed. B.Nordström and A.Ranis, August 2008, Götterburg – LNLI (Lecture Notes in Artificial Intelligence), Vol.5221, 2008, p.417-427, Springer-Verlag, Berlin Heidelberg.
- Pembe, F.C. and Gönğör, T., *Heading-Based Sectional Hierarchy Identification for HTML Documents*, 22nd International Symposium on Computer and Information Sciences (ISCIS 2007), November 2007, Ankara.
- Gönğör, T. and Çelik, A., *Developing Methods and Heuristics with Low Time Complexity for Filtering Spam Messages*, 12th International Conference on Applications of Natural Language to Information Systems (NLDB 2007), June 2007, Paris – LNCS (Lecture Notes in Computer Science), Vol.4592, 2007, p.35-47, Springer-Verlag, Berlin Heidelberg.
- Pembe, F.C. and Gönğör, T., *Automate Query-Based and Structure-Preserving Text Summarization on Web Documents*, International Symposium on Innovations in Intelligent Systems and Applications (INISTA 2007), June 2007, Istanbul.
- Merat, H.M., Sevinç, E., Önkar, E.J., Sankur, B., Özsoy, S. and Gönğör, T., *Türkçe Metin Çıkarma: Sözcüklerin Bir Modeli*, IEEE 15th Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SIU 2007), June 2007, Eskişehir.
- Sak, H., Gönğör, T. and Sarıgür, M., *Morphological Disambiguation of Turkish Text with Perceptron Algorithm*, 8th International Conference on Intelligent Text Processing and Computational Linguistics (Cicling 2007), Ed. A.Gelbukh, February 2007, Mexico City – LNCS (Lecture Notes in Computer Science), Vol.4394, 2007, p.107-118, Springer-Verlag, Berlin Heidelberg.
- Gönğör, O. and Gönğör, T., *Türkçe bir Sözlükteki Anımlardan Kavramlar Arasındaki Üst-Kavram İlişkilerinin Çıkarılması*, Akademik Bilim Konferansı, Cukurova Üniversitesi, February 2007, Adana.
- Merat, H.M., Sevinç, E., Önkar, E.J., Sankur, B., Özsoy, S. and Gönğör, T., *Syntactic Tools for Text Watermarking*, 19th Annual Symposium on Electronic Imaging Security, Stenography, and Watermarking of Multimedia Contents, January 2007, San Jose, CA.

April 4, 2016 Turkic Computational Linguistics - TurCling 99