

ITU  
NATURAL LANGUAGE PROCESSING  
RESEARCH GROUP  
2000-2014



Prof. Dr. Eşref Adalı



Asst. Prof. Dr.  
A. Cüneyd Tantuğ



Asst. Prof. Dr. Gülşen Eryiğit

2014

- 6 PhD students
- 20 Ms students

# ACTIVE RESEARCH INTERESTS

- Producing state of the art solutions for any layer of Turkish Natural Language Processing
- Machine Translation between Turkish – English
- Machine Translation between Turkic Languages
- Machine Translation between Turkish – Turkish Sign Language
- Parsing Web2.0 Sentences
- Mobile Personal Assistant  
for Turkish



6/11/2014

# ACTIVE RESEARCH PROJECTS

Project Name	Dates	Supporter	Budget	Project Coordinator
Machine Translation Between English-Turkish	2011-2015	DPT – Turkey MINISTRY OF DEVELOPMENT	1.800.000 TL	Prof.Dr.Eşref Adalı Asst.Prof.Dr. A. Cüneyd Tantuğ
Parsing Web 2.0 Sentences	2013-2015	ICT COST (European Cooperation in Science and Technology) Action IC1207 ve Tubitak 1001	164.000 TL	Asst.Prof.Dr. Gülşen Eryiğit
Machine Translation from Turkish to Turkish Sign Language and the Avatar Animation	2014-2016	TUBITAK 1003 (The Scientific and Technological Research Council of Turkey)	500.000 TL	Asst.Prof.Dr. Gülşen Eryiğit (NLP group coordinator)
Turkish Mobile Personal Assistant	2013-2014	Turkey MINISTRY OF Science, Industry and Technology	53.014 TL	Asst.Prof.Dr. Gülşen Eryiğit

6/11/2014

# FORMER RESEARCH PROJECTS

Project Name	Dates	Supporter	Budget	Project Coordinator
Machine Translation between Turkmen-Turkish	2006-2009	TUBITAK 1001	180.000 TL	Prof.Dr.Eşref Adalı A. Cüneyd Tantuğ
Dependency Parsing of Turkish	2004-2006	İTÜ Research Unit	6000TL	Prof.Dr.Eşref Adalı Asst.Prof.Dr. Gülşen Eryiğit
Machine Translation between Uyghur-Turkish	2007-2010	İTÜ Research Unit	10000TL	Prof.Dr.Eşref Adalı
CLARIN - Common language resources and technology infrastructure	2010-2011	EU 7th Framework	4.100.000 Euro	Asst.Prof.Dr. Gülşen Eryiğit (Turkey's Representative)

6/11/2014

# RESEARCH PROJECT PROPOSALS (under evaluation)

Project Name	Dates	Supporter	Budget	Project Coordinator
Machine Translation between Kazakh - Turkish	2014-2017	Kazak Ministry of Education and Science ITU ve L.N. Gumilev National Eurasian University	1.080.000 USD	Prof.Dr.Eşref Adalı (Turkey's Representative)
Computer Aided Language Education for Arabic – Turkish Language Pairs	2015-2017	TUBITAK & CNRST (Moroccan National Research Center)	150.000 TL	Asst.Prof.Dr. A. Cüneyd Tantuğ

6/11/2014

ITU Turkish Natural Language Processing Pipeline  
ITU Türkçe Doğal Dil İşleme Yazılım Zinciri

Login

Tokenizer  
Normalization  
Deasciifier  
Vowelizer  
Spelling Corrector  
isTurkish  
Morphological Analyzer  
Morphological Disambiguator  
Named Entity Recognizer  
Dependency Parsing  
Whole pipeline  
Datasets  
Simple API  
Changelog

This webpage provides the Turkish NLP Tools and APIs developed by our Natural Language Processing group at Istanbul Technical University.

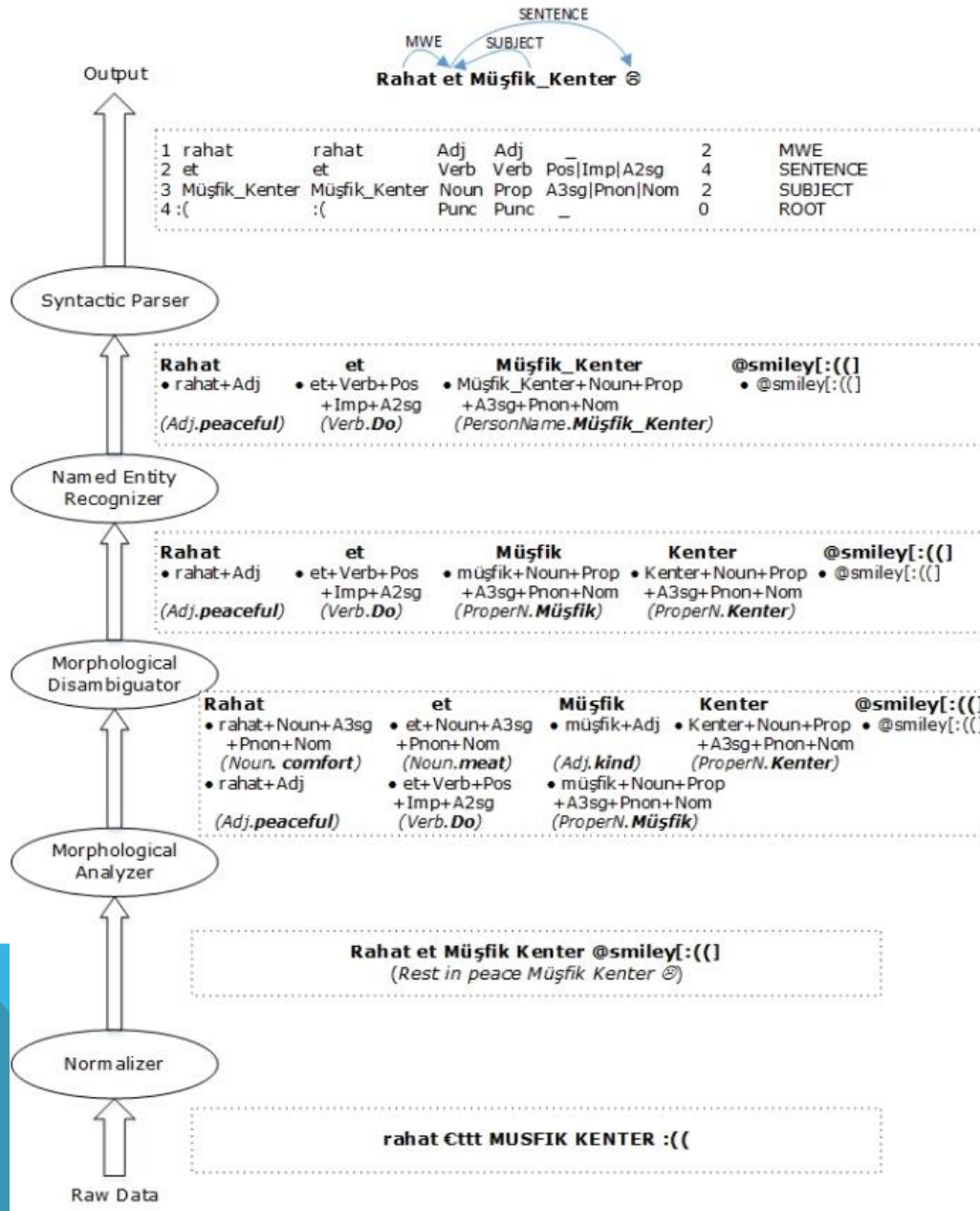
The platform operates as a SaaS (Software as a Service) and provides the researchers and the students the state of the art NLP tools in many layers: preprocessing, morphology, syntax and entity recognition. The users may communicate with the platform via three channels:

1. via a user friendly web interface,
2. by file uploads and
3. by using the provided Web APIs within their own codes for constructing higher level applications.

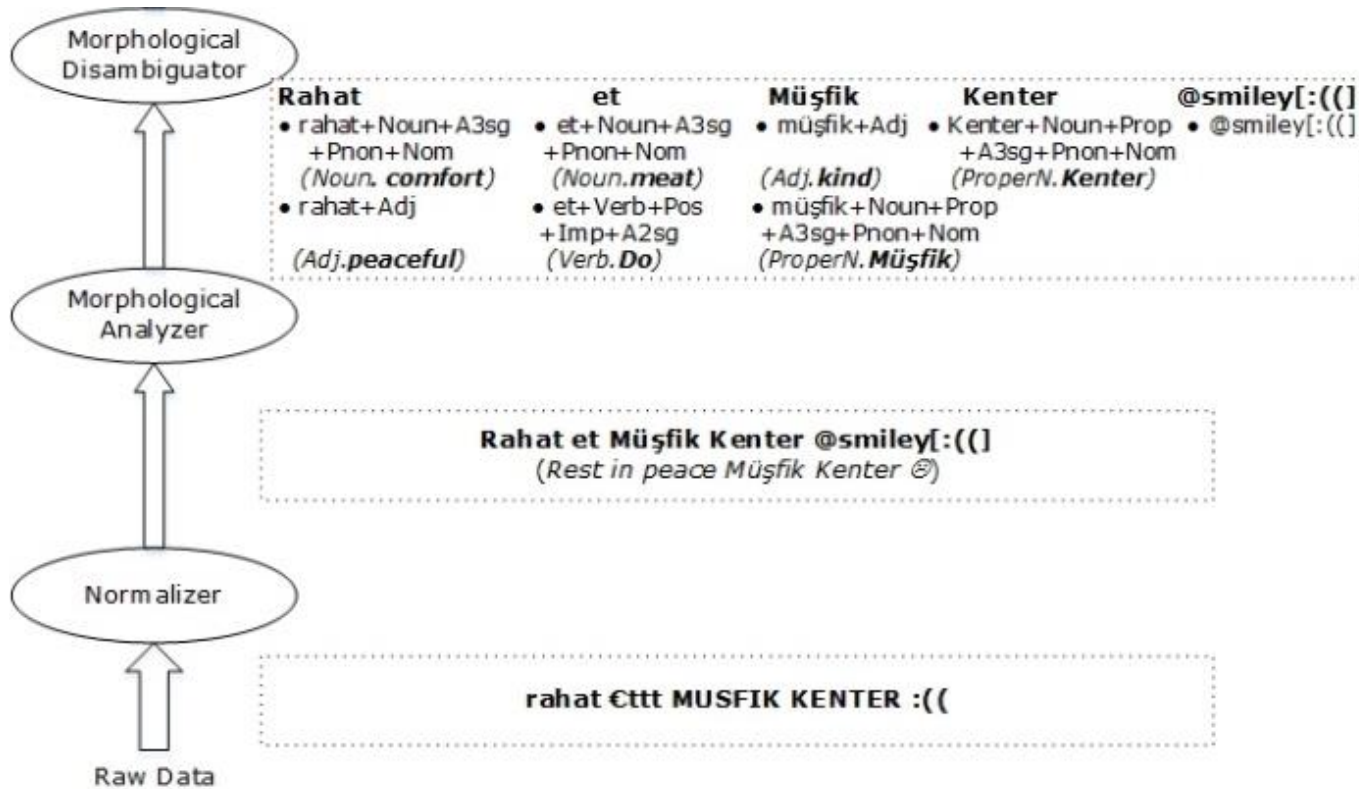
The details of this work are given in the following publication, please refer to it if you would use this tool in your research:

Gülşen Eryiğit . **ITU Turkish NLP Web Service** 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. (bibtex) (pdf) (poster)

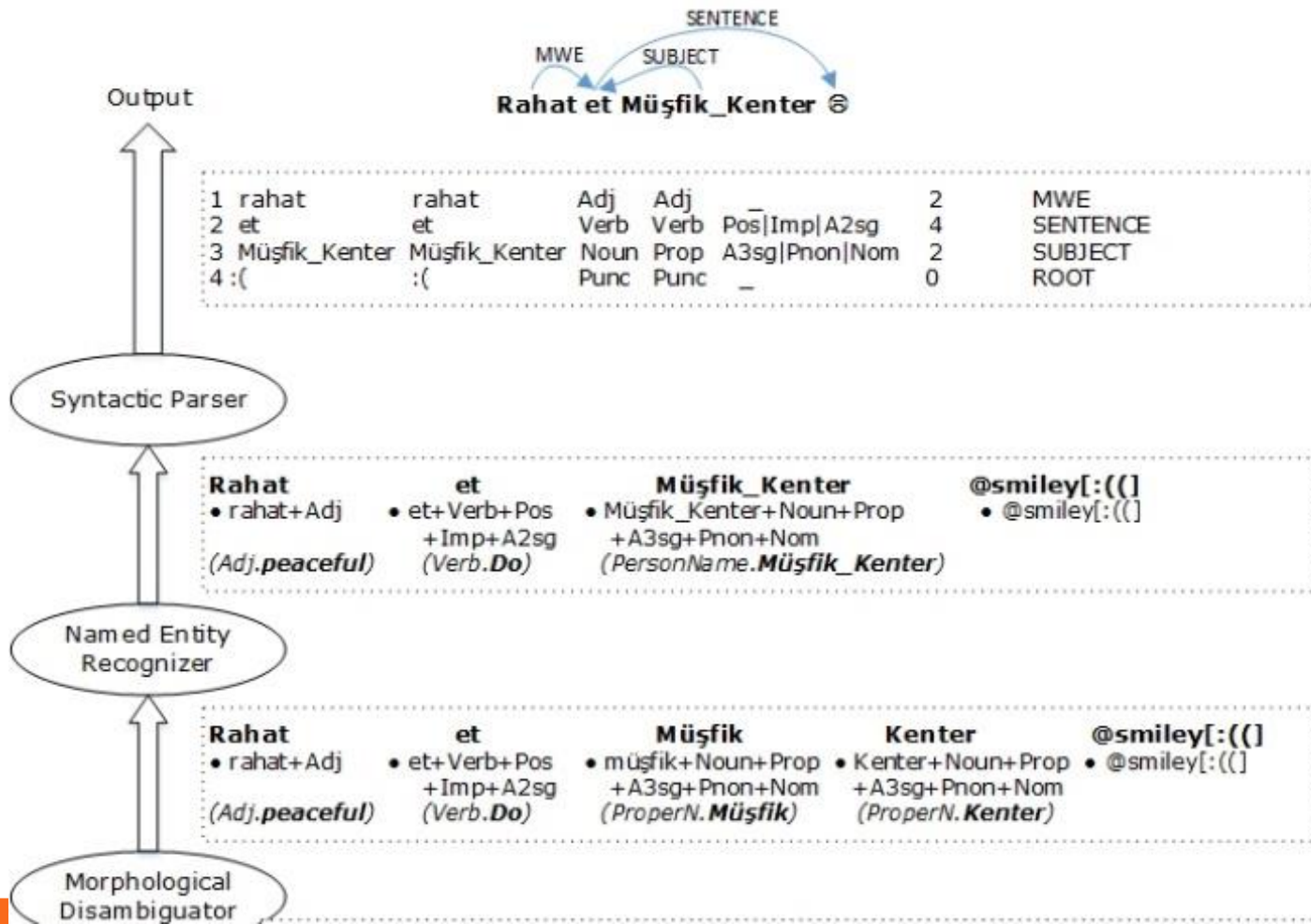
Report a bug



6/11/2014







6/11/2014

# INTERFACE AND APIS

Normalization  
tools.nlp.itu.edu.tr/Normalization

## ITU Turkish Natural Language Processing Pipeline

### Normalization

- Tokenizer
- Normalization
- Deasciifier
- Vocalizer
- Spelling Corrector
- isTurkish
- Morphological Analyzer
- Morphological Disambiguator
- Named Entity Recognizer
- Dependency Parsing
- Whole pipeline
- Datasets
- Simple API
- Changelog

@dida bi yere gitcem "bütüüüüüüüünnn insanlar hür,  
hysyt ve haklari bkmdn eşit doğarlaaaar.  
aaaaaakıl ve vicdana sahiptirler ve birbirlerine karşı  
kardeşlik zhnyt ile hrkt etmelidirler ." diye bağırcaam  
::) #insanhakları

Normalize

@mention[@dida] bir yere gideceğim " bütün insanlar hür  
, haysiyet ve hakları bakımından eşit doğarlar .  
Akıl ve vicdana sahiptirler ve birbirlerine karşı  
kardeşlik zihniyet ile hareket etmelidirler ." diye  
bağıracağım @smiley{:))} @hashtag[#insanhakları]

The details of this work

Ur

How to use the API

tools.nlp.itu.edu.tr/api\_usage.jsp

## ITU Turkish Natural Language Processing Pipeline

ITU Türkçe Doğal Dil İşleme Yazılım Zinciri

Login

- Tokenizer
- Normalization
- Deasciifier
- Vowelizer
- Spelling Corrector
- isTurkish
- Morphological Analyzer
- Morphological Disambiguator
- Named Entity Recognizer
- Dependency Parsing
- Whole pipeline
- Datasets
- Simple API
- Changelog

This page explain briefly how to use the API

First of all it's a web API, so you can access to it with a HTTP request. As described on the default API response, you can use GET or POST method, but keep in mind the POST method can be bigger than the GET method (Approximatly 2MB of POST and 64kB for GET)

To be able to use the API, you need a token. To get a token you first need an account. To obtain an account, please send an email to Gülşen Cebiroğlu Eryiğit (gulsen[dot]cebiroglu[at]itu.edu.tr). Explain briefly who you are, your affiliation and why you need to access the API. Also give your desired password. You will have access to your token information (which should be replaced to the position of mulpile XXXX's in the below code) by clicking on your username after logging in.

More details in the default API reponse: [Access to the API](#)

Here you can find a basic example in java:

```
1 // Using http://hc.apache.org: HttpClient (any version)
2 // and http://commons.apache.org/proper/commons-Logging/
3
4 String input = "Gülşen Gülşen+Noun+Prop+A3sg+Pnon+Nom\n" +
```

Report a bug

6/11/2014

# ITU TREEBANK ANNOTATION TOOL

Proje Yürütücüsü - 112E27 x ITU Web Annotation Tool x

tools.nlp.itu.edu.tr/Annotation/Disambiguation?sentenceId=2011

ITU Web Annotation Tool Welcome, gulsen [Logout]

Morphological Disambiguation Dependency Parsing Supervisor

## Morphological Disambiguation

gulsen

Skip F9 Skipped Saved

bigın varsa brz aydınlatılrmın Sentence ID: 2011

Bilgin varsa biraz aydınlatılabilir misin

Update

Bilgin	varsa	biraz	aydınlatılabilir	misin
<input type="radio"/> bilgi+Noun+A3sg+P2sg+Nom	<input checked="" type="radio"/> var+Adj*DB+Verb+Zero+Cond+A3sg	<input checked="" type="radio"/> biraz+Adj	<input type="radio"/> aydınlat+Verb+Pos+Verb+Able+Aor+A3sg	<input type="radio"/> mi+Noun+A3sg+Pnon+Nom*DB+Verb+Zero+Pres+
<input checked="" type="radio"/> bilgin+Noun+A3sg+Pnon+Nom	<input type="radio"/> var+Verb+Pos+Desr+A3sg	<input type="radio"/> biraz+Adverb	<input checked="" type="radio"/> aydınlatılabilir+Verb+Pos+Aor+A3sg	<input checked="" type="radio"/> mi+Ques+Pres+A2sg
<input type="radio"/> <input type="text"/>	<input type="radio"/> var+Noun+A3sg+Pnon+Nom*DB+Verb+Zero+Cond+A3sg	<input type="radio"/> <input type="text"/>	<input type="radio"/> <input type="text"/>	<input type="radio"/> mis+Noun+A3sg+Pnon+Gen
	<input type="radio"/> <input type="text"/>			<input type="radio"/> mis+Noun+A3sg+P2sg+Nom
				<input type="radio"/> <input type="text"/>

Save (F8) Save&Proceed (F10)

6/11/2014

# ITU TREEBANK ANNOTATION TOOL

The screenshot shows the ITU Web Annotation Tool interface. The browser address bar displays the URL `tools.nlp.itu.edu.tr/Annotation/DependencyParsing`. The page title is "ITU Web Annotation Tool" and the user is logged in as "gulsen". The navigation menu includes "Morphological Disambiguation", "Dependency Parsing" (which is highlighted), and "Supervisor".

The main content area is titled "Dependency Parsing" and shows the word "gulsen" in a green box. To the right of the word are buttons for "Skip (F9)", "Skipped", and "Saved". Below the word, there is a sequence of boxes: "Bilgin", "varsa", "biraz", "aydınlatabilir", "misin", and "Root". The "aydınlatabilir" box is highlighted in yellow. A tooltip is visible over the "aydınlatabilir" box, showing the morphological information "var+Adj" and "+Verb+Zero+Cond+A3sg".

Below the word sequence is a "Relation Table" with the following data:

Word 1	Word 2	Relation	Count	Action
aydınlatabilir	misin	ARGUMENT	0	Delete

At the bottom right of the interface, there are two green buttons: "Save (F8)" and "Save&Go Back (F10)".

6/11/2014

# LANGUAGE RESOURCES

- IMST – ITU METU SABANCI Turkish Dependency Treebank
- IWT – ITU Turkish Web Treebank
- IVS – ITU Validation Set
- Test sets for all introduced Turkish NLP Layers
- English – Turkish Parallel Corpus

# LIST OF PUBLICATIONS

<http://ddi.ce.itu.edu.tr/>

<http://web.itu.edu.tr/~gulsenc/>

<http://www.tantug.com/>

[www.adali.net/](http://www.adali.net/)

6/11/2014

**Thank you for listening  
you are all welcome to visit our lab**



E. Adalı

6/11/2014