



Department Of  
**CREATIVE  
TECHNOLOGIES**

**AIR UNIVERSITY**  
ISLAMABAD

Dr. Imran Ihsan  
[www.imranihsan.com](http://www.imranihsan.com)

# Linguistic Knowledge

---

Natural Language Processing

# Phonetics and Phonology

- The study of linguistic sounds and their relations to words

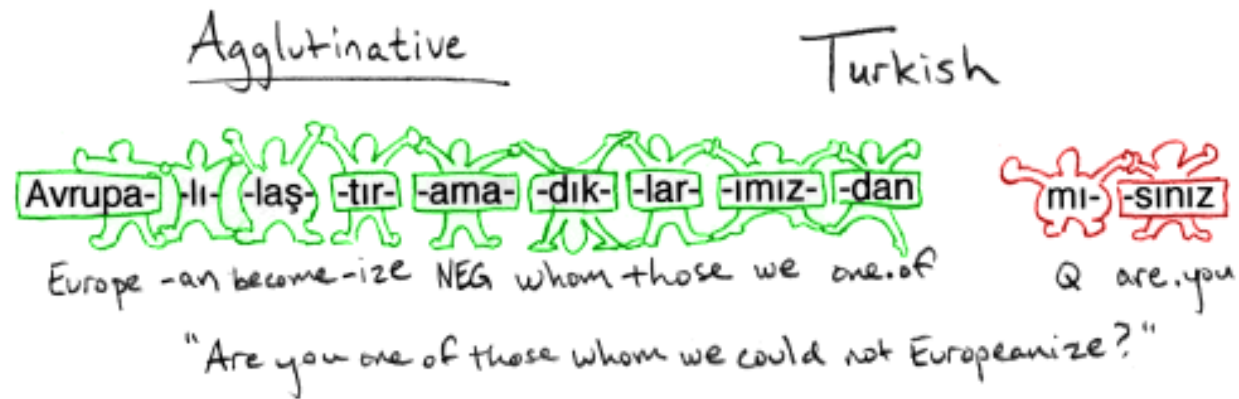
Symbol	Code Word	Morse Code	Phonic (pronunciation)
A	Alfa/Alpha	•—	AL FAH
B	Bravo	—•••	BRAH VOH
C	Charlie	—•—•	CHAR LEE
D	Delta	—••	DELL TAH
E	Echo	•	ECK OH
F	Foxtrot	••—•	FOKS TROT
G	Golf	—•—	GOLF
H	Hotel	••••	HOH TELL
I	India	••	IN DEE AH
J	Juliett	•—••	JEW LEE ETT
K	Kilo	—•—	KEY LOH
L	Lima	••—•	LEE MAH
M	Mike	—•	MIKE
N	November	—•	NO VEMBER
O	Oscar	—••	OSS CAH
P	Papa	••—•	PAH PAH

## NATO Phonetic Alphabet

Das Funkalphabet - German Phonetic Spelling Code compared to the international ICAO/NATO code Listen to AUDIO for this chart! (below)		
Germany*	Phonetic Guide	ICAO/NATO**
A wie Anton	AHN-tone	Alfa/Alpha
Ä wie Ärger	AIR-gehr	(1)
B wie Berta	BARE-tuh	Bravo
C wie Cäsar	SAY-zar	Charlie
Ch wie Charlotte	shar-LOT-tuh	(1)
D wie Dora	DORE-uh	Delta
E wie Emil	ay-MEAL	Echo
F wie Friedrich	FREED-reech	Foxtrot
G wie Gustav	GOOS-tahf	Golf
H wie Heinrich	HINE-reech	Hotel
I wie Ida	EED-uh	India/Indigo
J wie Julius	YUL-ee-oos	Juliet
K wie Kaufmann	KOWF-mann	Kilo
L wie Ludwig	LOOD-vig	Lima

# Morphology

- The study of internal structures of words and how they can be modified
- Parsing complex words into their components



<https://allthingslinguistic.com/post/50939757945/morphological-typology-illustrations-from>

# Syntax

- The study of the structural relationships between words in a sentence

## Parse

```
(ROOT
  (S
    (NP
      (NP (NNP Surgical) (NN resection) (NNS specimens))
      (PP (IN of)
        (NP
          (NP (CD 85) (JJ invasive) (JJ ductal) (NNS carcinomas))
          (PP (IN of)
            (NP
              (NP (CD 85) (NNS women))
              (SBAR
                (WHNP (WP who))
                (S
                  (VP (VBD had)
                    (VP (VBN undergone)
                      (NP (CD 3D) (NN ultrasound))))))))))
            (VP (VBD were)
              (VP (VBN included)))
            (. .)))
```

# Semantics

The study of the meaning of words, and how these combine to form the meanings of sentences

- **Synonymy:** fall & autumn
- **Hypernymy & hyponymy (is a):** animal & dog
- **Meronymy (part of):** finger & hand
- **Homonymy:** fall (verb & season)
- **Antonymy:** big & small

# Pragmatics

- Social use of language
- The study of how language is used to accomplish goals, and the influence of context on meaning
- Understanding the aspects of a language which depends on situation and world knowledge

*Give me the salt!*

*Could you please give me the salt?*

# Discourse

- The study of linguistic units larger than a single statement

*John reads a book. He borrowed it from his friend.*

## Pakistan

From Wikipedia, the free encyclopedia

*This article is about the country in South Asia. For other uses, see [Pakistan \(disambiguation\)](#)*

**Pakistan**,<sup>[c]</sup> officially the **Islamic Republic of Pakistan**,<sup>[d]</sup> is a country in South Asia. It is the world's fifth-most populous country, with a population exceeding 225.2 million, and has the world's second-largest Muslim population. Pakistan is the 33rd-largest country by area, spanning 881,913 square kilometres (340,509 square miles). It has a 1,046-kilometre (650-mile) coastline along the Arabian Sea and Gulf of Oman in the south, and is bordered by India to the east, Afghanistan to the west, Iran to the southwest, and China to the northeast. It is separated narrowly from Tajikistan by Afghanistan's Wakhan Corridor in the north, and also shares a maritime border with Oman.



Department Of  
**CREATIVE  
TECHNOLOGIES**

**AIR UNIVERSITY**  
ISLAMABAD

Dr. Imran Ihsan  
[www.imranihsan.com](http://www.imranihsan.com)

# Linguistic Knowledge

---

Challenges



# Paraphrasing

Different words/sentences express the same meaning

- Season of the year
  - Fall
  - Autumn
- Book delivery time
  - When will my book arrive?
  - When will I receive my book?

# Ambiguity

One word/sentence can have different meanings

- Fall
  - The third season of the year
  - Moving down towards the ground or towards a lower position
- The door is open.
  - Expressing a fact
  - A request to close the door

# Phonetics and Phonology

[http://worldsgreatestsmile.com/html/phonological\\_ambiguity.html](http://worldsgreatestsmile.com/html/phonological_ambiguity.html)



**Communication tip:**

**Phonological ambiguities** or **Give peas a chance!**

One of my favourite ways to have fun with communication are phonological ambiguities.

Phonological ambiguities are two or more words which sound the same and have different meanings.



Language can contain ambiguities - and more than one way to compose a set of sounds into words.

So listen to yourself: It is always good to notice a spoken sentence often contains many words which are (sometimes not) intended to be heard.

**English examples:**

- there - their
- here - hear
- plane - plain
- Hamburger (Citizens of Hamburg) - hamburger (burger, food)
- sea - see
- Friday - fry day
- weekend - weak end
- ice cream - I scream.
- new direction - nude erection
- new day - nude, eh?
- I don't know! - I don't - no!
- but - butt
- Wait - Weight
- psychotherapist - psycho the rapist
- You're unconscious now... - Your unconscious now...
- Your students... - You're students...
- Two - too - to

# Syntax and Ambiguity

- I saw the man with a telescope.  
– *Who had the telescope?*



# Semantics

- The astronomer loves the star.
  - Star in the sky
  - Celebrity



# Discourse analysis

- Alice understands that you like your mother, but she ...

*Does **she** refer to Alice or your mother?*



# Natural Language Processing

---

NLP Course

Dr. Imran Ihsan  
[www.imranihsan.com](http://www.imranihsan.com)

# NLP Course

- Google Classroom
  - Code:
- Lecture
  - Monday – 4:40 to 7:30
  - Credit Hours: 03



# Grading

- Assignments 05 %
- Quizzes 05 %
- Research Project 20 % (10% additional in Final Exam)
- Midterm 30 %
- Final Exam 40 %

# Course Outline

- Introduction to Natural Language Programming
- Regular Expressions and Automata
- N-Grams
- Part-of-Speech Tagging
- Syntactic Parsing
- Lexical Semantics
- Discourse
- Information Extraction
- Text Classification and Sentiment Analysis
- Information Retrieval
- Question Answering and Summarization
- Research Project Presentations



# Research Project

Development of an NLP application

- Information Retrieval
- Information Extraction
- Text Summarization
- Question Answering
- Sentiment Analysis
- Machine Translation
- Etc..

# Research Project

The application should include following components:

- Part-of-speech tagging
- Syntactic parsing
- Lexical semantics
- Discourse analysis
- Named-entity recognition

# Research Project

Any NLP or ML libraries

- Stanford Core NLP
- NLTK
- spaCy
- Gensim
- Ployglot
- Apache OpenNLP
- GATE
- SAP HANA (contact me)
- R
- Weka



# Research Project

Any language

- English, Urdu, German, etc.
- Check available NLP tools

Any text collections

- Social media, Web pages, publications, citations, Wikipedia, etc.
- Benchmarks or new collections

Any domain

# Research Project

- Teams (2 students)
- Send me an email with your proposal as soon as possible

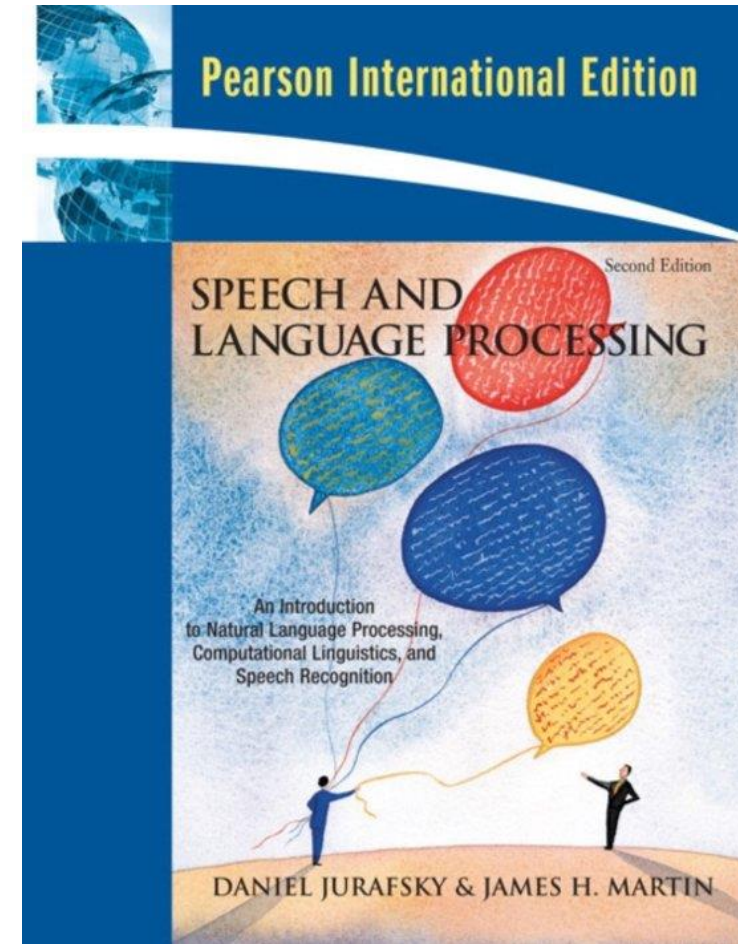
Updates (presentations) on the progress of the project

- Slots during the lectures
- Also considered for grading

# Course book

## Speech and Language Processing

– Daniel Jurafsky and James H. Martin





# Journal and conferences

## Journal

- Computational Linguistics

## Conferences

- ACL: Association for Computational Linguistics
- NAACL: North American Chapter
- EACL: European Chapter
- HLT: Human Language Technology
- EMNLP: Empirical Methods on Natural Language Processing
- CoLing: Computational Linguistics
- LREC: Language Resources and Evaluation

# NLP Course

## Contact

- Dr. Imran Ihsan – [imranihsan@mail.au.edu.pk](mailto:imranihsan@mail.au.edu.pk)
- Web: <http://www.imranihsan.com>