CSCI 2304
Intelligent Systems
Week 9

Natural Language Processing (NLP)

Human and Language

- Understanding language requires inferences about speaker's goals, knowledge and expectations of domain and reason effectively about them.
- Humans are commonly identified by the languages that they produce ("speak your mind").

One reason for studying language — and for me personally the most compelling reason — is that it is tempting to regard language, in the traditional phrase, as a "mirror of mind".

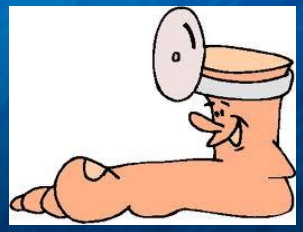
Chomsky, 1975
MIT

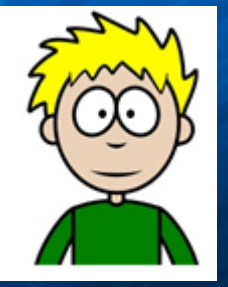
Human and Language

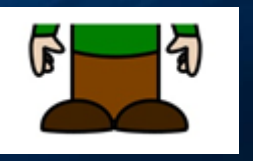
- There needs to be an ability to characterize and explain the multitude of linguistic observations around us in conversations, writing, media, etc…
- Understanding and processing language concerns:
  i. the cognitive side (information processing in the brain) of how human acquire, produce and understands language
  ii. understanding the relationship between linguistic utterances and the world
  iii. understanding the linguistic structures by which language communicates

Why is Language Hard?

- Ambiguities in Language
  - *"School dropouts cut in half this year"*
  - *"Hospitals Are Sued by 7 Foot Doctors"*
  - *"Include Your Children When Baking Cookies"*
  - *"Kids Make Nutritious Snacks"*







- Complexity in Language
  - *"There is a boy, who lost his toy, who jumped with joy, who drank the soy, who made everyone annoyed*

What is NLP?

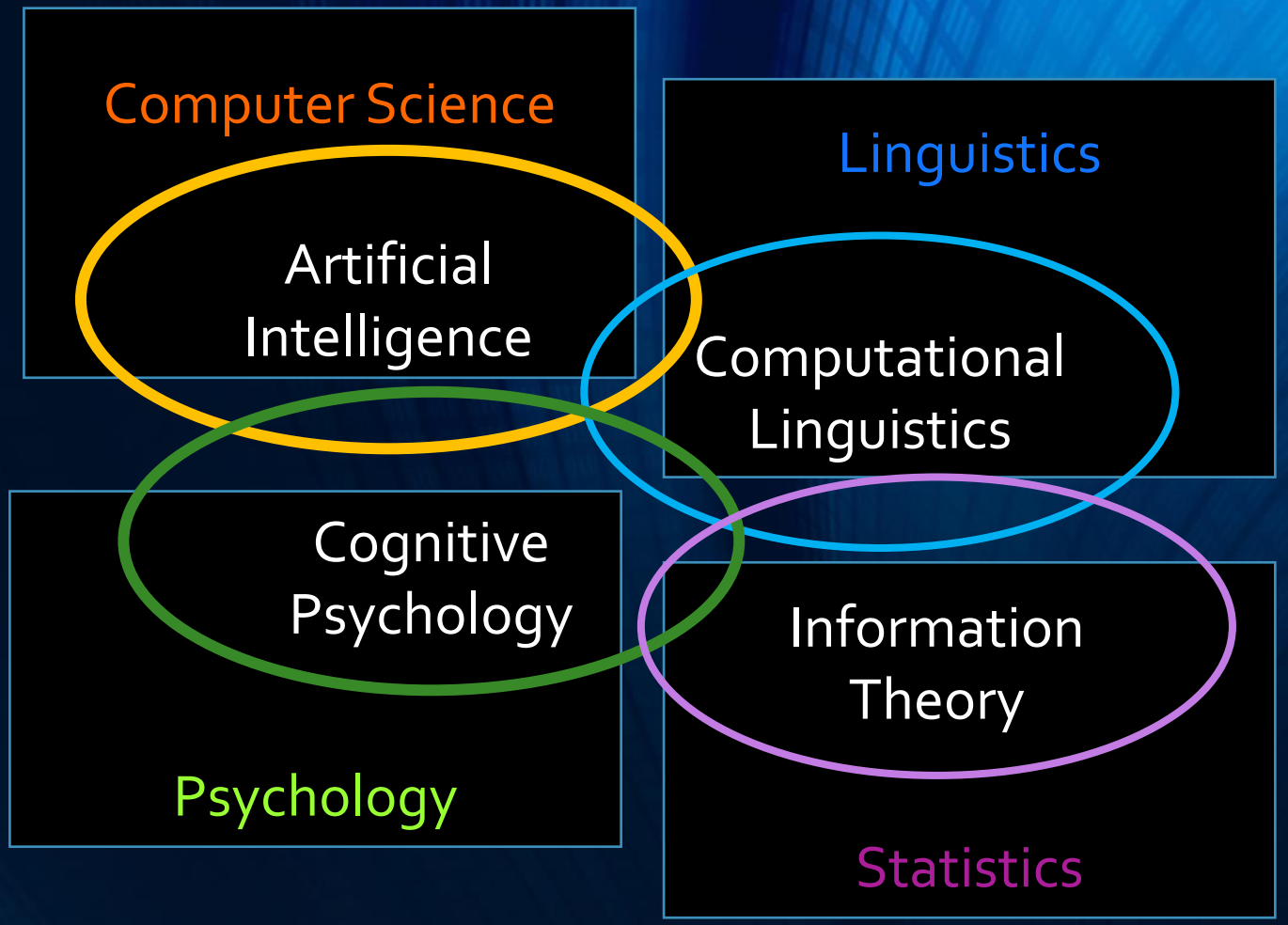
The process of building computational models for understanding natural language.

A coherent study of the human language from the point of views of several disciplines : Linguistics, Psychology, Cognitive Science, Computer Science, Statistics and Mathematics.

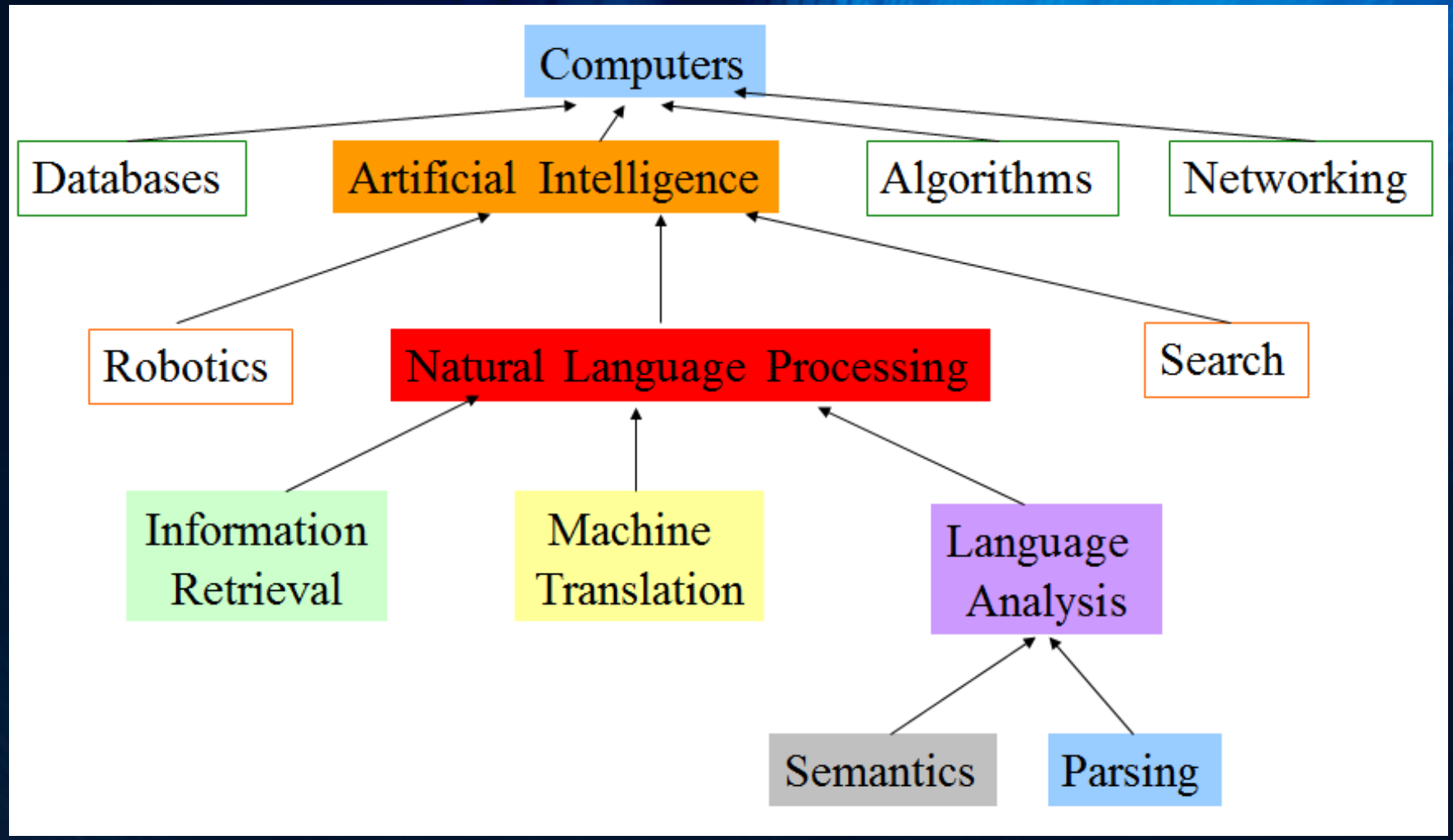
A theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications

Also known as *Computational Linguistics*

Multidisciplinary NLP



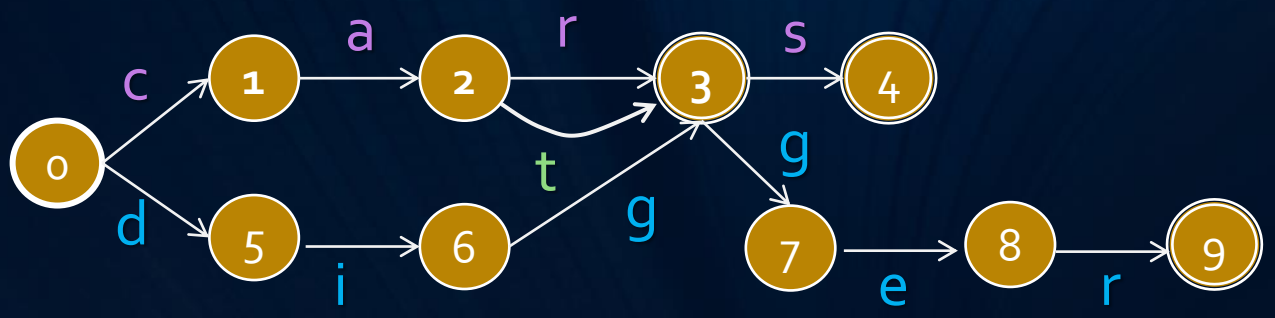
Where does NLP fit in CS?



Common NLP Tasks

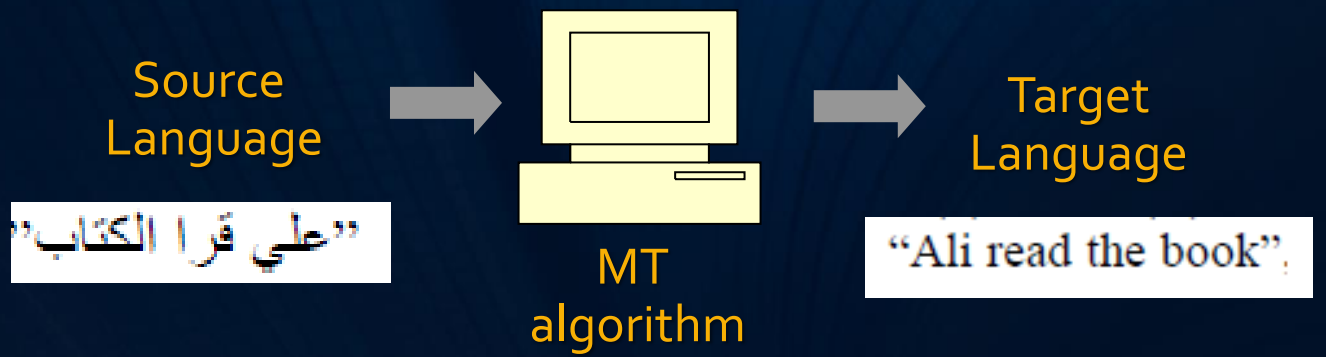
- Part-of-Speech (POS) Tagging
  - Assigning a part-of-speech to each word in a sentence

    "Malaysia/**N** has/**V** 23/**NUM** million/**N** people/**N**".
- Computational Morphology
  - Processing of words and word forms, in both their graphemic (written form) and their phonemic (spoken form)
  - Example: finite state morphology



Common NLP Tasks

- Sentence Parsing
  - Analysing a sentence into its component categories and functions (refer Slide 6)
- Machine Translation (MT)
  - An automated system that analyzes text from source language and produces “equivalent” text in the target language



Common NLP Tasks

- Sentiment Analysis
  - Identify, analyze and classify opinions in text into categories such as "positive" , "negative" or "neutral"

    "I love Macintosh." (Positive)

    "I hate Blackberry! " (Negative)

    "The food is not bad at all" (Neutral)

    "What a great car, it did not start the first day" (Ambiguous)

- Co-reference Resolution
  - Two textual entities that refers to the same object in the "real world" (Mitkov)

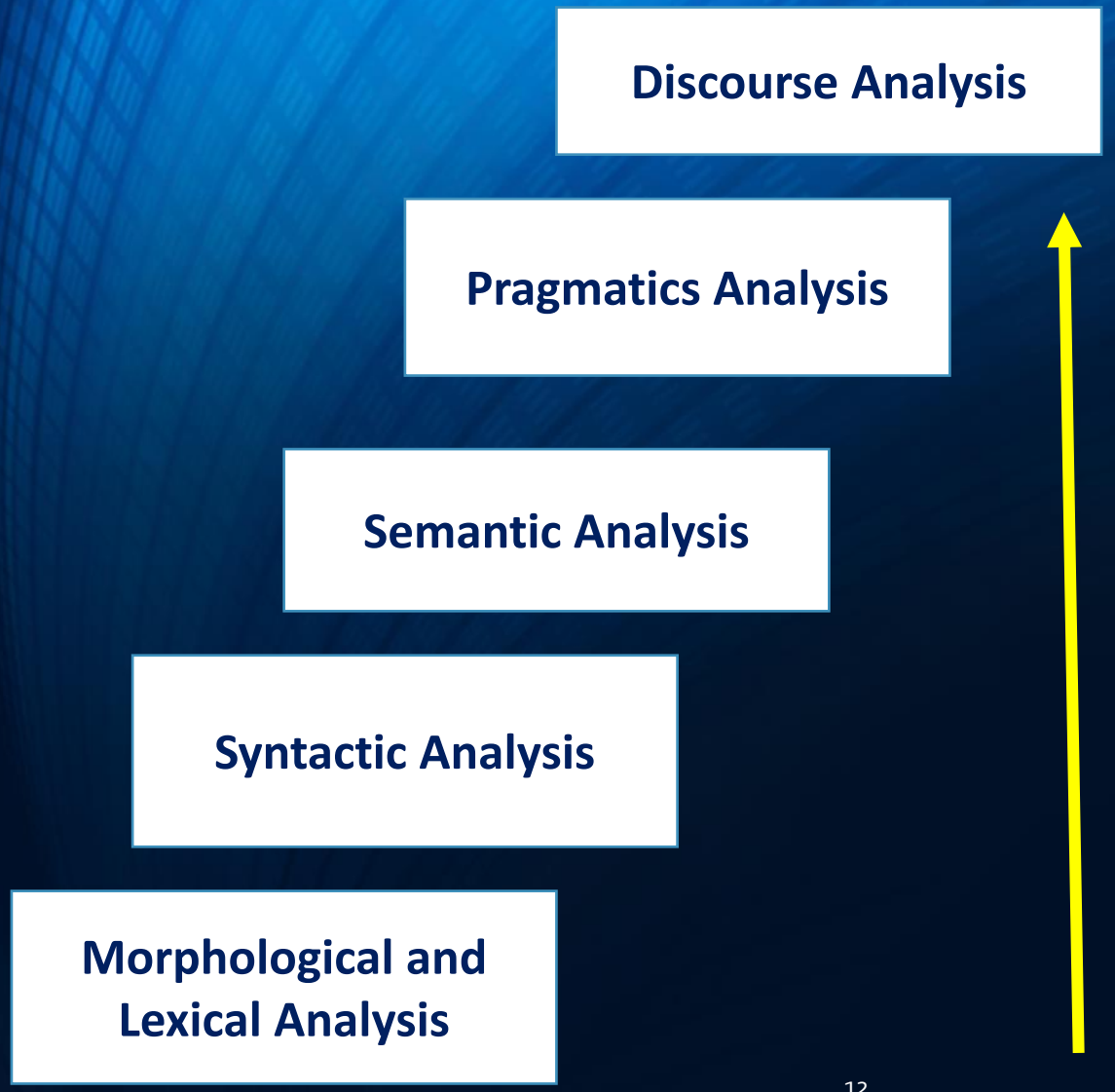
    Saha Hisham Ismail$_1$, 45, said poor drainage$_2$ in the village$_3$ was the main cause of the problem$_4$. "We$_{1,3}$ have reported it$_2$ to the authorities$_5$ and they$_5$ have promised to look into it$_2$, but nothing has been done to rectify the problem$_4$."

Common NLP Tasks

- Text Summarization
  - Automatically reducing a text document to create a summary that preserves the most important points of the original document.
  - Example : Given a single document, produce abstract, outline and headline.
- Topic Categorization
  - Classifies documents according to their topics
    - *"Serena and Nadal relieved after surviving tough opponents in Madrid"* [Sports]
    - *"Facebook eyes $1billion deal for GPS app Waze"* [Technology]
    - *"Property, constructions to lead stock market"* [Business]
    - *"All eyes on cabinet lineups"* [Politics]

Levels of Language Processing and Analysis

- Morphological and Lexical Analysis
- Syntactic Analysis
- Semantic Analysis
- Pragmatics Analysis
- Discourse Analysis



Syntactic Categories

- **Syntax** terms in NLP can be confusing for those with limited background in linguistics and grammar
- Terminals refer to words from the lexicon and are less important than non-terminals
- Non-terminals are critical in Context-Free Grammars (CFGs) for describing phrases
- Rules are used in CFGs to define non-terminals and create a framework for constructing phrases
- Understanding non-terminals is essential for NLP practitioners.
- Phrases are simply groups of words and the name of the phrase usually takes its name from some important word in the phrase, for instance:
  - **noun phrase** (often written as *NP* or *np*)
  - **verb phrase** (often written as *VP* or *vp*)
  - **prepositional phrase** (often written as *PP* or *pp*)
  - **sentence** (often written as *S* or *s*)

Syntactic Categories

There are other phrases that are sometimes used, such as verb_group and categories such as:

- **ADJ = adjective**
  A word that qualifies a noun - for instance tabby in tabby cat.
- **ADV = adverb**
  A word that qualifies a verb - for instance boldly in to go boldly.
- **AUX = auxiliary**
  A common verb occurs with another verb to add some further meaning - for instance will in will fly adds the notion of future action.
- **DET = determiner**
  A word which specifies - for instance the in the book.
- **NOUN**
  A word that refers to an object or entity - for instance cat.
- **PREP = preposition**
  A word that denotes position - for instance in in in the park.
- **PRONOUN**
  A word that refers to a person without using its name - for instance they, he, it.
- **VERB**
  A word that refers to an action - for instance flying, washing, thinking.

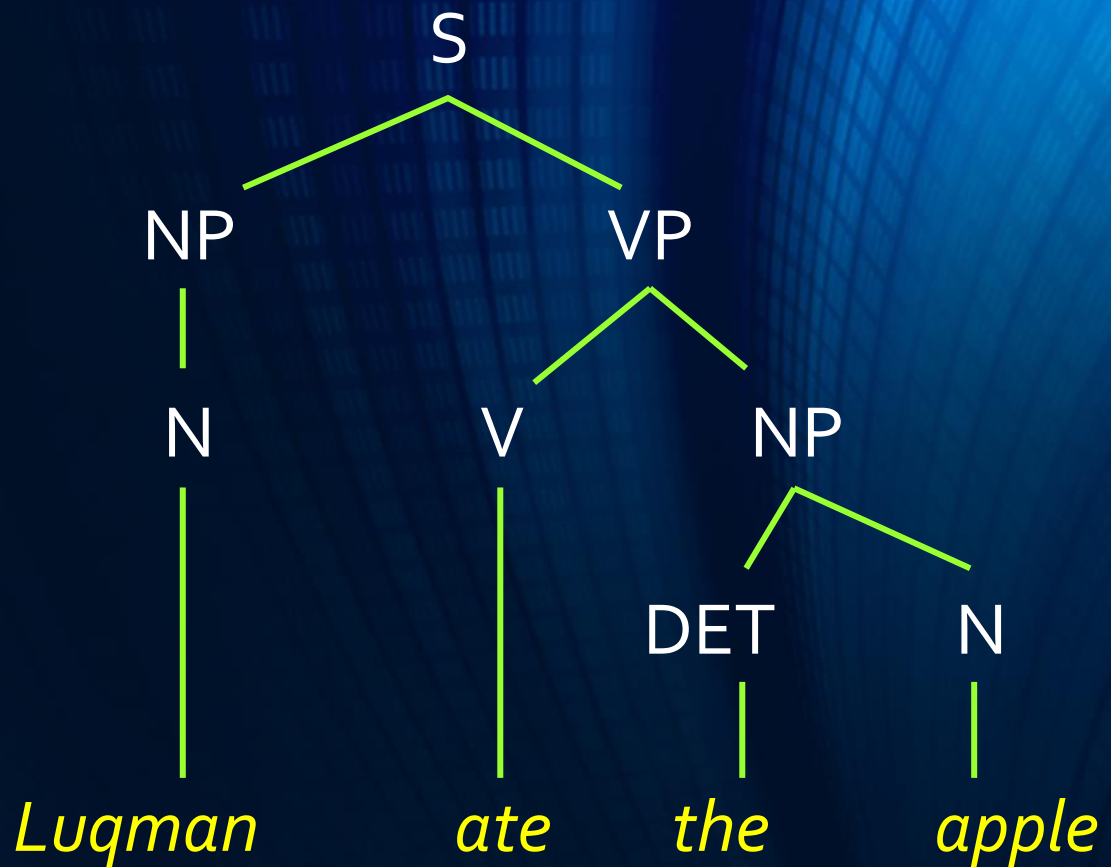
Syntactic Analysis- Phrases

- Noun Phrase, NP
  - He is a(ART) policeman(N).
  - My(N) house(N) is in Gombak.
- Verb Phrase, VP
  - They play(V) street(N) soccer(N)
  - Perform(V) your(N) prayers(N) 5 times a day
- Adjectival Phrase-AP
  - My grandfather is a slow(ADJ) driver(N)
  - The delicious(ADJ) meal(N) is gone.
- Prepositional Phrase, PP
  - The astronaut travels to(P) the(DET) moon(N)
  - The food on(P) the(DET) table(N) is contaminated.

Syntactic Analysis- Grammar

- sentence(S) -> noun_phrase (NP) verb_phrase (VP)
- noun_phrase(NP) -> proper_noun (N)
- noun_phrase (NP)-> determiner (DET) noun (N)
- verb_phrase (VP)-> verb (V) noun_phrase (NP)
- prepositional_phrase (PP) -> preposition(P) noun phrase(NP)
- proper_noun(N) -> [Luqman]
- noun(N) -> [apple]
- verb (V)-> [ate]
- determiner(DET) -> [the]

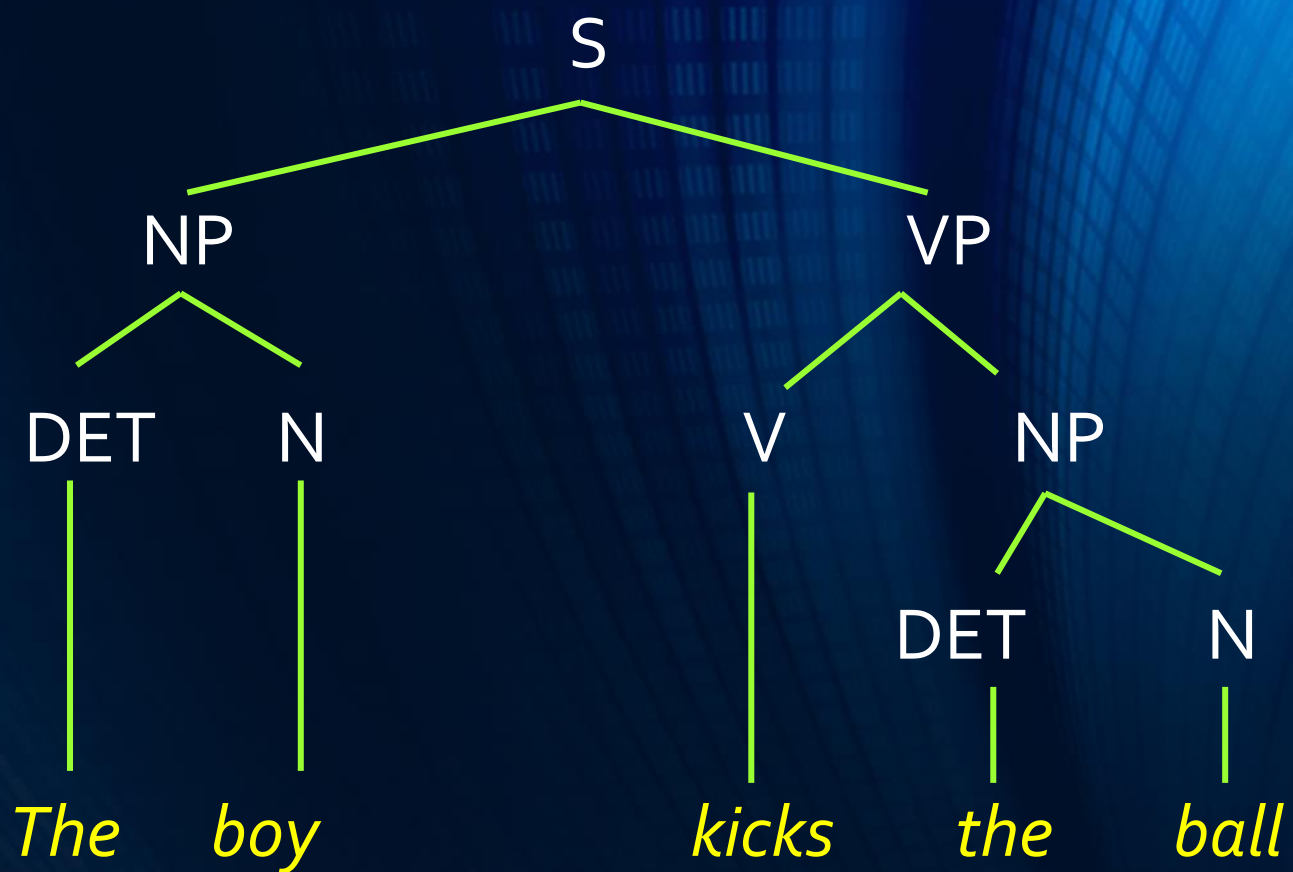
Syntactic Analysis- Parsing (Example 1)



Grammar rules used:

S -> NP VP
VP – > V NP
NP -> DET N
NP -> N

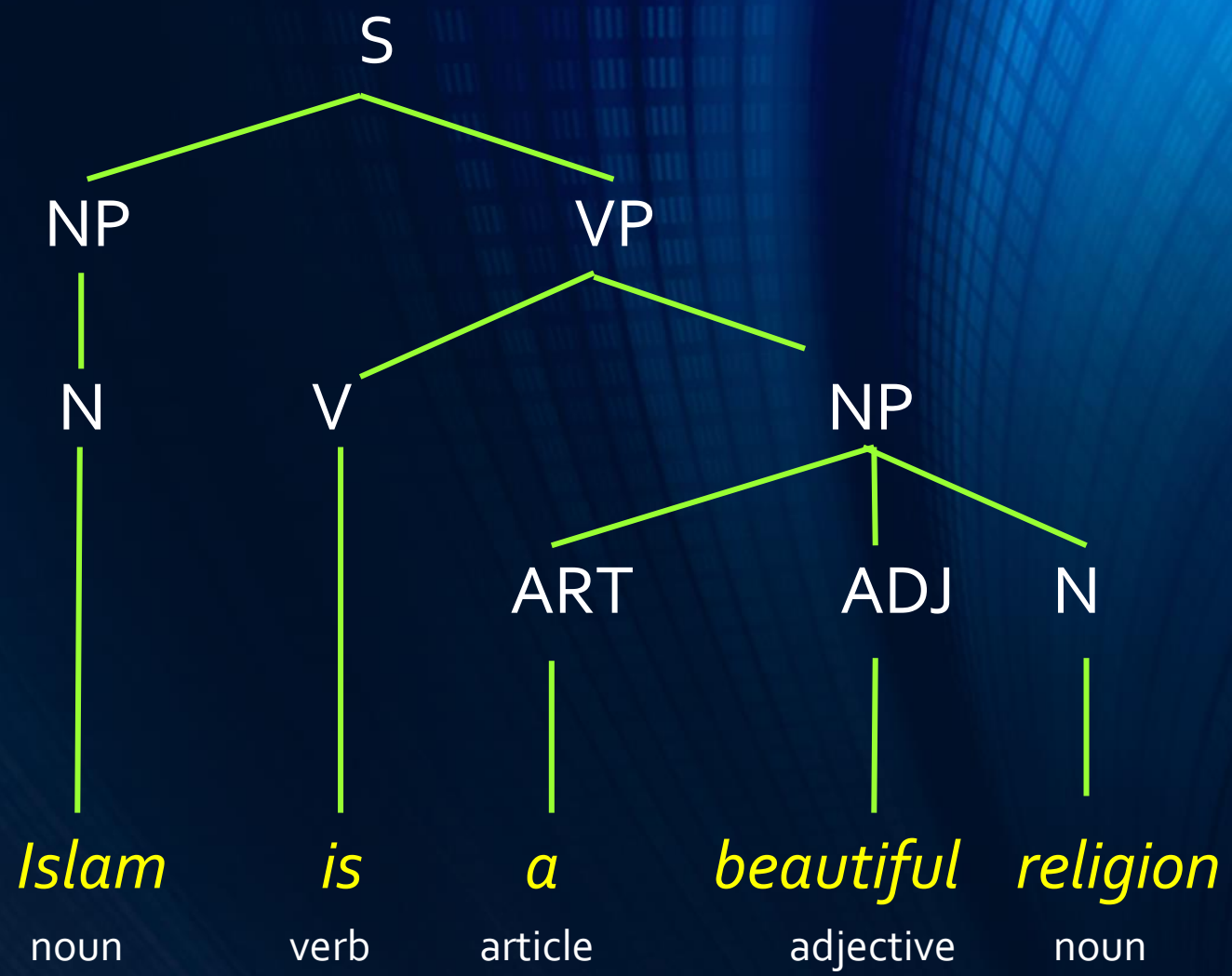
Syntactic Analysis- Parsing (Example 2)



Grammar rules used:

S -> NP VP
VP – > V NP
NP -> DET N
NP -> N

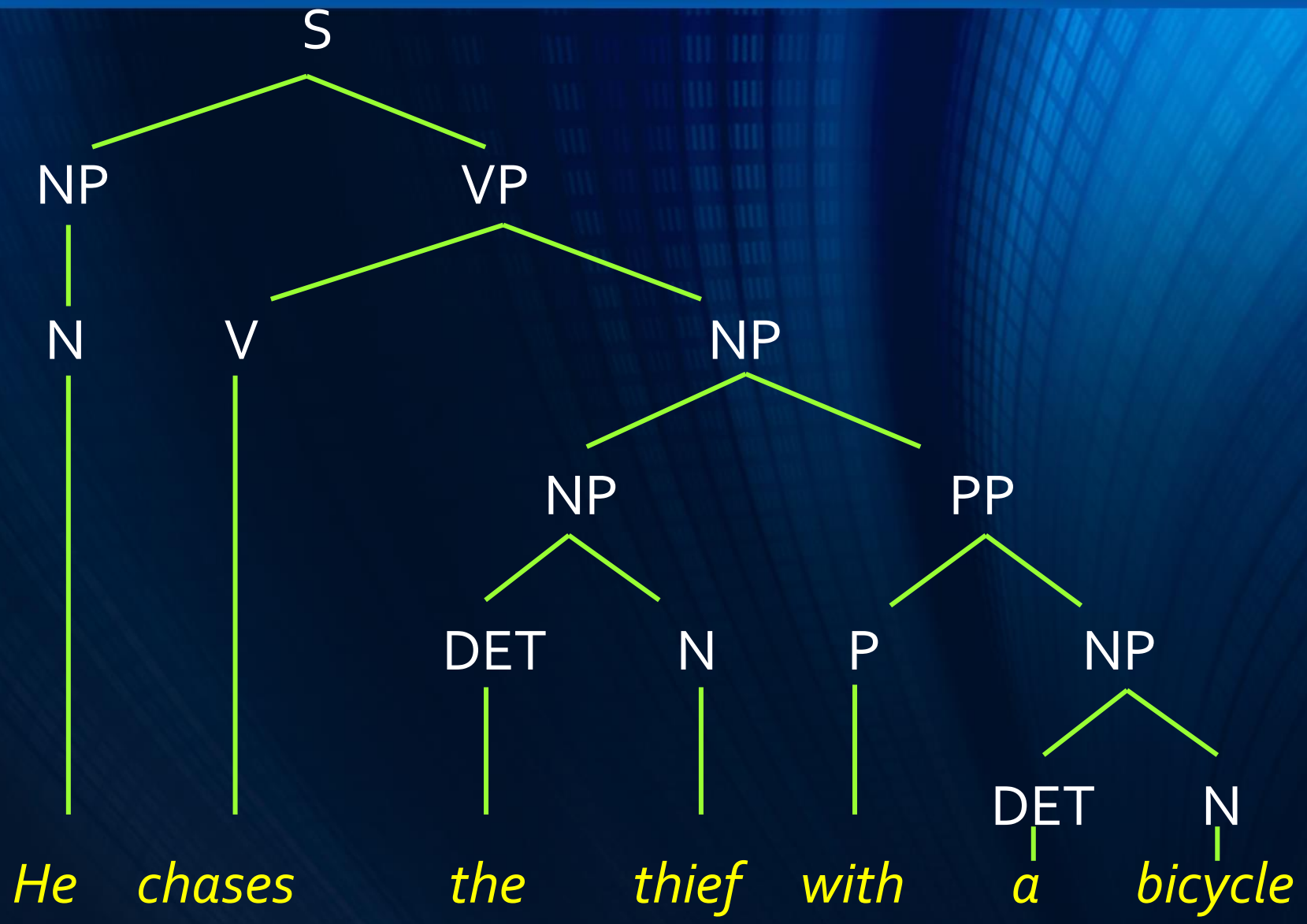
Syntactic Analysis- Parsing (Example 3)



Grammar rules used:

S -> NP VP
VP – > V NP
NP -> ART (ADJ) N
NP -> N
*Note : rules in ( ) are optional

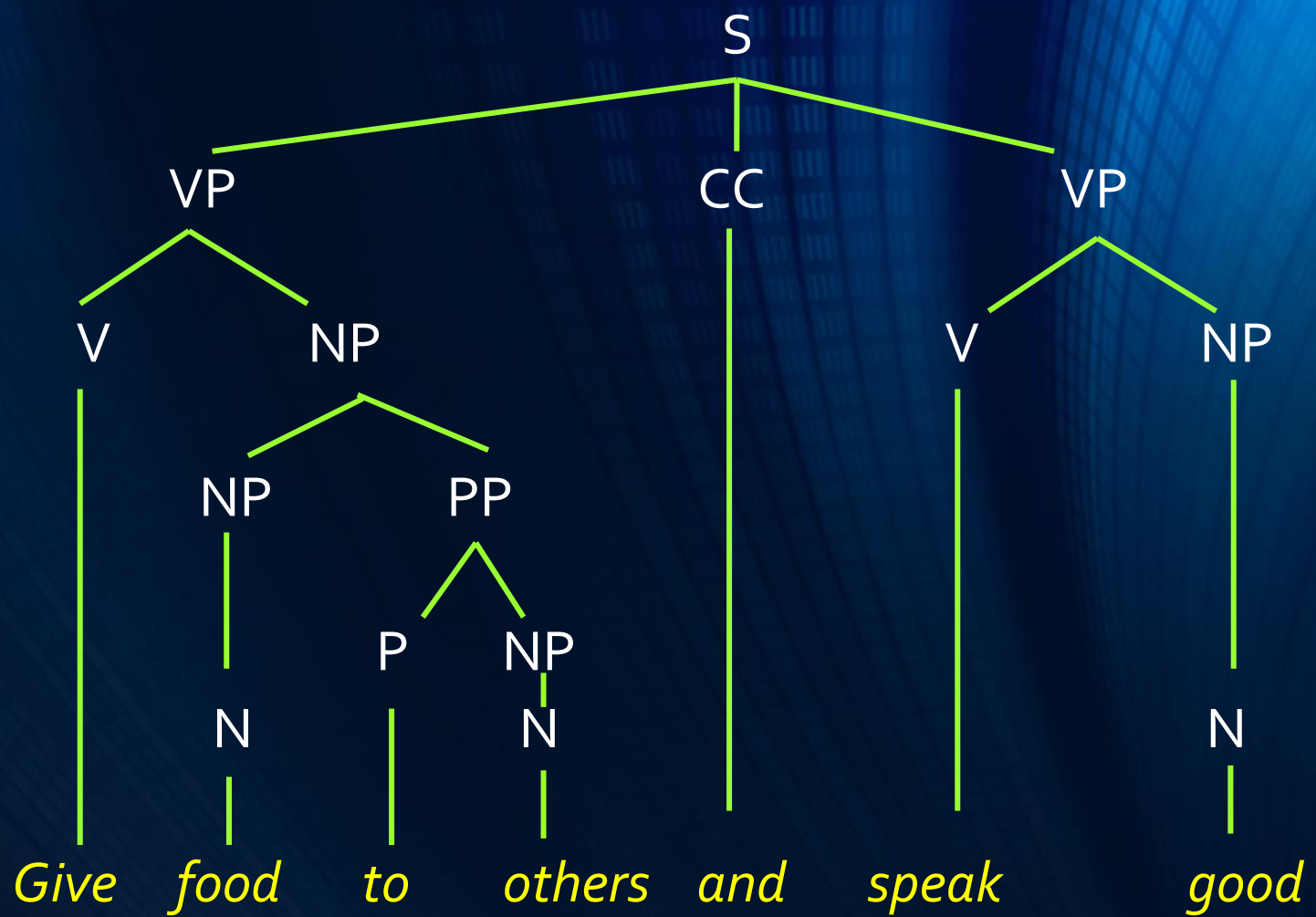
Syntactic Analysis- Parsing (Example 4)



Grammar rules used:

S -> NP VP
VP – > V NP
NP -> NP PP
NP -> DET N
NP -> ART (ADJ) N
PP -> P NP
NP -> N
*Note : rules in ( ) are optional

Syntactic Analysis- Parsing (Example 5)



Grammar rules used:

S -> VP CC VP

S- > NP VP

VP – > V NP

NP → NP PP

PP -> P NP

NP -> N

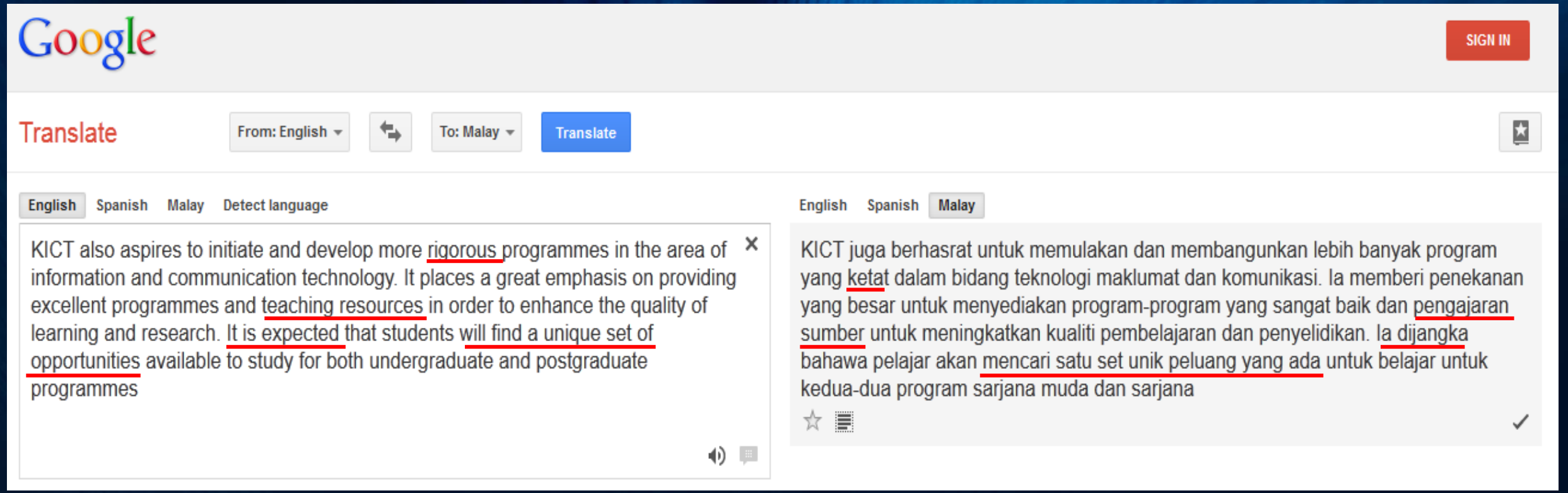
Issues in Syntactic Analysis

- Handling ambiguity
  - Syntactic ambiguity:
    - "Fruit flies like a banana"
    - "Fruit flies like a banana"

- Having to parse syntactically incorrect sentences
  - "(The boy) (the ball) kicks"
    (NP) (NP) (VP)

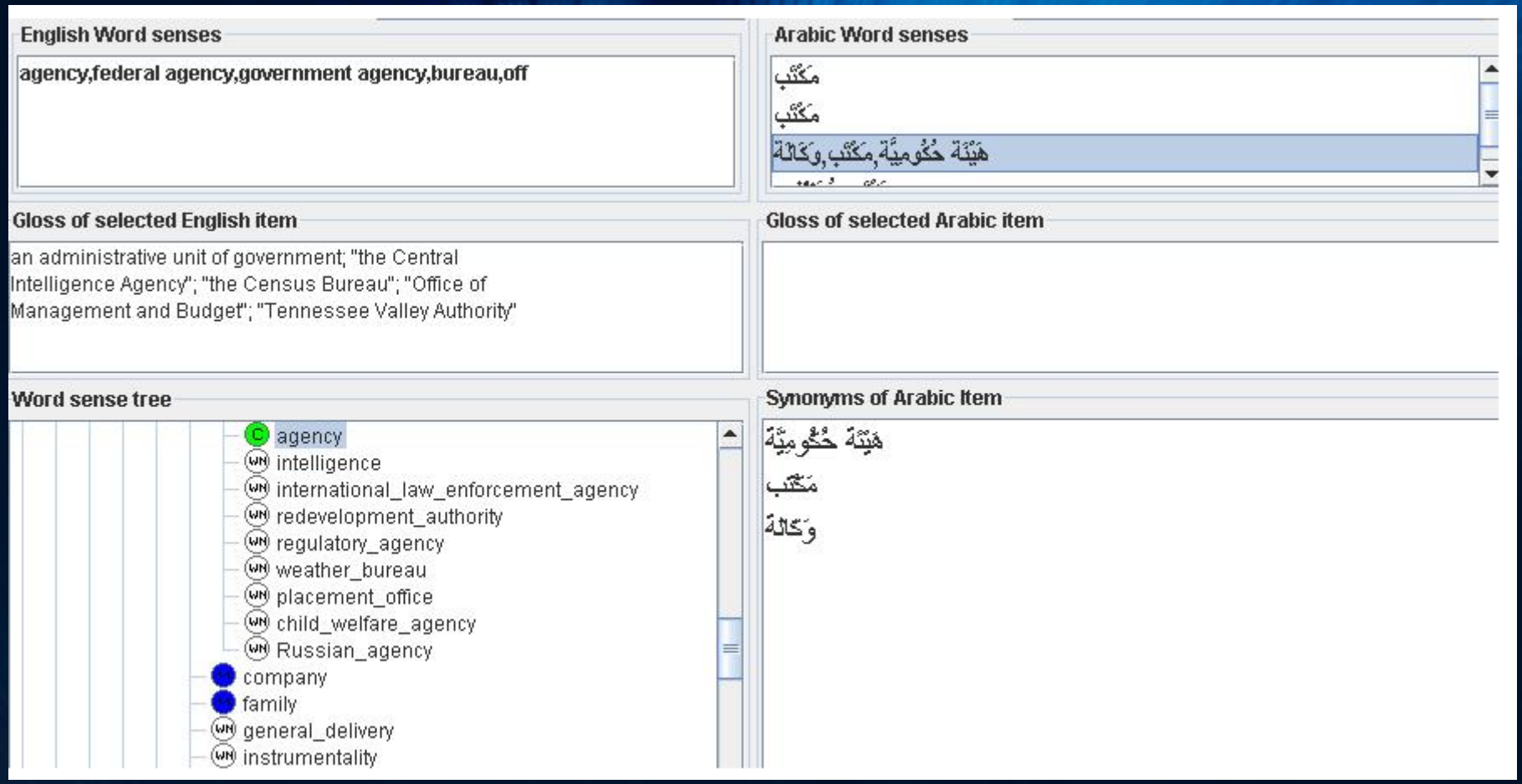
Examples of NLP Applications

Google Translate



Examples of NLP Applications

Arabic WordNet



Examples of NLP Applications

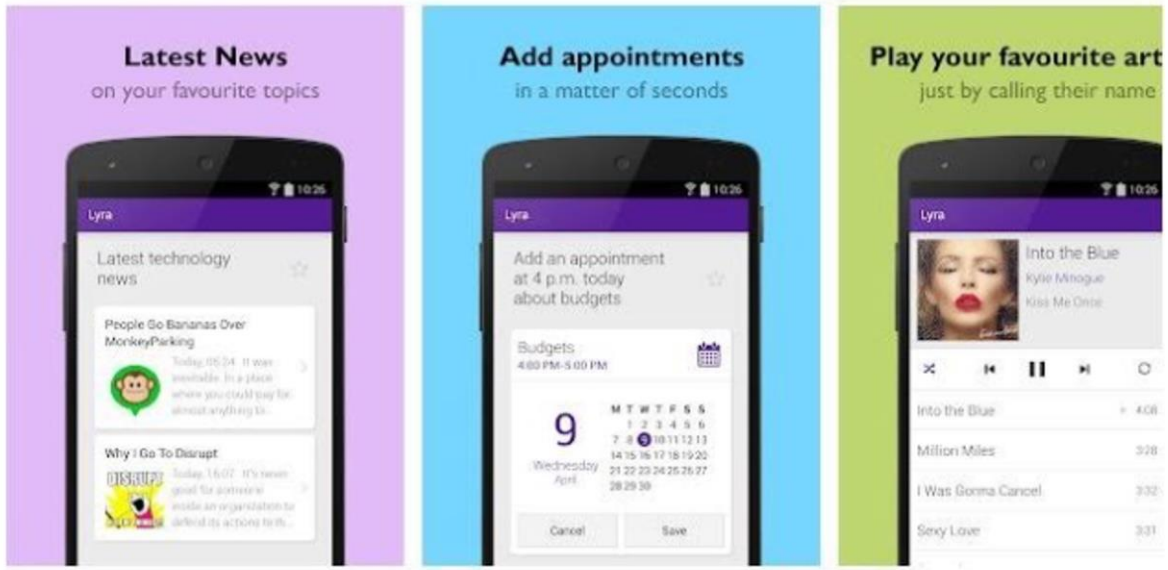
- Gmail Spell Checker

Examples of NLP Applications

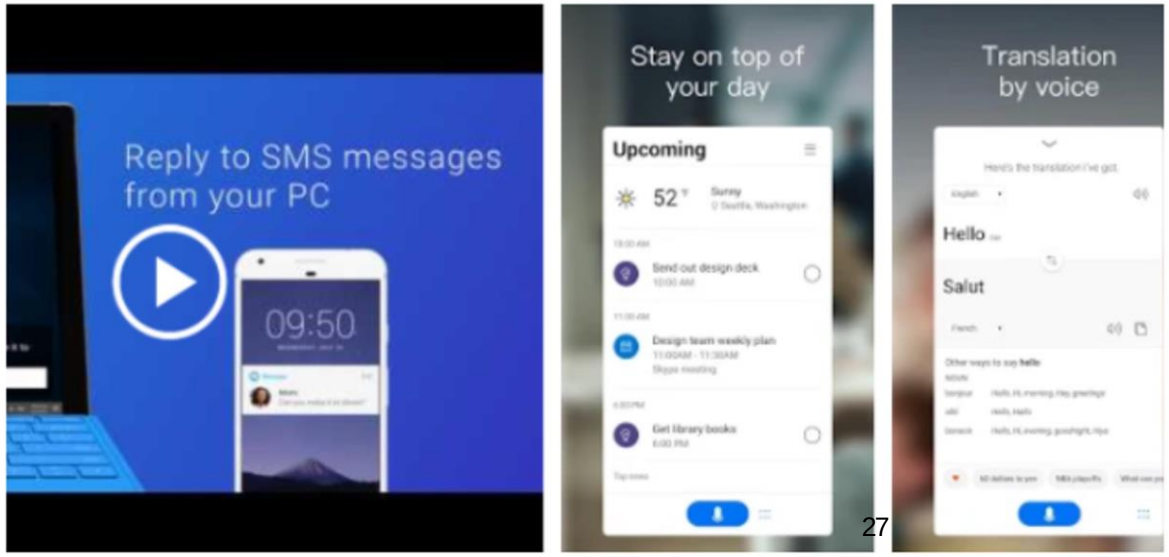
- Spam Classifier

- Personal Assistant Application with Speech/Voice Recognition

Lyra

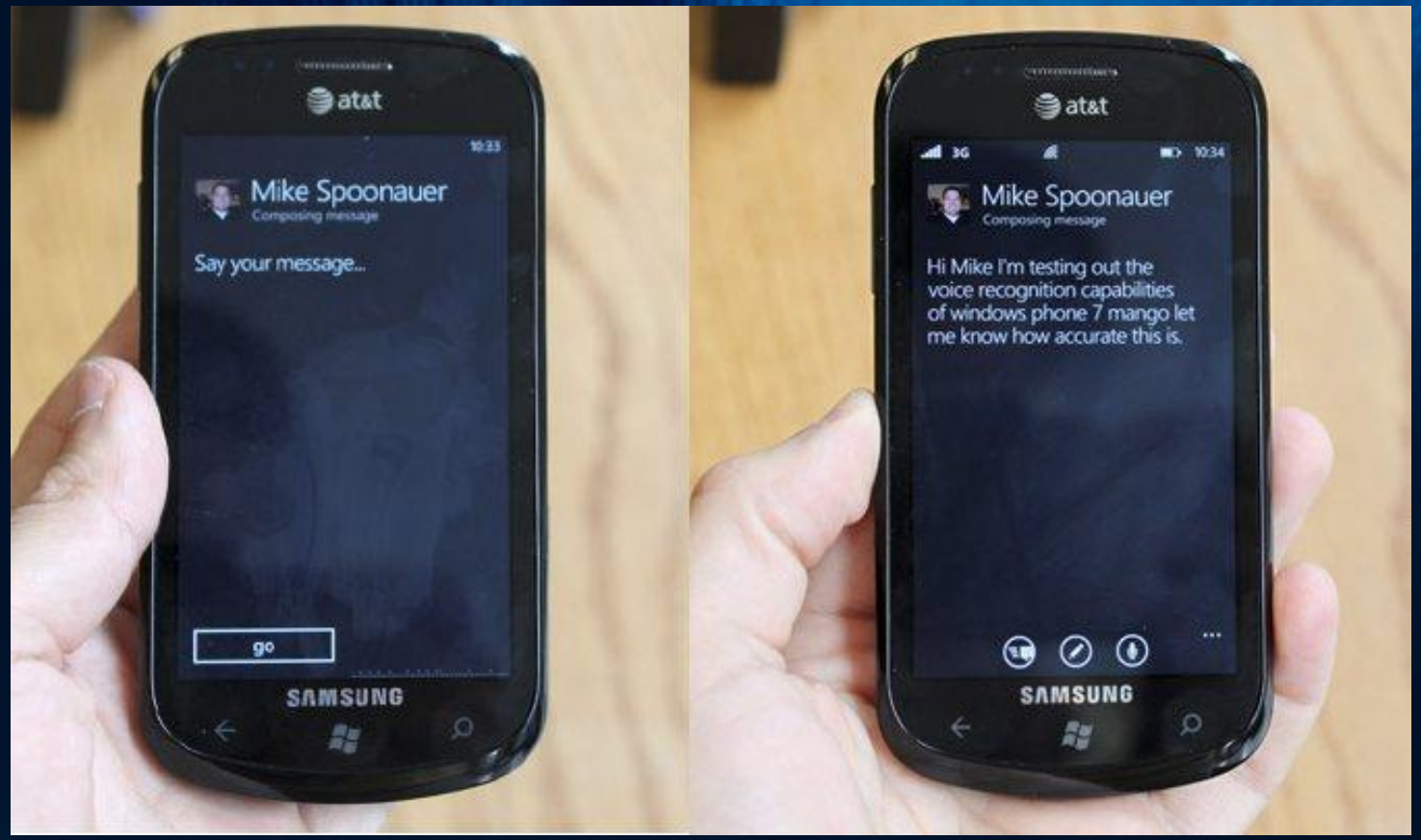


Cortana



Examples of NLP Applications

- Mango Speech Recognition



END OF WEEK 9 LECTURE

WHAT HAVE YOU LEARNED?

Questions ?