# *COMPUTER PROCESSING OF HUMAN LANGUAGE*

# *COMPUTATIONAL MORPHOLOGY*

- The processing of word structures by computers is computational morphology.
- The computer needs to understand the structure of words both to understand the words and to use the words in a grammatically correct way.

# *COMPUTATIONAL SYNTAX*

- The programming of computers of computers to analyze the structures of sentences.
- For example;a parser may contain the following rules:
- 1)A Sentence is often a *NOUN PHRASE*(or subject)followed by a VERB PHRASE (or predicate)
- *2)A NOUN PHRASE* is often DETERMİNER followed by a NOUN.
- FOR EXAMPLE; The child found the kittens.
- We can see that determiner (the)followed by a noun (child).rule 2 tells it that this phrase is a Noun Phrase. It would continue to process found,the,and kittens to construct a Verb Phrase.

# COMPUTATİONAL SEMANTİCS

- The programming of computers to determine the meanings of words phrases,sentences,and discourse

# COMPUTATİONAL PRAGMATICS

- The programming of computers to take context and situation into accound when determining the meanings of expressions.


QUESTION
TIME

1).....is the programming of computers to analyze the structure of words.

A)COMPUTATIONAL SYNTAX
B)COMPUTATIONAL MORPHOLOGY
C)COMPUTATIONAL LINGUISTICS
D)COMPUTATIONAL PRAGMATICS

2).....is the programming of computers to analyze the structures of sentences.

A)COMPUTATIONAL SYNTAX

B)COMPUTATIONAL  MORPHOLOGY

C)COMPUTATIONAL LINGUISTICS

D)COMPUTATIONAL PRAGMATICS

3).....is the programming of computers to determine the meanings of words,phrases,sentences,and discourse.

A)COMPUTATIONAL SYNTAX
B)COMPUTATIONAL MORPHOLOGY
C)COMPUTATIONAL SEMANTICS
D)COMPUTATIONAL PRAGMATICS

4).....is the programing of computers to take context and situation into account when determining the meanings of expressions.

A)COMPUTATIONAL PRAGMATICS
B)COMPUTATIONAL SEMANTICS
C)LİNGUİSTİC
D)SYNTAX

5Which of the following is the definition for "polysemy"?

A) Two senses of a word that do not have particular relations between them, for example, the "financial institution" and "sloping mound senses" of "bank."
B) Two senses of a word that are related semantically, for example, the "financial institute" and "the building belonging to a financial institution" senses of "bank."
C) Two senses of two different words that are (nearly) identical.
D) Two senses of two different words that are opposite to each other.
E) None of the above

# The End

**The Linguist Chicken**

Orhan Veli TEKİN
2013191055

Tugay KARACA
2012191065

# INFORMATION RETRIEVAL AND SUMMARIZATION

## ( BİLGİ ERİŞİMİ VE ÖZETLEME )

Information retrieval is the use of computers to locate and display data gathered from possibly very large databases. For example, the keyword *bird* might retrieve information based on *bird*, *birds*, *to bird* (verb infinitive), *bird feeders*, *water birds*, *avian*, *sparrow*, *feathers*, *flight*, *migration*. Through summarization programs, computers can eliminate redundancy and identify the most salient features of a body of information. These programs often use "concept vectors." It is a list of meaningful keywords whose presence in a paragraph is a measure of the paragraph's significance.

# SPELL CHECKERS
# (YAZIM DENETLEYİCİSİ)

- Spell checkers means checking the keywords to prevent misspellings from misleading the search. But spell checkers cannot replace careful editing in poems.

# MACHINE TRANSLATION

The first use of computers for natural language processing began in the 1940s with the attempt to develop automatic machine translation. The aim in automatic translation is to input a spoken utterance or a written passage in the source language and to receive a grammatical passage of equivalent meaning in the target language. Translation is more than word-for-word replacement. Often there is no equivalent word in the target language, and the order of words may differ, as in English *the red house* versus Spanish *la casa roja*. There is also difficulty in translating idioms, metaphors, slang, and so on. Human translators cope with these problems because they know the grammars of the two languages and draw on general knowledge of the subject matter and the intended meanings. There are challenges in morphology when translating between languages. For example, *mbuki-mvuki* is a Swahili word that means 'to shuck off one's clothes in order to dance.' English does not have a word for that practice. Professional translators make abundant use of computers to assist their translations, saving them much time and effort, but they are the ultimate arbiters of what is a grammatical, semantically faithful translation.

# COMPUTATIONAL FORENSIC LINGUISTICS

( ADLİ DİLBİLİM )

Computational forensic linguistics is a sub-area that concerns itself with computer applications in forensic linguistics. In this section we will look at three such applications: trademarks, interpreting legal terms, and speaker identification.

# TRADEMARKS

Forensic linguist Roger Shuy used a computer to search a huge corpus for words containing the precious morpheme. He found a large number of already accepted usages such as *McMansions*, *McArt*, *McCinema*, and *McPrisons*, and based on those data argued that the morpheme *Mc-* had entered the language with its own meaning, 'basic, inexpensive,' and was therefore available to the public at large. The judge did not agree and ruled against the hotel chain because market research showed that the public's perception of the morpheme *Mc-* was nonetheless strongly associated with McDonald's.

# INTERPRETING LEGAL TERMS

A recent case hinged on the legitimate use of the word *visa*, not as a credit card trademark, but as a legal term relevant to international travel. This analysis and many like it show the usefulness of a corpus-based, computer-driven approach to thorny legal problems.

# SPEAKER IDENTIFICATION

Many crimes involve anonymous recorded messages in which it is important to identify the speaker. Speaker identification is the use of computers to assist in such a task. Two computational tools are commonly applied to assist in speaker identification. One displays the waveform of an utterance, which shows the amplitude changes of the speech over time; the second displays a spectrogram which shows a breakdown of the frequencies of the speech signal over time.

# SUMMARY

- Computational linguistics is the study of how computers can process language, thus allowing natural language human-computer interfaces. As well, computers help scholars to analyze literature and language, to translate between languages, to extract useful information from large corpuses, and to assist with criminal and legal affairs.
- speech recognition is processing the speech signal into phonemes, morphemes, and words.
- Speech synthesis is a two-step process in which a text-to-speech program first converts text to phones or other basic units such as words or syllables.
- Computational phonetics and phonology relate phonemes to the acoustic signal of speech
- Computational morphology deals with the structure of words, so it determines that the meaning of *bird* applies as well to *birds*
- Computational syntax is concerned with the syntactic categories of words and with the larger syntactic units of phrases and sentences.
- Computational semantics is concerned with representing meaning inside the computer
- Computational pragmatics is the response of the computer by taking into account knowledge that the computer system has about the real world
- Computational lexicography is the use of computers both to construct "ordinary" dictionaries and to construct electronic dictionaries with far more information
- culturomics reveals many details of language change.
- Other applications of computational linguistics are found in the forensic fields, in which computers are used to examine huge corpuses to infer how people interpret trademarks such as the *Mc*in McDonald's; and speaker identification, where a computational analysis of speech used in a crime such as a bomb threat can assist in identifying a suspect.

‘’The government bans the smoking of children’’(sign in Istanbul)

What is the main problem, which makes the meaning incomplete, in this sentence?

A. Culturomics
B. Computational synthesis
C. Informational retrieval
D. Trademarks
E. Translation(machine)

The processing of word structures by computers is called .....

Choose the best answer below

A. semantics
B. speech recognition
C. foresnsic fields
D. computational morphology
E. pragmatics

Which one of the following is a subfield of culturomics ?

A. speaker identification
B. forensic fields
C. lexicography
D. twitterology
E. information retrieval

..... processes the speech signal into phonemes, morphemes, and words. Choose the best answer below

A. text-to-speech
B. computational syntax
C. speech recognition
D. speech synthesis
E. waveforms

What challenges do not seem in a translation?

A. Word choice
B. morphologic problems
C. syntactic problems
D. grammatical problems
E. information retrieval


UNIVERSAL
A Yusuf Hamza Company Presents

# *COMPUTER PROCESSING OF HUMAN LANGUAGE*

*Computational linguistics* is a subfield of linguistics and computer science that is concerned with the interaction of human language and computers.
*Computational linguistics* includes the analysis of written texts and spoken discourse,the translation of text and speech from one language to another,the use of human(not computer)languages for communication between computers and people.

# *COMPUTERS THAT TALK and LISTEN*

- The ideal computer is multilingual;it should “speak”computer languages such as Visual Basic and Java,and human languages such as French and Japanese.
- Computational linguistics is concerned with the interaction between language and computers in all dimensions,from phonetics to pragmatics,from producing speech to comprehending speech.

- *Computational phonetics and phonology* is concerned with processing speech.Its main goals are converting speech to text on the comprehension side,and text to speech on the production side.

- The areas of *computational morphology,computational syntax, computational semantics,and computantional pragmatics*,discussed below,are concerned with higher levels of linguistic processing.

# *COMPUTATIONAL PHONETICS AND PHONOLOGY*


# Computational Phonetics and Phonology

- The two sides of computational phonetics and phonology are speech recognition and speech synthesis. Speech recognition is the process of analyzing
the speech signal into its component phones and phonemes, and producing,
in effect, a phonetic transcription of the speech. Further processing may convert the transcription into ordinary text for output on a screen, or for further
processing such as a speech understanding application. (*Note*: Speech recognition is *not* the same as speech understanding, as is commonly thought. Rather,
speech recognition is a necessary precursor to the far more complex process of
comprehension.)

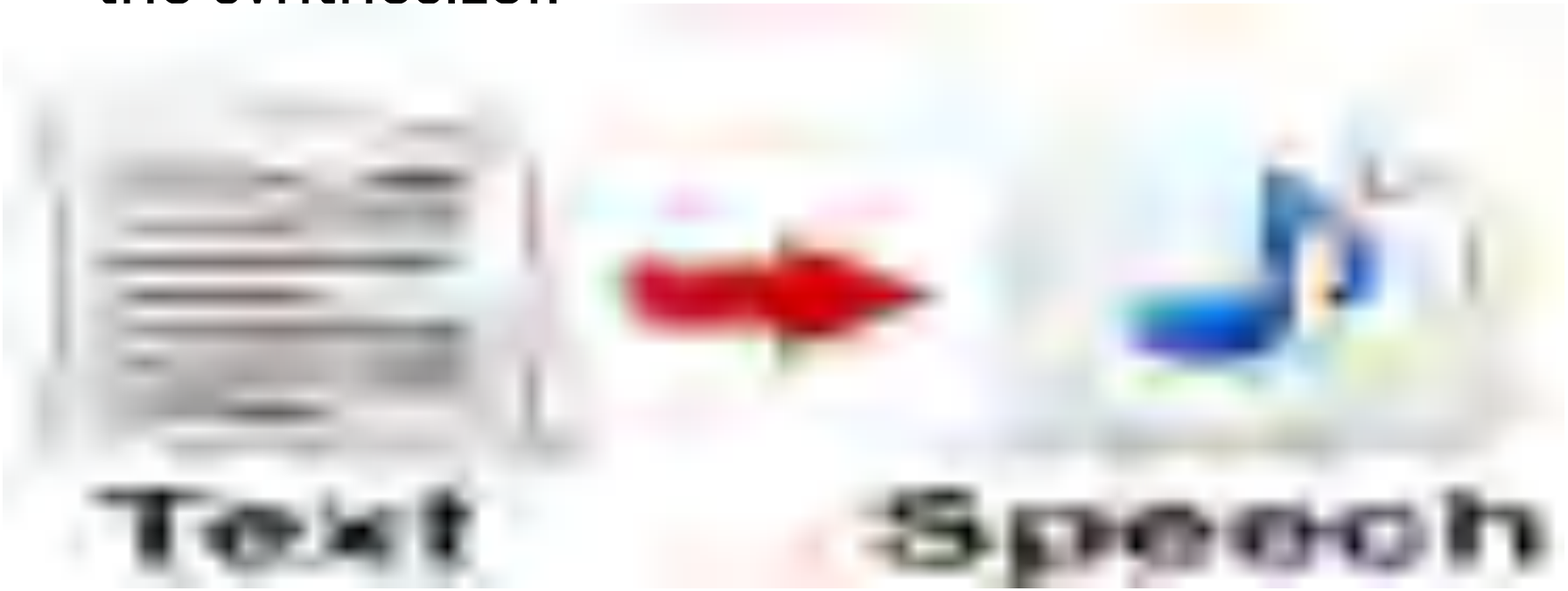
- ***SPEECH SYNTHESİS***;An electronic process that produces speech either from acoustically simulated sounds or from prerecorded units.
- *SPEECH RECOGNITION*;In computer processing,the ability to analyze speech sounds into phones,phonemes,morphemes, and words; the transcription of speech.

# Speech Recognition

- When Frederic was a little lad he proved so brave and daring,
  His father thought he'd 'prentice him to some career seafaring.
  I was, alas! his nurs'rymaid, and so it fell to my lot
  To take and bind the promising boy apprentice to a pilot—
  A life not bad for a hardy lad, though surely not a high lot,
  Though I'm a nurse, you might do worse than make your boy a pilot.
  I was a stupid nurs'rymaid, on breakers always steering,
  And I did not catch the word aright, through being hard of hearing;

# *TEXT-TO-SPEECH*

- To provide input to the speech synthesizer, a computer program called text-to-speech converts written text into the basic units of the svnthesizer.




QUESTION
TIME

1).....is an electronic process that produces speech either from acoustically simulated sounds or from prerecorded units.

A)PRAGMATİC
B)SPEECH RECOGNATION
C)PHONOLOGY
D)SPEECH SYNTHESIS

2)To provide input to do speech synthesizer,a computer program called.....converts written text into the basic units of the synthesizer.

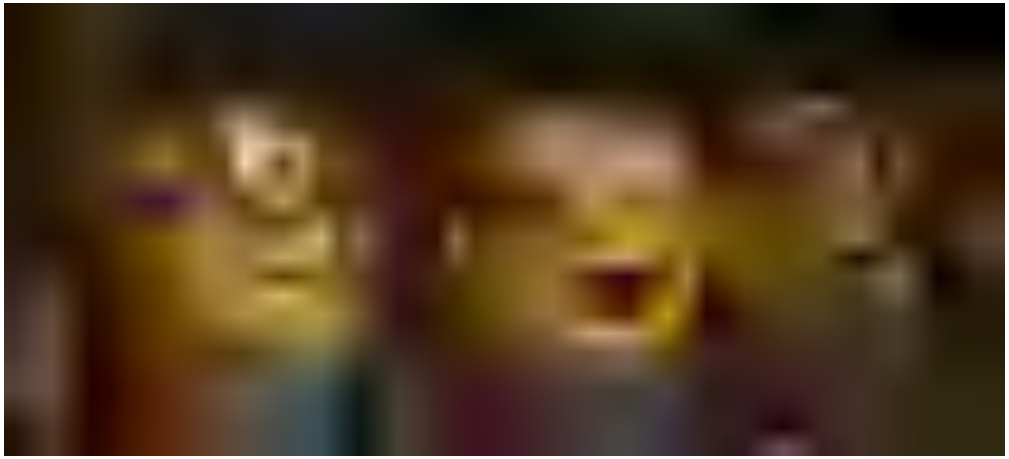
A)TEXT-TO-SPEECH
B)SPEECH RECOGNITION
C)COMPUTATIONAL MORPHOLOGY
D)SEMANTICS

3) Speech recognition- speech recognisers must be trained to the voice of a specific person- and speech synthesis- format synthesis and concatenative synthesis (application: benefiting visually impaired people)

A) COMPUTATIONAL SEMANTICS
B) COMPUTATIONAL FORENSIC LINGUISTICS
C)COMPUTATIONAL PHONETICS & PHONOLOGY
D)COMPUTATIONAL LEXICOLOGY

# Yusuf HORUZ

2013191019



# COMPUTER PROCESSİNG OF HUMAN LANGUAGE

*Applications of Computational Linguistics*

In this section we discuss some of the more common application areas, ranging from the use of computers to test a linguist´s grammar for faithfulness to the actual language ,to the use of computers to solve language crimes the field of computational forensic linguistics.

# Computer Models of Grammar

Computers may be programmed to model the linguist´s description of the (human) grammar of a language.

Because grammars are infinite in scope, prodigious amounts of testing are necessary to verify their linguistic description , and that´s where computers play an important role.

Computational linguists develop computer programs to generate the sentences of a language and simulate human processing of these sentences using the rules included in various current linguistic theories

# Frequency Analysis , Concordances , and Collocations

The relative frequencies of letters and sounds , morphemes , words , word categories , types of phrases , and so on may be swiftly and accurately computed for any corpus (body of language data) , whether textual or spoken.

A concardance takes frequency analysis one step further by specifying the location within the text of each word and its surrounding context.

A way to refine a concordance is through collocation analysis. A collocation is the occurrence of two or more words within a short space of each other in a corpus. This point is to find evidence that the presence of one word in the text affects the occurrence of other words.

# Computational Lexicography

The field of computational lexicography, is concerned not only with the making of standard dictionaries but also with the building of electronic dictionaries specifically designed for computational linguists.

Some of the information computational linguists need follows :

- Phonemic transcription
- Phonetic variants ( dialectal, societal )
- Syllabification
- Syntactic categories
- Semantic properties
- Number ,e.g , people is plural ,person is singular
- Gender , e.g , ship is female
- C-selection (murder requires a direct object )
- S-selection (murder requires a human subject and object )
- Stylistic level (ain´t is informal, rad is slang, fuck is taboo, etc.)
- Synonyms , antonyms ,possible homophones, etc.

# The Culturomic Revolution

Analysis of corpus , a study termed culturomics , is producing fascinating insights in the fields of lexicography and grammatical change as well as in many of the social sciences.

The culturomics revolution has produced results on the historical change of irregular verbs. Irregular past-tense verbs live side-by- side with their regular -ed rivals (e.g.,dove and dived) that threaten to supplant them.

Culturomic analysis informs us that in the period between 1800 and 2000 the six irregular past-tense verbs burn , chide , smell , spell , spill , and thrive ( burnt , chid , smelt , spelt , spilt , throve ) regularized , while at the same time two regular past tense verbs light /lighted and wake /waked became irregular : light/lit and wake/woke .

# Twitterology

Twitterology , the computer analysis of a vast corpus of microblogs known as tweets , is becoming a subfield of culturomics.

# The End