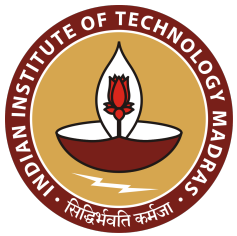
BSCS5002: Introduction to Natural Language Processing

Lecture 4: Levels of Language Processing and Linguistic Fundamentals

Parameswari Krishnamurthy



Language Technologies Research Centre
IIIT-Hyderabad

param.krishna@iiit.ac.in

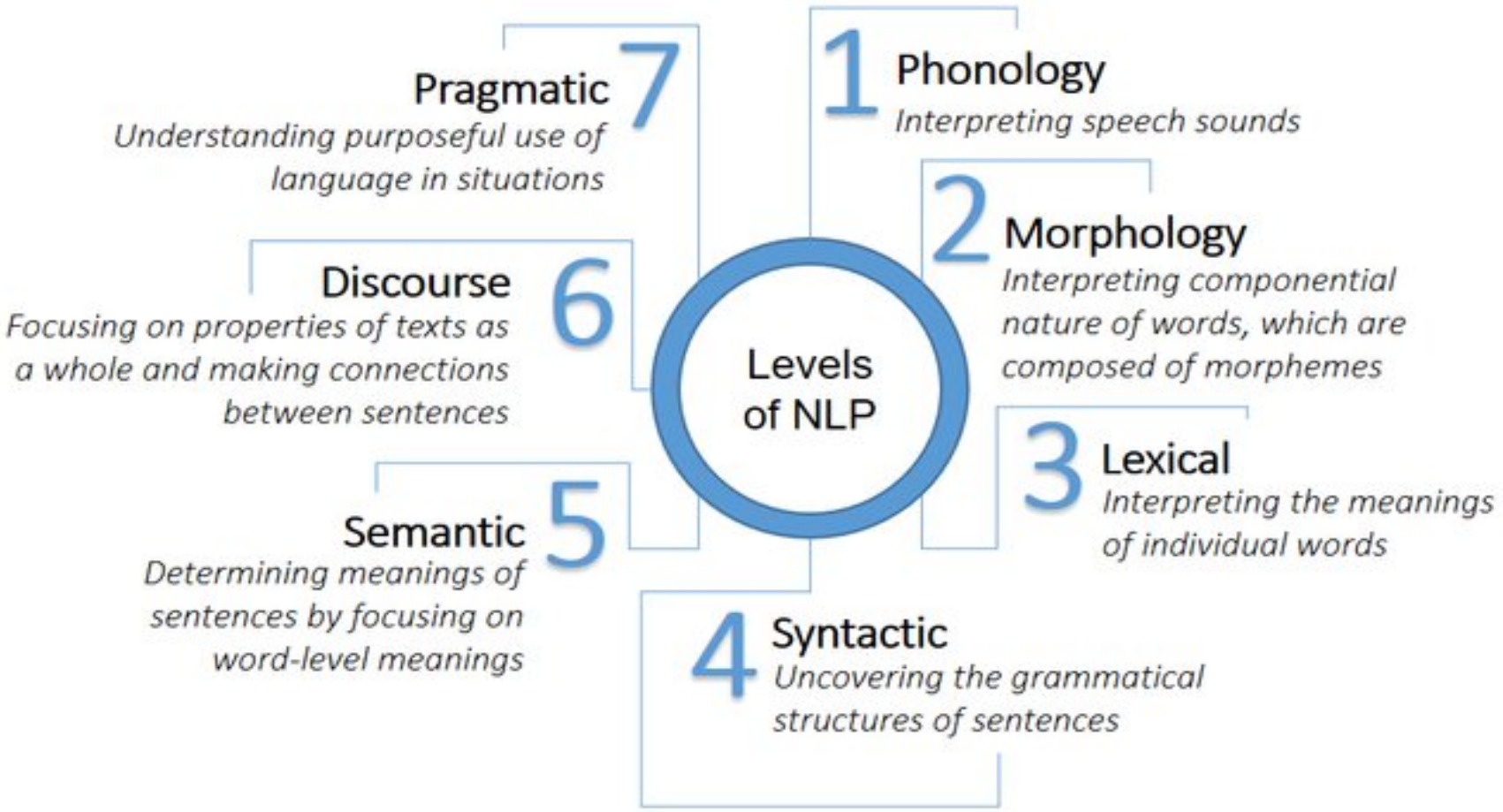
Levels of Language Processing in NLP

- Natural Language Processing (NLP) involves multiple levels of language processing.
- Each level addresses different linguistic features and contributes to understanding and generating natural language.

Levels of Language Processing:

- Phonological level
- Morphological level
- Lexical Level
- Syntactic Level
- Semantic Level
- Discourse Level
- Pragmatic Level

Levels of NLP



Seven levels of NLP (Feldman, 1999, p. 62-64; Liddy, 2010, p. 3867-3868)

Levels and Applications of NLP

PROCESSING LEVEL	TASKS AND APPLICATIONS
Character & strings level	Word tokenization, sentence boundary detection, gene symbol recognition, text pattern extraction
Word token level	POS-tagging, parsing, chunking, term extraction, gene mention recognition
Sentence level	Sentence classification and retrieval and ranking, question answering, automatic summarization
Sentence window level	Anaphora resolution
Paragraph & passages level	Detection of rhetorical zones
Whole document level	Document similarity calculation
Multi-document collection level	Document clustering, multi-document summarization

source: researchgate.net/profile/Martin-Krallinger

1. Phonological Level in NLP

Definition:

- The phonological level in NLP focuses on the processing of sounds in natural language, including their organization and patterns.

Significance in NLP:

- Essential for applications like speech recognition and text-to-speech systems.
- Helps in understanding pronunciation, accent variations and phonetic transcription.

1. Phonological Level in NLP

Key Components:

- **Phonemes:** The smallest units of sound that distinguish meaning.
- **Phonetic Transcription:** Representing sounds using phonetic symbols (e.g., IPA).
- **Prosody:** The rhythm, stress and intonation patterns of speech that convey meaning beyond the words themselves.

1. Phonological Level in NLP

Examples in NLP:

- **Speech Recognition:** Identifying spoken words by analyzing phonemes. For example, recognizing "cat" vs. "bat" based on initial phoneme differentiation.
- **Text-to-Speech (TTS):** Generating spoken language from text while considering prosody for natural-sounding speech. For example, varying intonation in questions vs. statements.
- **Accent Recognition:** Differentiating between various accents (e.g., American vs. British English) by analyzing phonetic patterns.

2. Morphological Level in NLP

Definition:

- The morphological level in NLP involves the study of the structure and formation of words, including their internal components.

Significance in NLP:

- Crucial for tasks such as morphological analysis, lemmatization and stemming and part-of-speech tagging.
- Helps in understanding word forms, variations and their meanings.

2. Morphological Level in NLP

Key Components:

- **Morphemes:** The smallest meaningful units of language (e.g., prefixes, suffixes, roots).
- **Inflection:** Modifications of words to express different grammatical categories (e.g., tense, number).
- **Derivation:** The process of creating new words by adding prefixes or suffixes (e.g., "happy" $\rightarrow$ "unhappy").

2. Morphological Level in NLP

Examples in NLP:

- **Morphological Analysis:** Analyzing the structure of words (prefixes, suffixes).
- **Lemmatization:** Reducing words to their base or dictionary form. For instance, "better" is lemmatized to "good."
- **Stemming:** Cutting words to their root forms. For example, "running," "runner," and "ran" may all stem to "run."
- **Part-of-Speech Tagging:** Identifying the grammatical category of words based on their morphology. For example, "running" can be tagged as a verb or noun depending on context.

3. Lexical Level in NLP

Definition:

- The lexical level in NLP deals with the vocabulary of a language, focusing on the meaning and usage of words.

Significance in NLP:

- Essential for understanding the meaning of text, word relationships and language generation.
- Important for tasks such as information retrieval, sentiment analysis and machine translation.

3. Lexical Level in NLP

Key Components:

- **Tokenization:** The process of breaking text into individual words or tokens.
- **Lexical Semantics:** The study of word meanings, relationships and nuances (synonyms, antonyms, hypernyms).
- **Vocabulary:** The set of words available in a language or a specific domain.

3. Lexical Level in NLP

Examples in NLP:

- **Tokenization:** The sentence *"The quick brown fox jumps over the lazy dog."* is tokenized into:

  [The, quick, brown, fox, jumps, over, the, lazy, dog]

- **Lexical Semantics:** Understanding that the word *"bank"* can mean a financial institution or the side of a river based on context.
- **Synonym Detection:** Identifying that *"happy"* and *"joyful"* convey similar meanings in sentiment analysis tasks.

4. Syntactic Level in NLP

Definition:

- The syntactic level in NLP involves the arrangement of words to form grammatical sentences and the study of sentence structure.

Significance in NLP:

- Crucial for understanding sentence meaning and structure.
- Important for tasks such as parsing, machine translation and information extraction.

4. Syntactic Level in NLP

Key Components:

- **Parsing:** The process of analyzing the grammatical structure of a sentence.
- **Constituency Parsing:** Identifying phrases within sentences based on hierarchical structure.
- **Dependency Parsing:** Establishing relationships between words in a sentence and their dependencies.

4. Syntactic Level in NLP

Examples in NLP:

- **Parsing:** For the sentence *"The cat sat on the mat."*, parsing can show:

  [NP The cat] [VP sat [PP on [NP the mat]]]

- **Dependency Parsing:** In the sentence *"She loves him."*, dependencies can be visualized as:

  loves $\leftarrow$ She
  loves $\rightarrow$ him

- **Grammar Rules:** Syntax can be governed by rules such as:
  - Subject-Verb-Object (SVO) structure in English.
  - Differences in structure like Subject-Object-Verb (SOV) in Hindi.

5. Semantic Level in NLP

Definition:

- The semantic level in NLP deals with the meaning of words, phrases and sentences, focusing on how meaning is constructed and interpreted.

Significance in NLP:

- Essential for understanding context and resolving ambiguities in language.
- Important for tasks such as sentiment analysis, information retrieval and machine translation.

5. Semantic Level in NLP

Key Components:

- **Word Sense Disambiguation (WSD):** Determining which meaning of a word is used in context.
- **Semantic Role Labeling (SRL):** Identifying the roles that words play in a sentence (e.g., agent, action, object).
- **Named Entity Recognition (NER):** Recognizing and classifying entities in text (e.g., names, dates, locations).

5. Semantic Level in NLP

Examples in NLP:

- **Word Sense Disambiguation:** The word *"bank"* can mean a financial institution or the side of a river and WSD helps determine the correct meaning based on context (e.g., "He went to the bank to deposit money." vs. "The river bank was eroded.").
- **Semantic Role Labeling:** In the sentence *"John gave Mary a book"*, SRL identifies:
  - Agent: John
  - Recipient: Mary
  - Theme: a book
- **Named Entity Recognition:** In the sentence *"Apple is looking at buying U.K. startup for dollar 1 billion"*, NER identifies:
  - Organization: Apple
  - Location: U.K.
  - Monetary Value: dollar 1 billion

6. Discourse Level in NLP

Definition:

- The discourse level in NLP deals with how sentences relate to one another in larger contexts, focusing on coherence and cohesion in language.

Significance in NLP:

- Essential for understanding context beyond individual sentences.
- Important for tasks such as text summarization, machine translation and dialogue systems.

6. Discourse Level in NLP

Key Components:

- **Coherence:** The logical flow of ideas in a text, ensuring that sentences connect meaningfully.
- **Cohesion:** The grammatical and lexical linking within a text that helps maintain the flow.
- **Anaphora Resolution:** Identifying which words refer back to others (e.g., resolving pronouns to their antecedents).

6. Discourse Level in NLP

Examples in NLP:

- **Coherence:** In the text *"She went to the bakery. She bought a loaf of bread."*, the ideas are coherent as the second sentence logically follows the first.
- **Cohesion:** The use of linking words such as *"however,"* *"therefore,"* and *"moreover"* helps create cohesion between sentences.
- **Anaphora Resolution:** In the sentences *"John loves his dog. It follows him everywhere."*, resolving *"It"* refers to *"his dog"*.

7. Pragmatic Level in NLP

Definition:

- The pragmatic level in NLP focuses on how context influences the interpretation of meaning in communication, considering the speaker's intent and the situational context.

Significance in NLP:

- Essential for understanding implied meanings, sarcasm and conversational dynamics.
- Important for tasks such as dialogue systems, sentiment analysis and context-aware machine translation.

7. Pragmatic Level in NLP

Key Components:

- **Speech Acts:** Actions performed via utterances (e.g., requests, promises, assertions).
- **Deixis:** Words and phrases (like pronouns and demonstratives) that require contextual information to convey meaning (e.g., "here," "you").
- **Contextual Understanding:** The ability to interpret meaning based on situational context and background knowledge.

7. Pragmatic Level in NLP

Examples in NLP:

- **Speech Acts:** The phrase *"Can you pass the salt?"* is a request, not just a question about capability.
- **Deixis:** In the sentence *"I will meet you here tomorrow,"* the interpretation of *"here"* and *"tomorrow"* depends on the speaker's location and the current day.
- **Contextual Understanding:** In a conversation where one person says, *"It's cold in here,"* they may be implying a request to close a window rather than just stating a fact.

Conclusion

- The levels of language processing in NLP are fundamental for interpreting human language effectively.
- Each level contributes uniquely in NLP applications.
- **Phonological Level:** Focus on sounds and forms, essential for speech recognition and text-to-speech systems.
- **Morphological, Lexical and Syntactic Levels:** Deal with meanings and grammatical relationships, crucial for parsing and information extraction.
- **Semantic and Discourse Levels:** Enhance understanding of meaning beyond sentences, vital for machine translation and dialogue systems.
- **Pragmatic Level:** Emphasizes context in communication, relevant for conversational agents and sentiment analysis.

Conclusion

- Integrating insights from these levels leads to more accurate and effective interactions between humans and machines.
- Advancements in these processing levels will enhance NLP capabilities, making them adept at understanding and generating language contextually and semantically.