

Unit 4: Semantic Parsing

i) **Word Senses and WordNet:** Word senses, relation between senses, WordNet, wordsense disambiguation, WSD algorithm and task,

1. Word Senses:

- Words are often **ambiguous**, meaning they can have multiple meanings or senses. For instance, consider the word “mouse,” which can refer to both a small rodent and a hand-operated device to control a cursor.
- A **sense** (or **word sense**) represents one specific aspect of the meaning of a word. For example, the word “bank” has different senses: (1) a financial institution and (2) a sloping mound.
- Some words are **polysemous**, having multiple related senses. For instance, “bat” can mean either the animal or a baseball bat.
- WordNet, a large online thesaurus, captures these word senses and their relationships.

2. WordNet:

- **Definition:** WordNet is a lexical database that organizes words into synsets (sets of synonyms) based on their meanings.
- **Senses Representation:** WordNet represents word senses by associating them with lemmas (base forms of words). For example, “bank1” and “bank2” represent different senses of the word “bank.”
- **Relations Between Senses:**
 - **IS-A Relation:** WordNet establishes an IS-A relation between senses. For instance, “dog” IS-A “mammal.”
 - **Part-Whole Relation:** WordNet also captures part-whole relations. For example, an “engine” is part of a “car.”
 - Understanding these relations is crucial for language comprehension.

3. Word Sense Disambiguation (WSD):

- **Task:** WSD involves determining the correct sense of a word in a given context.
- **Supervised Approaches:** These use labeled corpora where words are annotated with senses (e.g., SemCor, which labels sentences with WordNet senses).
- **Unsupervised Approaches:** These algorithms decide the intended sense without labeled data.
- **Applications:**
 - Enhancing dialogue agents: Accurate WSD helps agents understand user queries better (e.g., distinguishing between “bat care” for a vampire or baseball).
 - Improving machine translation: Different senses may have distinct translations (e.g., Spanish “murciélago” for the animal and “bate” for the baseball bat).

ii) **Word sense inductions Semantic role labelling:** semantic roles, diathesis alteration, problems with thematic roles, proposition bank, framenet, semantic role labelling, selection restrictions, decomposition of predicates

1. Semantic Roles:

- Semantic roles, also known as **theta roles**, represent the different functions that words or phrases play in a sentence with respect to the main verb (predicate).
- These roles help us understand the underlying meaning of a sentence by identifying who performs the action, what is affected by the action, and other relevant details.
- Common semantic roles include:
 - **Agent:** The entity that initiates or performs the action (e.g., "Mary" in "Mary baked a cake").
 - **Theme:** The entity affected by the action (e.g., "cake" in the same sentence).
 - **Experiencer:** The entity experiencing a certain state or emotion (e.g., "I" in "I feel happy").
 - **Instrument:** The means or tool used to perform an action (e.g., "knife" in "cut with a knife").
 - **Location:** The place where an action occurs (e.g., "park" in "played in the park").
 - **Time:** The temporal information associated with an event (e.g., "yesterday" in "She visited yesterday").

2. Diathesis Alteration:

- Diathesis alteration refers to changing the voice or grammatical structure of a sentence.
- In English, we commonly encounter two voices:
 - **Active Voice:** The subject performs the action (e.g., "The cat chased the mouse").
 - **Passive Voice:** The subject receives the action (e.g., "The mouse was chased by the cat").
- Diathesis alteration affects the assignment of semantic roles. For example:
 - In the active voice, the cat is the **agent** (performing the action), and the mouse is the **theme** (affected by the action).
 - In the passive voice, the mouse becomes the **theme**, and the cat is still involved but not explicitly mentioned.

3. Problems with Thematic Roles:

- While semantic roles provide valuable insights, they are not always straightforward:
 - **Ambiguity:** Some words can play multiple roles depending on context (e.g., "bank" as a financial institution or a riverbank).
 - **Implicit Roles:** Not all roles are explicitly stated in a sentence (e.g., "She laughed" lacks an explicit experiencer).
 - **Verb-Specific Roles:** Different verbs assign different roles (e.g., "give" has a giver and a recipient).

- **Frame Variation:** The same verb can have different roles in different frames (e.g., “run” in “run a race” vs. “run a business”).

4. Proposition Bank (PropBank):

- PropBank is a valuable linguistic resource that annotates text with information about basic **semantic propositions**. It adds **predicate-argument relations** to the syntactic trees of the Penn Treebank.
- The original PropBank project, funded by ACE, laid the foundation for this corpus. Since then, it has evolved and expanded through various funding sources.
- PropBank annotations are available for different genres, including broadcast news, webtext, conversation, and even clinical notes.
- [Notably, PropBank has been mapped to other resources like VerbNet and FrameNet as part of SemLink, creating a unified representation of these three valuable linguistic databases¹.](#)

5. FrameNet:

- FrameNet provides a rich lexicon of **frames**, which are semantic structures representing typical situations, events, or scenarios.
- Each frame includes a set of **lexical units** (words or phrases) associated with specific semantic roles.
- [By interconnecting FrameNet with PropBank and VerbNet, we enhance our understanding of how verbs behave syntactically and semantically².](#)

6. Semantic Role Labeling (SRL):

- SRL involves assigning semantic roles (such as **agent**, **theme**, **experiencer**, etc.) to words or phrases within a sentence based on their relationship to the main verb (predicate).
- SRL bridges the gap between syntax and semantics, enabling better comprehension of language and supporting downstream NLP tasks.
- [PropBank plays a crucial role in SRL by providing annotated corpora with semantic role labels, allowing researchers to build and evaluate SRL systems³.](#)

7. Selection Restrictions:

- Selection restrictions, also known as **selectional restrictions**, are constraints on the arguments that a verb can take.
- Verbs impose specific requirements on the types of entities that can fill their semantic roles. For example:
 - The verb “eat” expects a **theme** (e.g., “pizza,” “apple”).
 - The verb “give” requires both an **agent** (the giver) and a **theme** (the recipient).
- Understanding selection restrictions is essential for accurate SRL and natural language understanding.

8. Decomposition of Predicates:

- Decomposing predicates involves breaking down complex verb phrases into their constituent parts.

- For instance, the verb phrase “is playing soccer” can be decomposed into:
 - **Predicate:** “playing”
 - **Theme:** “soccer”
 - **Agent** (implicit): The person doing the playing
- Decomposition helps identify the semantic roles associated with each part of the predicate

iii) **Lexicon for sentiment, affect and connotation:** emotions, sentiment and affect lexicons, Creating Affect Lexicons by Human Labeling, Semi-supervised Induction of Affect Lexicons

1. Affective Meaning and Semantic Roles:

- Affective meaning encompasses various aspects, including **emotion, sentiment, personality, mood, and attitudes**.
- To explore these dimensions, we can refer to Scherer’s typology of affective states:
 - **Emotion:** A relatively brief episode of response to an external or internal event, characterized by emotions such as anger, sadness, joy, fear, pride, and elation.
 - **Mood:** A diffuse affect state of low intensity but relatively long duration, often without an apparent cause (e.g., cheerful, gloomy, irritable).
 - **Interpersonal Stance:** The affective stance taken toward another person in a specific interaction (e.g., warm, supportive, contemptuous).
 - **Attitude:** Relatively enduring, affectively colored beliefs, preferences, and predispositions towards objects or persons (e.g., liking, loving, hating).
 - **Personality Traits:** [Emotionally laden, stable personality dispositions \(e.g., nervous, anxious, reckless\)](#) ¹.

2. Sentiment Lexicons as a Subset of Emotion Lexicons:

- Sentiment lexicons focus on extracting the **positive or negative orientation** expressed by writers in text.
- For example, sentiment analysis identifies what people like or dislike based on affect-rich texts like consumer reviews, editorials, or social media posts.
- Detecting emotions and moods is useful for understanding student engagement, caller frustration, or even tracing fear in novels over time.
- Sentiment lexicons are a valuable resource for NLP tasks that involve opinion mining and sentiment classification.

3. Building Lexicons:

- Lexicons can be constructed through various methods:
 - **Human Labeling:** Annotators manually label words or phrases with affective information.

- **Semi-Supervised:** Leveraging existing labeled data and propagating labels to unlabeled data.
- **Supervised:** Training models using labeled examples.
- Publicly available lexicons serve as valuable starting points for research and applications.

iv) Supervised Learning of Word Sentiment, Using Lexicons for Sentiment Recognition

1. Supervised Learning of Word Sentiment:

- In the context of sentiment analysis, supervised learning involves training a model to predict the sentiment (positive, negative, or neutral) of a given text.
- One approach is to learn **word vectors** that capture both semantic and sentiment similarities among words. These vectors encode continuous similarities between words in a high-dimensional space.
- The model combines unsupervised techniques (such as probabilistic models of documents) with supervised sentiment components. The latter predicts sentiment annotations based on the vector representations of words in their contexts.
- [By leveraging both continuous and multi-dimensional sentiment information, this approach outperforms previous methods for sentiment classification¹.](#)

2. Using Lexicons for Sentiment Recognition:

- Lexicon-based sentiment analysis is a technique in natural language processing (NLP) that detects sentiment in text using predefined lists of words and phrases (lexicons or dictionaries).
- Here's how it works:
 - Lexicons contain sentiment scores or polarity labels associated with words. For example, "happy" might have a positive sentiment score, while "sad" could have a negative sentiment score.
 - Each word in a given text is labeled using a valence dictionary. Words like "great" (positive valence) or "terrible" (negative valence) contribute to the overall sentiment.
 - The sentiment score is calculated by counting positive and negative words and combining their values. A positive score indicates positive sentiment, a negative score indicates negative sentiment, and zero means neutral.
 - Lexicon-based approaches skip the process of building a machine learning model and provide on-the-fly sentiment analysis based on valence dictionaries.
- [Lexicons are language-dependent and can be sourced from linguistic departments of national universities².](#)

3. Benefits of Lexicon-Based Sentiment Analysis:

- Quick and efficient: Lexicon-based methods automate sentiment analysis for large volumes of text.

- Simple yet insightful: While context-dependent meanings can be challenging, lexicons provide fast insights into sentiment.
- No need for model training: Lexicon-based approaches avoid the complexity of building machine learning models.

iv) Other tasks: Personality, Affect Recognition, Lexicon-based methods for Entity-Centric Affect, Connotation Frames

1. Affect Recognition:

- **Emotion:** Emotion refers to a **relatively brief episode of response** triggered by the evaluation of an external or internal event as being of major significance. [Examples include feelings of anger, sadness, joy, fear, shame, pride, elation, and desperation¹.](#)
- **Mood:** Mood, on the other hand, is a **diffuse affect state** characterized by low intensity but relatively long duration. It often lacks an apparent cause and manifests as a change in subjective feeling. [Think of it as the overall emotional backdrop that colors our experiences¹.](#)

2. Personality:

- Personality encompasses various dimensions that define an individual's behavioral tendencies and emotional patterns. One widely studied model is the **"Big Five"** personality traits:
 - **Extroversion vs. Introversion:** Extroverts thrive on social interactions, while introverts recharge through solitude.
 - **Emotional Stability vs. Neuroticism:** [Emotional stability reflects resilience and calmness, whereas neuroticism involves emotional volatility and anxiety².](#)

3. Lexicons for Sentiment, Affect, and Connotation:

- Lexicons play a crucial role in understanding affective meaning. They provide information about sentiment, emotion, and connotation associated with words.
- Lexicon-based methods involve:
 - **Human Labeling:** Manually annotating words with affective labels.
 - **Semi-Supervised Approaches:** Combining labeled data with unlabeled data to build lexicons.
 - **Supervised Learning:** Training models to predict affective labels based on context.
- Beyond sentiment analysis, we explore:
 - **Entity-Centric Affect:** Detecting affect toward specific entities (e.g., products, people, events).
 - **Connotation Frames:** [Capturing nuanced connotations associated with words or phrases](#)