Module 1: Introduction to Conversational AI, NLP, and Deep Learning

Challenges – Language is Compositional

Language is inherently compositional, meaning that the meaning of sentences is derived not only from the meanings of individual words but also from how these words are combined. This presents a major challenge in Natural Language Processing (NLP). For example:

- "The dog bit the man" versus "The man bit the dog" – the words are the same, but the structure creates a completely different meaning.
- Ambiguity arises because words can combine in multiple ways depending on syntax and context.

This is why machines need both lexical knowledge (the meanings of words) and structural understanding (grammar, syntax, and composition rules).

Carefully Slide

This slide emphasizes the idea that human language cannot be treated as a simple bag of words. Care must be taken to preserve ordering, relationships, and context, which are crucial for meaning.

Challenges – Scale

Another difficulty in NLP is the **scale** of language data:

- Bible (King James Version): $\sim 700,000$ words
- Penn Treebank: ~ 1 million words from the Wall Street Journal
- Newswire collections: 500 million+ words
- Wikipedia (English): 2.9 billion words
- The Web: several billion words and growing

The enormous variation in scale requires systems that can handle both small, carefully curated corpora and massive, noisy, constantly updated datasets. Large-scale computation, distributed storage, and efficient algorithms become essential.

The Lingo of NLP

NLP has its own technical vocabulary:

Lexicography: What does each word mean?

Words can have multiple meanings (polysemy). Example:

- "He plays bass guitar." (instrument)
- "That bass was delicious!" (fish)

Semantics: How can we infer meaning?

- "I saw the man on the hill with the telescope." (ambiguous: Who has the telescope?)
- "The iPod is so small!" vs "The monitor is so small!" (pragmatic meaning differs depending on expectations).

Discourse: Meaning across multiple sentences

- "President Bush met with President-Elect Obama today at the White House. He welcomed him and showed him around." Who is "he" and who is "him"? Humans resolve this easily, but for a machine, it requires co-reference resolution.

Classic NLP Pipeline

Traditional NLP often follows a pipeline: tokenization, morphological analysis, parsing, semantic interpretation, discourse integration, and finally, applications like translation or information retrieval.

Digital Signal Processing Meets NLP

Speech processing combines signal processing (handling raw audio) with language modeling. For instance, recognizing spoken words requires analyzing sound waves, mapping phonemes to words, and then interpreting context.

Spoken Language Processing

Spoken language processing is a major area of NLP with applications such as:

- Speech recognition: automatic transcription, dictation software, accessibility tools, indexing video content.
- Studying intonation: prosody affects meaning (e.g., sarcasm).
- Emotion detection: distinguishing happiness, anger, or uncertainty.
- Deception detection: analyzing voice cues.

Challenges include variability in voices (male vs female, adult vs child), accents, speaking styles, background noise, and conversational dynamics (interruptions, turn-taking).

Text-to-Speech and Dialogue Systems

Beyond recognition, systems can also generate spoken language:

- Call center automation and virtual assistants.
- Tutoring systems that interact with students.
- Challenges: making artificial voices sound natural and ensuring dialogue feels human-like rather than robotic.

Part of Speech Tagging

Part of speech (POS) tagging assigns grammatical categories (noun, verb, adjective, etc.) to words in context.

Tag	Meaning	English Examples
ADJ	adjective	new, good, high, special, big, local
ADP	adposition	on, of, at, with, by, into, under
ADV	adverb	really, already, still, early, now
CONJ	conjunction	and, or, but, if, while, although
DET	determiner	the, a, some, most, every, no, which
NOUN	noun	year, home, costs, time, Africa
NUM	numeral	twenty-four, fourth, 1991, 14:24
PRT	particle	at, on, out, up, over, with
PRON	pronoun	he, their, her, its, my, us
VERB	verb	is, say, told, given, playing, would
.	punctuation	., ; !
X	other	ersatz, dunno, gr8, univeristy

Syntactic and Dependency Parsing

Parsing identifies structure in sentences:

- Constituency parsing: breaks sentences into nested phrases.
- Dependency parsing: shows relationships between words (e.g., subject, object).

These structures allow machines to infer meaning beyond individual words.

Semantic Analysis

Semantic analysis involves:

- Word sense disambiguation: determining the correct meaning of a word in context.
- Semantic role labeling: identifying roles such as agent, action, and object. Example: "Chris gave a book to Robin." → Agent: Chris, Action: gave, Recipient: Robin, Object: book.

Co-reference Resolution

Example text: "Christopher Robin is alive and well. He is the same person you read about in Winnie the Pooh. As a boy, Chris lived at Cotchfield Farm. When Chris was three, his father wrote a poem about him. The poem was printed in a magazine. Mr. Robin then wrote a book."

Here, "He", "Chris", and "Mr. Robin" all refer to the same person. Machines must link these mentions correctly to maintain coherence.

Artificial Intelligence: What is AI?

Artificial Intelligence (AI) is the attempt to mimic human intelligence or behavior in machines. AI involves reasoning, learning, perception, and decision-making. The AI revolution is driven by:

- Faster computing power (GPUs, TPUs).
- Large-scale data availability.
- Advances in algorithms (neural networks, transformers).

Deep Learning: No Rules!

Deep learning has shifted away from rule-based systems toward learning patterns directly from data. Instead of handcrafting features, deep neural networks automatically learn representations.

Modern Conversational AI

Conversational AI integrates speech recognition, natural language understanding, dialogue management, and text-to-speech. Examples include Siri, Alexa, and customer service chatbots.

Machine Learning Basics

Machine learning is a field that enables computers to learn from data without being explicitly programmed. Instead of following fixed rules, systems infer patterns.

Applications:

- Spam detection
- Fraud detection
- Predictive text
- Recommendation systems

Types of Learning

- **Supervised Learning:** Training with labeled data. Example: email classification with spam/non-spam labels.
- **Unsupervised Learning:** Discovering hidden patterns in unlabeled data. Example: clustering news articles by topic.
- **Reinforcement Learning:** Learning from feedback and rewards. Example: training an agent to play Go or control a robot.

Machine Learning vs Deep Learning

Traditional ML often depends on human-engineered features, while deep learning automatically learns representations from raw data.

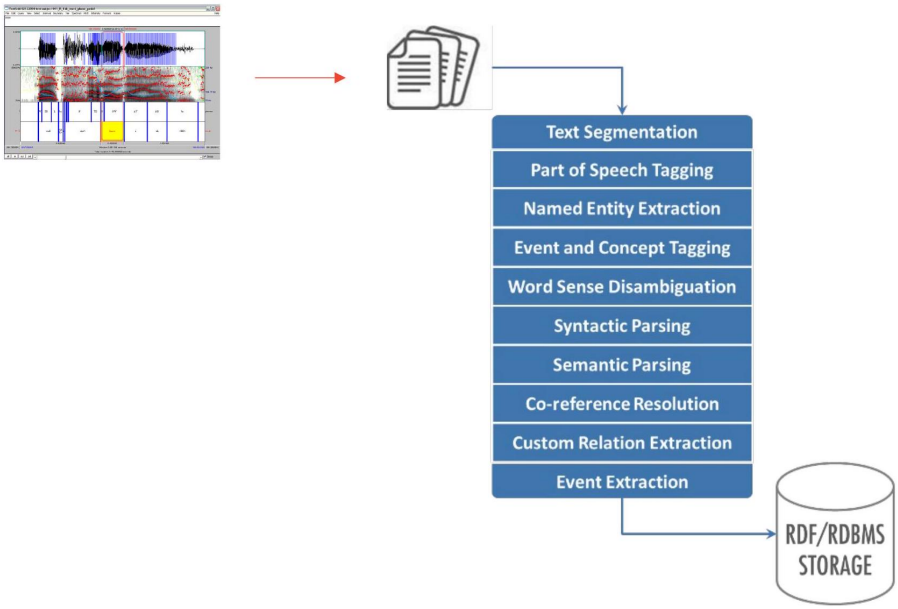
Features in NLP: Example from Named Entity Recognition (NER)

NER systems use many features:

Feature	Used in NER
Current Word	✓
Previous Word	✓
Next Word	✓
Character n-grams	all
Current POS Tag	✓
Surrounding POS Tag Sequence	✓
Current Word Shape	✓
Surrounding Word Shape Sequence	✓
Presence of Word in Left Window	size 4
Presence of Word in Right Window	size 4


CAUTION
WET FLOOR
小心地滑


Translate
English Spanish French Chinese - detected
English Spanish Arabic
Translate
小心地滑
Carefully slide
Wrong?
Xiǎoxīn dì huá



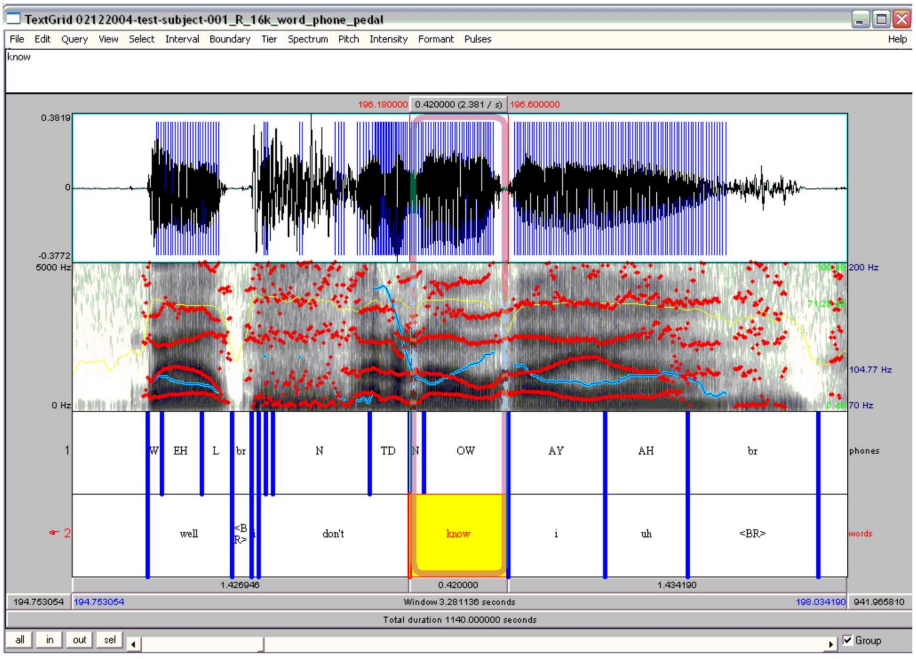
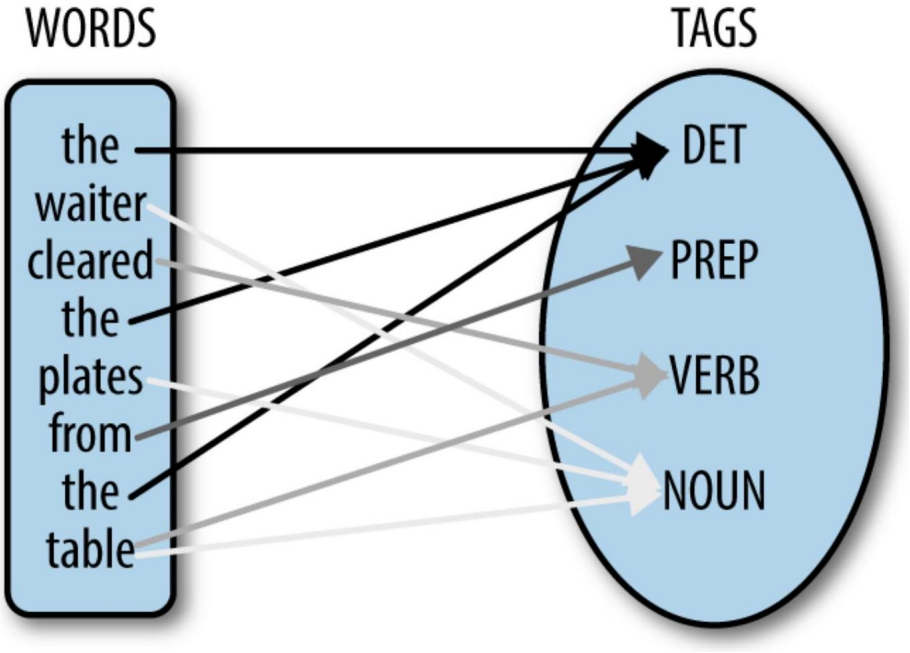
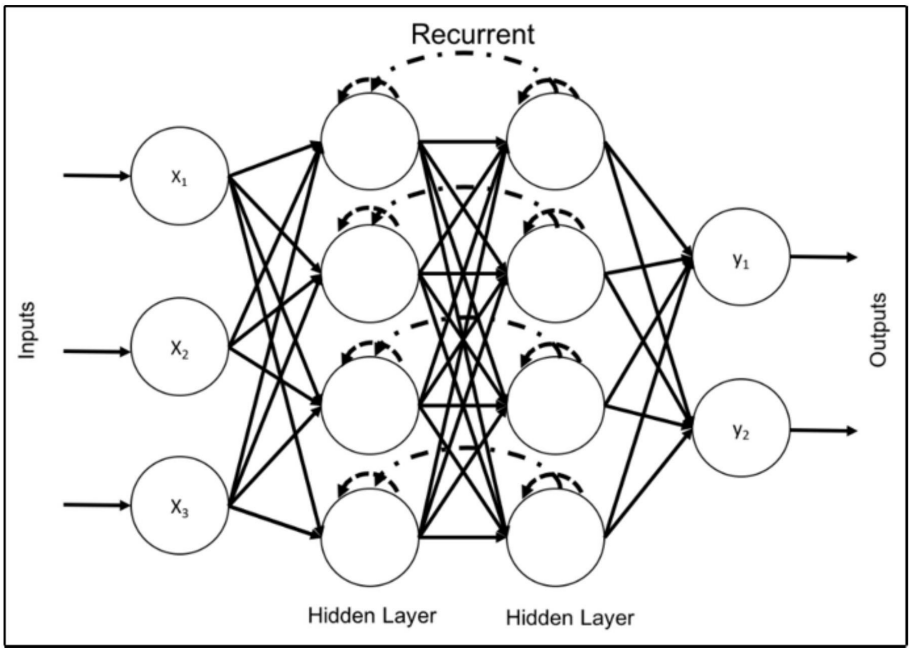
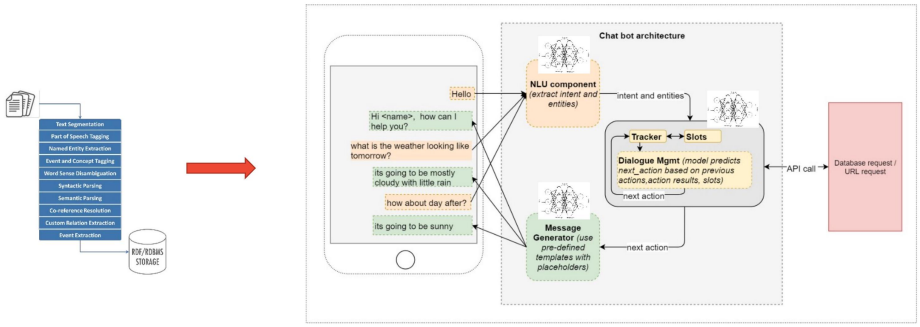
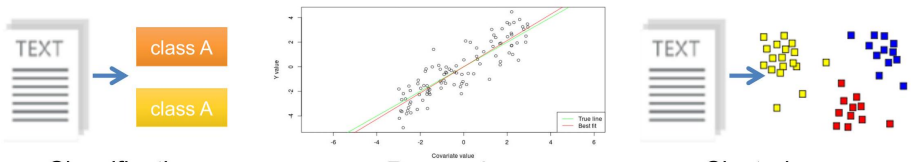


Figure 1: *
Examples from Prof. Julia Hirschberg's slides


WORDS
TAGS
the
waiter
cleared
the
plates
from
the
table
DET
PREP
VERB
NOUN


Recurrent
Inputs
x_1
x_2
x_3
y_1
y_2
Outputs
Hidden Layer
Hidden Layer





Classification Regression Clustering