

UNIT - 2 – NLP & NLU NOTES

NATURAL LANGUAGE PROCESSING (NLP) – OVERVIEW

What is NLP?

NLP (Natural Language Processing) is a field of AI that enables computers to understand, analyze, generate, and respond to human language.

Why NLP is Needed?

- Humans speak in natural language (English, Tamil, Hindi).
- Machines understand only binary or structured data.
- NLP acts as a bridge between humans and machines.

Where NLP is Used?

- Chatbots
- Auto-correct, auto-suggestions
- Google search
- Alexa/Siri voice assistants
- Sentiment detection (positive/negative reviews)
- Spam detection in emails

Real-time Example

You type "best restaurants near me" → Google understands your query using NLP → identifies your intention → shows local restaurants.

2. NLP EXPLAINED (CORE CONCEPTS)

1. Tokenization

Breaking a sentence into words or sentences.

Example:

Input: "AI makes life easier"

Tokens: [AI, makes, life, easier]

2. Stop Words Removal

Removing common words that do not add meaning.

Example:

"the", "an", "is"

Sentence: "This is a laptop"

After removing stopwords → "laptop"

3. Stemming

Reducing words to their base (root) form.

Example:

"playing", "played", "plays" → "play"

4. Lemmatization

Converts words to a meaningful dictionary root form.

Example:

"better" → "good"

"am", "are", "is" → "be"

5. Part of Speech (POS) Tagging

Labeling each word as noun, verb, adjective, etc.

Example:

"Student writes exam"

Student → Noun

writes → Verb

exam → Noun

6. Named Entity Recognition (NER)

Identifying important entities (person, location, company, date).

Example:

"Virat Kohli scored 70 runs in Mumbai."

Entities:

- **Person → Virat Kohli**
 - **Location → Mumbai**
-

7. Chunking

Grouping related words into phrases.

Example:

"the big dog" → noun phrase

8. Text Classification

Assigning category labels to text.

Example:

Email → Spam or Not Spam

Review → Positive or Negative

9. Sentiment Analysis

Finding emotion in text.

Example:

"I love this product" → Positive

"This is worst" → Negative

10. Summarization

Condensing a long text into short meaningful text.

Example:

News → Short summary (used in Inshorts app)

11. Machine Translation

Translating languages.

Example:

English → Tamil (Google Translate)

12. Topic Modeling

Finding hidden topics in large text.

Example:

Customer reviews grouped into:

- Delivery Issue
 - Price
 - Product Quality
-

13. Speech Recognition

Converting voice to text.

Example:

You say “Call Amma” → Phone converts speech to text → performs action.

14. Text Generation

Machine generates meaningful text.

Example:

ChatGPT writing essays or answers.

15. Keyword Extraction

Extracts the main words from text.

Example:

Resume → extracts skills like Python, SQL, Java.

NLU (NATURAL LANGUAGE UNDERSTANDING):

What is NLU?

Natural Language Understanding (NLU)** is a subfield of Natural Language Processing (NLP) that focuses on enabling computers to understand, interpret, and derive meaning from human language. It helps machines comprehend *what users mean* rather than just *what they say*.

❖ Text Input

The user gives an input in natural language — either typed or spoken.

Example: “Book a flight from Chennai to Delhi tomorrow.”

❖ Tokenization

The sentence is broken down into smaller parts — words or tokens.

Result: [Book], [a], [flight], [from], [Chennai], [to], [Delhi], [tomorrow]

❖ Part-of-Speech Tagging (POS)

Each token is labeled with its grammatical role (noun, verb, etc.).

Example: Book (verb), flight (noun), Chennai (proper noun).

❖ Named Entity Recognition (NER)

Important entities (places, dates, names, etc.) are identified.

Example: Chennai → Source City, Delhi → Destination City, Tomorrow → Date

❖ Intent Recognition

The system identifies the user's main goal or action.

Example: Intent = *Flight Booking*

❖ Context Understanding

The system uses previous conversation or user data to understand context.

Example: If the user said earlier “Make it business class,” the system keeps that in memory.

❖ **Semantic Parsing / Meaning Extraction**

Converts natural language into structured data the machine can use.

Example:

{

"intent": "book_flight",

"source": "Chennai",

"destination": "Delhi",

"date": "Tomorrow"

}

❖ **Action or Response Generation**

The system then takes appropriate action — showing flight options, confirming a booking, etc.

NLG(NATURAL LANGUAGE GENERATION):

What is NLG (Natural Language Generation)?

Natural Language Generation (NLG) is a branch of Artificial Intelligence (AI) and Natural Language Processing (NLP) that focuses on producing human-like text or speech from data. It helps computers communicate information in a natural and understandable way, similar to how humans write or speak.

Core Concepts of NLG:

1. Content Determination

- Decides *what information* from the data should be included in the output.
- Example: From weather data, choose “temperature” and “condition” to report.

2. Document Planning

- Organizes selected content into a logical flow or structure.
- Example: “Start with condition, then mention temperature.”

3. Sentence Aggregation

- Combines related information into meaningful sentences.
- Example: “It is sunny” + “The temperature is 30°C” → “It is a sunny day with a temperature of 30°C.”

4. Lexicalization

- Chooses the right words or phrases to express the information.
- Example: Choosing “warm” instead of “hot” for 28°C weather.

5. Referring Expression Generation

- Uses correct references like *he*, *she*, *it*, or *this* to make the text sound natural.
- Example: “The device is charging. It will be ready soon.”

6. Surface Realization

- Applies grammar, punctuation, and sentence structure rules to finalize the text.
- Example: Converts the structured plan into a grammatically correct sentence.

IBM Watson Natural Language Understanding (NLU):

What is Watson NLU?

IBM Watson NLU is a cloud API service that analyzes text to extract metadata such as entities, keywords, sentiment, emotion, categories, relations, concepts, semantic roles, and syntax. It's a managed service — you send text or a URL and get rich structured annotations in return.

Main features (common)

- Entities: people, companies, locations, products (with types and relevance).
- Keywords: important phrases and their relevance and sentiment.
- Sentiment: overall and per-entity sentiment.
- Emotion: scores for joy, sadness, anger, fear, disgust.

- **Categories:** high-level taxonomies (e.g., “technology > mobile phones”).
- **Concepts:** general topics/concepts related to the text.
- **Relations:** extracted relationships between entities.
- **Syntax:** tokens, POS, dependency info (for deeper parsing).
- **Semantic roles:** who did what to whom (subject, action, object).

How Watson NLU works (conceptual)

1. **Input ingestion:** send text or a URL to the NLU endpoint.
2. **Preprocessing:** the service cleans text, removes HTML if URL, normalizes.
3. **Model pipelines:** different models run in parallel — entity recognition, sentiment, categories, etc.
4. **Scoring & enrichment:** each extracted item is scored for relevance, sentiment, confidence.
5. **Return JSON:** you receive a JSON with sections per feature, ready for indexing or visualization.

IBM Watson Natural Language Understanding (NLU)

What is Watson NLU?

IBM Watson NLU is a cloud API service that analyzes text to extract metadata such as entities, keywords, sentiment, emotion, categories, relations, concepts, semantic roles, and syntax. It's a managed service — you send text or a URL and get rich structured annotations in return.

Main features (common)

- **Entities:** people, companies, locations, products (with types and relevance).
- **Keywords:** important phrases and their relevance and sentiment.
- **Sentiment:** overall and per-entity sentiment.
- **Emotion:** scores for joy, sadness, anger, fear, disgust.
- **Categories:** high-level taxonomies (e.g., “technology > mobile phones”).

- **Concepts:** general topics/concepts related to the text.
- **Relations:** extracted relationships between entities.
- **Syntax:** tokens, POS, dependency info (for deeper parsing).
- **Semantic roles:** who did what to whom (subject, action, object).

How Watson NLU works (conceptual)

1. **Input ingestion:** send text or a URL to the NLU endpoint.
2. **Preprocessing:** the service cleans text, removes HTML if URL, normalizes.
3. **Model pipelines:** different models run in parallel — entity recognition, sentiment, categories, etc.
4. **Scoring & enrichment:** each extracted item is scored for relevance, sentiment, confidence.
5. **Return JSON:** you receive a JSON with sections per feature, ready for indexing or visualization.

Applications & Use Cases of NLU :

1) Customer Support Ticket Triage & Prioritization

Description: Automatically classify incoming customer emails/chats into categories, route to the right team, and highlight urgent issues.

How it works: (1) Ingest message; (2) Extract intent, entities, sentiment, urgency; (3) Classify ticket category & priority; (4) Route to queue + add metadata for agent.

Benefits: Faster response times and reduced manual sorting; improves SLA compliance.

Features used: intent classification (custom classifier), entities, sentiment, keywords, categories.

Real-time example: An angry tweet mentioning “refund” + “payment failed” gets auto-priority and sends alert to billing team.

2) Social Media Monitoring & Brand Reputation

Description: Monitor social channels to measure brand sentiment, detect spikes in negative emotion, and find trending topics.

How it works: (1) Stream tweets/posts; (2) Extract entities (brand, product); (3) Score

sentiment & emotions; (4) Aggregate and visualize trends.

Benefits: Early detection of PR crises and insight into customer perception trends.

NLU features: entities, sentiment, emotion, keywords, concepts.

Real-time example: Sudden increase in “battery overheating” mentions with negative sentiment triggers product team investigation.

3) Product & Market Analytics (Feature Extraction)

Description: Extract frequent product features and opinions from reviews to guide product improvements.

How it works: (1) Ingest reviews; (2) Extract keywords and entity-based aspects; (3) Sentiment per aspect; (4) Rank features by negative/positive ratios.

Benefits: Data-driven product roadmap decisions; focus on features causing most complaints.

NLU features: keywords, entity sentiment, relations, concepts.

Real-time example: Many reviews mention “camera focus” with negative sentiment → prioritize camera fix.

4) Compliance & Contract Analysis (Legal / Finance)

Description: Automatically find clauses, obligations, and risky terms in contracts or communications.

How it works: (1) Parse contract text; (2) Detect entities (party names, dates, amounts); (3) Extract relationships and clause categories; (4) Flag non-standard terms.

Benefits: Faster contract review, reduce legal risk and manual effort.

NLU features: entities, relations, categories, concepts, semantic roles.

Real-time example: Flagging a contract clause that grants an unusual indemnity to the counterparty.

5) Conversational Agents & Chatbots (Enhancing NLU)

Description: Improve chatbot understanding by extracting entities, intent, and context from user utterances.

How it works: (1) User message → (2) Intent detection + entity extraction + sentiment → (3) Select dialog path or escalate.

Benefits: More accurate responses; better handover decisions to human agents.

NLU features: entities, intent (custom models), sentiment, keywords.

Real-time example: Chatbot detects user intent “change flight date” and extracts flight number and new date to start the workflow.

