Module-IV
Semantic & Discourse level Analysis

by:
Dr. Soumya Priyadarsini Panda
Sr. Assistant Professor
Dept. of CSE, SIT, Bhubaneswar

Semantic Analysis

Semantic Analysis

- Semantic analysis involves mapping of natural language utterances to some representation of meaning.
- Semantic analysis creates a representation of the meaning of a sentence,
  - where the basic building blocks of the system are: Entities, Concepts, Relations, Predicates
- Many language processing tasks requires non linguistic knowledge of the words besides, morphological and syntactic knowledge.
- A semantic representation bridges the gap between the linguist and common sense knowledge.
- The language used to describe the semantics representations is called a **meaning representation language**

Meaning Representation

- A meaning representation language bridges the gap between linguistic and common sense knowledge.

Example:

A cup on the table

On(cup, table)

Air India serves Hyderabad

Serves(Air India, Hyderabad)

Cont…

- Requires semantic knowledge
  - Word meanings
  - How word meanings combined in sentences to form sentence meanings

- Examples of meaning representation languages:
  - **First order predicate logic (FOPL)**
  - Semantic Nets
  - Conceptual dependency

Characteristics of a Meaning Representation Language

- A meaning representation language must be
  - **Verifiable**
  - **Unambiguous**
  - **Supports canonical forms**
  - **Support inference and variables**
  - **Expressiveness**

Cont…

Verifiability:

- Determines the truth of the representation in the knowledge base

- Example:
  - Does Kingfisher serve Hyderabad?
  - Serves(Kingfisher, Hyderabad)

Unambiguous:

- Support representations that have only one possible interpretation

Cont…

Support for canonical form:

Examples:

Does Air India **offers** a flight to Hyderabad

Does Air India **have** a flight to Hyderabad

Does Air India **serves** Hyderabad

Meaning representation :

Serves(Air India, Hyderabad)

Serves: {offers, have, serves}

Cont…

Support inference and variables:

Example:

Serves(Air India, Hyderabad}

R(x, y)

Expressiveness:

Should be able to address different domains

Example:

bats in the field

Approaches to Meaning Representation

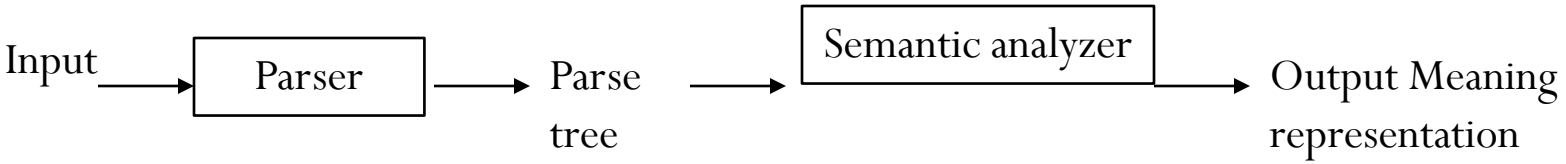
1.Syntax-driven semantic analysis

2. Semantic grammar

1. Syntax-driven Semantic Analysis

- Meaning of sentence are obtained from meanings of parts
  - E.g. groupings and relations from syntax

- It uses syntactic constituents of a sentence to build its meaning representation



Example:

President nominates speaker

N V N

[_NP___] [_____VP_____]

S -> NP VP

Cont…

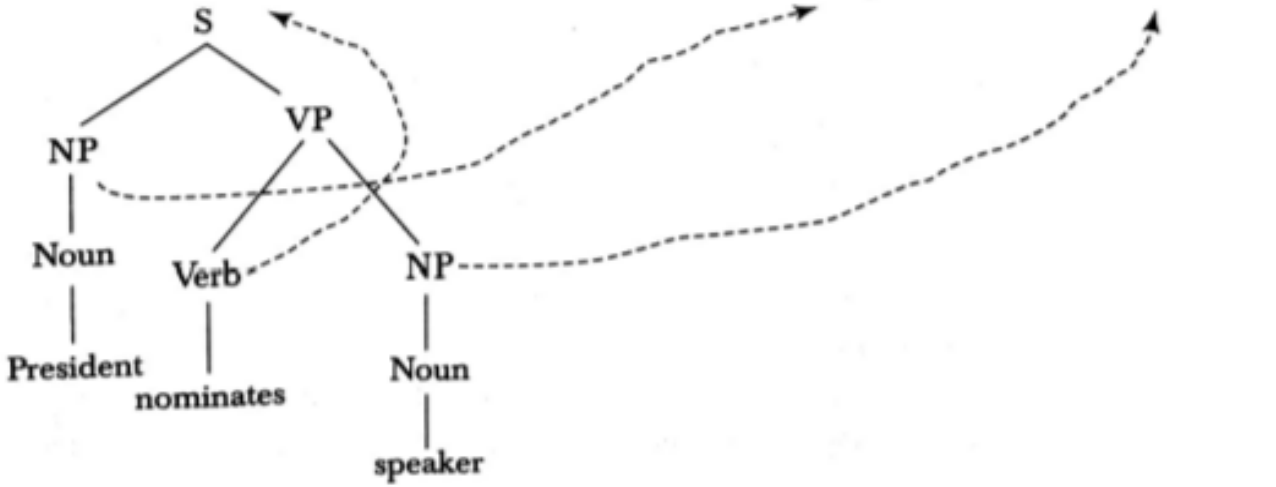
$\exists e, x, y$ is_a $(e,$ nomination$) \wedge$ nominator $(e, x) \wedge$ nominee (e, y)

verb $\rightarrow$ nominates $\{\exists e, x, y$ is_a (nomination$) \wedge$ nominator (e, x)
$\wedge$ nominee $(e, y)\}$

VP $\rightarrow$ verb NP $\{$verb$_{\cdot sem}$ (NP$_{\cdot sem}$)$\}$

President nominates speaker.

$\exists$e is_a (e, nomination) $\wedge$ nominator (e, President) $\wedge$ nominee (e, speaker)



2. Semantic Grammar

- Syntax-driven semantics analysis relies on traditional grammar rules which are not meant for semantic processing.

- Semantic grammar brings semantic capabilities into the grammar
  - A grammar developed with the intension of handling semantics

- Rules and constituents corresponds directly to entities and activities in the domain

Example:

I want to go from Delhi to Chennai on 24th April

- A rule can be created of the form:

  InfoRequest: User wants to go <**to city**> <**from city**> <**TimeExpr**>

Lexical Semantics

- It involves studying the meaning of individual words
- It includes words, sub-words, affixes (sub-units), compound words and phrases.
- Lexical semantics plays an important role in semantic analysis, allowing machines to understand relationships between lexical items (words, phrasal verbs, etc.)
- The relationships are useful in organizing words in lexical databases.
- **WordNet** is a widely known lexical database

Types of Word Relationships

- **Hypernym:**
  - A hypernym is a word with a more general sense.

    Example:
    - ‘Automobile’ is a hypernym for ‘car’, ‘truck’

- **Hyponym:**
  - A hyponym is a word with a more specific meaning.

    Example:
    - ‘car’ is hyponym of ‘automobile’

- **Polysemy:**
  - Word with different but related meanings

    Example: ‘chair’

    ‘chair person’ or ‘furniture’

Cont…

- **Homonyms**:
  - Two words that are sound the same and are spelled alike but have a different meaning.

    Example: bank (river bank), bank (Financial institution)

- **Synonyms:**
  - Words that have the same sense or nearly the same meaning as another

    Examples: happy, ecstatic, overjoyed

- **Antonyms:**
  - Words that have opposite meanings

    Example: happy, sad

WordNet

- **WordNe**t is a large lexical database for English language that lists all senses of a word.

Example:

"read"

Noun:

1. read (something that is read) "The article was a very good read"

Verb:

1. read(interpret something that is written or printed)

   "read the newspaper"

2. Read (scan) "this dictionary can be read by the computer"
3. Read (interpret the significance of; human behavior)

   "She read the sky and predicted rain";

   "I cant read his strange behavior"

….

11.

Internal Structure of words

Thematic role:

- Thematic role is the semantic relationship between predicate (a verb) and an argument (noun phrase) of a sentence

Example:

Nomination -> nominator

Singing -> singer

Cont…

Selectional Restrictions:

- Selectional restrictions are semantic constraints that are placed on arguments for a given word sense.

Example:

eat requires some eating object

A table eats grass X

$\exists$ e , x, y eating(e), $\wedge$

agent(e, x) $\wedge$

theme(e, y)

Ambiguity

- Ambiguity refers to a situation where an expression (words or phrases or sentence) can have more than one interpretation.

- Natural languages are highly ambiguous and resolving the varied interpretations in a reasonable time by computerized models is a difficult task.

- Ambiguity can occur at four different levels:
  - Lexical
  - Syntactic
  - Semantic
  - Pragmatic

Lexical Ambiguity

- Lexical ambiguity is the ambiguity of a single word.
- A word may be ambiguous with respect to its internal structure or its syntactic class.

Examples:

book (Noun, Verb)

plant, bat

Syntactic Ambiguity

- Syntactic ambiguity is the ambiguity in grammatical structure leading to different interpretations.

Examples:

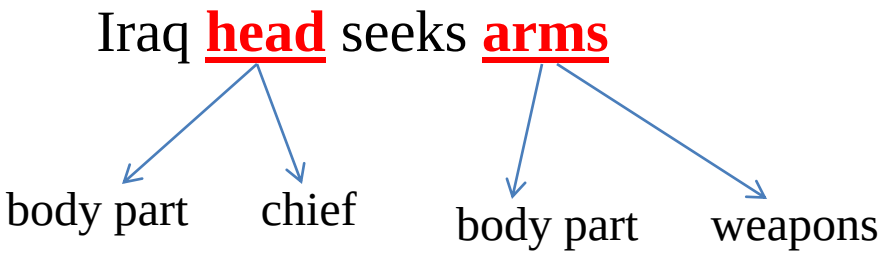
"He saw the man with a telescope"

"Stolen painting found by tree"

Semantic Ambiguity

- Semantic ambiguity occurs when meaning of the words can be misinterpreted.
- i.e. the meaning of the words in a phrase can be combined in different ways, leading to different interpretations.
- Different interpretations of the same sentence.

Example:



Pragmatic Ambiguity

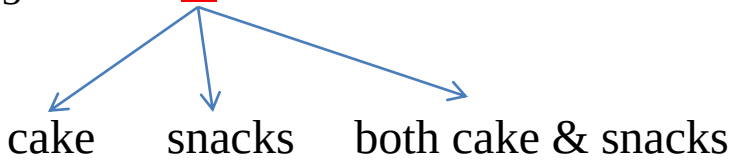
- Pragmatic Ambiguity refers to a situation where the context of a phrase gives it multiple interpretations.

Example:

S_1: Cake is on the table.

S_2: I have prepared some snacks

S_3: Give **it** to the kids



Cont…

S_1: Cake is on the table.

S_2: I have prepared some snacks

S_3: Give **it** to the kids

S_4: Kids enjoyed cake and snacks

=> 'it' refers to "both cake and snacks"

Word Sense Disambiguation (WSD)

- WSD is the process of identifying which sense of a word is used in a sentence when the word has multiple meaning.

- Given a word and its possible senses as defined by a dictionary, the objective is to classify an occurrence of the word in context into one or more of its sense classes.

Example:

- The word ***"orange"*** can refer to a color, a fruit, or even a city in Florida

- WSD has its applications in Machine Translation systems, question answering systems, text to speech systems, etc.

Approaches to WSD

- **Selectional Restriction-based WSD**
- **Context-based WSD**

Selectional Restriction-based WSD

- Selectional restrictions or preferences can be used in parsing to eliminate incorrect meaning representations

Example:

employ

The institute will employ new employees. ('to hire')

The committee employed her proposal. ('to accept')

- Human/organization -> human(to hire)
- Human/organization -> idea(to accept)

Context-based WSD

- Uses the context of ambiguous word

- Two types:
  - **Knowledge-based**
  - **Corpus-based**

Knowledge-based WSD

- It utilizes information from an explicit lexicon or knowledge base to disambiguate a word.

- The lexicon or knowledge base used may contain machine readable dictionary or hand coded rules to disambiguate.

Corpus-based WSD

- A corpus-based approach extracts information on word sense from a large sense tagged corpus.

- The information used to annotate an ambiguous word may be:
  - Distributional information (frequency distribution of a sense)
  - Context
  - Additional knowledge sources

- Two approaches to corpus-base WSD
  1. **Supervised WSD**
  2. **Unsupervised WSD**

Cont…

Supervised WSD:

- Rely on a sense-tagged training corpus for disambiguation.
- Classification algorithms like KNN may be applied for disambiguation.
- With a good quality tagged data set, the overall accuracy in WSD can be very high.

Unsupervised WSD:

- Uses raw or unlabelled text corpora for training and annotated data for evaluation.
- A completely unsupervised algorithm can not perform sense tagging. It can only perform sense discrimination.
- The unsupervised WSD algorithm creates a number of clusters based on the context of an ambiguous word and discriminate between them without labeling them.

Knowledge Sources in WSD

- Context of a word
- Frequency of a sense
- Selectional preferences
- Domain

Issues in Processing Indian Languages

- Indic scripts have a nonlinear structure
- Have a free word order
  - i.e. words can be moved freely within a sentence without changing the meaning of the sentence
- Have rich set of morphological variants
- Extensive and productive use of complex predicates
- Ambiguity

Discourse Analysis

Discourse Analysis

- Attempts to interpret structure and meaning of larger units
  - at paragraph and document level, in terms of words, phrases, clusters and sentences.
- Requires Discourse Knowledge
  - Knowledge of how the meaning of a sentence is determined by processing sentences
- Concerned with linguistic phenomena and extra linguistic phenomena
- Discourse level analysis uses: the situational context, cultural, social, interpersonal and linguistic knowledge

Cohesion

- Cohesion bounds text together.
- It's the textual phenomenon that involves linking elements

Example:

- "Yesterday, my friend invited me to my friend's house. When I reached, my friend was preparing coffee. My friend's father was cleaning dishes. My friend's mother was busy writing a book"

- This type of over specification is avoided through the use of "her"

- "Yesterday, my friend invited me to her house. When I reached, my friend was preparing coffee. Her father was cleaning dishes. Her mother was busy writing a book"
- "Her" is cohesive with "my friend"

Reference

- Link a referring expression to another referring expression in the surrounding expression.

Example:

- Suha bought a printer. It cost her Rs.20000.
  - ‘Her’ refers to a person named ‘Suha’
  - ‘It’ refers to an entity named ‘printer’

Ellipsis

- It’s a form of grammatical cohesion.
- It refers to the phenomenon when a part of a sentence is omitted or left un pronounced.
- The reader uses the surrounding text to recover the omitted text

Example:

- I know that lady. Do you?
- Here ‘Do You’ is an ellipsis stands for ‘Do you know that lady’

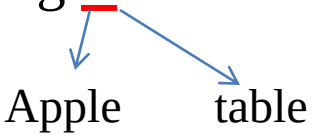
Reference Resolution

- It is the process of resolving references to earlier or later items in the discourse.

- Reference resolution techniques uses various constraints and preferences to identify the preferred reference.

Example:

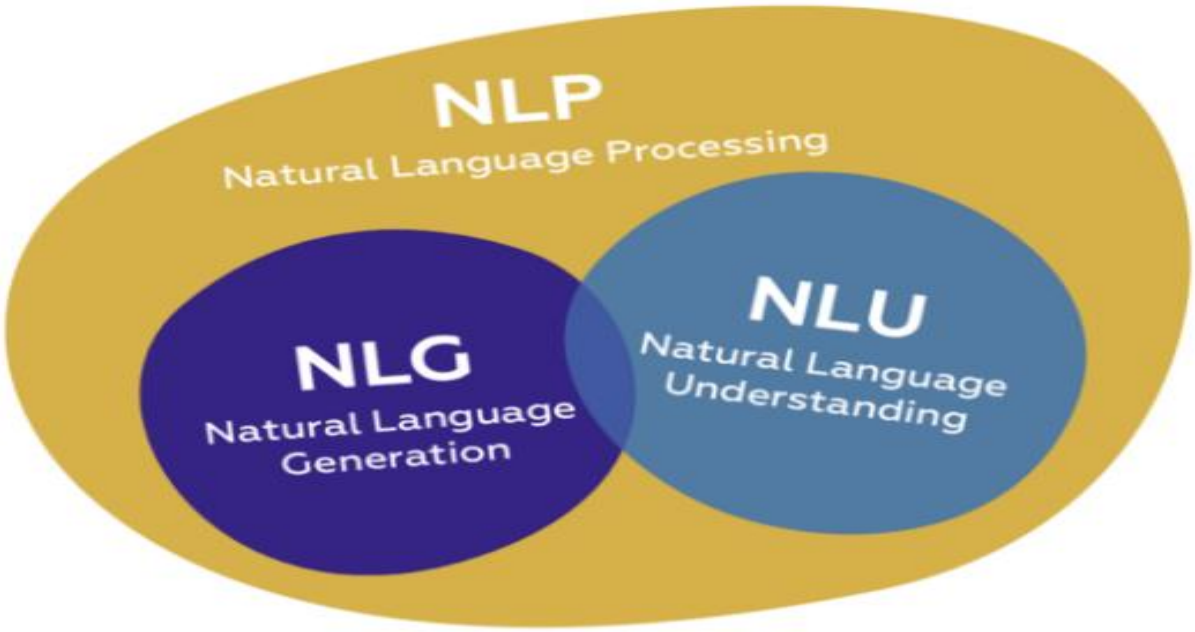
Apple on a table. She is eating **it**



- Using selectional restrictions ‘eat’ refers to ‘apple’ not the ‘table’

Natural language Generation (NLG)

The two Components of NLP



- NLU is focused on deriving analytic insights from textual data,
  - NLG is used to synthesize textual content by combining analytic output with contextualized narratives

Natural language Generation (NLG)

- NLG deals with automatic generation of natural language sentence using non linguistic inputs.
- The technology can actually tell a story– exactly like that of a human analyst – by writing the sentences and paragraphs for you.
- NLG is one of the fastest growing technologies being adopted in the industry.
- There many use-cases for NLG, but it is seen to be most effective in automating the time-intensive data analysis and reporting activities.

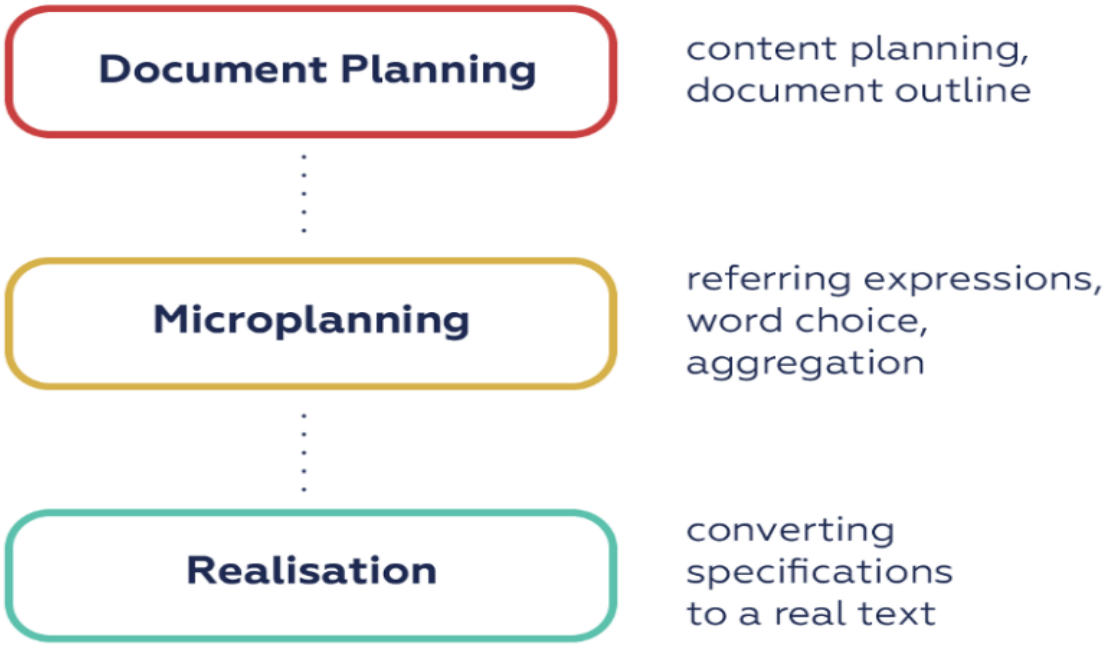
Usable Application of NLG

- Natural language databases: Airline schedule database, Accounting database or spreadsheets, expert system knowledge bases, etc.
- Summary generation
- Language Translator
- Social Media Monitoring
- Chat bots
- Survey Analysis
- Media content creation

NLG Cont…

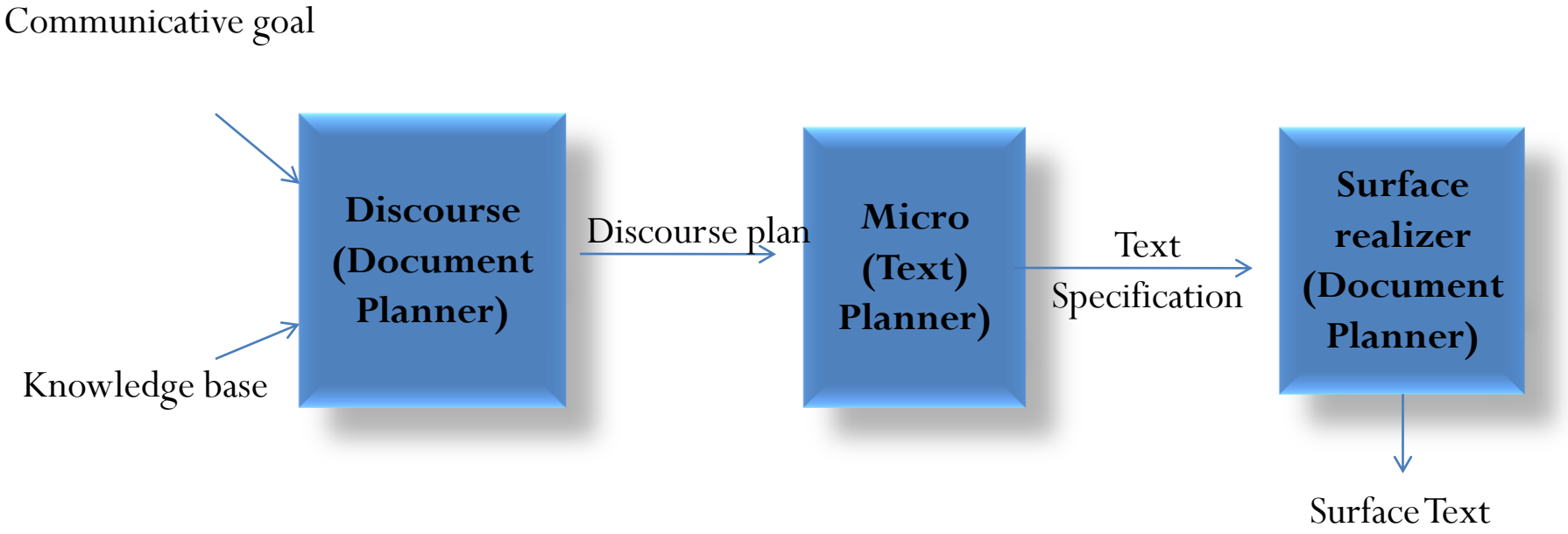
- The problems with most NLG platforms is that they **hard code intelligence into a template**.
- This makes for systems that are brittle and hard to change and are not able to accept new data without new coding.
- To generate text, the NLG model need to identify:
  - What to write
  - How to write it

Cont…



- There are two major approaches to language generation: using templates and dynamic creation of documents

Phases of NLG



Cont…

- **Discourse Planner:**
  - Decides the ordering and structure of the text that is to be generated
  - Output is represented in the form of a tree

- **Text Planner:**
  - Decides words and phrases to express the concepts and structure present in a discourse plan
  - Involves sentence aggregation, lexicalization and referencing expression generation

- **Surface Realization:**
  - Takes sentence plan produced by the text planner and generate individual sentences

Challenges

- Data availability and quality
- Originality and writing quality
- Bias