Why Sentence Structure Matters (CS224N –Lecture 4)

1. Why Sentence Structure Matters

- **Human Communication**
    - Language is received as a **linear stream**:
        - Written → sequence of words.
        - Spoken → continuous sound stream (no clear boundaries between words).
    - Despite this linear form, meaning requires identifying **relationships**:
        - Which words modify which (dependencies).
        - Grouping words into larger units (constituents).
- **Interpretation**
    - Human listeners resolve these structures automatically.
    - Models must also understand sentence structure to interpret meaning correctly.

---

2. Ambiguity in Natural Language

- Human language is **globally ambiguous**, unlike programming languages.
- Programming languages: ambiguities resolved by strict rules (e.g., `else` pairs with nearest `if`).
- Natural language: multiple interpretations may persist; humans resolve them using **context** and **world knowledge**.

---

3. Types of Ambiguities

a. Prepositional Phrase (PP) Attachment Ambiguity

- Example: *“Scientists count whales from space”*
    - Two readings:
        - Scientists in space count whales.
        - Scientists count whales that are from space.
- Ambiguity arises because **PPs can attach to different parts** of the sentence.
- **Ambiguity growth**
    - With multiple PPs, attachment possibilities multiply.
    - The number of interpretations grows according to **Catalan numbers**:
        - 4 PPs → 13 possible readings.
        - 5 PPs → 27 possible readings.
    - Ambiguity increases **exponentially**.

---

b. Coordination Scope Ambiguity

- Example: *“Shuttle veteran and longtime NASA executive Fred Gregory appointed to board”*
    - Reading 1: One person (Fred Gregory) has both roles.
    - Reading 2: Two people – one Shuttle veteran + Fred Gregory (NASA executive).
- Ambiguity in whether the **conjunction joins descriptors** or introduces **separate entities**.

---

c. Modifier Ambiguity

- **Adjectival/Adverbial Ambiguity**: Multiple possible nouns/verbs could be modified.
- **VP Attachment Ambiguity**: Infinitival or participial phrases may attach at different levels.

- Example: *“Mutilated body washes up on Rio beach to be used for Olympics beach volleyball”*
    - Could mean:
        1. The **beach** is to be used for volleyball.
        2. The **body** is to be used for volleyball (nonsensical but grammatically valid).

---

d. Coordination without Conjunctions

- Sometimes ambiguity arises with **juxtaposition + commas** instead of explicit conjunctions.
- Example: *“Doctor—No heart, cognitive issues.”*
    - Could mean:
        1. No heart **and** cognitive issues.
        2. No heart **or** cognitive issues.

---

e. Lexical & Structural Ambiguities

- Example: *“Students get first hand job experience.”*
    - Could mean:
        1. Hands-on job experience.
        2. (Ambiguous humorous interpretation).

---

4. Cross-Linguistic Perspective

- Ambiguities differ across languages:
    - Example: Chinese PP placement often avoids English-like PP ambiguity.
    - But every language has **its own ambiguity challenges**.

---

5. Why Ambiguity Matters for NLP

- Ambiguities affect **parsing** → essential for accurate understanding.
- Dependency structures **help extract meaning**:
    - Example: **Protein–Protein Interactions** (biomedical NLP) (slide page-18):
        - "KaiC interacts with SasA, KaiA, and KaiB."
        - Dependency parse reveals interaction applies to **all three** proteins.

---

6. Key Takeaways

- Sentence structure is essential to move from **linear input** → **meaningful interpretation**.
- Natural language is **highly ambiguous**:
    - PPs, coordinations, modifiers, VPs, appositions all contribute.
- **Humans resolve ambiguity** using context, knowledge, and reasoning.
- **NLP systems** must incorporate structural understanding (parsing, dependency trees, or implicit in neural models like Transformers).
- Parsing enables **practical tasks** (fact extraction, event detection, information retrieval).

✅ **Summary:**
Sentence structure is crucial because meaning depends on identifying **who modifies what**. Human languages allow multiple valid interpretations, which grow exponentially with complexity (Catalan growth). Humans resolve them effortlessly, but for NLP systems, handling ambiguity is central to tasks like parsing, translation, and information extraction.