Modeling Indirectness and Politeness in Natural Language Processing Using Centering Theory and Grice's Maxims

Your Name

August 2025

Abstract

This research outline explores the modeling of indirectness and politeness in natural language processing (NLP) by integrating Centering Theory, which tracks entity focus for discourse coherence, with Grice's Maxims, which govern conversational implicature. Aimed at aspiring scientists, this guide provides a theoretical framework, practical examples, visualizations, mathematical formulations, real-world applications, and actionable research steps. The approach ensures coherence in polite dialogues while capturing indirectness through intentional maxim violations, offering novel insights for NLP systems like chatbots and cross-cultural communication tools.

1 Introduction

As an aspiring scientist in linguistics and cognitive science, modeling **indirectness** (e.g., hinting rather than stating directly) and **politeness** (e.g., softening requests) is a promising research area for advancing natural language processing (NLP) and human-computer interaction. This paper combines **Centering Theory** (1), which ensures discourse coherence by tracking entity focus, with **Grice's Maxims** (2), which guide conversational implicature and politeness strategies. This beginner-friendly guide provides a comprehensive framework, examples, applications, and research steps to develop models that generate or interpret polite, indirect language, aligning with your goal of building a scientific career.

2 Theoretical Framework

2.1 Centering Theory

Centering Theory maintains discourse coherence by tracking entity focus across sentences. Key concepts include:

- **Forward-Looking Centers (Cf)**: Entities mentioned in a sentence, ranked by grammatical role (e.g., subject > object).

- **Backward-Looking Center (Cb)**: The entity linking to the previous sentence, ensuring focus continuity.
- **Transitions**: Continue (same Cb, high coherence), Retain (same Cb with new entities introduced), or Shift (new Cb, less coherent).

Role in Politeness: By maintaining focus on the listener (e.g., "you" as Cb), Centering enhances politeness, making requests feel less imposing. **Analogy**: A spotlight follows the listener, keeping them central to the discourse, as in polite requests like "Could you possibly help?"

2.2 Grice's Maxims

Grice's Maxims define principles for effective communication:

- **Quantity**: Provide just enough information (e.g., avoid bluntness in politeness).
- **Quality**: Be truthful (e.g., avoid misleading hints).
- **Relevance**: Stay on topic (e.g., indirect requests must fit the context).
- **Manner**: Be clear, avoid ambiguity (e.g., polite phrasing should be interpretable).

Role in Indirectness: Indirectness often violates Manner (e.g., obscurity in "It's cold in here" to imply "Close the window") to achieve politeness, relying on context for interpretation. **Analogy**: A diplomatic dance where hints convey meaning without direct confrontation.

2.3 Combining Centering and Grice

Combining these frameworks enables modeling of indirectness and politeness:

- **Indirectness**: Use Centering to maintain listener focus (e.g., "you" as Cb) while violating Manner for subtle hints (e.g., "It's chilly, isn't it?").
- **Politeness**: Prioritize listener-related entities (Centering) and adhere to Quantity/Relevance (Grice) to soften requests (e.g., "Could you possibly help?").

Example:

- Direct: "Close the window."
- Indirect/Polite: "It's a bit chilly, isn't it? Could you maybe close the window?"
- Centering: Cb = "you" across sentences for politeness.
- Grice: Violates Manner (obscure hint) but adheres to Relevance (cold room context).

3 Example Analysis

Consider the dialogue:

1. S1: “It’s quite cold in here, don’t you think?”
2. S2: “You might find the window open.”
3. S3: “Could you possibly close it?”

3.1 Centering Analysis

- **S1**: Cf = {it, you}, Cb = None (first sentence).
- **S2**: Cf = {you, window}, Cb = you (Continue transition).
- **S3**: Cf = {you, it}, Cb = you (Continue transition).

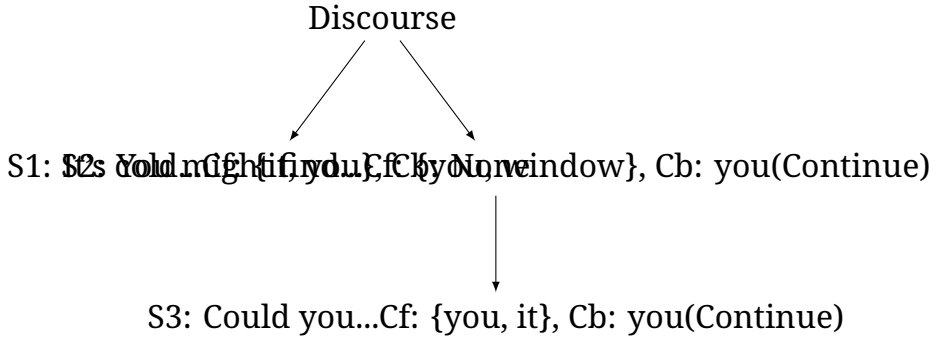
The consistent focus on “you” enhances politeness by centering the listener.

3.2 Grice’s Maxims Analysis

- **Quantity**: S1 provides minimal information (hinting coldness).
- **Quality**: Truthful (room is cold).
- **Relevance**: Relevant to the open window context.
- **Manner**: Obscure in S1 (indirect hint) but clear by S3 (explicit request).

4 Visualization

The discourse structure can be visualized as a tree:



Grice Analysis: S1 violates Manner (obscure), S2/S3 adhere to Relevance.

5 Mathematical Formulations

5.1 Centering Coherence Score

The coherence of a dialogue is quantified as the product of transition probabilities:

- Continue: $P = 0.9$
- Retain: $P = 0.7$
- Shift: $P = 0.5$

For the example (two Continue transitions):

$$\text{Coherence Score} = P(\text{Continue}) \times P(\text{Continue}) = 0.9 \times 0.9 = 0.81$$

5.2 Politeness Score

A politeness score is a weighted sum of features:

- Modals (e.g., "could"): 0.4
- Hedges (e.g., "possibly"): 0.3
- Indirectness (e.g., hints): 0.3

For S3 ("Could you possibly close it?"):

$$\text{Politeness Score} = 0.4(\text{modal}) + 0.3(\text{hedge}) + 0.3(\text{indirect}) = 1.0$$

6 Real-World Applications

- **Chatbots**: Polite, indirect responses improve user experience (e.g., "It looks like you're busy. Would you like a reminder?").
- **Customer Service**: Indirect suggestions (e.g., "You might prefer our premium plan") enhance satisfaction.
- **Cross-Cultural NLP**: Model politeness variations (e.g., Japanese honorifics vs. English indirectness).

7 Research Steps

1. **Literature Review**: Study **(author?)** (1) for Centering, **(author?)** (2) for Maxims, and **(author?)** (3) for politeness.
2. **Data Collection**: Gather corpora like the Stanford Politeness Corpus or Switchboard Corpus for dialogues.

3. **Model Development**:
    - *Centering*: Implement entity tracking (e.g., using spaCy for entity recognition).
    - *Grice*: Develop rules to detect maxim violations (e.g., obscurity for Manner) and politeness markers (e.g., modals).
    - Example: Train a transformer (e.g., BERT) to classify utterances as direct/indirect.
4. **Evaluation**:
    - Metrics: Coherence score (Centering transitions), politeness score (modal/hedge presence).
    - Example: Coherence = 0.81, Politeness = 1.0 (from above).
5. **Experimentation**: Test on dialogues to generate polite responses (e.g., input: "Close the window," output: "It's chilly, could you close it?").
6. **Publication**: Submit to NLP conferences (e.g., ACL, EMNLP).

8 Rare Insights

- **Centering for Politeness**: Maintaining listener focus (e.g., "you" as Cb) reduces perceived imposition, enhancing politeness across cultures.
- **Grice in NLP**: Modeling intentional Manner violations can improve chatbot empathy, a novel area for AI.
- **Cross-Linguistic Potential**: Indirectness varies (e.g., high in Japanese, low in German); Centering standardizes politeness modeling.

9 Research Notes

- **Challenges**: Detecting intentional maxim violations requires context-aware models.
- **Tools**: Use spaCy for entity tracking, Hugging Face Transformers for intent classification.
- **Future Work**: Extend to multimodal contexts (e.g., tone, gestures) for richer politeness models.

10 Next Steps

- Implement a prototype using the paragraph generator from `paragraph_generator.py`.
- Experiment with datasets like the Stanford Politeness Corpus.

- Develop a transformer-based politeness model for your research portfolio.
- Submit findings to conferences to advance your scientific career.

References

[1] Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2), 203–225.

[2] Grice, H. P. (1975). Logic and conversation. In *Syntax and Semantics*, 3, 41–58.

[3] Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.