1

Unified Research Proposal: Neural Networks and Cross-Linguistic Variability under Information Bottleneck Framework

1.1

Objectives

- Train a neural network (NN) to compress sensory inputs (colors) into discrete linguistic categories using the Information Bottleneck (IB) principle.
- Examine whether learned NN categories align with empirical human language categories.
- Investigate how the IB principle explains variability across languages, considering linguistic, geographic, cultural, and demographic factors.

1.2

Research Questions

- Can neural network models trained under IB constraints develop categories resembling human-defined semantic categories?
- Does the IB principle comprehensively explain both universal patterns and culturally-driven variations in color categorization across languages?

1.3

Data

- **Primary Source**: World Color Survey (WCS) dataset available at WCS](https://www1.icsi.berkeley.edu website.
- **Metadata**: Geographic location, language age, number of speakers, and cultural context of WCS languages.

1.4

Methodology

Neural Network Model Construction

- **Architecture**: Autoencoder with discrete bottleneck layers using Gumbel-Softmax discretization.
- **Objective Function**: Optimize IB objective

  $$\mathcal{L} = I\_q(M; W) - \beta I\_q(W; U) \quad (1)$$

  where M is the input sensory distribution, W is discrete category labels, and U is reconstructed meaning.

Cross-Linguistic Analytical Approach

- Compute IB efficiency scores for languages and correlate with metadata using regression models.
- Conduct comparative case studies (e.g., English vs. Amazonian languages) to elucidate cultural impacts.

1.5

Evaluation

- **Quantitative Analysis**: Measure category similarity using Normalized Information Distance (NID).
- **Qualitative Analysis**: Visualize learned category boundaries using PCA and t-SNE.
- **Statistical Analysis**: Regression analysis to identify factors affecting cross-linguistic variability.

1.6

Deliverables (Course Presentation)

- A trained and documented neural network model.
- Comparative analysis of NN-derived and empirical human categories.
- Comprehensive statistical analysis correlating IB efficiency with linguistic metadata.
- Visualizations clearly illustrating category structures and variability across languages.

1.7

Follow-Up Steps

- Expand analysis using historical data to trace the evolution of semantic categories.
- Generalize research methodology to additional perceptual domains beyond color (e.g., auditory or geometric stimuli).

2

Overall Significance This unified research integrates neural networks and Information Bottleneck theory to provide a comprehensive understanding of semantic categorization, cognitive constraints, and cross-cultural linguistic variations, offering valuable insights into language evolution.