

**Syntax vs NLU Preference Accuracy**  
Diverging comparison across models (T=0.0, Few-Shot)  
■ Syntax Accuracy ■ NLU Preference

