Step1.Transfer-based Pseudo Data Generation LLM prompts Fine-grained konwledge Transferring System prompt Original parallel sentences [1] Sentence Segmentation -Identify the {main sentence}, specifying that {the [Chinese sentence] System: **sole root node**} is in the main sentence {这是一张无效合同。} [Role]An expert in cross-lingual {<main> 这是一张无效{合同<root>}。} [Vietnamese sentence] dependency parsing. {<main> đây là một {hợp đồng<root>} vô hiệu .} {đây là một hợp đồng vô hiệu .} [Skill] Leverage source language [2] **Chunking** - Grouping words into meaningful phrases or units Annotator 1 knowledge to improve target LLM parser {<main> [这 Pronoun] [是 Copula] [一张 Quantifier Phrase] language parsing accuracy. [无效{合同<root>} Noun Phrase, NP][。Punctuation]} Pre-processing {<main> [đây *Pronoun*][là *Copula*][môt *Quantifier*] [{hop dong<root>} vo hieu Noun Phrase, NP][. Punctuation]} [3] Contrastive Alignment-Explore commonalities[The unchanged structure] and User prompt differences[The changed structure] The key five content of CoNLL-U {<main> 这 是 一张 无效 {合同<root>}。} User: Column[0]: word index Aligned: {<main> đây là một vô hiệu <swap> {họp đồng<root>} .} Use {Fine-grained konwledge Column[1]: word form Original: {<main> đây là một {hop đồng<root>} vô hiệu .} Transferring | method to parse Data Column[2]: part-of-speech tag [hợp đồng and vô hiệu change their position] {Original parallel sentences} extraction Column[3]: Headword index Commonalities: Subject + Copula + Predicate and fill {The key five content of Column[4]: Dependency relation CoNLL-U of the target language. **Differences:** Chinese determiners are prepositioned, Vietnamese determiners are [Chinese] [Vietnamese] Annotator 2 postpositioned; Chinese uses quantifier phrases, Vietnamese uses separate quantifiers. 1 đây _ _ _ _ 1 这 Traditional 2 là _ _ _ parsers 3 một _ _ _ Sets of pseudo data Traditional parser 4 hợp đồng Traditional parser-5 vô hiệu MLP based pseudo data LLM-baesd Input **BiLSTM** BiAffines layer Encoder pseudo data Arc score MLP Label score Step3. Synthetic Data Evaluation Step2. Pseudo Data Self-optimizing-Three typical cross-lingual models High-quality synthetic data Three LLMs Evaluation BiAffine BiAffine Task A Task B [Vietnamese] 1 đây PROPN 3 nsubj MLP MLP MLP 2 là VERB 4 cop MLP 3 môt EDT 4 det **Self-optimizing** 4 hợp đồng NOUN 0 root **BiLSTM BiLSTM BiLSTM** 5 vô hiệu ADJ 4 amod 6 . PUNCT 0 punct Manual assessment root Lan-emb Input Input punct **FulSha** LanEmb MulLea Training dataset một hợp đồng vô hiệu This invalid contract UD treebank Our synthetic data