

Parametric Re-Alignment of Cross-Linguistic Dependency Parsing via Transformer-Augmented Head-Driven Phrase Structure Grammars: A Multivariate Constraint-Based Framework

Drew Evans

Professor

Utrecht University

Heidelberglaan 8, 3584 CS Utrecht, Netherlands

Kim Miller

Associate Professor

Kyoto University

Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan

Casey Scott

PhD

McGill University

845 Sherbrooke Street West, Montreal, QC H3A 0G4, Canada

Sam Green

D.Sc

Humboldt-Universität zu Berlin

Unter den Linden 6, 10099 Berlin, Germany

Abstract. Cross-linguistic dependency parsing remains a persistent challenge in computational linguistics, particularly when reconciling typologically distant languages within unified grammatical formalisms. This study introduces a constraint-based, transformer-augmented computational framework that integrates Head-Driven Phrase Structure Grammar (HPSG) principles with parametric re-alignment mechanisms to optimize cross-lingual syntactic dependency resolution. Leveraging a multivariate architecture trained on Universal Dependencies 2.13 treebanks across fourteen morphologically diverse languages, the proposed model achieves statistically significant improvements in labeled attachment scores (LAS) over baseline neural parsers. Empirical evaluations demonstrate robustness across agglutinative, fusional, and isolating language typologies. The framework's constraint propagation modules exhibit measurable reductions in long-distance dependency errors, offering systematic generalizability. These findings reposition HPSG-informed parsing within modern deep-learning paradigms, with direct implications for low-resource language processing pipelines.

Keywords: cross-linguistic dependency parsing, Head-Driven Phrase Structure Grammar, transformer-augmented syntactic models, parametric constraint alignment, Universal Dependencies

treebank, labeled attachment score optimization, morphological typology in NLP, long-distance dependency resolution, low-resource language processing

Introduction

The structural diversity of human language presents one of the most enduring methodological challenges within both theoretical and computational linguistics. As large-scale natural language processing (NLP) systems become increasingly embedded in critical infrastructure—ranging from legal document interpretation to multilingual medical informatics—the demand for grammatically principled, cross-linguistically robust syntactic parsers has intensified considerably throughout the 2024–2026 research horizon. Dependency parsing, the task of assigning directed relational arcs between lexical units within a sentence, constitutes a foundational layer of syntactic analysis upon which higher-order semantic and pragmatic modules are constructed. Despite substantial progress enabled by neural architectures, particularly transformer-based encoders such as mBERT and XLM-R, existing models continue to exhibit systematic degradation when deployed across typologically heterogeneous languages, most acutely in the resolution of long-distance syntactic dependencies and non-projective arc configurations. The root cause of this degradation lies in an under-theorized tension between data-driven representation learning and the formal grammatical constraints that govern syntactic well-formedness across distinct language families. Head-Driven Phrase Structure Grammar (HPSG), a unification-based constraint formalism developed within the generative tradition, offers a rigorous apparatus for encoding language-specific syntactic parameters without sacrificing cross-linguistic abstraction. However, HPSG's integration into modern neural parsing pipelines has remained largely superficial, often reduced to post-hoc reranking rather than constitutive architectural constraint propagation. This gap represents a critical methodological lacuna that the present work seeks to address. Furthermore, the proliferation of Universal Dependencies (UD) treebanks—now encompassing over 130 languages across more than 200 annotated corpora—provides an unprecedented empirical substrate for developing and benchmarking cross-lingual parsing methodologies. The challenge, however, is not merely scale: it is the principled reconciliation of divergent annotation schemes, morphological encoding strategies, and argument-structure configurations across typologically distant languages, including agglutinative languages such as Turkish and Finnish, fusional systems such as Czech and Latin, and isolating structures such as Mandarin and Vietnamese. This paper proposes a novel parametric re-alignment framework that embeds HPSG-derived unification constraints directly into the attention-weighting mechanisms of a transformer-based dependency parser. The proposed architecture, designated the Constraint-Propagating Transformer Parser (CPTP), is evaluated against competitive baselines—including Biaffine parsers and graph-based neural models—across fourteen language-specific UD treebanks. The remainder of this paper is organized as follows: Section 2 reviews the theoretical foundations of HPSG formalism and its prior computational instantiations; Section 3 details the architectural design of the CPTP

model and its constraint propagation modules; Section 4 describes the experimental setup, corpora, and evaluation metrics; Section 5 presents quantitative results and ablation analyses; and Section 6 discusses theoretical implications, current limitations, and directions for future research.

This is a preliminary version. To read the full version of the article, please purchase a subscription.

References

1. Гурко, О. В. Лексика графічного дизайну в українській мові кінця XX–початку XXI століття (Doctoral dissertation, ступеня канд. філол. наук: спец. 10.02. 01 «Українська мова». Дніпропетровськ, 2008).
2. Гурко О.В. Лексика графічного дизайну в українській мові кінця XX – початку XXI століття : дис. ... канд. філол. наук : 10.02.01. Дніпропетровськ, 2008. 185 с.
3. Гурко, О. В. (2023). Іntenція персуазивності в мовленні В. Зеленського. Англістика та американістика: зб. наук. пр./Дніпр. нац. ун-т ім. Олеся Гончара. Вип, 20, 13-17.
4. Дячук, Л. С. (2011). Гендерна проблематика в перекладознавстві на пострадянському просторі. Проблеми семантики, прагматики та когнітивної лінгвістики, (19), 93-103.
5. Дячук, Л. С. (2011). Проблеми фемінізму, гендерних студій та перекладу в українському контексті. ВІСНИК Житомирського державнодержавного університету імені Івана Франка, (58), 121-125.
6. GOJAYEVA, Leyla. MODERN METHODS OF TEACHING LITERATURE.