Typology-Aware Joint Modeling for Multilingual Morphosyntactic Parsing with Abstract Nodes

Kutay Acar
Istanbul Technical University
acarku18@itu.edu.tr

Relevant UniDive working groups: WG3, WG4, WG1

1 Introduction

The UniDive Morphosyntactic Parsing (MSP) shared task (Goldman et al., 2025b) introduces a unified representation that departs from the standard assumption of token-to-node correspondence by incorporating abstract nodes for non-overt syntactic elements and emphasizing content-based dependency structures. This formulation enables the explicit modeling of linguistically central phenomena such as implicit arguments and dropped elements, which are typically absent from conventional dependency parsing frameworks, and reflects ongoing efforts to extend morphosyntactic annotation schemes toward greater typological adequacy.

At the same time, this shift exposes a fundamental limitation of existing multilingual parsers: their reliance on surface-aligned representations. As a result, they struggle to generalize across languages with varying degrees of non-overt structure. This limitation becomes particularly salient in typologically diverse settings, where the frequency and realization of such phenomena differ substantially.

As illustrated in Figure 1, MSP departs from surface-aligned representations by structuring dependency trees over content elements while encoding functional material as features. This representational shift motivates the need for modeling non-overt and layered morphosyntactic structure.

2 Approach

We present a typology-aware, joint modeling framework for multilingual morphosyntactic parsing developed within the context of the UniDive MSP shared task. The approach is designed to directly address the representational and cross-lingual challenges introduced by abstract nodes.

The model jointly captures multiple interdependent aspects of morphosyntax within a unified architecture: dependency structure over content nodes, morphological features, word-type distinctions, and explicit abstract node predictions. The architecture combines a shared multilingual XLM-R encoder with typology-conditioned adapters inspired by UDapter (Üstün et al., 2020, 2022). Separate task heads are trained jointly for word-type prediction, content-only dependency parsing, morphological tagging, and abstract-node modeling. Rather than treating abstract nodes as auxiliary or post-processing artifacts, they are modeled as first-class structural elements anchored to tokens. For each token, the model predicts the presence of an abstract node together with its structural and morphological properties, allowing the system to reconstruct non-overt syntactic structure in a principled way.

To support multilingual generalization, we incorporate typology-conditioned adaptation mechanisms that modulate model behavior across languages. Training is performed in a multilingual setting covering nine typologically diverse languages: Czech, English, Hebrew, Italian, Polish, Portuguese, Serbian, Swedish, and Turkish (Goldman et al., 2025a). This setup enables the model to capture both shared regularities and language-specific patterns, while also providing a concrete setting for examining how multilingual variation and abstract morphosyntactic phenomena are internalized by the model.

3 Results and Insights

Evaluation follows the UniDive MSP shared task protocol, comparing against strong multilingual MSP baselines using MSLAS, LAS, and FEATS metrics. On the MSP test set, the proposed framework improves over the leading shared-task submission by +3.23 MSLAS, +3.35 LAS, and +1.78 FEATS macro F1 across languages. The gains are especially strong in morphologically rich languages with frequent non-overt morphosyntactic phenomena, particularly Turkish, where improvements exceed +10 LAS and MSLAS points.

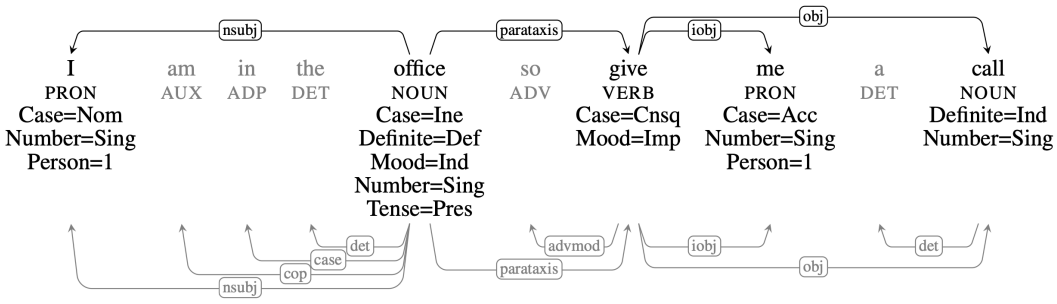

Figure 1: An example comparing a morphosyntactic (MSP) representation (top) with a standard Universal Dependencies (UD) tree (bottom), adapted from Goldman et al. (2025a). In the MSP representation, only content-bearing elements are realized as nodes, while functional elements are not represented as independent nodes but instead encoded as grammatical features. This shift from surface-aligned, word-based structure to content-centered representation underlies the need for modeling non-overt syntactic structure in multilingual settings.

The results show that multilingual parameter sharing alone is insufficient to capture morphosyntactic variation, while typology-conditioned adapters consistently improve performance across typologically diverse languages. Explicit abstract-node modeling contributes most strongly in languages where unrealized arguments are frequent and structurally informative. Analyses of learned language representations further suggest that typological information is reorganized through task-driven multilingual learning rather than simply preserved.

The experiments also highlight several limitations. Abstract-node supervision remains sparse and unevenly distributed across languages, limiting gains in languages with very low frequencies of abstract phenomena. In addition, typological information is currently limited to URIEL-style features and learned language embeddings, leaving more fine-grained morphosyntactic and language-specific variation underexplored. Extending the framework with richer typological representations and more detailed linguistic signals remains an important direction for future work.

4 UniDive Context and Contributions

This work is deeply embedded in UniDive WG3 activities, while also contributing to WG4 through its focus on typological diversity and cross-lingual variation, and to WG1 through its engagement with unified morphosyntactic representations. We contribute to the UniDive 2025 Morphosyntactic Parsing Shared Task as both a data provider for Turkish and a system participant (Acar and Eryiğit, 2025). The proposed model builds directly on the MSP framework developed within UniDive and provides an empirical validation of its computational tractability.

An earlier version of this work was presented as a poster at the UniDive 2nd Training School (Acar, 2026). The current study extends these efforts into a multilingual system with explicit abstract-node modeling and typology-aware adaptation.

More broadly, this work aligns with WG3 objectives by addressing underrepresented morphosyntactic phenomena, leveraging multilingual learning for data-scarce settings, and integrating typological information into model design. It also contributes to the broader UniDive agenda by showing that abstract, non-surface-aligned representations can be operationalized within modern neural architectures.

5 Conclusion

We propose a unified, typology-aware approach to multilingual morphosyntactic parsing that treats abstract nodes as integral components of syntactic structure. By jointly modeling dependency structure, morphology, word type, and abstract-node information within a multilingual architecture, the proposed framework directly addresses the representational challenges introduced by MSP.

The results demonstrate that combining structural abstraction with typological conditioning leads to improved performance and more robust cross-lingual generalization, particularly in morphologically rich languages with frequent unrealized arguments. These findings support the broader UniDive vision that unified linguistic representations can be both theoretically meaningful and practically effective across typologically diverse languages.

References

Kutay Acar. 2026. Typology-aware joint morphosyntactic parsing with functional and abstract node integration in multilingual settings. Accepted project and poster, UniDive 2nd Training School on Linguistic Diversity in NLP.

Kutay Acar and Gülşen Eryiğit. 2025. Typology-aware multilingual morphosyntactic parsing with functional node filtering. In *Proceedings of The UniDive 2025 Shared Task on Multilingual Morpho-Syntactic Parsing*, pages 27–33, Ljubljana, Slovenia. Association for Computational Linguistics.

Omer Goldman, Leonie Weissweiler, Kutay Acar, Diego Alves, Anna Baczkowska, Gülşen Eryiğit, Lenka Krippnerová, Adriana Pagano, Tanja Samardžić, Luigi Talamo, Alina Wróblewska, Daniel Zeman, Joakim Nivre, and Reut Tsarfay. 2025a. Findings of the UniDive 2025 shared task on multilingual morpho-syntactic parsing. In *Proceedings of The UniDive 2025 Shared Task on Multilingual Morpho-Syntactic Parsing*, pages 1–18, Ljubljana, Slovenia. Association for Computational Linguistics.

Omer Goldman, Leonie Weissweiler, and Reut Tsarfaty, editors. 2025b. *Proceedings of The UniDive 2025 Shared Task on Multilingual Morpho-Syntactic Parsing*. Association for Computational Linguistics, Ljubljana, Slovenia.

Ahmet Üstün, Arianna Bisazza, Gosse Bouma, and Gertjan van Noord. 2020. UDapter: Language adaptation for truly Universal Dependency parsing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2302–2315, Online. Association for Computational Linguistics.

Ahmet Üstün, Arianna Bisazza, Gosse Bouma, and Gertjan van Noord. 2022. UDapter: Typology-based language adapters for multilingual dependency parsing and sequence labeling. *Computational Linguistics*, 48(3):555–592.