Green Structured Prediction in NLP

Ana Ezquerro, Diego Roca, Carlos Gómez-Rodríguez, and David Vilares

Universidade da Coruña, CITIC
{ana.ezquerro,diego.roca,carlos.gomez,david.vilares}@udc.es

Abstract

We summarize our efforts in green NLP, showing how representations traditionally modeled with complex transition-, graph-, or seq2seq-based methods—such as trees, dependency graphs, and other semantic structures—can be reformulated as sequence labeling, a simpler and faster paradigm. Using compact encodings and lightweight taggers, we achieve competitive accuracy across diverse formalisms while reducing computational and energy costs. This shows that accurate multilingual NLP is possible with smaller models, lower carbon footprints, and broader accessibility.

1 Introduction

Structured prediction refers to NLP tasks where the goal is to compute rich structural representations—such as trees or graphs—from input sentences. These structures capture syntactic or semantic relations that are essential for both human and machine language analysis, supporting applications that rely on understanding meaning or relations between entities, events, or concepts across domains. However, traditional approaches—transition-based (i.e., shift-reduce), graph-based, or seq2seq—tend to be computationally expensive. We explore an alternative: recasting parsing as a *sequence labeling* problem. By assigning one label per token, decoding becomes linear in time (exactly n tagging actions), architecture-agnostic, and highly parallelizable.

2 From Trees to Graphs: The Evolution of Encodings

Our first contribution (Gómez-Rodríguez et al., 2023) [1] introduced **4- and 7-bit encodings** for dependency trees. Each word receives a compact bit-vector encoding its role as head or dependent, the presence of siblings, and the direction of dependents. These bounded encodings reduced label space and achieved strong accuracy with linear-time decoding.

Building on this, Ezquerro et al. (2024) [2] generalized sequence labeling to dependency graphs. By defining both unbounded (bracketing-based) and bounded ($4k$-bit and $6k$-bit) linearizations, their work enabled the modeling of reentrancies and cycles, which had not been explored before. Results showed that such taggers rivaled more complex decoders while being faster and easier to train.

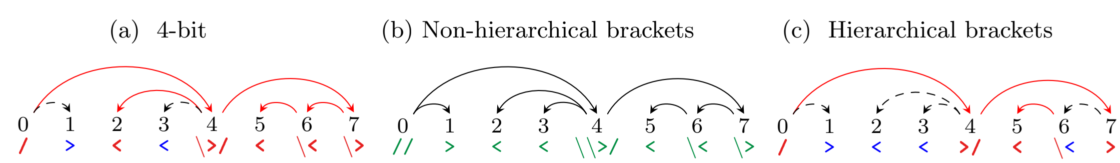

Figure 1: Example from [3] showing the same tree encoded with the standard (non-hierarchical) bracketing encoding (Figure 1b), and hierarchical bracketing encodings (Figures 1a and 1c). In the latter, structural arcs are shown in red and solid, whereas auxiliary arcs are black and dashed. Figure 1c corresponds to the optimal hierarchical bracketing.

Ezquerro et al. (2025a) [3] introduced the notion of **hierarchical bracketing encodings**. This framework demonstrated that previous encodings are suboptimal in symbol usage. The proposed hierarchical variant minimized label cardinality (12 vs. 16 labels for trees), yielding competitive accuracy with fewer tags. Finally, Ezquerro et al. (2025b) [4] extended hierarchical bracketing to dependency graphs. The new encoding retained full graph coverage, reduced label entropy, and improved exact-match accuracy across multilingual benchmarks, all while preserving linear-time decoding. Figure 1 shows a didactic tree linearized with different encodings.

3 Results Overview

Our models achieved accuracy close to state-of-the-art graph-based parsers. The 4- and 7-bit encodings reached about 94% LAS on English UD. Graph encodings obtained roughly 90% LAS on Enhanced UD. Hierarchical bracketing further improved exact match by 1–2 points.

4 Discussion

The main advantage of these contributions lies in their efficiency and simplicity. Sequence labeling parsers operate in linear time, using only a single forward pass and eliminating the need for complex decoding algorithms. This dramatically reduces memory and compute costs, enabling smaller models to achieve high accuracy. The compact encodings also simplify training and inference, making structured prediction faster, more reproducible, and easier to deploy in resource-constrained environments. The proposed encodings demonstrate that structural prediction can be both accurate and sustainable. Each reformulation—from 4-bit trees to hierarchical graphs—replaces complex decoders with efficient taggers. Compact label spaces also reduce entropy and training time, showing that smaller models can yield competitive results when paired with well-designed representations.

5 Conclusion

Linearizing trees and graphs as sequences reveals that linguistic structure can be learned without complex decoders. Compact encodings paired with lightweight taggers achieve accurate and sustainable structured prediction. Future work will explore applying hierarchical bracketing to other structured tasks and integrating it with multilingual small language models.

Acknowledgments. We acknowledge grants GAP (PID2022-139308OA-I00) funded by MICIU/AEI/10.13039/501100011033/ and ERDF, EU; LATCHING (PID2023-147129OB-C21) funded by MICIU/AEI/10.13039/501100011033 and ERDF, EU; and TSI-100925-2023-1 funded by Ministry for Digital Transformation and Civil Service and "NextGenerationEU" PRTR; as well as funding by Xunta de Galicia (ED431C 2024/02). Additionally, it is co-financed by the EU through the FEDER Galicia 2021-27 operational program (Ref. ED431G 2023/01).

References

[1] C. Gómez-Rodríguez, D. Roca, and D. Vilares. "4- and 7-bit Labeling for Projective and Non-Projective Dependency Trees." In *EMNLP*, 2023.

[2] A. Ezquerro, D. Vilares, and C. Gómez-Rodríguez. "Dependency Graph Parsing as Sequence Labeling." In *EMNLP*, 2024.

[3] A. Ezquerro, D. Vilares, A. Yli-Jyrä, and C. Gómez-Rodríguez. "Hierarchical Bracketing Encodings for Dependency Parsing as Tagging." In *ACL*, 2025.

[4] A. Ezquerro, C. Gómez-Rodríguez, and D. Vilares. "Hierarchical Bracketing Encodings Work for Dependency Graphs." In *EMNLP*, 2025.