

HOLINET: Holistic Knowledge Graph for French

Graphe de connaissances holistique pour le français

Jean-Philippe Prost
LPL, CNRS – Aix-Marseille Université
Jean-Philippe.Prost@univ-amu.fr

Credits : detail from "Le Roi, bleu", 2021, by Kristin DeGeorge, photography by Jacqueline Poggi

How to integrate grammar knowledge and lexical semantic knowledge within a homogeneous Knowledge Graph ?

Résumé

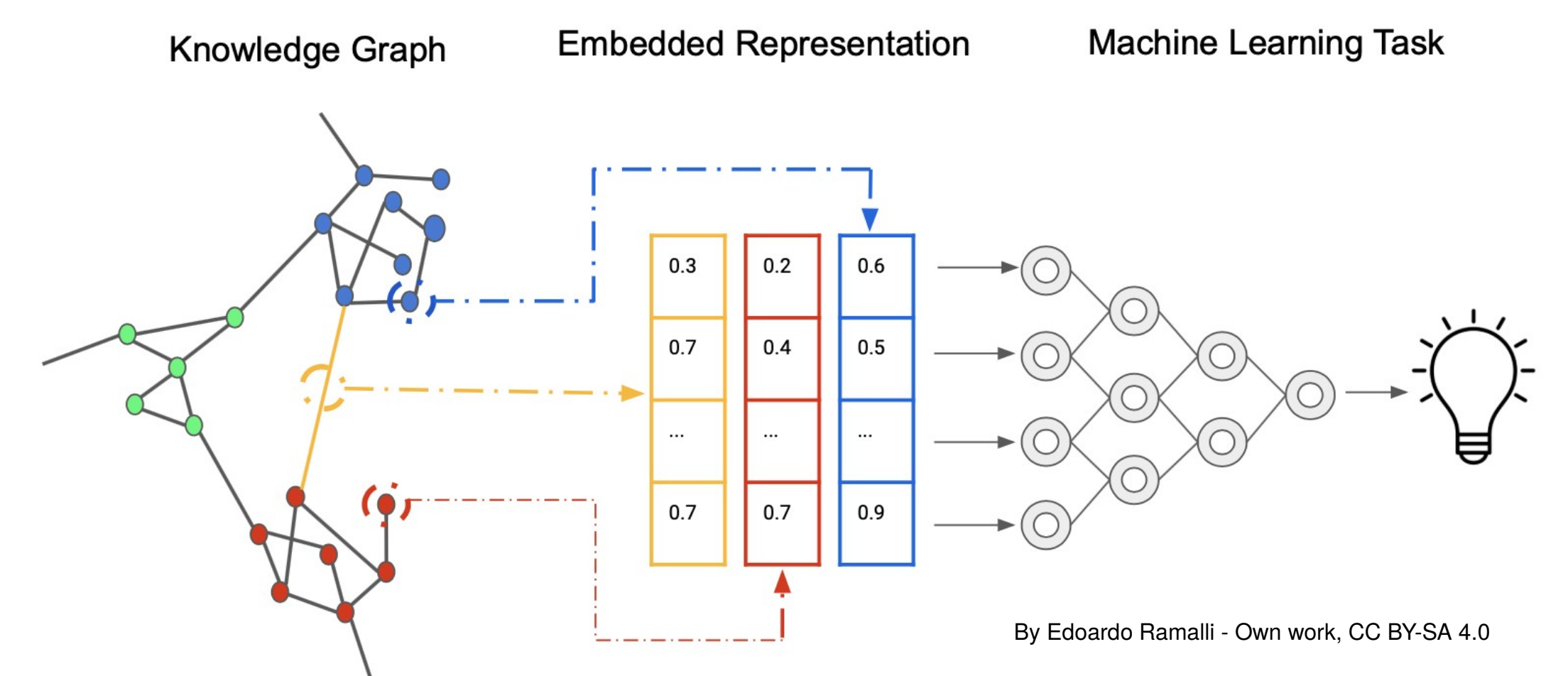
HOLINET est un graphe de connaissances pour le français, qui vise à fournir une perspective holistique de la représentation des connaissances linguistiques. Ainsi, il approche la langue à la fois comme tout, et comme la somme de ses parties sur les différentes dimensions linguistiques. Nous formulons l'hypothèse qu'une telle modélisation holistique des connaissances linguistiques facilitera le traitement automatique du langage et améliorera les performances des applications en aval. HOLINET ouvre de nouvelles pistes de recherche en tant que graphe de connaissances qui intègre des connaissances syntaxiques de référence et des connaissances lexico-sémantiques, et qui combine constituance et dépendance. L'encodage de HOLINET en un modèle de plongement de graphe de connaissances reste une perspective saillante à explorer.

Abstract

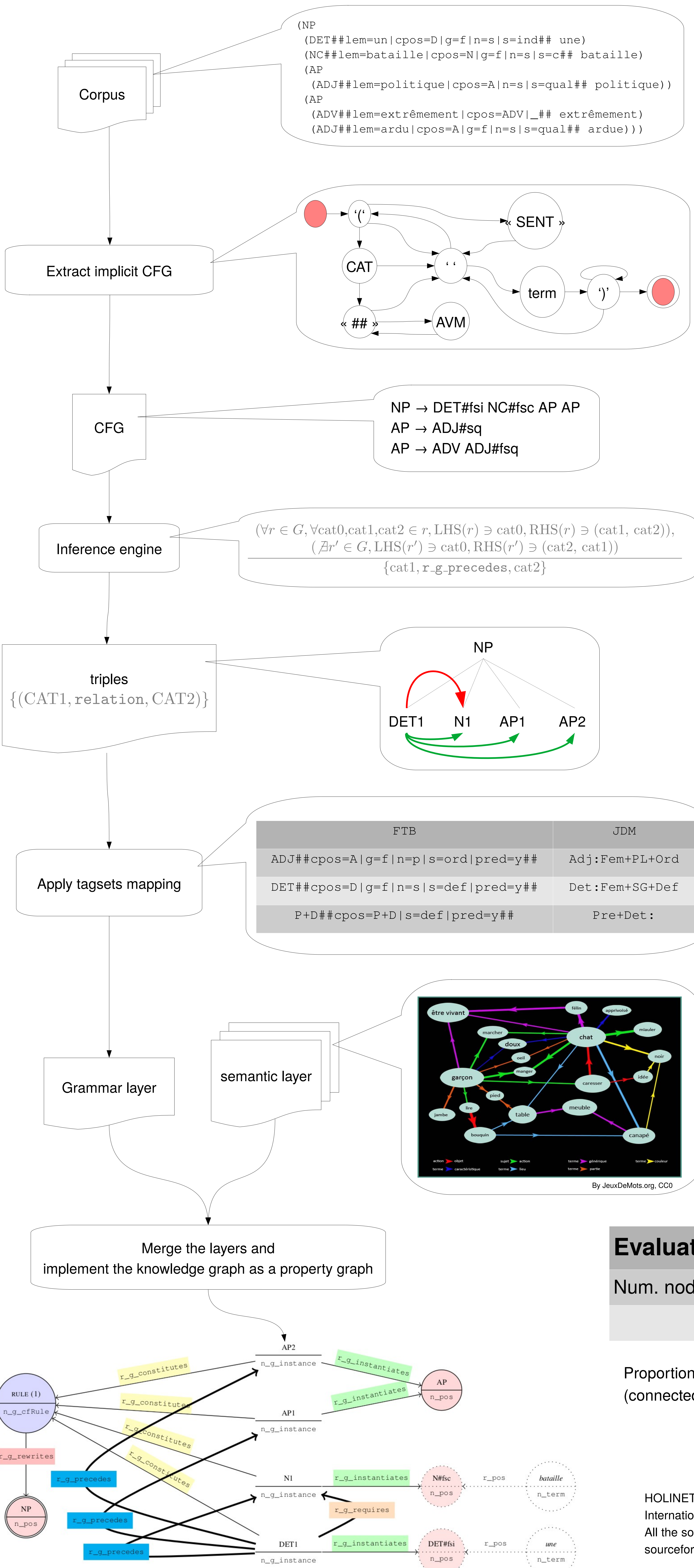
HOLINET is a knowledge graph (KG) for French, which aims to provide a holistic perspective on language knowledge representation. As such, it approaches language as a whole, as well as a sum of its parts on various linguistic dimensions. We hypothesize that a holistic modelling of language knowledge will ease its processing and improve the performance of downstream applications. HOLINET opens new avenues of research as a KG which integrates gold-standard syntactic knowledge along with lexical semantic one, and which is open to combining constituency and dependency information. The computation of a KG embedding model, for instance, is a salient option to investigate.

Motivation and further works

- Would syntax-semantics Knowledge Graph embeddings based on HOLINET outperform traditional approaches in downstream applications ?



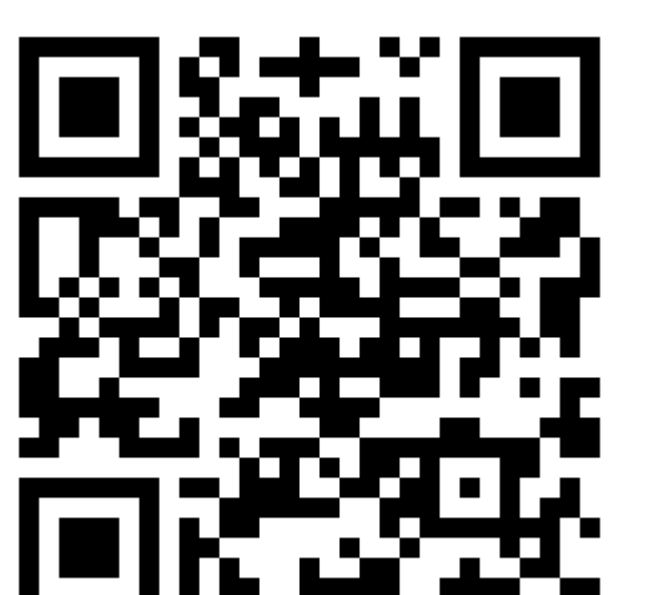
- Could HOLINET integrate Dependency Grammar ? Construction Grammar ? Gesture ? etc.



Evaluation	Connected	Disconnected	Null	Total
Num. nodes	22 742	43 361	9 408	75 511
%	30.1 %	57.4 %	12.5 %	100 %

Proportion of the POS categories found in FTB grammar rules, that are actually found in JDM (connected), or not (disconnected).

HOLINET v1.0 is distributed by ORTOLANG under a Creative Commons Attribution 4.0 International licence (CC-BY 4.0). All the software involved in the creation process is publicly available as git repositories on sourceforge.net. See (Prost, 2022) for more details.



<https://holinet.lpl-aix.fr/>