

# Enrichissement d'ontologie avec des règles SWRL à partir de texte pour la culture sous serre

Pauline Armary<sup>1,2</sup>, Antoine Spicher<sup>2</sup>, Cheikh-Brahim El-Vaigh<sup>1</sup>, Christine Lahoud<sup>3</sup>,  
Hajer Bazaoui<sup>4</sup>, Ouassila Labbani Narsis<sup>1</sup>, Christophe Nicolle<sup>1</sup>

<sup>1</sup> Université Bourgogne Europe, CIAD UR 7533

<sup>2</sup> Anabasis-Assets

<sup>3</sup> Université de Technologie de Belfort-Montbéliard, CIAD-LAB

<sup>4</sup> ETIS UMR 8051, CY University, ENSEA, CNRS

pauline.armary@u-bourgogne.fr

## Résumé

*Pour permettre l'inférence de recommandations riches pour la culture sous serre, nous présentons une méthode pour enrichir l'ontologie AllInOneGreenhouseOntology avec des règles SWRL extraites de texte. Cette méthode s'appuie sur l'identification de motifs grammaticaux au sein des phrases, qui sont traduits en règles logiques. Notre approche a extrait et traduit 240 règles issues d'un corpus de recommandations agronomiques. Notre approche obtient un score ROUGE de 44.9 sur les phrases identifiées comme règles, en comparant entre les phrases initiales avec un langage naturel contrôlé généré automatiquement.*

## Mots-clés

*Construction automatique d'ontologie, Serre, ontologie, SWRL*

## Abstract

*To enable the inference of rich recommendations for greenhouse cultivation, we present a method for enriching the AllInOneGreenhouseOntology with SWRL rules extracted from text. This method relies on identifying grammatical patterns within sentences, which are translated into logical rules. Our approach extracted and translated 240 rules from a corpus of agronomic recommendations. Our approach achieves a ROUGE score of 44.9 on the sentences identified as rules, comparing the initial sentences with automatically generated controlled natural language.*

## Keywords

*Ontology Learning, Greenhouse, ontology, ontology enrichment, SWRL*

## 1 Introduction

Pour répondre aux enjeux de structuration et partage des connaissances des méthodes de cultures sous serre, l'ontologie AllInOneGreenhouseOntology [4] a été proposée comme une ontologie générique et modulaire conçue pour capturer l'ensemble des aspects multifacettes de ces systèmes. Son objectif est de servir de colonne vertébrale sé-

mantique pour le suivi et le contrôle d'une serre intelligente agrémentée de capteurs, qui peut être complétée par d'autres connaissances et ontologies existantes, et permettre à des amateurs de cultiver leurs plants dans un environnement contrôlé, soutenu par les connaissances d'agronomes experts du domaine.

Un des enjeux majeurs de la construction d'une telle ontologie est d'intégrer l'ensemble des règles liées à la culture sous serre pour permettre des recommandations précises sur la culture et le traitement des plantes. Ce travail, principalement manuel, peut être soutenu par des techniques de construction automatique d'ontologie (en anglais *ontology learning*) et d'enrichissement d'ontologie, pour permettre d'accélérer le temps de construction en proposant des règles candidates à intégrer à l'ontologie [1].

L'objectif spécifique de la présente recherche est d'enrichir l'ontologie AllInOneGreenhouseOntology avec des règles SWRL (*Semantic Web Rule Language*)<sup>1</sup> dérivées de manuels de recommandations agronomiques à destination de particuliers. Ces règles représentent des connaissances expertes exploitables et, une fois intégrées, améliorent les capacités de raisonnement de l'ontologie en permettant d'inférer des recommandations basées sur l'état de la serre et des plantes qui s'y trouvent.

Notre article présente les contributions suivantes :

1. une approche neuro-symbolique s'appuyant sur les arbres de dépendances syntaxiques générés par un llm pour identifier des motifs grammaticaux traduits en règles SWRL pour l'enrichissement d'ontologie.
2. une méthode d'évaluation basée sur un Langage Naturel Contrôlé pour valider la qualité des règles générées dans un contexte d'absence d'un Gold Standard pré-existant.

La section 2 présente l'approche neuro-symbolique développée pour extraire les règles du corpus textuel et les rapprocher de l'ontologie, la section 3 présente l'expérimentation réalisée sur un corpus de recommandation en agro-

1. <https://www.w3.org/submissions/SWRL/>

nomie. La section 4 présente les résultats obtenus et la méthode d'évaluation des règles générées.

## 2 Méthodologie

### 2.1 Génération des règles

L'algorithme *DeLIR* [2] est une approche neuro-symbolique conçue pour extraire des structures logiques à partir du langage naturel via une analyse grammaticale basée sur les *Universal Dependencies (UD)*. L'objectif central est de réaliser une traduction grammaticale du texte d'entrée qui respecte à la fois la structure syntaxique et la signification sémantique. Ce processus de traduction est structuré autour de trois tâches principales :

1. **Reconnaissance des Entités Nommées (NER)** : identifier les entités clés au sein de chaque phrase ;
2. **Résolution des Coréférences** : associer les pronoms et les expressions référentielles à leurs variables ou individus correspondants ;
3. **Extraction des Motifs Logiques** : transformer les motifs grammaticaux en motifs de logique du premier ordre (en anglais *First-Order Logic(FOL)*) en utilisant des algorithmes de parcours d'arbre appliqués à l'*Arbre de Dépendance Syntaxique*.

Pour relier les structures logiques extraites aux ontologies formelles, nous avons établi une correspondance entre les motifs FOL générés et les règles *SWRL (Semantic Web Rule Language)*. Ces règles sont ensuite structurées de manière récursive en fonction des motifs grammaticaux identifiés lors de l'analyse de chaque phrase. L'algorithme cible spécifiquement les motifs de règles exprimés de manière explicite dans chaque phrase, et ne capture pas les relations logiques entre phrases.

### 2.2 Alignement avec l'ontologie

Pour garantir l'interopérabilité sémantique, les règles générées sont ensuite alignées avec le vocabulaire spécifique au domaine de l'ontologie *AllInOneGreenhouseOntology*.

Cet alignement est effectué sur la base d'une correspondance exacte entre les labels des termes identifiés dans les règles et des éléments de l'ontologie. Les concepts et relations existants de l'ontologie sont utilisés pour ancrer les structures logiques, tandis que les nouvelles relations définies à partir des règles extraites sont ajoutées à l'ontologie. Pour chaque nouvelle relation, le domaine et le co-domaine sont également inférés à partir des variables présentes dans la règle. Initialement, chaque relation se voit attribuer le type le plus générique `owl:Thing` et est ensuite affinée lorsque la variable associée est également attribuée dans la même règle à une classe connue au sein de l'ontologie.

## 3 Données et expérimentation

Le corpus utilisé dans cette étude a été extrait d'un manuel agricole open-source en ligne<sup>2</sup>. Au sein de ce manuel

2. Wiki du site MyFood, <https://wiki.myfood.eu/docs/getting-started>, extraction automatique réalisée le 8 novembre 2024

	Classes	Propriétés Objet	Règles
Ontologie Initiale	98	85	0
Ontologie Extraite	356	329	240

TABLE 1 – Métrique d'extraction des règles et composants de l'ontologie

de recommandation, nous avons extrait 26 sections couvrant différents sujets de la culture sous serre : la présentation des principales étapes de culture, les recommandations en fonction des saisons et des variations qu'elles entraînent (lumière, eau, chaleur), la gestion des nuisibles et des traitements de régulation.

Initialement rédigé en français, le contenu a été traduit en anglais afin de s'aligner avec la langue de préférence de l'ontologie *AllInOneGreenhouseOntology*.

Pour générer des règles logiques, nous avons appliqué l'algorithme *DeLIR* [2] à l'ensemble du jeu de données.

Une étape de classification des phrases a été ajoutée à l'algorithme pour distinguer les phrases susceptibles d'être traduites en règles, suivant les motifs de règles identifiés en section 2, de celles ne comportant que des motifs neutres. Cette étape est nécessaire pour distinguer les connaissances traduites en format avec une certaine fiabilité et celles plus incertaines, que nous utilisons alors comme corpus contrôle pour l'évaluation.

L'étape d'enrichissement de l'ontologie s'effectue sur la base des règles candidates identifiées, pour les associer aux classes et propriétés de l'ontologie *AllInOneGreenhouseOntology*.

## 4 Résultat et Evaluation

### 4.1 Résultats

Notre approche a permis d'identifier et de traduire 240 règles parmi les 755 phrases issues du corpus. Parmi celles-ci, 127 ont été associées à une classe présente dans l'ontologie initiale. Le tableau 1 présente les principales métriques d'enrichissement de l'ontologie, avec le nombre de classes et propriétés initiales (98 classes et 85 propriétés objet), celles identifiées après extraction (356 classes et 329 propriétés objet) et le nombre de règles identifiées et traduites (240).

### 4.2 Evaluation

Évaluer la qualité et la pertinence des règles et axiomes générés présente un défi fondamental en raison de l'absence d'une référence Gold standard validée par des experts. L'objectif de l'enrichissement de l'ontologie avec des règles suppose que celle-ci ne les intègre pas initialement et ne puisse donc pas servir de référence à l'évaluation.

À cette fin, nous proposons d'adapter une stratégie d'évaluation présente dans la littérature en construction automatique d'ontologie [6], pour évaluer les règles générées. Cette stratégie vise à générer un texte proche du langage naturel à partir de la formalisation logique pour effectuer la comparaison entre le texte généré et le texte initial. Nous

Type de phrases	Nombre	BLEU	ROUGE
Règles	240	2.41	<b>44.99</b>
Non règles (base)	515	2.09	23.27

TABLE 2 – Evaluation de la traduction en logique avec CNL

proposons d'utiliser le Langage Naturel Contrôlé (en anglais *Controlled Natural Language (CNL)*) comme représentation intermédiaire, pour générer des phrases en anglais formalisé à partir des règles SWRL générées. Ces phrases proches du langage permettent d'effectuer une comparaison avec le texte d'origine à l'aide des métriques de distance BLEU [7] et ROUGE [5], qui font référence pour la traduction, en évaluant le nombre de bigrammes et n-grammes communs entre le texte source et le texte cible. BLEU se concentre sur la précision et donne une grande importance à l'ordre des mots, là où ROUGE cible plus particulièrement la complétude de l'information, indépendamment de l'ordre. Parmi les CNL disponibles, nous avons choisi d'utiliser *Attempto Controlled English (ACE)*, un CNL qui a des correspondances avec OWL et SWRL [3].

Le tableau 2 présente les résultats des expériences effectuées sur l'ensemble du corpus, en comparant le texte initial en anglais avec le texte en CNL généré automatiquement à partir des règles SWRL extraites du texte. Les scores présentés sont effectués sur les phrases identifiées comme règles dans le corpus ainsi que sur les phrases non identifiées comme règle qui servent de base de comparaison.

Nous pouvons constater que le score ROUGE est grandement amélioré dans le contexte des phrases identifiées comme règles. Le score de 44.99 est presque deux fois supérieur au score obtenu sur les phrases non-règles (23.27), et s'approche des scores obtenus dans le cadre de l'évaluation initiale de l'algorithme sur un jeu de données moins complexe (59.79) [2].

En revanche, le score BLEU reste bas à la fois pour les phrases identifiées comme règles (2.41) et celles qui ne le sont pas (2.09). Ainsi, BLEU ne semble pas être un score adapté à ce type d'approche, ce qui s'explique par le fait qu'il soit très sensible à l'ordre des mots et que, par construction, le CNL généré à partir des règles SWRL ne correspond pas à l'ordre initial de la phrase.

## 5 Conclusion

Notre travail propose la mise en application dans le domaine agricole pour la construction de l'ontologie des serres, d'une méthode d'enrichissement automatique d'ontologie par extraction de règles à partir d'un corpus de recommandations agricoles. Nous avons appliqué l'algorithme DeLIR pour traduire des motifs grammaticaux identifiés dans des arbres de dépendance syntaxique en logique du Premier Ordre et en règles SWRL. L'un des enjeux majeurs de ce travail est l'évaluation de l'exactitude des règles dans un contexte d'accès restreint aux experts et d'absence de source de référence (Gold Standard). Nous proposons ainsi une méthodologie basée sur la traduction des règles SWRL par un langage naturel contrôlé pour effectuer l'éva-

luation, avec un score ROUGE qui obtient 44.99, proche des résultats sur un jeu de données moins complexe.

En perspectives, de nombreuses améliorations sont en cours d'investigation, notamment en termes de motifs grammaticaux intégrés au sein de l'algorithme pour relever les nombreux défis du corpus : prise en compte des informations numériques (mesures, temporalité), de structures verbales complexes (avec de nombreux compléments), ou encore de formules lexicales de recommandation indirecte. Nous investiguons également les manières de vérifier la cohérence des règles entre elles et avec l'ontologie, aussi bien au niveau de la consistance logique que de l'alignement des propriétés qui sont exprimées avec différentes tournures.

## Remerciements

Nous remercions l'Association Nationale de la Recherche et de la Technologie (ANRT) pour son soutien financier fourni par le biais de la bourse CIFRE [Numéro 2023/0276]. Nous remercions également Anabasis-Assets d'avoir contribué à soutenir ce travail par le salaire de la doctorante.

## Références

- [1] Pauline Armary, Cheikh Brahim El-Vaigh, Ouassila Labbani Narsis, and Christophe Nicolle. *Ontology learning towards expressiveness : A survey*. *Computer Science Review*, 56 :100693, May 2025.
- [2] Pauline Armary, Cheikh-Brahim El-Vaigh, Antoine Spicher, Ouassila Labbani Narsis, and Christophe Nicolle. *Identifying Logical Patterns in Text for Reasoning*. October 2024.
- [3] Tobias Kuhn. *The understandability of OWL statements in controlled English*. *Semantic Web*, 4(1) :101–115, 2013.
- [4] Christine Lahoud, Mariam Gawich, Hajer Baazaoui, and Ihab Jomaa. *Aiogo : A modular ontology for managing greenhouses*. In Leonard Barolli, editor, *Advanced Information Networking and Applications*, pages 332–343, Cham, 2025. Springer Nature Switzerland.
- [5] Chin-Yew Lin. *ROUGE : A package for automatic evaluation of summaries*. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain, July 2004. Association for Computational Linguistics.
- [6] Luis Palacios Medinacelli, Yue Ma, Chantal Reynaud, and Gaëlle Lortal. *Knowledge based situation discovery for avionics maintenance*. In *Proceedings of the 10th International Conference on Knowledge Capture, K-CAP '19*. Association for Computing Machinery, 2019.
- [7] Matt Post. *A call for clarity in reporting BLEU scores*. In *Proceedings of the Third Conference on Machine Translation : Research Papers*, pages 186–191, Belgium, Brussels, October 2018. Association for Computational Linguistics.