HAL open science

Thesaurus Enrichment via Coordination Extraction

Anna Chepaikina, Robert Bossy, Catherine Roussey, Stephan Bernard

To cite this version:

Anna Chepaikina, Robert Bossy, Catherine Roussey, Stephan Bernard. Thesaurus Enrichment via Coordination Extraction. 16th International Conference on Metadata and Semantics Research (MTSR 2022), Nov 2022, London, United Kingdom. pp.191-202, 10.1007/978-3-031-39141-5_16 . hal-03933526

HAL Id: hal-03933526

https://inria.hal.science/hal-03933526v1

Submitted on 10 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

Thesaurus Enrichment via Coordination Extraction*

Anna Chepaikina[1][0000−0001−7494−7748], Robert Bossy[2][0000−0001−6652−9319], Catherine Roussey[3][0000−0002−3076−5499], and Stephan Bernard[3][0000−0001−9694−1443]

[1] Inria, 2 rue Simone Iff, 75012 Paris, France anna.chepaikina@inria.fr
[2] Université Paris-Saclay, INRAE, UR MAIAGE, 78350 Jouy-en-Josas, France robert.bossy@inrae.fr
[3] Université Clermont Auvergne, INRAE, UR TSCF, F-63000 Clermont–Ferrand, France catherine.roussey@inrae.fr, stephan.bernard@inrae.fr

Abstract. We advance a method of thesaurus enrichment, based on the extraction of coordinations in a domain-related corpus. Our hypothesis is that there is a semantic homogeneity between the conjuncts located in a coordination. We conducted an experiment that allowed us to evaluate the effectiveness of our method. This experiment aims to enrich the concept hierarchy of a French agricultural thesaurus named French Crop Usage (FCU), thanks to the texts of the Plant Health Bulletins (PHB). The FCU thesaurus is published on the Web using the SKOS model.

Keywords: thesaurus enrichment · agricultural thesaurus · SKOS model · term extraction · coordination · hyperonym · text mining

1 Introduction

The thesaurus French Crop Usage (FCU) normalises the crop names in French. Moreover, it organises the crop names in categories according to their uses in France territory. The uses represent the agricultural sectors. This thesaurus is not complete and evolves according to related projects. FCU was first built to annotate French agricultural alert bulletins in the Vespa project [14]. It is also reused to identify crops when French agricultural reference datasets are published on the Web using semantic Web technologies, like the plant phenological scales [15]. The enrichment of the thesaurus was based on a manual review of reference agricultural documents. The goal is to identify new crop names or new categories that should be included in FCU thesaurus. We propose a semi-automatic method for the enrichment of the FCU thesaurus based on the extraction of coordinations in a corpus of agricultural alert bulletins. Our hypothesis is that there is a semantic homogeneity between elements in an coordination structure. We conducted an experiment that allowed us to evaluate the effectiveness of our method.

* Supported by the French ANR D2KAB Project

The paper is organised as follows: section 2 presents the content of FCU, the manual update method and access methods. Section 3 describes the method we advance, as well as the corpus used for the extraction. Section 4 describes the evaluation setting and results. In section 5 we review existing works on automatic acquisition of new terms, using in particular coordination structures. Finally, in section 6 we conclude and provide insights for future work.

2 French Crop Usage Thesaurus

The FCU thesaurus organises the plants based on their roles in agriculture, in other words the agricultural plant uses. The thesaurus hierarchy has two main branches as shown in figure 1. The branch named *Multiusages* contains all the cultivated plants that have several uses in agriculture. For example, *carotte* (carrot) may be used as vegetable or as fodder. When a cultivated plant has several uses, it is represented by several crop names. First a crop name represents all uses. For example, *carotte* at the top of Figure 1 represents all the uses of carrot. Another crop name represents the specific use of the plant in agriculture. As shown in Figure 1, *carotte potagère* (vegetable carrot) represents the human consumption of carrot and *carotte fourragère* (fodder carrot) represents animal fodder. Thus, *carotte* is linked to *carrote potagère* and *carrote fourragère* by a hierarchical relationship. The branch *Usages_plantes_cultivees* organises the cultivated plants according to their uses. This branch represents agricultural sectors. In this branch the crop name *carotte potagère* is linked to the crop category *légume racine* (root vegetable).

2.1 Content Description

The FCU thesaurus is formalized using the Simple Knowledge Organisation System (SKOS) vocabulary proposed by W3C [8]. The FCU thesaurus is published on the Web using Linked Data principles. Each crop or crop category is represented by an instance of *skos*:*Concept* as shown in Figure 2.

This work uses the version 2.1 of FCU, which contains 521 instances of *skos*:*Concept*. The maximum depth of the hierarchy is 6 levels. Each *skos*:*Concept* is defined by several properties as shown in Figure 2. The description of a crop or crop category contains the following:

- *skos*:*prefLabel* property contains the crop label name in French. The term is the common name of the crop or crop category. To avoid ambiguity in the case of cultivated plant with different uses, the crop label name is the combination of the common name of the plant and its use. For example in Figure 2, the crop label name is "carotte potagère".
- *skos*:*altLabel* property contains other possible label names that can be used for the crop. For example in Figure 2, an alternative crop label name is "carotte cultivée".
- *skos*:*definition* contains the definition of the crop in French. The definition accounts for the crop position in the hierarchy.

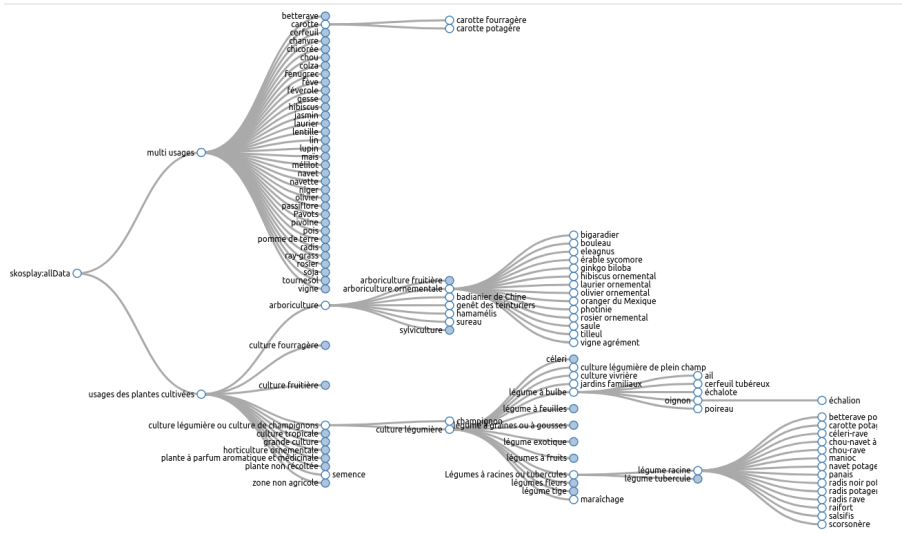

Fig. 1. An extract from the FCU thesaurus, visualised with the SKOS Play tool

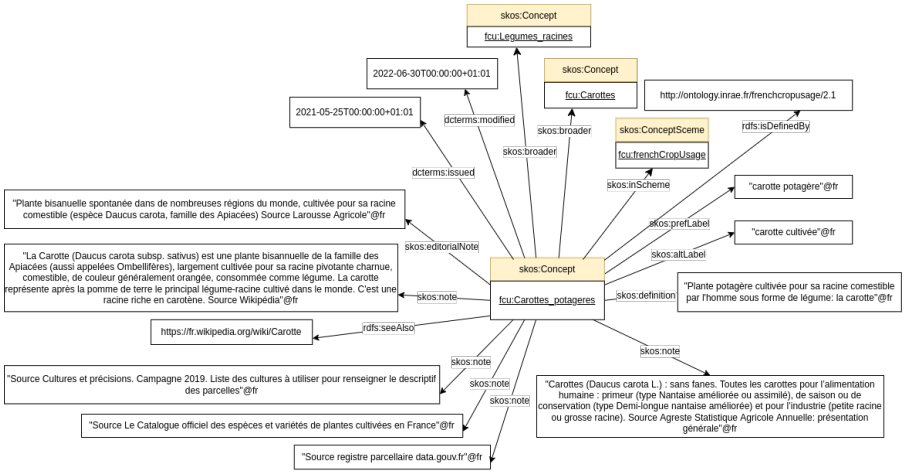

Fig. 2. The instance of *skos*:*Concept* that describes the "vegetable carrot" crop

- *skos*:*note* property contains at least one definition from another source, such as the French Wikipedia. The definition always ends by the indication of the Source. If the crop label name was found in a document without definition only the document name is mentioned. For example in Figure 2, The crop *carotte potagère* was found in the source the official catalogue of species and varieties of cultivated crops.
- *rdfs*:*seeAlso* property contains a Web link to the web site where one definition of the crop was found. For example, *skos*:*Concept* of Figure 2, is linked to the French Wikipedia page.
- *skos*:*inScheme* property expresses that the *skos*:*Concept* belongs to the thesaurus French Crop Usage.
- *rdfs*:*isDefinedBy* property indicates the thesaurus version when the concept was added.
- *dcterms*:*issued*/*dcterms* : *modified* property indicates the date when the concept was added / modified in the thesaurus.
- *skos*:*broader*/*skos*:*narrower* property links the crop to its parent category.

2.2 Sources

To build the thesaurus French Crop Usage, we have studied the terms contained in several documents that are reference sources of information in the agricultural sectors. First, generic sources were studied. More specific sources dedicated to a specific agricultural sector were also reviewed. Compared to [14], we mention only new reference documents manually reviewed :

- The documentation notice of telepac web service proposes a list of crop names to be used to fill in the parcel description. Telepac service enables farmers to obtain funds from the common agricultural policy[4].
- The official catalogue of species and varieties of cultivated crops in France produced by the French variety and seed study and control group (GEVES)[5].
- The catalog of plant protection products and their uses, fertilizing materials and growing media authorized in France[6]. The associated database is entitled Ephy. This catalog is published by the French Agency for Food, Environmental and Occupational Health and Safety (ANSES).
- Regarding fodder sector, two web sites were studied: The herbe-book web site[7] and the plantesfourrageres web site[8] describe different varieties of fodder crops. Those web sites are published by the French interprofession of seeds and seedling named SEMAE.

[4] `https://www1.telepac.agriculture.gouv.fr/telepac/pdf/tas/2019/Dossier-PAC-2019_notice_cultures-precisions.pdf.`
[5] `https://www.geves.fr/catalogue-france/`
[6] `https://ephy.anses.fr/`
[7] `https://www.herbe-book.org`
[8] `http://www.plantesfourrageres.org/pages/caracteristiques.htm`

– Regarding vegetables there is no crop reference documentation. Thus, the common crops between several data sources were sought: Wikipedia, Bonduelle, FranceAgriMer, Encyclopedia Universalis, the Bec Hellouin organic farm.

2.3 Manual Update Method

The RDF file of the FCU thesaurus is updated using Protégé ontology editor v5.1.0 [9] with Cellfie plugin[9]. When a new reference source of information is identified, the FCU thesaurus manager creates a new CSV file dedicated to this source. The reference source is then studied to extract crops cultivated in France. If possible, some definitions of the crop and their use description are also identified in the source. The choice of the label name of each crop concept, their definition and their position in the hierarchy is discussed by at least one expert of the agricultural sector. Finally, the thesaurus manager proposes a new reference definition in French in order to justify the position of the crop concept in the thesaurus hierarchy. All the information dedicated to new crop concepts in the source is stored in this CSV file. Each line corresponds to one crop concept. The line contains the labels, the definitions, notes, web link, date of creation, FCU version and parent crops. Cellfie is used to generate individuals of *skos*:*Concept* class using transformation rules applied on the CSV file. To improve the consistency of the final RDF dataset, we used some SWRL rules to infer all inverse properties using SWRL Protege Tab. A final check is performed using the SKOS Play! tool[10]: it enables to visualise and control the SKOS model and detects errors.

2.4 Access Description

The thesaurus is published into a git repository[11], the Agroportal repository[12] and a dedicated SPARQL endpoint[13]. The git repository is the reference storage system. It is divided into branches. The *dev* branch is updated by the thesaurus manager. This branch contains the input CSV files, the Cellfie transformation rules and the output OWL-XML file. The *master* branch contains the last version published on the web. Each old version has a dedicated branch. Note that the branch named *enrichissementThesaurus* contains files and documentation, used for the method proposed in this paper.

3 Automatic Detection of New Terms

For the moment the information sources, used to update the FCU thesaurus, are all agricultural reference documents that somehow define crops. We decide

[9] `https://github.com/protegeproject/cellfie-plugin`
[10] `http://labs.sparna.fr/skos-play`
[11] `https://gitlab.irstea.fr/copain/frenchcropusage`
[12] `http://agroportal.lirmm.fr/ontologies/CROPUSAGE`
[13] `http://ontology.inrae.fr/frenchcropusage/sparql`

to use another type of information sources: a corpus of french agricultural alert bulletins. This corpus can be categorised as usage documents, not reference documents. Due to the fact that the corpus contains a large number of files, we need an automatic extraction process to identify new crop names.

3.1 Corpus Description

In France, the Grenelle Environment and Ecophyto 2018 program strengthened national surveillance networks of crops and agricultural practices. Plant Health Bulletins are one of the modalities established by these surveillance networks in all regions and French overseas departments. A Plant Health Bulletin (PHB)[14] is an information document, both technical and regulatory in nature, written under the responsibility of a regional epidemiological surveillance committee. A PHB gathers information about the health status of crops. It first presents pest risk analyses, and it is also used to disseminate regulatory information (mandatory control order, national notes, regulation evolution, etc.) and non-regulatory information (description of pest biology or prophylactic methods, such as management of intercropping, tillage, choice of cultivars, etc.). Since the beginning of their publication, PHBs are freely available in PDF format on the websites of the Regional Chambers of Agriculture or the websites of the DRAAF (the regional agency of the French Ministry of Food and Agriculture). Therefore, PHBs are disseminated on different websites (one per region). Depending on the health issues or crop development, the frequency of PHB publications is variable, ranging from monthly to weekly. Nearly 15,000 plots are observed each year to edit approximately 3400 PHBs per year [14].

3.2 Method Description

Corpus preprocessing We rely on a test corpus of the PHBs for automatic extraction of new crop concepts and label names. Our test corpus holds 880 PHB files with an average of 2548 tokens per file, each covering different sectors of French agriculture. The PHB files have been collected during various periods from 2009 [15] and transformed from their initial PDF format into an HTML format using a conversion tool named `pdf2blocks`[16]. Our system applies on standard methods of text preprocessing (sentence segmentation, word tokenization, part-of-speech tagging) and also takes into account issues after the conversion from PDF.

Assumption In our method, we focus on detecting coordination structures in paragraphs of the bulletins. Coordinations are complex syntactic structures composed of several conjuncts, one or more of which are preceded by a coordinator (e.g. "et") or a comma [6]. Coordinations are often used to enumerate, exemplify

[14] PHB are named Bulletin de Santé du Végétal (BSV) in French.
[15] `https://gitlab.irstea.fr/copain/d2kab/`
[16] `https://doi.org/10.5281/zenodo.4067965`

and compare. [4] observes that the conjuncts share the same semantic relations with their environment. Hence, the assumption at the heart of our method is that if a term which is already present in the thesaurus appears to be a conjunct of a coordination, then the other conjuncts from the same coordination can share similar semantic relations, and serve therefore as candidates for enrichment of this resource. For instance, a coordination "pittospore, acacia, eucalyptus et eleagnus" consists of 4 conjuncts. One of them, namely "eucalyptus", is already featured in the FCU thesaurus. Therefore, the other conjuncts are to be considered as candidate terms for the enrichment. Our method is decomposed into three parts : extraction of coordinations, selection of candidate terms, suggestion of hyperonyms and similar label names.

Extraction of coordinations To detect coordinations, we use the latest[17] version of `Stanza` [12], a natural language processing toolkit that provides a neural pipeline for text analysis. Stanza offers a state-of-the-art performance for French in regards to the tokenization, lemmatization, part-of-speech, morphological feature labelling, and dependency parsing. We use its default lemmatization model with adjustments for outputs of a few domain-specific words. For example "radis", a french crop name for radish, should be lemmatized as "radis" and not "radi" (a form proposed by the default model) and thus adjusted by our system.

Regarding coordination structures, Stanza treats them as symmetric relations and uses the relation *CONJ* to connect the first conjunct of the coordination to all subsequent conjuncts. The language model used by Stanza is based on Universal Dependencies v2 [11]. According to this version, the coordinators *CC* and punctuation marks *PUNCT* of the coordination are attached to the conjunct *CONJ*, mentioned right afterwards. Figure 3 shows a dependency analysis of a sentence containing a coordination structure.

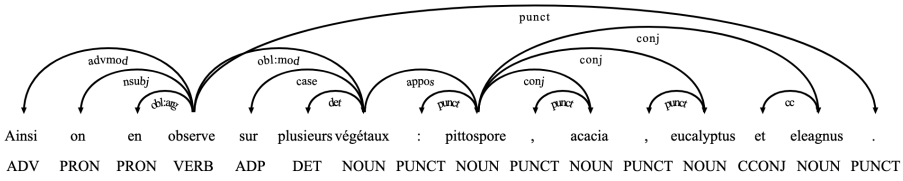

Fig. 3. Stanza's Dependency Analysis

The extraction process starts with a dependency parsing of the sentences, obtained from the corpus. In particular, we look for coordinations whose conjuncts have *NOUN* as a part-of-speech tag. These conjuncts may be accompanied by one or more expansions such as determiners, qualifying adjectives, prepositions

[17] `https://github.com/stanfordnlp/stanza/releases/tag/v1.4.0`

or noun complements. Due to frequent mentions of phenological stages of plants inside the coordinations, we impose that a conjunct must not contain any digit (e.g. the conjunct equal to *"BBCH55"* would represent a phenological stage in the BBCH scale [7] rather than a crop label name). Moreover, names of pests, plant parts, tools and scientific names of plants are filtered. Those terms are very common in the text of the bulletins but are not useful for enrichment.

Selection of candidates As already mentioned, we assume that conjuncts share same semantic relations with their environment. In order to find candidates for enrichment, we need to find at least one conjunct that already corresponds to a crop label name from the thesaurus. The other conjuncts from the same coordination can share similar semantic relations and serve as potential candidates for enrichment.

So, firstly, we search for crop concepts inside the coordination. We align each conjunct of each coordination with existing crop label names (*skos*:*prefLabel*, *skos*:*altLabel*) and link this conjunct to the respective crop concept. The alignment is done by comparing lemmatized forms of a conjunct and of a crop label name, with no regard for their case and punctuation marks. By the end of the alignment, we characterise conjuncts into the following groups:

- **exact match** : if the comparison between the conjunct and the crop label name is exact (e.g.: "choux chinois" is a label name of the crop concept *Choux_chinois* and is also found in the text).

- **partial match** : if the comparison between the conjunct and the crop label name is partial (e.g. "cerise de Cayenne" is found in the corpus and the crop concept *Cerise* has a label name "cerise").

- **null match** : if the comparison between the conjunct and the crop label name has not been made.

Secondly, we filter out coordinations which are out of topic, such as the coordinations whose conjuncts have made only null matches. We also remove coordinations whose conjuncts have made only exact matches, since they will neither enrich the thesaurus.

Finally, among the remaining coordinations, we retrieve the conjuncts which do not fully correspond to the entries of the thesaurus, that is those with a partial or a null match. Such conjuncts will serve as candidates for enrichment.

Suggestion of hyperonyms and similar label names We anticipate that some candidates will be considered as new crop concepts and that subsequently they will need to be placed in the hierarchical structure of the thesaurus. Therefore, we suggest a set of hyperonym concepts for each candidate, hoping that this will help to determine the best position for their integration. Hyperonyms are computed by extracting *skos*:*broader* concepts for conjuncts from the same coordination as the candidate, once these conjuncts have been linked to a crop concept.

On the other hand, we presume that some candidates will be added as new label names for already existing crop concepts. In this case, we suggest a set of

similar label names which would help to update an associated concept. Similar label names are retrieved by computing the cosine similarity between the vector of an existing label name and the vector of a candidate. In order to establish a vector, either for a label name or a candidate, we chunk their respective word sequence and find average of all the word embeddings in the chunk. We use `fastText` [1] to determine the word embeddings.

On average, we suggest 3 hyperonyms and 5 similar label names for a candidate. The choice of the most appropriate suggestion is ultimately made by the thesaurus manager.

4 Evaluation

The evaluation of the candidates is conducted manually by the thesaurus manager. We provided a document containing candidate terms, their respective coordination, context, linked crop concepts, suggested hyperonym concepts and similar label names. The first part of the evaluation concerned the relevance of the candidate terms and their typifying. The second part was focused on the accuracy of the suggested hyperonyms for those candidates, typified as new concepts, and on the accuracy of the suggested similar label names for those, typified as new label names. In addition, the coordinations were also examined for their well-formedness.

As a result, we have obtained 138 coordinations in our test corpus. 70% of coordinations are well-formed. For example, our system helped to find a coordination "saule, tilleul, marronnier, platane, cerisier, éléagnus" inside the sentence : "De façon moins fréquente les espèces suivantes ont aussi été touchées : le saule (Salix babylonica) , le tilleul (Tilia sp), le marronnier (Aesculus hippocastanum), le platane (Platanus sp,) le cerisier (Prunus cerasus), l'éléagnus (Eleagnus sp)". This coordination contains two conjuncts that were identified as FCU concepts (*Tilleuls*, *Cerisiers*) and the rest of the conjuncts were proposed as candidates for enrichment.

Regarding badly formed coordinations, they either lack conjuncts or at least one of the conjuncts is incomplete or grammatically incorrect. For example, in the sentence "Des taches sporulantes sont observées sur les feuilles de la couronne extérieure des laitues rouges et des laitues iceberg", we found a coordination "laitue rouge, laitue". It has all two conjuncts, nevertheless one of them lacks its nominal modifier "iceberg". This would be considered as an example of an incomplete extraction and a potential source of improvement.

With the help of the retrieved coordinations, we have submitted 121 candidate terms, most of which have been considered relevant by the thesaurus manager (Table 1). The irrelevant candidates were generally either ill-formed or outside the scope of the thesaurus, representing scientific names of plants. The thesaurus manager also concluded that approximately two-thirds of the relevant candidates have been added as new concepts and one-third as new label names. In regards to the integration of these relevant candidates, in 67% of cases the suggested hyperonyms have been considered helpful for placing a new crop con-

Relevency	*relevant*			*irrelevant*		
Type	new concepts	new label names	already existing label names	scientific names	ill-formed	unspecified
Count	54 45 %	32 26 %	4 3 %	13 11 %	12 10 %	6 5 %
Total	90 74 %			31 26 %		

Table 1. Candidate terms

cept in the hierarchy. The same goes for the suggested similar label names (84% of cases).

Note that the detection of a new term can lead to the creation of three new crop concepts: one in the branch *Multiusages* (Frênes) and two in the branch *Usages_plantes_cultivees* (Frênes forestiers, Frênes ornementaux). Thanks to this experiment, the version 3.0 of the FCU thesaurus was build.

5 Related Work

Our work pertains to the field of automatic acquisition or enrichment of lexicons from text. The acquisition of new labels is typically divided into two sub-tasks: the extraction of candidate labels from the text, and the prediction of the relation between candidate labels and existing labels in the lexicon. The two tasks can be tackled either in sequence, or jointly.

In the present work, we focus on noun phrase candidates, and on hypernym/hyponym relations. On the one hand, terminology labels are rarely adjectival or verbal phrases. On the other hand, the hierarchy of hyponyms is the most common denominator in lexical resources, such as terminologies or ontologies.

Since [5], hand crafted syntactic patterns have been used for automatically extracting information from text, in particular the extraction of hyperonyms. The syntactic patterns take form of regular expressions with placeholders for the extracted labels. Typical hyperonymy extraction patterns look like "$Label_{NP}$ *is a* $Hyper_{NP}$" or "$Hyper_{NP}$ *such as* $Label_{NP}$".

The SEXTANT system [3] introduced the use of distributional semantics for extracting labels from corpora. The core hypothesis of distributional semantics is that words that occur in similar contexts must have similar meanings. This family of methods have received a lot of traction with the availability of word embeddings trained on very large corpora [16, 10].

However [13] showed that pattern-based methods provide better and more robust results than embeddings-based methods. [18] demonstrates that patterns based on coordination structures are particularly powerful to extract new hyponym labels for domain-specific lexicons. Other works, like [2] demonstrated that coordination structures may be used to improve supervised extraction of hyponyms.

[17] is a method that aims at extracting a taxonomy in a specific domain (medical) from text using coordination. This system is also based in the *Lowest*

Common Ancestor in WordNet of coordinated nouns, which is very similar to our strategy for suggesting candidate hyperonyms.

To the best of our knowledge all methods extract coordination structures using lexical patterns. Our system uses the result of automatic parsing. We believe that parsers have reached sufficient performance to point to coordinations without limitations on the cardinality (number of elements) or the complexity of each element (noun phrases).

6 Conclusion and Perspectives

This paper presents an experiment of enriching French Crop Usage thesaurus by extracting coordinations in a corpus of Plant Health Bulletins. This experiment has improved the FCU thesaurus by adding 54 new crop concepts and update 32 crop concepts with new label names. The improvement concerns mainly arboriculture sector, which was poorly represented in FCU. All the results are available on a git repository.

In the future, we plan to ameliorate our extraction system to reduce the number of badly formed coordinations. We would like also to repeat this experiment on a new corpus of Plant Health Bulletins, to update regularly the FCU thesaurus, and other agricultural resources like plant phenological scales or pest lists.

References

1. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. arXiv preprint arXiv:1607.04606 (2016)
2. Cederberg, S., Widdows, D.: Using LSA and noun coordination information to improve the recall and precision of automatic hyponymy extraction. In: Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003. pp. 111–118 (2003), `https://aclanthology.org/W03-0415`
3. Grefenstette, G.: Exploration in Automatic Thesaurus Discovery. Kluwer Academic Publishers (01 1994). https://doi.org/10.1007/978-1-4615-2710-7
4. Haspelmath, M.: Coordination. In: Shopen, T. (ed.) Language Typology and Syntactic Description. vol. 2, p. 1–51. Cambridge University Press (2007). https://doi.org/10.1017/CBO9780511619434.001
5. Hearst, M.A.: Automatic acquisition of hyponyms from large text corpora. In: COLING 1992 Volume 2: The 14th International Conference on Computational Linguistics (1992), `https://aclanthology.org/C92-2082`
6. Maier, W., Kübler, S., Hinrichs, E., Krivanek, J.: Annotating coordination in the Penn Treebank. In: Proceedings of the Sixth Linguistic Annotation Workshop. pp. 166–174. Association for Computational Linguistics, Jeju, Republic of Korea (Jul 2012), `https://aclanthology.org/W12-3624`
7. Meier, U., Bleiholder, H., Buhr, L., Feller, C., Hack, H., Heß, M., Lancashire, P., Schnock, U., Stauß, R., Boom, T., Weber, E., Zwerger, P.: The bbch system to coding the phenological growth stages of plants-history and publications. Journal für Kulturpflanzen **61**, 41–52 (01 2009). https://doi.org/10.5073/JfK.2009.02.01

8. Miles, A., Bechhofer, S.: SKOS Simple Knowledge Organization System Reference. W3C Recommendation, World Wide Web Consortium, United States (Aug 2009)
9. Musen, M.A.: The protégé project: a look back and a look forward. AI Matters **1**(4), 4–12 (June 2015)
10. Nguyen, K.A., Köper, M., Schulte im Walde, S., Vu, N.T.: Hierarchical embeddings for hypernymy detection and directionality. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. pp. 233–243. Association for Computational Linguistics, Copenhagen, Denmark (Sep 2017). https://doi.org/10.18653/v1/D17-1022, `https://aclanthology.org/D17-1022`
11. Nivre, J., de Marneffe, M.C., Ginter, F., Hajič, J., Manning, C.D., Pyysalo, S., Schuster, S., Tyers, F., Zeman, D.: Universal Dependencies v2: An evergrowing multilingual treebank collection. In: Proceedings of the 12th Language Resources and Evaluation Conference. pp. 4034–4043. European Language Resources Association, Marseille, France (May 2020), `https://aclanthology.org/2020.lrec-1.497`
12. Qi, P., Zhang, Y., Zhang, Y., Bolton, J., Manning, C.D.: Stanza: A Python natural language processing toolkit for many human languages. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations (2020), `https://nlp.stanford.edu/pubs/qi2020stanza.pdf`
13. Roller, S., Kiela, D., Nickel, M.: Hearst patterns revisited: Automatic hypernym detection from large text corpora. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). pp. 358–363. Association for Computational Linguistics, Melbourne, Australia (Jul 2018). https://doi.org/10.18653/v1/P18-2057, `https://aclanthology.org/P18-2057`
14. Roussey, C., Bernard, S., Pinet, F., Reboud, X., Cellier, V., Sivadon, I., Simonneau, D., Bourigault, A.L.: A methodology for the publication of agricultural alert bulletins as lod. Computers and Electronics in Agriculture **142**, 632–650 (2017). https://doi.org/https://doi.org/10.1016/j.compag.2017.10.022, `https://www.sciencedirect.com/science/article/pii/S0168169917306361`
15. Roussey, C., Delpuech, X., Amardeilh, F., Bernard, S., Jonquet, C.: Semantic description of plant phenological development stages, starting with grapevine. In: Garoufallou, E., Ovalle-Perandones, M.A. (eds.) Metadata and Semantic Research. pp. 257–268. Springer International Publishing, Cham (2021)
16. Shwartz, V., Santus, E., Schlechtweg, D.: Hypernyms under siege: Linguistically-motivated artillery for hypernymy detection. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers. pp. 65–75. Association for Computational Linguistics, Valencia, Spain (Apr 2017), `https://aclanthology.org/E17-1007`
17. Widdows, D., Toumouh, A., Dorow, B., Lehireche, A.: Ongoing developments in automatically adapting lexical resources to the biomedical domain. In: Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06). European Language Resources Association (ELRA), Genoa, Italy (May 2006), `http://www.lrec-conf.org/proceedings/lrec2006/pdf/489_pdf.pdf`
18. Ziering, P., van der Plas, L., Schütze, H.: Bootstrapping semantic lexicons for technical domains. In: Proceedings of the Sixth International Joint Conference on Natural Language Processing. pp. 1321–1329. Asian Federation of Natural Language Processing, Nagoya, Japan (Oct 2013), `https://aclanthology.org/I13-1188`