



HAL
open science

Using a Lexical Semantic Network for the Ontology Building

Nadia Bebeshina-Clairret, Sylvie Despres, Mathieu Lafourcade

► **To cite this version:**

Nadia Bebeshina-Clairret, Sylvie Despres, Mathieu Lafourcade. Using a Lexical Semantic Network for the Ontology Building. RANLP: Recent Advances in Natural Language Processing, Sep 2019, Varna, Bulgaria. , 12th International Conference Recent Advances in Natural Language Processing, 2019. hal-02269128

HAL Id: hal-02269128

<https://hal.science/hal-02269128v1>

Submitted on 22 Aug 2019

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.

Using a Lexical Semantic Network for the Ontology Building

Nadia Bebishina-Clairet
LIRMM / LIMICS

Sylvie Despres
LIMICS
74 rue Marcel Cachin
93017 Bobigny

Mathieu Lafourcade
LIRMM
860 rue de St Priest
34095 Montpellier

Abstract

In the present article, we focus on mining lexical semantic knowledge resources for the ontology building and enhancement. The ontology building is often a descending and manual process where the high level concepts are defined, then detailed by human experts. We explore a way of using a multilingual or monolingual lexical semantic network to make evolve or localize an existing ontology with a limited human effort.

1 Introduction

Today's termino-ontological resources are increasingly rich in terms of data they rely upon. The scientific community works intensively on data acquisition for the ontology building. For instance, the NeOn project¹ has been set up to provide a methodology for the ontology engineering by integrating preexisting knowledge resources into an ontology building process. The NeOn methodology contains a consistent framework for modular ontology building as well as for setting up ontology networks. Here we focus on exploiting Lexical Semantic Networks (LSNs) to enrich an ontology or accompany the ontology building process. We assume that LSNs represent knowledge as it is expressed through the human language whereas ontologies provide a formal description (specification) of a conceptualization (concepts and relationships between those con-

cepts) shared by a community of agents. A concept corresponds to a set of individuals sharing similar characteristics and may or may not be lexicalized. Thus, the ontology labels cannot be polysemous. The strength of ontologies is in their formal consistency. The weaknesses are linked to their coverage, size (as stated in (Raad and Cruz, 2015), "large ontologies usually cause serious scalability problems"), and human effort needed for their building. The potential of the LSNs is linked to the large amount of explicit semantic information they contain. However, a filtering process is needed to discriminate irrelevant information (polysemy, noise).

2 State of the art

The opportunity of ontology construction empowered by the use of Natural Language Processing (NLP) techniques and tools has been explored for more than 20 years. Among the achievements, one can distinguish the tools which take into account the difference between the lexical term and the ontology concept (differentiated tools) and those that do not make such distinction. Differentiated tools and methods suggest extracting the terminological units from texts and organizing them as a network using a set of hierarchical and equivalence relation types. Such network guides the ontology expert through the conceptualisation and ontology building process. Such process relies on an intermediary structure, a termino-ontology (for instance, the *Terminae* suite, (Szulman, 2012)). Undifferentiated tools

¹http://neon-project.org/nw/About_NeOn.html

use some statistical information to suggest the candidate concepts. They exploit such methods as formal concept analysis (Mondary, 2011) or knowledge based methods (for instance, *TextToOnto*²). "Lexical ontologies" (Abu Helou et al., 2014) are successfully used for ontology building. Numerous approaches targeted at high level ontology or information retrieval ontology based on general knowledge (such as (Marciniak, 2013)) rely on PWN. Others use PWN and domain specific semantic lexicons for forming the concepts (Turcato et al., 2000). Many other ontology learning techniques use distributional semantics to learn lightweight ontologies, for example, (Wong, 2009).

In the framework of corpora based approaches to the ontology building such as described in (Kietz et al., 2000), the idea of *notable (salient, relevant) element* or *relevant piece of knowledge* (RPK) has been introduced. It corresponds either to the frequent terms appearing in a corpus and to the tacit knowledge contained in texts. Such tacit knowledge corresponds to the semantic relationships (subsumption relationship and other specialized relationships). Their presence in texts may take the form of "indices". In contrast, the explicit elements may reveal the presence of concepts. The main drawback such definition of RPKs is that it relies on the features defined or recorded for a particular language. In addition, statistical criteria are often preferred and it is difficult to qualify such RPKs from the semantic point of view and in a language independent manner. In this paper, we will detail the experiments we conducted to provide a new definition of the RPKs based on the structured lexical semantic information and describe the way so defined RPKs can be used to help the top down ontology building process.

3 Ressources

In the present section, we will describe the resources we used in our experiments.

The **RezoJDM** (Lafourcade, 2007) is a lexical semantic network (LSN) for French built using crowdsourcing methods and, in particular, games with a purpose (GWAPS) such as JeuxDeMots³ and additional games. This commonsense network has been built since 2007. It is a directed, typed and weighted graph. At the time of our writing, RezoJDM contains 2.7 millions of terms that are modeled as nodes of the graph and 240 millions of relations (arcs).

The **MLSN** (Bebeshina-Clairet, 2019) is a multilingual LSN (it covers French, English, Spanish, and Russian) with an interlingual pivot built for the cuisine and nutrition domain. This network is inspired by the RezoJDM in terms of its model. Structurally speaking, the MLSN is a directed, typed, and valuated graph. It contains k sub-graphs corresponding to each of the k languages it covers and a specific sub-graph which fulfills the role of the *interlingual pivot*. Similar to the RezoJDM, we call *terms* the nodes of the MLSN and *relations* - its typed, weighed, and directed arcs. The MLSN nodes may correspond to one of the following types : **lexical items** (*garlic*), **interlingual items** (pertaining to the interlingual pivot and also called *covering terms*), **relational items** (i.e. relationship reifications such as *salad[r_has_part]garlic*), and **category items** parts of speech or other morpho-syntactic features (i.e. *Noun:AccusativeCase*).

As it has been difficult to set up the pivot using a multilingual embedding (joining multiple spaces, one per language) as well as to avoid pairwise alignment based on combinatorial criteria, the pivot has been started as a natural one using the English edition of DBNary (Sérasset, 2014). It incrementally evolves to become interlingual. The pivot evolution relies on sense-based alignments between the languages of the MLSN and aims at taking into account the difference of sense "granularity" in different languages. For example, as *stew* in English can be translated as into French as *pot-au-feu* and *ragoût*. It

²<https://sourceforge.net/p/texttoonto/wiki/>

³<http://www.jeuxdemots.org>

reflects the conceptualization discrepancy as *ragoût* refers to sauté the ingredients and then add water whereas *pot-au-feu* refers to boiling them together in a larger amount of water than used for the *ragoût*. In the MLSN pivot we have the interlingual term corresponding to stew (which covers the English term) with two hyponyms corresponding to Orench terms. The alignments are progressively obtained through external resources or by inference. Thus it can be considered as a union of word senses lexicalized or identified in the languages covered by the MLSN. . Even though we assume the pivot as being interlingual, it is still close to a natural one. A relation $r \in R$ is a sextuplet $r = \langle s, t, type, v, l_s, l_t \rangle$ where s and t correspond respectively to the source and the target term of the relation. The relation $type$ is a typical relation type. It may model different features such as taxonomic and part-whole relations (r_isa , r_hypo , r_has_part , r_matter , r_holo), possible predicate-argument relations (typical object r_object , location $r_location$, instrument r_instr of an action), "modifier" relations (typical characteristic r_carac , typical manner r_manner) and more⁴. The relationship valuation v corresponds to the characteristics of the relation which are its weight, confidence score, and annotation. The relation weight may be negative in order to model noise and keep the information about erroneous relations easy to access programmatically so they could not affect the inference processes. The confidence score is a score attributed to a particular data origin (external resource, inference process). In practice, this feature is an array as different origins may provide the same relation. The confidence information is provided as an argument to the function that maps from some external knowledge resource to the MLSN. In case of relation calculated by an inference process, it corresponds to the precision evaluated on a sample of candidate relations returned by this process. To *annotate* a relation we add a complementary information that allows qualifying this relation. The figure 1

details and exemplifies the annotation scheme.

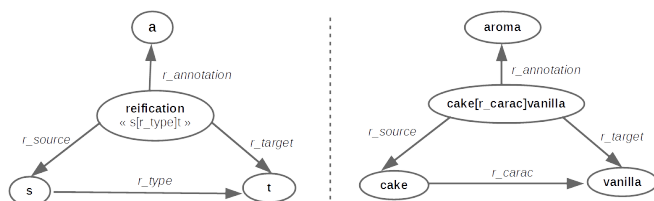


Figure 1: MLSN: relationship annotation scheme.

The labels l_s and l_t correspond to the language (sub-graph) labels. At the time of our writing, the MLSN contains 821 781 nodes and 2 231 197 arcs. It covers 4 languages : English, French, Russian, and Spanish.

MIAM (Desprès, 2016)⁵ is a modular terminology for the digital cooking. It provides knowledge necessary for the elaboration of general nutritional suggestions. The knowledge model of this ontology gathers expert knowledge on food, food transformation, cooking actions, relevant dishes that reflect french culinary tradition, recipes necessary to cook such dishes. MIAM contains about 7 000 nodes and 30 000 semantic (non subsumption/ontological is-a) relations.

4 Method: immersion - projection

4.1 Summing up the Method

Our method is built upon the idea of projecting a model (the MIAM model) onto a multilingual or monolingual LSN (respectively MLSN and Rezo-JDM) in order to extract an intermediary resource that can be used by ontology or domain experts in the scope of information retrieval or validation of the automatically suggested pieces of knowledge.

Our method differs from others by the definition of the RPK and by the use of a non ontological semantically structured resource for ontology building. We

⁴We also introduced more specific relation types such as $r_entailment$, r_cause , r_telic_role , $r_incompatible$, r_before , r_after etc.

⁵<http://www-limics.smbh.univ-paris13.fr/ontoMIAM/>

define RPK as follows : "a relevant piece of knowledge is either a term or a relation or a semantic structure which is known as qualified and qualifying". *Qualified* refers to the possibility to describe the RPK in a discrete way (i.e. by enumerating the typed relations). If the RPK is a term, it needs to have a high in-degree (which reveals its conceptual role as it is used to define other terms of the network). If the RPK is a relation, it needs to be contextualized (through the annotation mechanism represented on the figure 1 or through the constraints put on source and/or target terms of the relation). If the presumed RPM is a graph structure (path, sub-graph), it needs to possess a certain number of occurrences in the network. *Qualifying* refers to the possibility to use the candidate RPK for endogenous inference process. If the RPK is a term, it needs to have hypernyms, hyponyms, synonyms among its neighbours. It has to be aligned with other terms pertaining to the other languages of the LSN (if such LSN is multilingual). If the RPK is a relation, it must not be unique (other real or potential⁶ relations must exist in the network). If the candidate RPK is a structure, its terms and relations must be qualifying.

Here we detail the experiments that have been conducted on the basis of the MLSN in order to propose "pseudo-class" and "pseudo-property" candidate RPKs to enhance the MIAM ontology and those concerning the enrichment of an ontology draft using the monolingual LSN, RezoJDM (Lafourcade, 2007). These experiments rely on lexical knowledge. Therefore, the resulting RPKs have no pretension to the ontological validity. The decision pertains to the human expert.

4.2 Immersion

The projection of a given ontology model onto a LSN starts by the immersion of such model. The immersion mechanism uses a set of manually defined mapping rules. It is possible to generate them automatically for the ontological resources that ex-

ploit standard vocabularies (such as RDFS, SKOS and other machine readable formats). The input of the immersion algorithm is the reference ontology (MIAM) and the set of mapping rules whereas its output is the action of inferring terms and relations in the target LSN (MLSN, RezoJDM).

In their general form, the mapping rules state: "If x and y are respectively domain and range of an Object Property p of the ontology to be immersed and y is a subclass of C , then x has a relation R with y and y has a relation *is-a* with C in the receiving (target) LSN". Such rules have been defined for the multilingual experiment for two reasons. First, for each of the 93 MIAM properties, we determined relevant MLSN semantic relation types (or set of types). Thus, the *Object Property* `aPourProduitInitial` (*hasInitialProduct*) corresponds to the substance and part-whole meronymy (MLSN relations typed *r_has_part* and *r_matter*). Second, we mapped the ontology labels to the MLSN terms by coincidence (3 930 terms; i.e. *poulet basquaise* formally denotes a MIAM concept, a lexical item with the same label already exists in the MLSN) or by composition (4 135 terms, i.e.: *unité mesure capacité* doesn't correspond to any existing MLSN term because it doesn't correspond to any commonly used collocation in French; this label is split and integrated into MLSN with the semantic relations that link its parts.

As part of the monolingual experiment, 115 descriptors have been automatically expressed in French on the basis of their Uniform Resource Identifier (URI) strings. All the terms except one were already present in the RezoJDM network. We exploited the relations typed *r_carac* (typical characteristic) for this experiment. This relation has been annotated using the URIs of the ontology properties `aPourDescripteurBruit` (*hasSoundDescriptor* is immersed as follows: *croûte r_carac :: bruit croustillant* ("croûte has typical characteristic linked to the noise *croustillant*"⁷). The premises of the

⁶Relations that can be calculated using inference.

⁷*crust* has typical characteristic linked to the sound *crusty*

mapping rules rely on the contextualization of the LSN relations. Such contextualization is possible when using sets of hypernyms and neighborhood semantic relations of the source and target terms of a relation. Meta-information attached to the LSN relations (annotations, weight) may also be used. For instance: $p\acute{e}trir\ r_object\ p\acute{a}te \wedge p\acute{e}trir\ r_isa\ technique\ de\ base \wedge p\acute{a}te\ r_isa\ pr\acute{e}paration$ ⁸.

As part of the immersion process the ontology labels become LSN terms that can be polysemous.

4.3 Projection

4.3.1 Inference in the LSN Context

In the MLSN context we set up several algorithms to discover relevant pieces of knowledge (RPK) of the types "class/individual" (ci) and "ontology property" (op). To discover the RPK(ci) we compare the neighborhood terms inside an hierarchical chain which goes up to a high level MIAM concept immersed into the LSN. For the RPK(op) we look for (real or possible) MLSN relations similar to the immersed MIAM properties. The inference scheme we use is the abduction scheme. When we have two similar terms (such as cohyponyms) the relations detained by one of them can be proposed for the other. For a term T , the abduction implies selecting a set of similar terms (according to some criteria) in order to propose the relations detained by those similar terms to T .

4.3.2 Discovering the RPK(ci)

In MIAM, the **general axioms** concern the disjunction between the MIAM classes which is the basis of the ontology consistence. To translate them in terms of MLSN, we considered the labels of the classes listed in the axioms in order to identify the criteria that could have determined the disjunction. We manually analyzed a subset of the MIAM axioms and came up with the following criteria: *affiliation*

(*r_has_part* i.e. a specific label (organic)), *transformation* (*r_carac* i.e. boiled mixture, cubed vegetable), *composition* (*r_matter* i.e. *produit à base de poisson* "fish based product"), *category based distinctions* (*r_hypo* i.e. *volaille type dinde* "turkey type poultry"). This analysis allowed selecting a subset of relation types to consider during the experiment. The RPK(ci) inference includes two steps: validation of the hierarchical chain (figure 2) and RPK(ci) candidate suggestion.

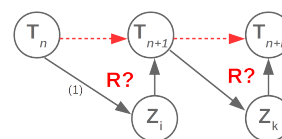


Figure 2: Hierarchy chain validation. T_i are the terms of the hierarchy chain. We check by triangulation their semantic relatedness and use a subset of relations types (such relations are noted R) for that.

We calculated and validated hierarchical chains corresponding to 1 322 top MIAM concepts pertaining to the *Aliment* module. First we obtained 132 213 chains. After filtering them by weight, the set has been reduced to 53 749 chains (40% of the initial set). Still, a certain number of redundancies may exist inside this statistically pre-filtered set since a long chain may include several shorter ones. The logical filtering by triangulation left us with a set containing 9 600 hierarchical chains (18% of the number of statistically filtered chains and 7% of the initial number of candidates).

Hierarchical chain examples after filtering:

- (1) "*baguette complète* → *pain complet* → *pain* → *ingrédient de recette de cuisine* → *aliment*"
- (2) "*angélique* → *confiserie* → *bonbon*".

RPK(ci) examples:

truffe chocolat subClassOf chocolat
pomme à cidre subClassOf pomme
sucré de pomme subClassOf confiserie

The analysis and validation of the hierarchical chains corresponds to the important memory load. The complexity of the algorithm depends on the im-

⁸ $knead\ r_object\ dough \wedge knead\ r_isa\ basic\ technique \wedge dough\ r_isa\ mixture$

portance of the MIAM concept being processed as well as on the length of the hierarchical chains that are considered. The use of a subset of the semantic relations types available in the MLSN reduces the number of combinations to process. Thus, given that the highest in-degree typed r_isa in the French sub-graph of the MLSN is 5 264 (for the term *aliment*) and that the maximum length $l = 9$, the complexity in the worst case would be $O(d_{isa}^l)$ or $O(5\,264^9) = 3,103436942 \times 10^{33}$.

The table 1 introduces the results obtained for the discovery of RPK(ci) related to the top level concept *Aliment* in the French sub-graph of the MLSN. The

#candidates	#valid	% valid	#new	% new
11 520	11 289	98%	4 741	42%

Table 1: The RPK(ci) discovery.

automatic evaluation of the proposed RPK(ci) would mostly rely on similarity measures. However, the projection step implicitly relies on relatedness and similarity between the LSN terms. Thus, in our future work, the RPK(ci) evaluation will need human expert decisions.

4.3.3 The RPK(op) Discovery

The RPK(op) discovery seems to be particularly useful in the context of multilingual ontology building or localization of an existing ontology. In our experience, each module of the MIAM ontology has its own hierarchy of properties. While immersing them into the MLSN, these properties have been expressed in terms of semantic relations contextualized using annotations. The choice of the MLSN semantic relation type made for these properties allows us to distinguish the following cases for the MIAM Object Properties (OPs): *composition based* (aPourProduitInitial, "hasInitialProduct"); *related to processes* (aPourMethodeDeConservation, "hasConservationMethod"); *temporal and spatial relation based* (aPourMoisPrimeur, "hasEar-

lyMonth"); *characteristic based* (aPourEtat, "hasState"); *OPs with a specific sub-graph*⁹ (aPourAlimentAmi, "hasFriendlyFoodItem").

The MIAM ontology we try to enrich counts 21 565 instances of *Object Properties*. Once they are immersed into the MLSN, one could consider that we have the same number of inference rule instances that can be used for the cross-lingual RPK(op) discovery. A naive approach would be setting up a cross-lingual inference mechanism. However, such approach would be error-prone due to the potential alignment and polysemy issues. In addition, as MIAM has been built according to the top-down methodology by a community of domain experts, it contains a variable number of instances per property. The naive approach would reiterate this imbalance.

To refine the RPK(op) discovery, we experimented a rule based approach. First, the validity of the rule for the source language is calculated. Second, structures similar to those specified by the rule are being discovered in the MLSN (in other languages).

The rule has the following form:

```
property=aPourEtatPhysique
(property name)
src=?s (domain, set of terms)
reltype=r_carac (relation type)
tgt=?o(range, set of terms)
source_isa=aliment(src hypernyms)
target_isa=etat physique
(tgt hypernyms)
annotation=int:physical state
(meta-information)
src_feat=OUT/r_pos/int:Noun
(in and out relations that characterize terms in the
source set)
tgt_feat=OUT/r_pos/int:Adj (in and out
relations that characterize terms in the target set)
```

If a given rule allowed detecting enough structures in the source language (at least, 2 structures), it is considered as a valid one and can generate the qualifying object. Thanks to this object, candidate struc-

⁹A subset of MLSN terms connected through semantic relations.

tures are detected in the other language sub-graphs of the MLSN. The mechanism of RPK(op) discovery reveals the following elements that allowed discovering new pieces of knowledge : **possibly annotated semantic relation** (case of the properties such as `aPresenceLactose`, `aPresenceGluten`); **specific pattern** (defined by rules); **complex structure** for properties related to processes. The results obtained using possibly annotated relations (in particular, *Data Properties*) are presented in the table 2. the potential improvement is estimated as an increase compared to the initial number of property instances (impr.%).

#DP	#MIAM	#RPK(op)	filt.	+%
aTeneurLipide	0	4 741	3 271	-
aPresenceLactose	2 593	530	408	+16%
aPresenceGluten	289	820	762	+263%

Table 2: The RPK(op) discovery on the basis of simple semantic relations.

Example of the output of the rule-based algorithm: *ru:jarkoje* `aPourProduitDiscriminant` *podlivka* ("stew hasDiscriminatingProduct sauce") and *en:stew* `aPourProduitInitial` *en:vegetable (produce)* ("stew hasInitialProduct vegetable"). Our rule-based RPK(op) experiment (given the actual state of the MLSN) yielded the results listed in the table 3.

Fully automated structure-based evaluation such as described in (Fernández et al., 2009) may be chosen to compare to other resources available on the Web such as (Dooley et al., 2018). To address the ontology accuracy, completeness, conciseness, efficiency, consistency, and other features (Raad and Cruz, 2015), a combination of methods is needed. In particular, gold standard ontology, specific tasks and corpora may be used for evaluation. A task-based evaluation such as semantic analysis (Bebeshina-Clairet, 2019), dietary conflict detection from dish titles (Clairet, 2017) have been used for the MLSN. To evaluate the output of the immersion- projection method, we need to organize our triples into a fully structured ontology. This will be one of the priorities

of our future work.

4.3.4 Towards the Automatic Suggestion of the RPKs(op)

To extend the RPK discovery experiment, we tried to automatically suggest pseudo ontology properties to be submitted to the domain and ontology human experts. We considered the ontology *SensoMIAM*¹⁰ for this experiment. This ontology is a MIAM module but we considered it as a "draft" ontology as the sensory aspect modeling is a flourishing research and development area and the *SensoMIAM* could be improved. We used the monolingual LSN (RezoJDM). The *SensoMIAM* contains sensory descriptors such as `DescripteurTact` ("TactileDescriptor") = {*astringent*, *filandreux*, ..., *nerveux*}; `DescripteurSubstance` ("SubstanceDescriptor") = {*aéré*, *dense*, ..., *épais*}

To calculate the RPK descriptors RPK(desc), we explored the semantics of the source terms of the relations typed *r_carac*. If the set of outgoing relations of such terms connects them to a food item and if they have a set of typical characteristics shared with other terms with an hypernym \approx "food", the target term of their outgoing relations typed *r_carac* that is not present in the *SensoMIAM* can be suggested as a potential RPK(desc). The relation typed *r_carac* is annotated. The process is represented on the figure 3. The experience allowed to suggest the RPK(desc) such as: `DescripteurArome`={*sucré-salé*, *miellé*, ..., *vinaigré*} or `DescripteurTact` = {*écailleux*, *spongieux*, ..., *floconneux*}.

We automatically suggested and semi-automatically validated 342 RPKs(desc). We explored the possibility of suggesting relevant RPKs to human experts. We defined 3 pseudo-properties for testing: *aPourComposantFlaveur* ("hasFlavourComponent"), *aPourComposantToucher* ("hasTouchComponent"), and *aPourComposantAspect* ("hasAspectComponent"). To populate them, we explored the

¹⁰www-limics.smbh.univ-paris13.fr/sensoMIAM/

m	prop	#in	en	fr	es	ru	#out	%aug
A	aPourProdInit.	2031	292	1208	203	2 245	3 039	+149%
A	aPourEtatPhys.	543	30	29	10	53	85	+16%
A	aPourForme	39	77	78	5	37	132	+338%
A	aPourLabel	114	15	11	3	1	29	26%
A	aPourMethodC.	115	94	101	13	156	309	+269%
A	aPourMois	116	117	221	23	28	116	+288%
A	aPourRegion	289	98	71	2	57	216	+75%
A	aPourProdCon.	98	256	302	143	103	570	+582%
A	aPourProdInitialA.	41	94	147	12	567	259	+633%
P	aPourTypeDeCuis.	23	155	124	80	285	686	+2 986%
P	aPourDomCul.	82	112	92	120	1313	1276	+1 557%
P	aPourDecoupe	82	82	78	56	77	272	+332%
S	aPourSaveur	752	51	78	47	98	232	+31%
S	aPourDescripteurBr.	119	67	80	10	6	159	+134%
S	aPourCouleur	233	192	451	59	423	911	+391%
S	aPourAspectSurf.	176	40	35	12	52	101	+58%
S	aPourSensationT.	54	84	77	21	12	155	+287%
-	Total	5388	2384	3960	937	4953	9531	+177%

Table 3: Rule-based approach. **m** -name of module (*Aliment* (A), *Preparation* (P), *Sensory* (S)), **prop** - property, **#in** - MIAM triples, **en**, **fr**, **es**, **ru** - MLSN sub-graphs. **#out** - overall number of suggested RPK(op) after filtering, **%aug** - potential improvement.

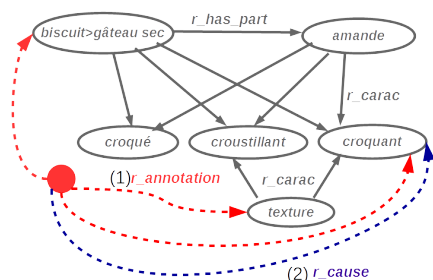


Figure 3: (1) relation annotation. (2) RPK(op) suggestion.

RezoJDM relations typed *r_has_part* and *r_matter* and considered the characteristics that can be shared by a "whole" and its "parts". We tried to generalize to the "whole" some of the characteristics of its parts. Automatically suggested toy triples: *veau Orloff* aPourComposantFlaveur *lard* from {*gras*, *viande*} "veal Prince Orloff has a component that influences its flavour *lard* because they share *fat*, *meat*"; *gratin* aPourComposantAspect *fromage* from {*gratiné*, *gras*, *brûlé*} "gratin has a component that influences its aspect *cheese* as they share characteristics *grilled*, *fat*, *burned*". We automatically suggested 1 709 RPKs(op) for the pseudo-

properties we explored. They have been automatically validated by constraining the range of the pseudo-property (it must be related to the terms "flavour", "touch", and "aspect") and by checking the sufficient intersection size between the relation sets typed *r_carac* of the "whole" and its "part".

Conclusion and perspectives

We described a method that attaches an intermediary resource containing relevant pieces of knowledge harvested from semantically structured resources in different languages to the main building process. The method suits for other domains of knowledge and the amount of work necessary for the immersion process is proportional to the size and to the type of the ontological resource to be enhanced. It is also possible to use such resources as WordNets¹¹ as the basis for the intermediary resource. Among the difficulties linked to our method appear the differences between formal representation paradigms as well as the availability of well structured and semantically rich resources.

¹¹<http://globalwordnet.org/>

References

- Mamoun Abu Helou, Mustafa Jarrar, Matteo Palmonari, and Ch Fellbaum. 2014. Towards building lexical ontology via cross-language matching. pages 346–354.
- Keith Allan. 2001. *Natural Language Semantics*. Blackwell.
- Nadia Bebeshina-Clairret. 2019. *Construction d'une ressource termino-ontologique multilingue pour les domaines de la cuisine et de la nutrition*. Theses, Université Paris 13.
- Christian Biemann. 2005. Ontology learning from text: A survey of methods. *LDV Forum* 20:75–93.
- Nadia Clairret. 2017. Dish classification using knowledge based dietary conflict detection. In *RANLP 2017*.
- Sylvie Desprès. 2016. Construction d'une ontologie modulaire. application au domaine de la cuisine numérique. *Revue d'Intelligence Artificielle* 30(5):509–532. <https://doi.org/10.3166/ria.30.509-532>.
- Zhendong Dong, Qiang Dong, and Changling Hao. 2010. [Hownet and its computation of meaning](#). In *Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations*. Association for Computational Linguistics, Stroudsburg, PA, USA, COLING '10, pages 53–56. <http://dl.acm.org/citation.cfm?id=1944284.1944298>.
- Damion Dooley, Emma Griffiths, Gurinder S. Gosal, Pier Luigi Buttigieg, Robert Hoehndorf, Matthew C. Lange, Lynn M. Schriml, Fiona S. L. Brinkman, and William W. L. Hsiao. 2018. Foodon: a harmonized food ontology to increase global food traceability, quality control and data integration. *npj Science of Food* 2. <https://doi.org/10.1038/s41538-018-0032-6>.
- Miriam Fernández, Chwhynny Overbeeke, Marta Sabou, and Enrico Motta. 2009. What makes a good ontology? a case-study in fine-grained knowledge reuse. In Asunción Gómez-Pérez, Yong Yu, and Ying Ding, editors, *The Semantic Web*. Springer Berlin Heidelberg, Berlin, Heidelberg, pages 61–75.
- E. Gaillard, J. Lieber, and E. Nauer. 2015. [Improving ingredient substitution using formal concept analysis and adaptation of ingredient quantities with mixed linear optimization](#). In *Computer Cooking Contest Workshop*. Frankfurt, Germany. <https://hal.inria.fr/hal-01240383>.
- J.U. Kietz, A. Maedche, and R. Volz. 2000. A method for semi-automatic ontology acquisition from a corporate intranet. *EKAW-2000 Workshop Ontologies and Text, Juan-les-Pins, France, October 2000*.
- Mathieu Lafourcade. 2007. [Making people play for Lexical Acquisition with the JeuxDeMots prototype](#). In *SNLP'07: 7th International Symposium on Natural Language Processing*. Pattaya, Chonburi, Thailand, page 7. <https://hal-lirmm.ccsd.cnrs.fr/lirmm-00200883>.
- Mathieu Lafourcade. 2011. *Lexique et analyse sémantique de textes - structures, acquisitions, calculs, et jeux de mots. (Lexicon and semantic analysis of texts - structures, acquisition, computation and games with words)*. <https://tel.archives-ouvertes.fr/tel-00649851>.
- Ora Lassila and Deborah McGuinness. 2001. The role of frame-based representation on the semantic web. Technical report, Knowledge Systems Laboratory Report KSL-01-02, Stanford University, Stanford (USA).
- Jacek Marciniak. 2013. Building wordnet based ontologies with expert knowledge. In *LTC*.
- Thibault Mondary. 2011. *Construction d'ontologies à partir de textes. L'apport de l'analyse de concepts formels.* Theses, Université Paris-Nord - Paris XIII. Equipe RCLN. <https://tel.archives-ouvertes.fr/tel-00596825>.
- Joe Raad and Christophe Cruz. 2015. [A Survey on Ontology Evaluation Methods](#). In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development, part of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*. Lisbonne, Portugal. <https://doi.org/10.5220/0005591001790186>.
- Lionel Ramadier. 2016. *Indexation and learning of terms and relations from reports of radiology*. Theses, Université de Montpellier. <https://hal-lirmm.ccsd.cnrs.fr/tel-01479769>.
- Christophe Roche. 2007. [Le terme et le concept : fondements d'une ontoterminologie](#). In *TOTh 2007 : Terminologie et Ontologie : Théories et Applications*. Nancy, France, pages 1–22. 22 pages. <https://hal.archives-ouvertes.fr/hal-00202645>.
- Gilles Sérasset. 2014. [DBnary: Wiktionary as a Lemon-Based Multilingual Lexical Resource in RDF](#). *Semantic Web – Interoperability, Usability, Applicability* pages –. To appear. <https://hal.archives-ouvertes.fr/hal-00953638>.
- Robyn Speer and Catherine Havasi. 2012. Representing general relational knowledge in ConceptNet 5. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*. European Languages Resources Association (ELRA), Istanbul, Turkey, pages 3679–3686.
- Sylvie Szulman. 2012. [Logiciel Terminae - Version 2012](#). TERMINAE est une plateforme d'aide à la construction de ressources termino-ontologiques à partir de ressources textuelles. <https://hal.archives-ouvertes.fr/hal-00719453>.
- Andon Tchechmedjiev. 2016. *Semantic Interoperability of Multilingual Lexical Resources in Lexical Linked Data*. Theses, Université Grenoble Alpes. <https://tel.archives-ouvertes.fr/tel-01681358>.

Davide Turcato, Fred Popowich, Janine Toole, Dan Fass, Devlan Nicholson, and Gordon Tisher. 2000. [Adapting a synonym database to specific domains](#). In *ACL-2000 Workshop on Recent Advances in Natural Language Processing and Information Retrieval*. Association for Computational Linguistics, Hong Kong, China, pages 1–11. <https://doi.org/10.3115/1117755.1117757>.

Piek Vossen. 2012. [Ontologies](#). *The Oxford Handbook of Computational Linguistics* <https://doi.org/10.1093/oxfordhb/9780199276349.013.0025>.

Wilson Wong. 2009. *Learning lightweight ontologies from text across different domains using the web as background knowledge*. Ph.D. thesis.

Manel Zarrouk. 2015. *Endogeneous Consolidation of Lexical Semantic Networks*. Theses, Université de Montpellier. <https://hal-lirmm.ccsd.cnrs.fr/tel-01300285>.