Réduire l'effort humain d'amélioration des ressources lexicales grâce aux inférences

Nadia Bebeshina-Clairet, Mathieu Lafourcade

▶ **To cite this version:**

Nadia Bebeshina-Clairet, Mathieu Lafourcade. Réduire l'effort humain d'amélioration des ressources lexicales grâce aux inférences. 27e édition du Traitement Automatique des Langues Naturelles (TALN), Jun 2020, Nancy, France. pp.146-154. hal-02784762v3

HAL Id: hal-02784762
https://hal.science/hal-02784762v3

Submitted on 23 Jun 2020

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 4.0 International License

Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition),
Traitement Automatique des Langues Naturelles (TALN, 27e édition),
Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition)
Nancy, France, 08-19 juin 2020

Réduire l'effort humain d'amélioration des ressources lexicales grâce aux inférences

Nadia Bebeshina[1, 2] Mathieu Lafourcade[1]
(1) LIRMM, 860 rue de St Priest, 34095 Montpellier, France
(2) Praxiling, Université Paul Valéry, Route de Mende 34199, Montpellier, France
nadia.clairet@gmail.fr, mathieu.lafourcade@lirmm.fr

RÉSUMÉ

Les inférences translingues représentent une piste intéressante pour la construction des ressources lexico-sémantiques multilingues. Cependant, la validation des éléments candidats nécessite un effort humain considérable. Nous décrivons une façon de construire des ressources lexico-sémantiques via des inférences monolingue et translingue. Son intérêt principal consiste à implémenter dans le contexte d'une ressource lexico-sémantique multilingue une approche où le processus de construction est un processus auto-apprenant car l'évaluation participe à la construction de celle-ci.

ABSTRACT

Reducing the Knowledge Resource Enhancement Human Effort through Inferences.

The inference based knowledge resource enhancement mechanism allows building lexical semantic resource incrementally and, presumably, minimises the human effort necessary for the resource building. Unfortunately, the knowledge parts obtained by inference have to be semi-automatically double checked to avoid the propagation of errors. The overall human effort in the inference process may be very important. In the present paper, we describe a resource building pipeline based on monolingual inference and cross-lingual inference. Its main contribution is proposing a method where the resource building process appears as a self learning process.

MOTS-CLÉS : inférence, ressource multilingue, relations sémantiques, validation.

KEYWORDS: inference, multilingual knowledge resource, semantic relations, validation.

1 Introduction

Les ressources lexico-sémantiques sous forme de graphe sont fréquemment utilisées pour l'analyse sémantique et la désambguïsation. Dans le contexte de construction de ces ressources, inférer de nouveaux éléments à partir des éléments existants dans la ressource semble permettre de réduire l'effort humain nécessaire à celle-ci. Cependant, les éléments candidats doivent être validés de façon supervisée afin d'éviter la propagation des erreurs. Ainsi, la part de l'effort humain nécessaire à la mise en place et au fonctionnement des méthodes à base d'inférence endogène peut être très importante (élaboration des règles d'inférence, validation, mise à jour des éléments existants). Après avoir introduit l'état de l'art des techniques de construction des ressources lexico-sémantiques, nous détaillons les ressources qui ont servi pour l'expérience. Puis, nous décrivons notre méthode d'inférence et d'évaluation par inférence des éléments candidats et ses résultats.

Volume 2 : Traitement Automatique des Langues Naturelles, pages 146–154.
HAL : hal-02784762.
Cette œuvre est mise à disposition sous licence Attribution 4.0 International.

2 État de l'art

Construire ou améliorer le ressources de connaissance multilingues peut s'appuyer sur un mécanisme d'inférence qui permet d'exploiter les différentes partitions monolingues de ces ressources et les liens entre ces partitions. Ainsi, dans le cadre des bases de connaissance fondées sur les entités et les faits sur les entités telles que NELL (Carlson *et al.*, 2010), plusieurs auteurs ont proposé d'établir les équivalences entre les entités et les relations des différentes partitions. Les auteurs dans (Hernández-González *et al.*, 2017) ont proposé de fusionner le différentes éditions monolingues de NELL. (Nickel *et al.*, 2015) détaillent l'apprentissage statistique dans le contexte multilingue en mettant l'accent sur la transitivité et les contraintes de type de relation. De même que les auteurs dans (Wang *et al.*, 2015), ils fondent leur méthode sur des bases à large couverture telles que NELL, KnowItAll (Etzioni *et al.*, 2005), YAGO (Suchanek *et al.*, 2007) ou DeepDive (Shin *et al.*, 2015).

Des travaux ont également été menés pour enrichir les modèles des ressources telles que Princeton WordNet (Fellbaum, 1998) et construire des ressources ayant une expressivité comparable dans d'autres langues, plus pauvres en ressources telles que le basque (Agirre *et al.*, 2002). Ces auteurs mettent l'accent sur l'importance de l'alignement concept à concept et mot à mot. Le projet BabelNet (Navigli & Ponzetto, 2012) a été la première expérience à grande échelle de construction des ressources non supervisée et intégrant plusieurs ressources pré-existantes et construites manuellement. Les processus d'inférence endogènes à base de règles a été étudié notamment par (Zarrouk, 2015) et (Ramadier, 2016) dans le cadre des travaux sur la construction endogène du réseau lexico-sémantique pour le français RezoJDM. Leurs méthodes reposent sur les relations sémantiques et annotations (méta-informations) de ces relations déjà présentes dans cette ressource pour proposer de nouveaux éléments grâce aux inférences déductives, inductives, abductives (exploitant la similarité sémantique) et fondées sur les raffinements de sens. (Gelbukh, 2018) a introduit un mécanisme d'inférence comparable pour enrichir une base de connaissances collocationnelle et proposer de nouvelles collocations par abduction à partir de celles déjà présentes dans la base. Cette méthode utilise WordNet pour calculer la similarité sémantique.

3 Ressources

RezoJDM (Lafourcade, 2007) est un réseau lexico-sémantique du français construit par peuplonomie dont, en particulier, les jeux avec un but tels que JeuxDeMots [1] ainsi que d'autres jeux [2]. Il s'agit d'un graphe orienté, typé et pondéré. Au moment où nous écrivons RezoJDM contient 4,2 millions de termes (noeuds) et 310 millions de relations (arcs). Le modèle de RezoJDM offre trois caractéristiques importantes pour notre expérience. Premièrement, il contient des relations à poids négatif (relations sémantiquement fausses) ce qui rend possible la détection des relations fausses par inférence. Deuxièmement, les termes de ce réseau peuvent être raffinés afin de représenter les distinctions de sens. Troisièmement, les relations peuvent être annotées et comporter plusieurs annotations. Nous nous référerons à RezoJDM comme *ressource de référence*.
RLSM (Bebeshina-Clairet, 2019) est un réseau lexico-sémantique multilingue (français, anglais, russe, espagnol) avec un pivot interlingue. Au moment où nous écrivons RLSM contient 821 781 noeuds (termes) and 2 231 197 arcs (relations). Il a été construit pour les domaines de la cuisine et

1. http://www.jeuxdemots.org
2. http://imaginat.name/JDM/Page_Liens_JDMv2.html

de la nutrition, mais contient également de la connaissance générale conformément à l'hypothèse sur la non-séparation de la connaissance générale et de spécialité vérifiée par (Ramadier, 2016). Inter-opérable avec RezoJDM (types de relations, raffinements de sens), le RLSM est un graphe orienté, typé et valué. Il contient k sous-graphes correspondant à chacune des k langues du réseau ainsi qu'un sous-graphe spécifique, le pivot interlingue. *Valué* se réfère au fait que le RLSM n'a pas été construit par peuplonomie et a nécessité d'autres moyens de valuation complémentaires à la pondération. L'ajout des méta-informations et l'attribution des scores de confiance ont été exploités pour permettre de renforcer ou non la relation entre les termes.

4 Expérience

4.1 Observations générales

L'expérience s'est appuyée sur les observations suivantes : (1) dans certaines langues les informations sémantiques peuvent être découvertes à partir des traits morpho-syntaxiques tandis que dans d'autres langues la même information est plus difficile à obtenir ; (2) dans le contexte des ressources lexico-sémantiques multilingues, une partition lexicalisée (sous-graphe d'une langue) peut fournir l'information sémantique manquante à ses autres partitions ; (3) dans le cas de deux ressources lexico-sémantiques interopérables, les composants (par exemple, termes, relations sémantiques) d'une ressource peuvent être validés en s'appuyant sur le contenu réel ou inférable de l'autre ressource. Compte tenu de ces observations, nous supposons que les relations sémantiques extraites depuis le corpus de langue riche en traits sémantico-syntaxiques (comme la langue russe) peuvent être exploitées pour enrichir une ressource sémantique dans une autre langue. La condition serait alors l'existence d'un lien (traduction, pivot) entre la ressource à évaluer et la ressource de référence. L'interopérabilité du RLSM avec la ressource plus riche et plus stable RezoJDM a permis la mise en place de la procédure d'évaluation automatique. L'exploitation des résultats de la procédure d'évaluation automatique rapproche l'expérience du processus auto-apprenant. L'expérience comprend trois étapes : (1) **identification et extraction des relations sémantiques** à partir du corpus de textes en russe pré-étiqueté en partie de discours en utilisant un ensemble de règles ; (2) **inférence translingue** des relations extraites et création de nouvelles relations dans le sous-graphe français du RLSM à partir des relations intégrées dans le sous-graphe russe suite à l'extraction ; (3) **validation translingue des relations inférées** grâce à la ressource de référence RezoJDM.

4.2 Extraction

Corpus et portée de l'extraction. L'extraction a été effectuée à partir d'un corpus des instructions de cuisine (2 473 654 mots) collecté sur le Web [3]. Ce corpus a été étiqueté en utilisant le parseur *Russian Malt parser* (Sharoff & Nivre, 2012). L'extraction a concerné les types de relations sémantiques suivants : (a) *caractéristique* avec quelques distinctions entre les caractéristiques de composition, et des caractéristiques qualitatives pouvant être observées à travers les traits morpho-syntaxiques en russe ; (b) *manière* avec distinctions observées entre les expressions adverbiales liées à l'instrument (décorer *avec des fruits tranchés*), manière partie-tout (*couper en cubes*) et d'autres réalisations ; (c) *lieu* avec distinction entre les lieux de type « surface plane » et les lieux de type « contenant ».

3. Principalement, à partir du portail https://www.gotovim.ru/

Règles. Les règles pour l'extraction des caractéristiques reposent sur une série de traits propres aux adjectifs en russe. Comme remarqué par (Corbett, 2004), les adjectifs du russe couvrent les aspects typologiques inhabituels. Un sous-ensemble de règles (*carac-r1*) correspond aux adjectifs canoniques qui donnent des caractéristiques de dimension, âge, valeur, et couleur. Un autre sous-ensemble de règles (*carac-r2*) sur les modèles de dérivation saillants pour les adjectifs du russe tels que décrits par (Zemskaja *et al.*, 1981). Plus spécifiquement, nous avons exploré les adjectifs dérivés des noms et verbes. La dérivation à partir de noms peut révéler la relation partie-tout (composant ou membre) tandis que la dérivation à partir des verbes peut révéler les changements d'état. Il est à remarquer que le rôle des adjectifs dans la représentation des caractéristiques typiques nécessaires catégorisation ont fait objet de nombreux travaux, notamment dans le cadre de la terminologie fondée sur les cadres, *Frame-Based Terminology* (Faber & Cabezas-García, 2019). Dans (Altmanova *et al.*, 2018), les adjectifs renfermant des relations sémantiques sont utilisés pour extraire d'autres adjectifs porteurs des mêmes types de relation (fonction, composition, localisation, cause et forme).

Sur la base des appariements entre les structures de surface (lexicalisation du mouvement) et les composants du sens nécessaires pour représenter le mouvement (objet mouvant, point de référence, chemin) la typologie de (Talmy, 1985) distingue deux groupes essentiels de langues : les langues S (langues qui exploitent des mots-satellites)[4] et les langues V où les verbes de mouvement sont directionnels et l'indication de manière est optionnelle comme détaillé dans (Strömqvist & Verhoeven, 2004). Dans notre expérience il s'agit de l'acquisition des relations de manière et trajectoire à partir d'une *langue S* pour les transférer vers la *langue V* sous forme de relations sémantiques explicites.

Les prémisses des règles sont des traits syntaxiques, grammaticaux et morphologiques tels que les cas des noms, « satellites » (préfixes, suffixes, mots-outils). La conclusion est la création d'une relation candidate dans le sous-graphe russe du RLSM.

Exemples des règles :
(*)`if X Y` (*contexte*) `and X Verb` (*prémisse 1*) `and Y Nom:CaseAccusative` (*prémisse 2*)
`X r_object Y` (*conclusion*)
(**)`if X` "na" `Y and X Verb and Y Noun:CaseLocative`
`X r_has_part::essential component Y` (*conclusion - une relation annotée*)

Pour donner un exemple d'une règle fondée sur les traits morphologiques, nous pouvons citer le cas des adjectifs avec une suffixe `[-yann]` qui permettent d'extraire les informations sur la matière (bois, verre) qui constitue l'objet.

L'ensemble des règles comporte moins de 30 règles pour les types de relations abordés par l'expérience. L'étape d'extraction reste dépendante de la sortie du parseur.

Les types d'information sémantique extraits sont présentés dans le tableau 1. Une mesure de confiance est attribuée à chaque instance de règle. Initialement, le score c est fixé à $c = 1$. Au fur et à mesure que les relations inférées sont évaluées grâce à la ressource de référence, ce score évolue et contribue à refléter la fiabilité de la règle *pour la langue cible*. Il est affecté par le nombre de relations justes inférées grâce à cette règle, le nombre de sens des termes polysémiques (dans le cas d'une relation source dont l'inférence ascendante produit n relations interlingues, la confiance pour chacune de ces relations est de $c_{int} = \frac{c_s}{n}$ *etc.*) et le statut suite à l'évaluation.

4. Finno–ougriennes, germaniques, sino–tibétaines et slaves.

type-règle	**#ext**	**%ext**
manner-r1 (adverbes)	8 229	25,4%
manner-r2 (instrument, forme : *en cubes*, *avec les mains*)	1 879	5,8%
manner-r3 (avec peu d'intensité)	295	0,9%
manner-r4 (action récurrente)	154	0,4%
carac-r1	12 488	38%
carac-r2 (dérivation nominale, *partie-tout*)	543	1,6%
carac-r3 (dérivation verbale, *état*)	756	2%
place-r1	5 679	17,5%
place-r2 (lieu plan ou contenant)	2 298	7%
Total	32 321	100%

TABLE 1 – Détails concernant les règles d'extraction et le nombre des instances obtenues **#ext**.

4.3 Inférence et évaluation

Pivot « interlingue ». Les schémas d'inférence dans le RLSM sont fondées sur l'utilisation du pivot interlingue. Ce pivot a été amorcé en tant que pivot naturel à partir des liens de traduction présents dans DBnary (Sérasset, 2014) (édition anglaise et russe) et évolue actuellement vers le pivot interlingue. La ressource DBNary est orientée vers l'alignement par sens (Tchechmedjiev, 2016). Le pivot a été enrichi grâce à des processus d'acquisition exogènes (à partir des ressources monolingues et multilingues pré-existantes) ainsi que d'extractions à partir des textes de spécialité (Bebeshina-Clairet, 2019). Le pivot peut être vu comme l'union de tous les sens disponibles dans le RLSM. Les termes interlingues qui appartiennent au pivot sont reliés aux termes lexicalisés grâce à la relation typée *r_covers*. Un terme lexicalisé peut avoir plusieurs termes couvrants. Un lien translingue correspond à un chemin qui traverse le pivot, processus également détaillé dans (Bebeshina-Clairet & Lafourcade, 2019). Ainsi, les relations extraites à partir des textes en russe sont créées dans le sous-graphe du russe. Lorsque le terme source t_s et le terme cible t_c d'une relation r_type_{ru} ainsi créée sont couverts par le pivot interlingue, la relation r_type_{inter} est inférée dans le pivot et relie le terme couvrant du t_s à celui du t_c. il est ensuite possible de produire des relations candidates en français afin de permettre la validation.

Inférence. Le processus d'inférence actuel détaillé dans (Bebeshina-Clairet & Lafourcade, 2019) combine la phase ascendante (langue source → pivot) et la phase descendante (pivot → langue cible). Pendant la phase ascendante, les relations sémantiques du sous-graphe de la langue source sont inférées dans le pivot grâce à un ensemble de contraintes qui dépendent du type de relation à inférer. Pour former une relation typée « caractéristique », il est nécessaire que le terme source soit un nom tandis que le terme-cible soit un adjectif. Vérifier si les termes sont sémantiquement liés (triangulation) ainsi que le calcul de similarité sémantique sont également utilisés. Pendant la phase descendante, le processus d'inférence est complété par le processus d'évaluation fondé sur la recherche et l'inférence dans la ressource de référence (RezoJDM).

Outil d'évaluation. Premièrement, les relations inférées sont ajoutées dans le sous-graphe français du RLSM. Deuxièmement, l'évaluation est effectuée grâce à l'outil Helix [5] destiné à vérifier la

5. http://www.jeuxdemots.org/rezo-ask.php

présence des relations et à compléter RezoJDM. Troisièmement, le retour de l'évaluation de la relation candidate affecte le poids et le statut des relations ayant contribué à son inférence (relations des sous-graphes russe, interlingue et français).

Les résultats d'inférence ascendante reflètent la couverture de la langue source par le pivot. L'inférence descendante montre la couverture de la langue cible par le pivot en ce qui concerne un type de relation particulier. Grâce à l'évaluation, cette phase descendante permet l'acquisition des relations sémantiquement fausses qui sont maintenues dans le RLSM afin de prévenir leur acquisition ultérieure par d'autres moyens (processus exogènes, inférences). Les résultats d'inférences en deux phases sont listés dans la table 2a. Les processus d'inférence génèrent de nombreuses relations candidates à cause de la présence des raffinements de sens dans le RLSM, (en particulier, les distinctions de sens acquises à partir de WordNet (Fellbaum, 1998), RezoJDM, ConceptNet (Speer & Havasi, 2012) qui sont reliées à des termes polysémiques mais non raffinés (notamment, dans le sous graphe russe). Lorsque les relations candidates sont automatiquement validées, les termes en langue source sont désambiguïsés car un poids négatif est attribué aux relations qui concernent le sens erroné $t > s_1$ du terme t à un terme p lorsqu'un autre sens de t, $t > s_2$, a une relation évaluée comme *vrai* avec p.

type	**#ext**	**#asc**	**#desc**
carac-r1	12 488	22 631	40 762
carac-r2	543	3 371	2 983
carac-r3	756	1 318	440
manner-1	8 229	6 000	8 929
manner-2	1 879	1 413	1 655
manner-3	295	283	192
manner-4	154	177	169
place-r1	5 679	2 548	2 969
place-r2	2 298	5 551	6 267
Total	32 321	43 292	66 366

(a) Nombre des relations extraites (**#ext**), inférences ascendante (**#asc**) et descendante (**#desc**).

type	**#inf**$_d$	**#true**$_r$	**#true**$_{inf}$	**#und**
carac-r1	40 762	815	2 454	37 493
carac-r2	2 983	64	283	2 636
carac-r3	440	44	264	132
manner-r1	8 929	35	2 419	6 475
manner-r2	1 655	17	83	1 555
manner-r3	192	13	27	152
manner-r4	169	6	13	150
place-r1	2 969	74	453	2 442
place-r2	6 267	49	3 181	2 777
Total	66 366	1 117	9 177	53 812

(b) Évaluation des relations inférées. Par souci de simplicité, seulement les relations acceptées par recherche (**#true**$_r$) ou inférence (**#true**$_{inf}$) et les relations « indécidables » (**#und**) soit « ne sait pas », « terme absent », « faux par inférence » sont listées.

TABLE 2 – Inférence et évaluation des relations

Évaluation et ajustement. Le processus d'évaluation est un processus fondé sur l'inférence monolingue dans le réseau lexico-sémantique RezoJDM destiné à améliorer automatiquement le RLSM et le processus d'extraction des relations sémantiques. L'évaluation avec l'outil Helix retourne les valeurs suivantes pour les relations testées : (1) « vrai » (la relation est présente dans RezoJDM) ; (2) « vrai par inférence » (la relation est inférable dans le RezoJDM) ; (3) « ne sait pas » (la relation est absente et le processus d'inférence ne parvient pas à valider la relation, cette réponse est quasi-équivalente à « faux ») ; (4) « faux » (la relation a un poids négatif) ; (5) « faux par inférence » (la relation est inférable et fausse dans RezoJDM) ; (6) « terme inconnu » (terme source ou cible de la relation à tester est absent de RezoJDM). L'expérience s'intéresse principalement aux relations validées par recherche, par inférence et aux relations invalidées (classification à 3 termes). Une analyse ultérieure plus approfondie des relations candidates invalidées se situera dans un cadre plus large qui concerne la ressource de référence.

Les résultats d'évaluation montrent le rôle significatif d'inférence monolingue. Les relations sémantiques qui exploitent les propriétés telles que *plan*, *contenant* (place-r2), action faite « un par un » (manner-r4) peuvent ne pas être explicitement présentes dans la ressource de référence mais elles sont inférables. Le pourcentage des relations validées automatiquement comme « vraies » correspond au pourcentage généralement observé dans le contexte d'évaluation humaine : **#true**$_{inf}$ indiqué dans la table 2b est d'environ 5% à 10% des relations inférées par inférence translingue ce qui est similaire aux résultats de validation manuelle dans (Zarrouk, 2015) et (Bebeshina-Clairet, 2019).

5 Discussion

L'expérience a été menée sur la base des réseaux lexico-sémantiques et s'est appuyée sur des processus d'inférence conçus pour ce type de ressource. Est-elle facile à reproduire ? Pour une paire de langues donnée, la réponse à cette question dépend de la disponibilité de la ressource de référence (RezoJDM dans le cadre de l'expérience) dans une des langues. La ressource de référence est celle où la proportion des termes polysémiques désambiguïsés (raffinement de sens), la couverture et le nombre de relations sémantiques sont importants. Cette ressource doit également disposer d'un nombre suffisant de relations typées entre les termes et les raffinements pour permettre des inférences. Une des affirmations défendues à travers l'expérience présentée est que, pour un réseau lexico-sémantique multilingue, il suffit de disposer d'une ressource de référence dans une seule langue pour améliorer et raffiner l'ensemble de la ressource de façon faiblement supervisée. Cette expérience est ainsi un encouragement à la conception de ressources raisonnablement interopérables avec une ressource de référence afin de pouvoir de réduire l'effort humain de leur construction.

Un intérêt primordial des réseaux lexico-sémantiques pour le TAL est de contenir des termes et des relations entre ces termes représentés de façon explicite aussi bien pour un usage humain que pour un usage machine. La ressource que nous avons exploitée en tant que référence est le RLS le plus important construit pour le français. L'état actuel du RezoJDM permet d'apprécier son modèle polyvalent. Par ailleurs, l'expérience de construction d'un RLSM avec un pivot interlingue a démontré que, dans le contexte d'un RLS multilingue, les différences de granularité de sens (le terme anglais « stew » correspond à la fois à *ragoût* et *pot-au-feu* en français) peuvent aisément être représentés.

6 Conclusion

L'impact d'extraction à base de critères morpho-syntaxiques peut être limité car de nombreux traits sont « polysémiques » (correspondent à plusieurs relations sémantiques potentielles). Les inférences translingues et monolingues apparaissent comme un moyen de résoudre ces ambiguïtés et réduire l'effort humain nécessaire à la construction et l'amélioration des ressources de façon significative. Nos expériences montrent également qu'il suffit de disposer de la ressource de référence (pouvant supporter les processus d'inférence) dans une seule langue couverte par la ressource multilingue pour construire et améliorer les partitions des autres langues de façon moins supervisée. L'expérience d'évaluation a montré que le recours à une base de connaissance importante semble être le bon chemin vers la construction non supervisée des ressources lexico-sémantiques multilingues.

Références

AGIRRE E., ANSA O., ARREGI X., ARRIOLA J. M., DE ILARRAZA A. D., POCIELLO E. & URIA L. (2002). Methodological issues in the building of the Basque WordNet : quantitative and qualitative analysis. In *Proceedings of First International WordNet Conference*, p. 32–40.

ALTMANOVA J., GRIMALDI C. & ZOLLO S. (2018). Le rôle de l'adjectif dans la catégorisation des déchets. *SHS Web of Conferences*, **46**, 05004. DOI : 10.1051/shsconf/20184605004.

BEBESHINA-CLAIRET N. (2019). *Construction d'une ressource termino-ontologique multilingue pour les domaines de la cuisine et de la nutrition*. Theses, Université Paris 13.

BEBESHINA-CLAIRET N. & LAFOURCADE M. (2019). Inférence des relations sémantiques dans un réseau lexico-sémantique multilingue. In *TALN : Traitement Automatique des Langues Naturelles*, volume PFIA, Toulouse, France. HAL : hal-02269113.

CARLSON A., BETTERIDGE J., KISIEL B., SETTLES B., JR. E. R. H. & MITCHELL T. M. (2010). Toward an architecture for never-ending language learning. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, USA, July 11-15, 2010*, p. 1306–1313.

CORBETT G. G. (2004). *The Russian Adjective : A Pervasive Yet Elusive Category*, In *Adjective classes : a cross-linguistic typology*, p. 199–222. Oxford University Press.

ETZIONI O., CAFARELLA M., DOWNEY D., POPESCU A.-M., SHAKED T., SODERLAND S., WELD D. S. & YATES A. (2005). Unsupervised named-entity extraction from the Web : An experimental study. *Artificial Intelligence*, **165**(1), 91 – 134. DOI : http ://dx.doi.org/10.1016/j.artint.2005.03.001.

FABER P. & CABEZAS-GARCÍA M. (2019). Specialized knowledge representation : From terms to frames. *Research in Language*, **17**, 197–211. DOI : 10.2478/rela-2019-0012.

FELLBAUM C. (1998). *WordNet An Electronic Lexical Database*. Cambridge, MA ; London : The MIT Press.

GELBUKH A. F. (2018). Inferences for enrichment of collocation databases by means of semantic relations. *Computación y Sistemas*, **22**(1), 103–117. DOI : http ://dx.doi.org/10.13053/cys-22-1-2923.

HERNÁNDEZ-GONZÁLEZ J., HRUSCHKA JR. E. R. & MITCHELL T. M. (2017). Merging knowledge bases in different languages. In *Proceedings of TextGraphs-11 : the Workshop on Graph-based Methods for Natural Language Processing*, p. 21–29, Vancouver, Canada : Association for Computational Linguistics. DOI : 10.18653/v1/W17-2403.

LAFOURCADE M. (2007). Making people play for Lexical Acquisition with the JeuxDeMots prototype. In *SNLP'07 : 7th International Symposium on Natural Language Processing*, Pattaya, Chonburi, Thailand. HAL : lirmm-00200883.

NAVIGLI R. & PONZETTO S. P. (2012). BabelNet : The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, **193**, 217–250.

NICKEL M., MURPHY K., TRESP V. & GABRILOVICH E. (2015). A review of relational machine learning for knowledge graphs : From multi-relational link prediction to automated knowledge graph construction. In *CBMM Memos*. CBBM.

RAMADIER L. (2016). *Indexation and learning of terms and relations from reports of radiology*. Theses, Université de Montpellier. HAL : tel-01479769.

SÉRASSET G. (2014). DBnary : Wiktionary as a Lemon-Based Multilingual Lexical Resource in RDF. *Semantic Web – Interoperability, Usability, Applicability*, p. 355–361. DOI : 10.3233/SW-140147.

SHAROFF S. & NIVRE J. (2012). The proper place of men and machines in language technology. processing russian without any linguistic knowledge. In *Dialogue 2011, Russian Conference on Computational Linguistics*, p. 657–670.

SHIN J., WU S., WANG F., DE SA C., ZHANG C. & RÉ C. (2015). Incremental knowledge base construction using deepdive. *Proc. VLDB Endow.*, **8**(11), 1310–1321. DOI : 10.14778/2809974.2809991.

SPEER R. & HAVASI C. (2012). Representing general relational knowledge in ConceptNet 5. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, p. 3679–3686, Istanbul, Turkey : European Languages Resources Association (ELRA).

STRÖMQVIST S. & VERHOEVEN L. (2004). *Relating events in narrative : Typological and contextual perspectives*. Mahwah, NJ : Lawrence Erlbaum.

SUCHANEK F., KASNECI G. & WEIKUM G. (2007). Yago : A Core of Semantic Knowledge. Unifying WordNet and Wikipedia. In *Proceedings of the 16th International Conference on World Wide Web*, p. 697 – 697, Banff, Canada. DOI : 10.1145/1242572.1242667.

TALMY L. (1985). *Lexicalization patterns : Semantic structure in lexical forms.*, In *Language typology and syntactic description.*, volume 3, p. 57–149. Cambridge : Cambridge University Press.

TCHECHMEDJIEV A. (2016). *Semantic Interoperability of Multilingual Lexical Resources in Lexical Linked Data*. Theses, Université Grenoble Alpes. HAL : tel-01681358.

WANG Q., WANG B. & GUO L. (2015). Knowledge base completion using embeddings and rules. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, p. 1859–1865 : AAAI Press.

ZARROUK M. (2015). *Consolidation endogène de réseaux lexico-sémantiques : Inférence et annotation de relations, règles d'inférence et langage dédié*. Theses, Université de Montpellier. HAL : tel-01300285.

ZEMSKAJA E., KITAJGORODSKAJA M. & SIRJAEV E. (1981). *Russkaja razgovornaja rec' : Obcie voprosy, slovoobrazovanie, sintaksis*. Moskva : Nauka.