Vers une ontologie de la nomination et de la référence dédiée à l'annotation des textes

Agata Jackiewicz, Nadia Bebeshina-Clairet, Manon Cassier, Francesca Frontini, Anais Anais Lefeuvre-Halftermeyer, Julien Longhi, Giancarlo Luxardo, Damien Nouvel

Vers une ontologie de la nomination et de la référence dédiée à l'annotation des textes

Agata Jackiewicz*, Nadia Bebeshina*, Manon Cassier*** **** Francesca Frontini*, Anais Halftermeyer**, Julien Longhi***, Giancarlo Luxardo*, Damien Nouvel****

*PRAXILING, Route de Mende 34199 Montpellier cedex 5
prénom.nom@univ-montp3.fr
http://www.praxiling.fr
**Laboratoire d'Informatique Fondamentale et Appliquée de Tours (LIFAT), 64, Avenue Jean Portalis 37200 TOURS
prénom.nom@univ-tours.fr
https ://lifat.univ-tours.fr/
***Laboratoire AGORA, 33 Boulevard du port 95011 Cergy-Pontoise
prénom.nom@u-cergy.fr
https ://www.u-cergy.fr/fr/laboratoires/agora.html
****ERTIM-InaLCO, 3 bis rue Taylor 75010 Paris
prénom.nom@inalco.fr
http://www.er-tim.fr

Résumé. Le présent article introduit un thesaurus enrichi relatif aux phénomènes de la nomination et de la référence, construit pour la linguistique, l'analyse de discours et le TAL. Nous détaillons les étapes et les méthodes employées lors de son élaboration, la ressource de type « folksonomie » adossée, ainsi que les expériences d'intégration partielle des connaissances à partir de ressources de connaissance pré-existantes, afin de réduire l'effort humain nécessaire à la construction du thésaurus.

1. Introduction

L'étude de la construction et de la stabilisation du sens en discours, au cœur des recherches en analyse de discours (AD), s'avère également pertinente pour de nombreuses applications en traitement automatique des langues (TAL) et ingénierie linguistique (veille sociale, analyse d'opinion, recherche d'information…). Comme l'ont remarqué (Siblot 1997, 2001), (Frath, 2015),

(Calabrese et Mistaen 2016) et plus récemment (Jackiewicz et Pengam 2018), la notion de nomination, dynamique et contextuelle, permet de renseigner non seulement sur le sens actualisé en discours, mais aussi sur la prise de position des locuteurs sur l'entité nommée. Ainsi, *musulmans modérés, appropriation culturelle*, *réfugiés climatiques*… sont des exemples d'expressions linguistiques relativement instables, rattachées à des questions socialement vives, dont la signification semble se négocier essentiellement en discours.

Considérés dans le cadre discursif, l'acte de nommer et l'usage des nominations se manifestent par un ensemble de traces pouvant être observées dans les corpus. Cependant, le référentiel terminologique relatif à la nomination dont le rôle est de déterminer et de conceptualiser la nature de ces traces semble nécessiter une systématisation et une harmonisation. Les ressources existantes (imprimées ou numériques) souffrent, selon les cas, d'incomplétude, d'obsolescence, de redondance terminologique, de généralité des définitions ou de leur absence (pour des termes dits « orphelins »[1]), ce qui rend difficile leur utilisation pour caractériser les phénomènes repérés dans les textes[2].

Dans le présent article, nous introduisons un thésaurus enrichi constitué dans le but de regrouper et d'harmoniser – au sein d'un système notionnel cohérent et opératoire – les termes issus des travaux en AD, en linguistique et en TAL, relatifs à la nomination et à la référence. Un tel référentiel permet d'éclairer conceptuellement ces deux phénomènes langagiers. Le travail réalisé dans ce cadre permet également d'amorcer une réflexion sur l'appariement entre une sémantique de la nomination ainsi rendue opératoire et les besoins réels en veille (politique, sociétale…).

2. Domaine et finalités du thesaurus

La construction d'un thésaurus dédié plus spécifiquement à l'étude des phénomènes de nomination et de référence a été motivée par le besoin de mieux cerner les rapports complexes entre les entités référentielles et les expressions choisies pour les désigner.

1 Les termes « orphelins » sont des termes introduits dans les textes de spécialité sans définition explicite.

2 Ressources telles que, notamment, *Thesaulangue* et *TermTLF* intégrées dans *Termsciences*, portail terminologique développé par l'INIST en association avec le LORIA et l'ATILF (http://www.termsciences.fr/).

2.1. Les mots et les choses

Le terme « nomination » renvoie à l'acte d'attribution d'un nom à une entité, ainsi qu'au résultat de cet acte. La nomination est une opération linguistique et cognitive, indissociable des processus d'appréhension et de catégorisation des réalités[3]. Elle possède une dimension discursive et dialogique, car l'expression choisie pour nommer un référent reflète la position que le sujet parlant adopte à son égard. Les nominations s'inscrivent enfin dans une dynamique des relations sociales et révèlent des représentations que les locuteurs construisent, négocient et font circuler.

Sur le plan lexical et discursif, la problématique de la nomination touche à la question d'ajustement (adéquation) entre termes ou expressions (dénominations, désignations…) et référents (réalités perçues, vécues, construites…). De nombreux linguistes, dont Authier-Revuz (1995 : 507-520), Culioli (1991), ont étudié la non-coïncidence, le non-un constitutif du rapport de la langue et du monde, en insistant sur l'illusion de la transparence des mots et de l'évidence des choses. Cet écart est d'autant plus sensible que la réalité à verbaliser est complexe ou problématique : instable ou évolutive, émergente, hypothétique ou seulement visée, chargée d'enjeux contradictoires.

Selon les cas, la réponse à ce besoin sera apportée par une innovation lexicale, un emprunt à une autre langue, une néologie de sens, une spécialisation ou une généralisation sémantique. Ainsi, la nomination « patriotisme économique » issue du langage de l'extrême droite française a acquis un sens plus général et une polarité neutre voire positive avec la mise en avant du *made in France* par le ministre du Redressement productif Arnaud Montebourg en 2012.

2.2. Un thésaurus et une méthodologie

Le thésaurus TNR est destiné à être articulé à un modèle linguistique, un schéma d'annotation et un outil d'annotation manuelle.

L'un des objectifs du projet ANR TALAD vise, en effet, à construire une méthodologie générale de repérage et d'analyse des expressions à valeur géné-

3 « La propriété première de la nomination qui, en même temps qu'elle catégorise l'objet nommé, positionne l'instance nommante à l'égard de ce dernier. » (Siblot, Paul, 1997, « Nomination et production de sens : le praxème », *Langages*, n° 31 (127), p.42)

ralisante [4]susceptibles de mettre en évidence des catégories. Ce sont des entités émergentes et relatives, plutôt que des réalités ultimes et absolues, qui sont visées, dans la mesure où le travail d'élaboration est explicitement marqué dans les discours. L'attention est portée tout particulièrement sur la labilité discursive catégorielle des locuteurs. La démarche repose sur l'observation systématique du cotexte de ces expressions afin d'y identifier les traces des différentes formes d'élaboration discursive (intralocutive, interlocutive et interdiscursive) lesquelles fondent et accompagnent l'acte de nomination. Le cotexte est un contexte riche articulant plusieurs catégories de marques linguistiques. Les termes du thésaurus sont destinés à annoter et à caractériser ces marques, suivant un parcours interprétatif défini par la méthode.

3. Thésaurus de la nomination et de la référence (TNR)

3.1. Acquisition des connaissances expertes

Les études de la nomination et de la référence font appel à plusieurs domaines et courants d'analyse linguistique : analyse de discours, sémantique, néologie, lexicologie et terminologie, stylistique.

L'étape initiale de l'acquisition des connaissances linguistiques issues de l'analyse de discours (AD) consiste à étudier l'usage et la définition des termes (lorsque cette définition est fournie) par les spécialistes du domaine. Outre l'étude de l'état de l'art, deux approches ont été employées en ce sens.

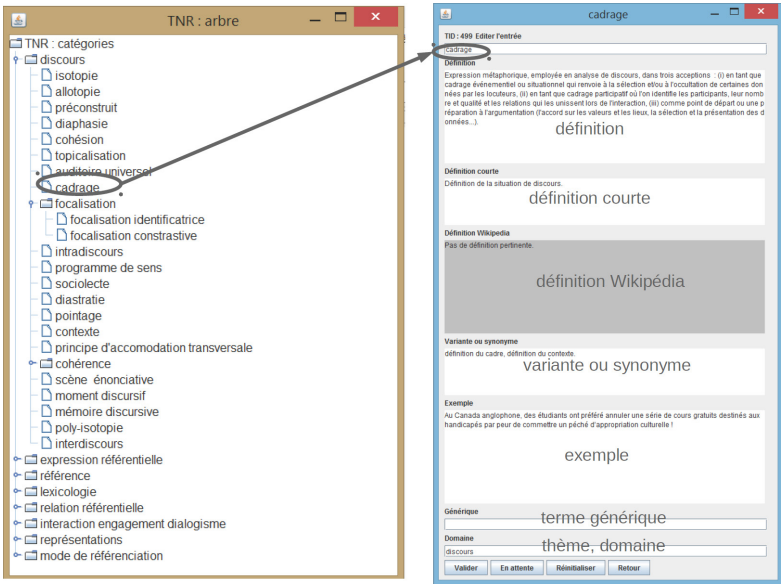
La première approche déployée de façon incrémentale consiste à enquêter auprès des membres de la communauté AD concernés par les problématiques de la référence et de la nomination. Elle permet d'identifier les modèles notionnels et les termes utilisés au sein de la communauté et de faire émerger puis de consolider la structure taxonomique de la ressource. Commencée par de simples listes de termes, cette approche contributive a donné lieu à la création d'un outil dédié qui permet de visualiser et d'enrichir les entrées du thésaurus (figure 1).

À l'heure où nous écrivons, la ressource construite grâce à l'approche experte comporte 372 entrées reparties en 9 catégories générales : *Discours, Expression référentielle, Interaction Engagement Dialogisme, Lexicologie, Mode de référenciation, Référence, Relation référentielle, Représentations,*

4 La *valeur généralisante* est entendue comme une valeur qui permet de rendre général, d'intégrer dans un ensemble d'idée les cas similaires.

Relation de discours. Pour chaque entrée, le TNR contient les informations suivantes : discipline ou aire d'emploi, variante formelle (terme équivalent ou terme approché), statut, définition synthétique, définition étendue, propriétés, relation terme → terme de tête, terme de tête (catégorie, sous ensemble, hyperonyme ou méronyme), terme associé, antonyme (ou terme complémentaire), exemple, catégorie, terme anglais, auteurs et références, publication de référence.

L'interface de complétion par les experts (vue partielle *figure 1*) permet de choisir des champs à afficher (les champs sur lesquels on souhaite travailler). En termes d'expressivité, le thesaurus apparaît comme une ressource riche en informations structurées qui tend vers un dictionnaire collaboratif. La présence des termes vedettes et variantes, des hyperonymes, des termes similaires permet son exploitation ultérieure dans le cadre d'analyse sémantique des textes de spécialité.



Les liens vers les autres ressources (notamment, thesaurus bilingues) sont en train d'être construits sachant que les ensembles des termes communs à ces ressources externes et au TNR sont de taille assez réduite et qu'il s'agit principalement des ressources de terminologie linguistique généralistes.

À ce jour, nous avons exploré les ressources pré-existantes suivantes :

- les ressources répertoriées sur le portail Termsciences (ressources terminologiques génériques) telles que Lexique383, OpenLexicon, Thesaulangue ;
- « Guide terminologique pour l'analyse des discours » (de Nuchèze et Colette, 2002), ressource livresque ;
- *SIL Glossary of Linguistic Terms*[5], glossaire généraliste de termes linguistiques bilingue français – anglais. Ce glossaire contient 8 600 entrées pour le français dont assez peu concernent l'analyse de discours. Pour le terme « discours » :
- nous avons repéré 14 termes composés ayant le terme discours comme tête syntaxique : « discours + *expansion* », *expansion* ∈ *{argumentatif, authentique, cité, d'exhortation, d'exposition, d'instructions, descriptif, dialogique, direct, indirect, indirect libre, monologique, narratif, rapporté}*.
- Parmi les termes ayant « discours » comme expansion « *tête*+(Préposition)?+ discours », l'on trouve *tête* ∈ {*type, genre, grammaire, analyse*}. Par ailleurs ce glossaire inclut les termes *interdiscours*, *métadiscours*.

L'amorçage et la construction collaborative du thesaurus ont révélé que la production terminologique importante est caractéristique des pratiques de la communauté d'analyse de discours. Cette observation a induit la démarche semi-automatique d'extraction des termes candidats à partir des textes de spécialité afin de se donner les moyens d'observer l'activité terminologique des linguistes et des notions qui sont « en chantier » au sein de cette communauté. L'abondance des termes, la spécification des termes existants témoignent de la spécialisation du domaine d'analyse de discours. Cependant, cette abondance peut également compter des cas de redondance inutile.

Outre la construction par les experts du domaine, l'approche semi-automatique a été expérimentée sur un corpus de 40 articles scientifiques traitant de la nomination et de la référence, un sous-ensemble de corpus constitué pour l'état de l'art. Ce sous-corpus de 1 115 115 mots a permis de procéder à une extraction terminologique avec l'outil *TermSuite* (Rocheteau et Daille, 2011) afin d'obtenir des termes-candidats. Cette extraction a eu pour objectif :

- la preuve d'adéquation (notamment, en termes de couverture) du thesaurus constitué par les experts dans une démarche descendante ;

5 https://feglossary.sil.org/

- l'identification des termes-candidats qui pourraient enrichir le thesaurus.

Dans le cadre de cette approche alternative, de nombreux candidats ont pu être extraits automatiquement. Après une pré-validation semi-automatique (par règles définies manuellement qui concernent l'inclusion lexicale et la structure des termes), un ensemble de candidats pré-validés (**#pré-validés**) qui correspond à un pourcentage du nombre des candidats issus de l'extraction automatique (**%pré-validés**) a été obtenu. Parmi ces termes pré-validés 91 % (**validés**) ont pu être retenus pour étude en vue de leur intégration dans le thesaurus (Tableau 1).

extraits	**# pré-validés**	**% pré-validés**	**#validés**	**%validés**
13 885	542	3 %	493	91 %

Tableau 1. Acquisition semi-automatique des termes candidats

Parmi les termes extraits validés, on trouve par exemple *applicabilité référentielle, allocutaire, propriété événementielle, saillance du cadrage* etc.

Dans la suite de l'expérience d'acquisition automatique, nous avons tenté de caractériser ces termes-candidats du point de vue de leur distribution, les rapprocher des termes déjà contenus dans le thesaurus et des autres termes candidats. Pour cela, nous avons considéré une méthode distributionnelle fondée sur le calcul des plongements des mots (*word embeddings* (Mikolov, 2013)). Les mesures *cosinus*[6] calculées pour explorer la similarité existant entre les vecteurs des mots obtenus à partir du sous-corpus qui a servi pour l'extraction terminologique ont permis de faire les rapprochements détaillés dans le tableau 2.

6 $\cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$ où A et B sont les vecteurs obtenus par plongement lexical. Il n s'agit pas de cosinus d'angle, les valeurs négatives (mots opposés) ont été peu considérés dans le cadre de la présente expérience.

Terme 1	Terme 2 (mesure cosinus)	**Qualification et remarques**
applicabilité référentielle	charge dialogique (0,84) inapplicabilité référentielle (0,88) statuer (0,95), allocutaire (0,88), congruence (0,87), degré (0,85), enchaînement (0,80)	Rapprochement d'un terme candidat et terme déjà dans le TNR. Proposition d'un terme au sens opposé. Description du terme candidat.
formant signalétique	description définie (0,82) d'agent (0,86), différenciateur (0,86) adverbe (0,85)	Proposition du terme associé (termes 1 et 2 déjà dans le thésaurus). Spécification du terme (pistes)
polarité	déviance (0,88), idéalités (0,87), spécifiante (0,81), morale (0,84) Gosselin (0,87)	Introduction du terme candidat (termes similaires). Suggestion d'une référence bibliographique
onomastique (nom)	sémantique (0,92), mono-référentielle (0.87), formants (0,83), d'entités (0,82)	Introduction du terme candidat.
rhétorique éristique	dénotation (0,88), étiquetage (0,86), intension-extension (0,85)	Termes associés à un terme issu du TNR.
étiquetage accusateur	diabolisant (0,97), fasciste (0,91), nazisme (0,89), supprimer (0,86), régime (0,86), abus (0,84), dictature (0,82), ethnocide (0,82), meurtre (0,81), victimes (0.80)	Exemplification du terme déjà présent dans le thésaurus.

Tableau 2. Exemples d'informations obtenues à partir de l'analyse de similarité distributionnelle.

Par ailleurs, la distribution des termes permet d'étudier les descripteurs déjà présents ou pouvant être intégrés dans le TNR : les plongements des adjectifs « catégoriel », « sémantico-syntaxique », « sémantico-référentiel » sont assez proches dans l'espace vectoriel considéré dans le cadre de l'expérience.

Même si la taille du corpus des articles scientifiques est insuffisante pour permettre des résultats fiables strictement issus des méthodes d'apprentissage automatique et des approches neuronales, cette première étude combinant l'extraction et la pré-caractérisation automatique semble permettre la réduc-

tion de charge de travail des experts humains dans le cadre de la construction d'une ressource de spécialité. Son application est envisagée pour le traitement de corpus plus vastes d'articles scientifiques en support à la construction experte.

L'intersection entre les résultats obtenus issus des méthodes semi-automatiques est de 83 termes (repérés automatiquement et déjà présents dans le thesaurus). Cette intersection a été exclue de l'ensemble des termes-candidats. Une expérience sur un corpus plus vaste serait nécessaire pour se rendre compte de la couverture du thesaurus en cours de construction.

Le TNR est destiné à guider la mise en place d'un environnement qui l'inclut comme ressource terminologique, mais dispose également d'un modèle linguistique, d'un schéma d'annotation et d'un outil d'annotation par les humains ; des ressources complémentaires de type « folksonomie » pour stocker la connaissance non terminologique qui permet la pré-annotation automatique des textes.

4. Folksonomie TNR : structuration et exploitation

4.1. Motivation

En termes d'expressivité, le type de ressource recherché est celui d'une ressource notionnelle riche et originale, qui n'est spécifiquement ni un dictionnaire, ni un thésaurus, ni une ontologie, mais – en puissance – un peu tout cela. Les niveaux de structuration de la connaissance sur la référence et la nomination (rendre la connaissance opératoire) et de représentation (permettre des sorties en utilisant de différents formats dont ceux du Web sémantique) ont été séparés.

Le but de cette démarche est de permettre un maximum de souplesse quant à l'acquisition et à la structuration des connaissances pertinentes pour la tâche visée, afin de pouvoir répertorier et caractériser finement les traces des phénomènes linguistiques et discursifs à l'œuvre dans les processus de référence et de nomination.

4.2. Mise en œuvre

Pour la systématisation des traces des opérations concernées (définition, prise en charge, relations sémantiques…), nous avons construit une base de connaissances (« folksonomie ») sous forme de graphe. Les nœuds de ce

graphe sont des termes du thésaurus et les items lexicaux (patrons lexicaux et méta-discursifs, segments textuels pertinents pour l'étude de la nomination, items lexicaux). Les relations sont

- des relations faiblement typées entre les nœuds représentant les termes du thesaurus et les nœuds représentant les items lexicaux (traces pertinentes pour l'étude de la nomination et de a référence repérés dans les textes) ;
- des relations sémantiques et discursives associées aux items lexicaux.
- Pour réduire le coût de l'acquisition des termes et relations des items lexicaux, nous avons exploré l'utilisation des ressources telles que :
- réseau lexico-sémantique de connaissance générale pour le français RezoJDM (Lafourcade 2007)[7] ;
- ASFALDA[8], FrameNet (Ruppenhofer *et al*. 2016) pour le français ;
- corpus annoté ANNODIS[9] (exploitation des relations discursives actuellement à l'étude).

La folksonomie$_{TNR}$ est parcourue dans le cadre de test d'annotation automatique afin de détecter les traces potentielles de construction de sens des nominations émergentes à partir des traces connues et déjà répertoriées et rattacher ces traces potentielles aux entrées du thesaurus. La structure de données obtenue est celle d'un graphe avec différents items pertinents pour l'annotation en termes des phénomènes de nomination et de référence : *terme_source, type_de_relation, poids, origine, terme_cible*. Les exemples de relations obtenues sont donnés ci-dessous.

(1) information d'ordre taxonomique et variantes

acronyme, r_definition, 111, tnr, Mot formé des initiales ou éléments initiaux de plusieurs mots, prononcé comme un mot ordinaire.
acronyme, r_example, 111, tnr, SNCF
acronyme, r_isa, 111, tnr, expression référentielle
acronyme, r_isa, 148, jdm, sigle
pseudonyme, r_isa, 111, tnr, anthroponyme
pseudonyme, r_syn, 84, jdm, faux nom

(2) termes polysémiques (relation de raffinement r_raff _sem explicite les sens possibles)

7 http://www.jeuxdemots.org
8 http://asfalda.linguist.univ-paris-diderot.fr/frameIndex.xml
9 http://redac.univ-tlse2.fr/corpus/annodis/

type d'entité, r_domain, 111, tnr, TAL (traitement automatique du langage)
type d'entité, r_raff_sem, 43, jdm, type d'entité >informatique
type d'entité, r_raff_sem, 45, jdm, type d'entité>caractéristique

(3) relations sémantiques

pronom personnel, r_lieu, 98, jdm, phrase *(relation lieu typique)*
point de vue, r_carac, 30, jdm, partagé>commun *(relation caractéristique typique)*
point de vue, r_holo, 17, jdm, débat d'idées *(relation partie-tout)*

(4) lien vers les cadres («frames») correspondants

admettre, r_frame, 50, framenet, FR_Agree_or_refuse_to_act
admettre, r_frame, 50, framenet, FR_Awareness-Certainty-Opinion
admettre, r_frame, 50, framenet, FR_Being_in_favor_of
admettre, r_frame, 50, framenet, FR_Statement-manner-noise

Ces exemples montrent que la représentation sous forme de graphe permet d'harmoniser les informations issues des différentes ressources et processus. Le poids des relations est actuellement fixé par défaut, il fera l'objet d'une harmonisation ultérieurement. La folksonomie$_{TNR}$ contient 12 515 relations dont 2 048 relations sont issues du TNR (restructurent le TNR sous forme relationnelle), 8 101 - du RezoJDM à ce jour, 1 574 - du FrameNet français, 792 - des contributions et listes (il s'agit, en particulier, des patrons méta-discursifs).

4.3. Exploitation

L'exploitation de la folksonomie$_{TNR}$ concerne l'analyse et la pré-annotation des textes. Ces processus sont en train d'être construits à l'heure où nous écrivons. La pré-annotation exploite les patrons méta-discursifs et les cadres (*frames*).

Exemple 1 : **représentation d'un patron méta-discursif** (segment textuel «réflexion à mener en France sur cet islam modéré»).

Patron méta-discursif : [$X :Nom r_isa acte de penser] en [$Y :Nom r_isa pays] sur [adj :dem] $Z :Nom
Patron (forme relationnelle) :
acte de penser, *r_pattern*, [$X :Nom r_isa acte de penser] en [$Y :Nom r_isa lieu] sur [adj :dem] $Z :Nom
(dans la folksonomie$_{TNR,}$ le patron apparaît comme une chaîne de caractères, poids de la relation r_pattern est fixé par défaut)

Exemple 2 : **pré-annotation automatique du segment** «on observe avec dédain la prolifération de cette pratique» :

patron : quand on observe $X :Nom :	**cadre** : Judgment	**relations** : regarder—r_instr-->mépris regarder--r_instr-->dédain

À titre exploratoire et grâce à un ensemble des patrons lexico-sémantiques et méta-discursifs classés par catégorie pour l'expérience, il a été possible de quantifier les éléments présents dans un corpus (66 359 mots) constitué dans l'objectif d'étudier l'élaboration de la nomination « musulmans modérés », corpus décrit dans (Pengam et Jackiewicz, 2019).

catégorie	**#occurrences**	**%occurrences**
Signalement minimal (présence des guillemets)	29	18,2 %
Signalement ou avertissement	35	22 %
Désignation	7	4 %
Élaboration (définition ou explicitation)	5	3 6%
Reprise, citation, adhésion, interaction	6	3,6 %
Distanciation critique et reformulation	40	25 %
Cadre de validité	16	10 %
Rejet et renomination polémique	21	13,2 %
Total	159	100 %

Tableau 3. Pré-annotation

La pré-annotation automatique grâce à une ressource de connaissance riche est un outil puissant de structuration notamment en ce qui concerne l'élaboration des modèles linguistiques et des schémas d'annotation. Les intuitions qui ont émergé de l'expérience de pré-annotation ont appuyé la mise en place d'une campagne d'annotation par les annotateurs humains. Cette campagne a été axée sur le phénomène de nomination émergente suscitant ou non la controverse (*musulman modéré, appropriation culturelle, mobilité douce, écofascisme*). 230 segments textuels (co-textes) de nomination émergente ont été annotés en termes d'analyse de discours. Le schéma d'annotation a intégré 3 aspects : le plan ontologique ou linguistique (par exemple, controverse sur le phénomène ou sur le terme utilisé pour le nommer), le procédé (introduction, ajustement ou rejet) et l'attitude (prise en charge, interaction, cadrage). Outre ces angles d'analyse, les relations sémantiques (en particulier, les relations statiques au sens de Desclès (2013)) ont été intégrées dans le schéma d'annotation.

5. Conclusion : vers une ontologie de la nomination et de la référence

Nous avons décrit les différentes expériences qui ont été menées dans le but de structurer, enrichir et harmoniser les connaissances théoriques liées à l'étude de la nomination et de la référence, puis de proposer une ressource de connaissance intermédiaire permettant de lier les termes qui servent à nommer les phénomènes à la réalisation de ces phénomènes dans les textes.

Les perspectives de ces travaux concernent la stabilisation des ressources que nous avons décrites ainsi que l'interopérabilité à la fois du TNR appelé à évoluer vers une ressource termino-ontologique et des données structurées issues des campagnes d'annotation que nous avons évoquées. L'effort d'intégrer les ressources de connaissance pré-existantes témoigne de la volonté d'aboutir à une ressource dotée d'une interopérabilité de contenu. L'interopérabilité de format du TNR peut être obtenue grâce à l'utilisation des formats interopérables tels que :

- TBX[10] (glossaire dans un format d'échange des terminologies, ISO 30042) ;
- Lemon OntoLex[11] (modèle de représentation dans le format du web sémantique qui étend le format OWL afin de permettre de capturer les informations lexicales et linguistiques associées à des concepts d'une ontologie) ;
- SKOS[12] (format d'organisation des connaissances de type thésaurus) .

Celle des données de sortie des campagnes d'annotation est atteinte grâce au développement d'une ressource dédiée à l'annotation qui permet le format de sortie interopérable (brat, format compatible avec l'outil d'annotation brat). L'outil permet d'ajouter d'autres formats de sortie interopérables.

Bibliographie

Calabrese, Laura et Valériane Mistaen, « La nomination des migrants dans Le Monde et Le Figaro. Analyse d'une catégorisation polémique », REFSICOM [en ligne], Médias et migrations/immigrations 1. Des

10 https ://www.iso.org/standard/45797.html
11 https ://www.w3.org/community/ontolex/wiki/Main_Page
12 https ://www.w3.org/2004/02/skos/

représentations aux traitements des médias traditionnels, mis en ligne le 23 novembre 2018, consulté le mercredi 04 décembre 2019.

Cance, Caroline, et Danièle Dubois. « Dire notre expérience du sonore : nomination et référenciation », *Langue française*, vol. 188, no. 4, 2015, p. 15-32.

Cislaru, Georgeta (dir.) *et al*. *L'acte de nommer : Une dynamique entre langue et discours*. Nouvelle édition [en ligne]. Paris : Presses Sorbonne Nouvelle, 2007 (généré le 8 février 2019).

Desclés Jean-Pierre. Interactions entre langage, perception et action. In : *Faits de langues*, n° 1, mars 1993. Motivation et iconicité. p. 123-127.

Fellbaum, Christiane, 1998, WordNet : An Electronic Lexical Database. Cambridge, MA : MIT Press.

Frath, Pierre, 2015, « Dénomination référentielle, désignation, nomination », Langue française, n° 188, p. 33-46.

Firas Hmida. Identification et exploitation de contextes riches en connaissances pour l'aide à la traduction terminologique. Informatique et langage [cs.CL]. Université de Nantes, 2017. Français.

Hoffart, Johannes, Suchanek, Fabian M., Berberich, Klaus, and Weikum, Gerhard. 2013. YAGO2 : A spatially and temporally enhanced knowledge base from Wikipedia. Artif. Intell. 194 (January 2013), 28-61.

Jackewicz, Agata, Pengam, Manon, 2018, « Des *musulmans modérés* dans les discours médiatiques. Etude linguistique d'une expression controversée », *Colloque international Les représentations médiatiques de l'islam et des musulman.e.s,* 19-20 juin 2018, Versailles Saint-Quentin-en-Yvelines, France.

Lafourcade, Mathieu, 2007, *Making people play for Lexical Acquisition*. In Proc. SNLP 2007, 7th Symposium on Natural Language Processing. Pattaya, Thailande, 13-15 December 2007, 8 p.

Longhi Julien, « Stabilité et instabilité dans la production du sens : la nomination en discours », Langue française, 2015/4 (N° 188), p. 5-14.

Mikolov, Tomas, Sutskever, Ilya, Chen, Kai, Corrado, Greg, and Dean, Jeffrey. 2013. Distributed representations of words and phrases and their compositionality. In Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'13), C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Eds.), Vol. 2. Curran Associates Inc., USA, 3111-3119.

Rocheteau, Jérôme et Daille, Béatrice, 2011. TTC TermSuite : A UIMA Application for Multilingual Terminology Extraction from Comparable Corpora. Proceedings of the 5th International Joint Conference on Natural Language Processing.

Ruppenhofer, Josef ; Ellsworth, Michael ; Petruck, Miriam R. L. ; Johnson, Christopher R. ; Baker, Collin F. ; Scheffczyk, Jan, 2016, FrameNet II : Extended Theory and Practice (revised ed.). Berkeley, CA : International Computer Science Institute.

Siblot, Paul, 1997, « Nomination et production de sens : le praxème », *Langages*, n° 31 (127), p. 38-55.

Abstract

The present article introduces a rich thesaurus built for the discourse analysis domain and focused on nomination and reference issues. We detail the stages and the methods that have been used in the framework of the thesaurus building process. Our experiences of structuring the thesaurus as well as a « folksonomy » supporting it aim at building an ontology for nomination and reference. Such ontology will be designed for natural langage annotation in terms of nomination phenomena. Thus, human annotation campains and automatic annotation tests accompany our experiences of knowldge structuring and representation. We also explored different possibilities in order to reduce human effort necessary for designing a specialized resource.