Comment extraire des définitions des textes

Marc Bertin, Jean-Pierre Desclés, Taouise Hacène

ISSN: 1824-7482

Publif@rum 11, 2010

Autour de la définition

Marc BERTIN, Jean-Pierre DESCLÉS, Taouise HACÈNE

Comment extraire des définitions des textes ?

Farum.it

Farum è un gruppo di ricerca dell'Università di Genova

Pour citer cet article :

Marc BERTIN, Jean-Pierre DESCLÉS, Taouise HACÈNE, *Comment extraire des définitions des textes ?*, Autour de la définition, Publifarum, n. 11, pubblicato il 2010, consultato il 13/12/2018, url: http://publifarum.farum.it/ezine_pdf.php?id=133

Editore Publifarum (Dipartimento di Lingue e Culture Moderne - Univerità di Genova)

http://www.farum.it/publifarum/

http://www.farum.it

Documento accessibile in rete su:

http://www.farum.it/publifarum/ezine_articles.php?art_id=133

Document généré automatiquement le 13/12/2018.

Comment extraire des définitions des textes ?

Marc BERTIN, Jean-Pierre DESCLÉS, Taouise HACÈNE

Table

L'étude de la définition ne peut se faire sans avoir connaissance des travaux de Blaise PASCAL. En effet, dans *De l'Esprit Géométrique* et de *l'Art de Persuader*, il pose le problème de la méthode géométrique qui devrait être la méthode de toute science. Nous pouvons concevoir cela comme un idéal puisqu'il ne faut employer aucun terme dont on n'a pas auparavant expliqué nettement le sens et qu'il puisse prouver toutes les propositions qui sont avancées. La définition réelle ou de « *chose* » permet donc l'analyse de contenus d'une formulation linguistique préexistante (lui rapportant tel ou tel définissant). Il reconnait les définitions nominales (ou de « *noms* ») comme seules impositions de noms aux choses. Il existe de nombreuses autres définitions à prendre à compte, comme la définition au sens aristotélicien ou les définitions en logique combinatoire. Nous nous intéresserons donc aux applications qui permettent ou permettraient de retrouver des définitions selon différents points de vue afin de faciliter les travaux de recherche. Au travers de deux scénarios, nous considérerons des besoins et analyserons le service Define de Google. Puis, suite à une étude linguistique de corpus, nous présenterons les ressources construites et intégrées au sein d'EXCOM2, une plate-forme d'annotation sémantique, afin de proposer divers services innovants.

I. L'application « Define » de Google

Google Define est un service qui propose la recherche de définitions à partir des sites web indexés. Pour accéder à ce service, l'utilisateur doit rajouter « *define* » en tant que premier terme de la requête. Par exemple, « *define: définition* » donnera les définitions selon Google pour le terme « *définition* ». Le principe de « Google define » repose sur la recherche de champs spécifiques dédiés comme des marqueurs HTML ou des descripteurs de glossaires. Cette approche ne repose sur aucune analyse linguistique ou de recherche en plein texte. Il en résulte une carence pour l'utilisateur. En effet, comme nous allons le constater, beaucoup de réponses ne sont pas pertinentes. Le besoin est pourtant présent. Considérons les deux scénarios suivants :

Le cas d'un chercheur

Supposons qu'il s'agisse d'un chercheur voulant rapprocher les deux sens du mot « *définition* », trouvés dans le Dictionnaire HACHETTE de l'édition 2003, à savoir :

- Explication de ce qu'un mot signifie.
- *AUDIOV Nombre de lignes balayées par le spot pour composer une image de télévision.*

Une recherche du terme « *définition* » dans différents ouvrages, comme des dictionnaires techniques, montre que la définition varie d'un domaine à l'autre. Dans le cas présent, deux domaines vont être confrontés : la linguistique et l'audiovisuelle. Certains termes sont directement liés aux évolutions technologiques. En effet, Aristote n'évoquera pas la définition au sens de « *norme caractéristique physique déterminée principalement par la densité des points* », définition trouvée dans le *Dictionnaire d'informatique et d'Internet : anglais-français* par Jean-Guy Grenier de 2000. Dans le cadre de notre recherche, la définition proposée ci-dessus sera donc à mettre en relation avec la première définition du dictionnaire HACHETTE.
Le chercheur continue sa recherche en s'appuyant sur un autre dictionnaire, Le Nouveau Larousse Elémentaire, une édition de 1972 qui est un peu plus ancienne. La seule définition existante de « *définition* » est «*Explication claire et précise de la nature d'une chose, du sens des mots* ». À cette époque, la définition du point de vue technique n'existait pas. Le problème qui se pose ici est celui de la science actuelle opposée aux encyclopédies. Ainsi, notre chercheur se voit dans l'obligation de rechercher cette notion dans d'autres documents. La tâche qu'il doit accomplir est d'autant plus difficile puisqu'il doit non seulement chercher la notion «*définition*» dans des textes mais aussi classer les significations retrouvées en deux domaines pour pouvoir les mettre ensuite en parallèle. Il teste maintenant la requête « *define: définition* » sur Google. Certes, plusieurs documents, qui selon Google nous offrent une définition du mot « *définition* » s'affiche, mais la relation entre le definiens et le definiendum existe-t-elle dans les segments textuels extraits ? Quelle est la pertinence des résultats ainsi obtenus ? Le tableau ci-dessous commente les définitions retrouvées à l'aide de la commande « define ».

Définitions extraites par Google à titre indicatif	Commentaires
Une définition est un discours qui dit ce qu'est une chose ou ce que signifie un nom.	Enoncé définitoire
Nombre fixé de lignes, formées de pixels, qui constituent l'image sur un écran.	Dictionnaire
la définition d'une image numérique s'exprime par le nombre de pixels dans la largeur et par le nombre de pixels dans la hauteur.	Enoncé définitoire
qualité de perception des détails de l'image, qui dépend de sa granularité, sa densité et sa résolution.	Glossaire du site
Certaines définitions peuvent contenir des illustrations et des liens vers d'autres termes du dictionnaire.	Enoncé non définitoire
Aptitude d'une image à reproduire des détails fins. La définition s'exprime sous forme de nombre de points par ligne (définition horizontale) et nombre de lignes par image (définition verticale).	Enoncé définitoire
Formule lexicographique qui énonce les traits sémantiques distinctifs d'un concept. Exemple : définition terminologique. Sur la fiche terminologique, justification textuelle permettant d'établir le crochet terminologique.	Glossaire du site
Vidal a réalisé une base de données constituée de fiches d'informations sur les médicaments. Le service proposé par Vidal (ci-après dénommé « le Service ») a pour objet de permettre à l'Utilisateur d'accéder à ces fiches.	Hors-sujet
la modalité la plus fréquente	Hors-sujet
La définition d'une image est le nombre de pixels qui la composent. Une image en 1024x768 affiche 1024 pixels horizontaux sur 768 pixels ...	Enoncé définitoire
Nouvel hominidé, cousin nain de l'Homo sapiens, retrouvé dans la grotte de Liang Bua sur l'île indonésienne de Florès. ...	Non trouvé
Fête annuelle d'origine anglo-saxonne, célébrée le 31 octobre, à l'occasion de laquelle les enfants masqués et déguisés font en soirée la tournée des maisons de leur quartier pour quêter des friandises. Hors-sujet	Hors-sujet
L'impact est toléré par l'élève.	Hors-sujet
Capacité de reproduire le maximum de détails d'un document. La définition est liée à la distance d'observation de l'image : une affiche, lisible à plusieurs mètres, ne nécessite pas une définition aussi fine qu'une photo d'un magazine.	Lexique du site
Terme servant à indiquer la qualité de l'image, sa finesse, son nombre de pixels.	Lexique du site
Nombre de points lumineux (pixels) composant une image.	Glossaire du site
Le Management par la Qualité Totale est la méthode de gestion de l'organisation pour aboutir à l'excellence opérationnelle. ...	Hors-sujet
Indique la précision d'analyse ou de représentation d'un écran d'une imprimante ou d'une image. Par abus de langage, on parle aussi de résolution.	Lexique numérique

Tableau 1:*Résultat commenté de la requête 'define: définition'*

L'application « Define » de Google extrait des énoncés définitoires se rapportant au mot «*définition* » ainsi que des *définitions* de mots. Ces derniers sont signalés par « *Hors-sujet* ». Les segments textuels relevés ici ne sont pas tous des définitions pertinentes pour notre recherche. De même, nous avons indiqué dans la colonne commentaire par « *Dictionnaire* », « *Lexique du site* » ou «*Glossaire du site*» les extraits où la relation entre definiens et definiendum est inexistante. Le commentaire « *Non trouvé* » indique que la mise à jour vers le site n'est pas effective, la définition n'est donc plus accessible. Enfin, nous trouvons dans le tableau quatre extraits que nous qualifierons d' « *Enoncé définitoire* » qui correspondent à une définition ou une facette définitoire sous réserve que l'ambigüité du marqueur de définition soit levée. Notre chercheur doit donc continuer sa fouille pour espérer trouver par exemple la définition d'Aristote : « *une définition est une formule qui exprime l'essentiel de l'essence d'un sujet* » (*Topique*s, I, 4, 101b37-38).

Le cas d’un étudiant en Économie

Considérons le sujet d’un devoir d'étudiant en 3ème année d’Administration Économique et Sociale :
« *Pour Marx, les classes sociales ne sont pas seulement un outil de description sociologique, elles sont au coeur de son explication du mouvement de l'histoire. L'appartenance de classe façonne les valeurs et les pratiques des individus (1). À l'opposé, la tradition webérienne suppose que les classes sociales sont des groupes d'individus semblables, partageant une même dynamique (Max Weber parle de Lebenschancen ou « chances de vie »), sans qu'ils en soient forcément conscients. Pour lui, la classe sociale est constituée par des individus (2) rassemblés en fonction des critères que l'on juge les plus discriminants (le diplôme, le revenu, le patrimoine, etc.); c'est une construction sociale et non une donnée tangible. » Louis CHAUVEL, Issu de l’article « Comment les classes se distinguent ?* » Alternatives Economiques - n°207 - Octobre 2002. En s’appuyant sur la précédente citation, montrer que dans la définition des classes sociales, deux grands courants s’affrontent.
Afin de mener à bien son devoir, l'étudiant doit dégager les notions fondamentales des différents courants. Pour cela, il devra donc fouiller à travers une littérature conséquente afin de définir ces deux grands courants. Il recherchera d’abord les définitions de cette notion chez Marx puis chez Weber. Face à la difficulté de cette tâche, il souhaite contourner le problème en trouvant à travers d'autres textes la notion sans pour autant que l’auteur soit Marx ou Weber. Son analyse intuitive de documents consistera à relever toutes les phrases contenant par exemple les mots ou expressions « Selon Marx », « D’après Weber », et « classes sociales ». Bien entendu, il faut qu’une relation définitoire subsiste entre le défini et le définissant.
Bien que ce type de repérage s'avère pertinent, l'utilisation d'un moteur de recherche classique lui fournira des milliers de documents contenant ces expressions sans pour autant pouvoir les exploiter dans leur intégralité. Ne pouvant parcourir les documents dans leur ensemble, l'étudiant ne sera pas forcément en mesure d'appréhender les notions essentielles à une bonne compréhension des courants. Cette présente étude nous permet de comprendre en quoi l’étude des positions de différents auteurs est utile dans les définitions. En effet, dans le cas présent, « *classe sociale* » ne se définit pas de la même manière chez Marx et chez Weber. Il en va de même pour de nombreuses autres notions du devoir.
Ces scénarios mettent en évidence d'une part les besoins pour ce type d'application, et d'autre part la nécessité de développer de nouveaux types d'applications comme l'annotation automatique de définition. Pour cela, il est nécessaire de prendre en considération les facettes définitoires et d'utiliser les incises par exemple « *en tant que* » ou « *c'est-à-dire* » en tant qu'indices complémentaires. Nous distinguerons les deux catégories suivantes : « *Définitions »* et « *Facettes définitoires »*. Pour cela, nous allons à travers une étude linguistique de la définition construire une carte sémantique. Nous disposerons alors des ressources nécessaires à l'annotation de corpus avec le point de vue de la définition.

II. Analyse linguistique de la définition

Cette étude linguistique a pour but de mieux comprendre la distinction entre définition et facette définitoire afin de pouvoir envisager la constitution de ces ressources linguistiques. Une facette définitoire est un énoncé dans lequel seules les propriétés essentielles sont définies. L'étude que nous proposons s'appuie sur des textes issus de l’Encyclopédie Universalis (Catégories, Espace, Phonologie). Ils nous ont permis de dégager trois catégories en ce qui concerne les facettes définitoires : l’identification, la catégorisation déterminée et la pseudo-définition.
La première catégorie des facettes définitoires est l’identification. Elle permet d’assimiler une première expression à une seconde plus précise. Par exemple : « *Une corrélation est un ensemble d'oppositions présentant un même rapport entre deux traits distinctifs binairement opposés.* » Nous ne nous intéresserons pas ici aux spécifications de l’identification, c’est-à-dire à la relation d’égalité et à la relation d’identité. Nous parlerons de catégorisation déterminée lorsque le processus d’identification est déterminé. Pour cela, nous pouvons donner l'exemple suivant : « *L'espace est cette communication qui reprend et englobe la singularité de tous les constituants par synergie de leurs charges.* » Enfin, l’emboîtement de catégorisations déterminées sera défini par la notion de « *pseudo-définition* ». Nous pouvons considérer l'exemple suivant pour mieux saisir cette catégorie : « *La phonétique est une science expérimentale qui traite des caractéristiques physiques du signal phonique ainsi que des conditions de sa production et de sa réception.* »
Nous soulignerons également l’existence d’énoncés qui ne s’inscrivent pas dans le cadre des définitions de par l’absence ou la présence d’indices qui permettent de lever l’ambiguïté d’un indicateur donné, par exemple, « les…sont + préposition à » qui ne relève plus du cas de l’identification : « *Les catégories sont au service de la différenciation et en cela consiste leur vocation critique, antidogmatique.* ».
De même, l'étude des textes précédemment cités, nous a permis de dégager deux grandes catégories de la définition à savoir les définitions générales (définition rapportée, définition engagée, définition contextualisée, définition contextualisée concessionnelle) et les définitions axiomatiques (axiomes, définition d'entités mathématiques).
Jusqu'à présent, nous n'avons pris en considération que la première relation entre le definiendum (ce qui est défini) et le definiens (ce qui définit) de la définition. Or, il existe une seconde relation entre cette première relation d'une part et d'autre part l’agent qui l’érige. L’omniscience de cet agent se trouve généralement dans les définitions rapportées : « *Cela est manifeste dans l'oeuvre de C. S. Peirce, où les catégories se révèlent être, selon les mots de Peirce lui-même, « des idées si vastes qu'elles doivent être entendues comme des états (moods) ou des tonalités (tones) de la pensée, plutôt que comme des notions définies... Envisagées en tant que numéraux, susceptibles d'être appliqués à tous les objets que l'on veut, elles constituent en*

réalité de minces squelettes de pensée, ou même de simples mots » (Collected Papers, vol. I, pp. 353-355). »
La présence de l'agent n'est pas toujours manifeste dans la trame discursive aussi nous ne trouverons pas de trace effective. Dans ce cas précis, nous parlerons donc de définition contextualisée puisqu'elle est marquée dans un contexte par un déictique, limité dans un cadre, un domaine ou dans le temps voire introduite par un collectif diffus « *on* ». Nous pouvons donner pour exemple : «*Le rapport sourd/sonore qui caractérise ainsi six oppositions est appelé corrélatif, car il oppose une qualité sonore à son absence ou, ce qui revient au même, à la seule qualité qui peut lui être opposée, à savoir sourd* ». Nous relevons ce type de définition dans les dictionnaires où les lexicographes sont effacés mais reste existant par le biais d'un collectif large comme le montre les instructions précédant les définitions : Math., Litt., … Si cette définition repose sur l'établissement d'une condition nécessaire et suffisante en d'autres termes précédée d'une condition, généralement introduite par « *si* », il s'agira d'une définition contextualisée concessionnelle. Nous donnerons pour exemple : « *Si les Américains se contentent le plus souvent de définir les phonèmes comme des classes d'allophones, les Européens poursuivent, quant à eux, l'analyse pour les identifier à des ensembles de traits distinctifs.*» Par ailleurs, l'énonciateur/locuteur qui reprend une définition rapportée choisit ou non de l'attester, de s'en servir pour élaborer une démonstration, d'insérer une nouvelle notion. Ce type de définition s'inscrit dans l'évolution du texte par le biais de modalités et nous parlerons de définition engagée. Pour clore notre analyse des différentes catégories de la définition, nous terminerons par la définition axiomatique, en d'autres termes un énoncé permettant de désigner une vérité première, c'est-à-dire une proposition évidente en soi.
« *Par exemple, si l'on examine les cinq postulats de Peano pour l'axiomatisation de l'arithmétique -à savoir : (1) 1 est un nombre ; (2) le successeur de tout nombre est un nombre ; (3) deux nombres ne peuvent avoir le même successeur ; (4) il n'est le successeur d'aucun nombre ; (5) toute propriété qui appartient à 1 et également au successeur de tout nombre qui la possède appartient à tous les nombres -, il apparaît que (1), (2) et (4), ensemble, ont pour fonction d'indiquer les objets de la théorie (les nombres naturels), alors que (5) a une fonction constitutive et que la fonction de (3) est d'individuation.* »
La définition d'entités mathématiques permet d'introduire un nouveau mot ou symbole associé à un objet abstrait décrit par un assemblage d'autres mots ou symboles dont le sens a déjà été précisé ou dont on connait intuitivement sa signification (les primitives).
Essayons à présent de considérer, avec l'aide de notre précédente analyse, les marqueurs propres à la définition, tel qu'ils pourraient être définis et entrevoir ce que pourrait être une règle d'exploration contextuelle.
Pour l'Identification, l'analyse de nos textes de l'Encyclopédie Universalis nous a permis de ne trouver qu'une seule structure «*article… être + article…* ». Ainsi, en indice positif avant, nous aurons une liste d'articles définis/indéfinis singulier/pluriel. En tant qu'indicateur, le verbe « *être* » est directement suivi d'un article. C'est pourquoi nous avons décidé d'inscrire les indices positifs après (articles définis/indéfinis singulier/pluriel) dans la liste de la copule « *être* ». Nous avons également choisi d'éviter la présence d'incises introduites par « *c'est* » ainsi que les séquences introduites par le démonstratif « *ce* » suivant l'indicateur par une relative telle que «*ce qui, ce que, ce qu'* ». Pour exemple, «*Les catégories sont des notions « stratégiques » ; susceptibles d'usages divers, elles circulent entre le syntaxique et le sémantique et témoignent d'une pensée constructive »*.
Pour la catégorisation déterminée, nous utilisons une liste d'indices positifs avant et la liste d'indicateurs de la précédente règle. La catégorisation déterminée se construit de la même manière que l'identification mais il faut cependant lui ajouter une liste de relatives (qui, que, sous lequel, où) placées après la liste d'indicateurs (indices positifs après). « *L'espace est ce mode sous lequel un homme et le réel s'explicitent l'un l'autre par échange de leur structure.* »
De même, la règle permettant de créer l'annotation « *pseudo-définition* » reprend la règle de la catégorisation déterminée mais en transformant quelque peu la liste d'indices positifs placée après l'indicateur à l'aide d'une expression régulière. La première proposition relative doit être suivie d'une autre proposition relative. Ainsi pour définir cette construction, nous pourrons par exemple écrire « qui (.)* ainsi que (.)* ». Nous pouvons donner comme exemple : «*La phonétique est une science expérimentale qui traite des caractéristiques physiques du signal phonique ainsi que des conditions de sa production et de sa réception.* » Bien entendu, ces listes de marqueurs seront étendues à l'aide de notion équivalente comme par exemple, « est identifié à » ou « est équivalent à», « qui (.)* (ainsi|et) que (.)* ».
Reprenons maintenant chaque sous-concept sur les marqueurs de la définition générale. Pour la définition rapportée, les indices positifs placés avant l'indicateur marquent cette relation entre la définition et son auteur ainsi que leurs importances. C'est pourquoi, nous réaliserons une liste d'indices positifs avant contenant une expression régulière désignant des prépositions telles que « *d'après* », «*chez* » ; des expressions telles que «*dans l'oeuvre de* », «*ces derniers* »,… Cette liste pourrait par exemple s'ajouter à la structure de l'identification et servir d'indices positifs après : « « *L'espace est lui-même une notion temporelle », dira bientôt Paul Klee.* ». La constitution de règles repose donc sur l'identification d'indicateur comme « *définir* », «*désigner* », ‹*appeler*». Pour indiquer un énoncé définitoire, nous devons prendre en compte que :

- Le verbe « *définir* » doit être suivi de préposition telle que «*comme* », «*par* », «*avec* », ‹ *grâce à* » et «*au moyen de* ».
- Le verbe « *se définir* » est directement suivi des préposions «*comme* » ou «*par* ».
- Les verbes « *désigner* » ou «*signifier* » fonctionnent a priori tout seuls (sans l'aide de préposition).
- Le participe passé « *appréhendé* » est toujours suivi de la préposition «*comme* ».
- Le verbe « *se révéler* » est suivi directement de la copule ‹ *être* ».
- Les verbes de dénomination « *appeler* » et «*nommer* » sont suivis ou précédés d'articles définis ou indéfinis.

En ce qui concerne la définition contextualisée, elle s'inscrit dans un contexte, un domaine, un cadre spatial et/ou temporel.

Une fois les structures identifiées, nous ajouterons des marqueurs faisant référence au cadre temporel comme « *siècle* », «*aujourd'hui*», au cadre spatial comme «*en + pays* », à un domaine «*en + domaine* », au contexte avec les déictiques. On retrouve de nouveau les verbes « *définir* » et «*appeler* » en tant qu'indicateurs. «*Appeler* » sous forme passive doit être précédé d'un article. De plus, un troisième indicateur se manifeste « *se présenter* » qui doit être suivi nécessairement de l'expression « *comme étant* » pour qu'il n'y ait pas d'ambigüité.
Nous sommes donc à même de pouvoir proposer une règle du type :

- **Soit** un indicateur de définition, par exemple le verbe « *appeler* »

- **Si** un collectif diffus, par exemple « *on* » le précède

- **Alors** attribuer l'annotation « *définition contextualisée* ».

Pour la définition contextualisée concessionnelle, seules deux structures sont retrouvées dans nos textes :

- « *Si* » en tant qu'indice positif avant, «*définir* » en tant qu'indicateur et «*comme* » en tant qu'indice positif après.
- « *Si* » en tant qu'indice positif avant et «*signifie* » en tant qu'indicateur.

Nous remarquerons que la précision des deux règles ne conduira pas, a priori, à l'extraction d'énoncés non définitoires. À partir des différentes considérations établies précédemment, il devient possible de regrouper et classer les énoncés définitoires selon différents critères et en différentes sous-catégories. Chaque valeur de l'ontologie linguistique que nous présentons ici représente une des étiquettes sémantiques avec lesquelles nous pouvons annoter un énoncé définitoire. Elle est présentée ici sous forme d'arbre :

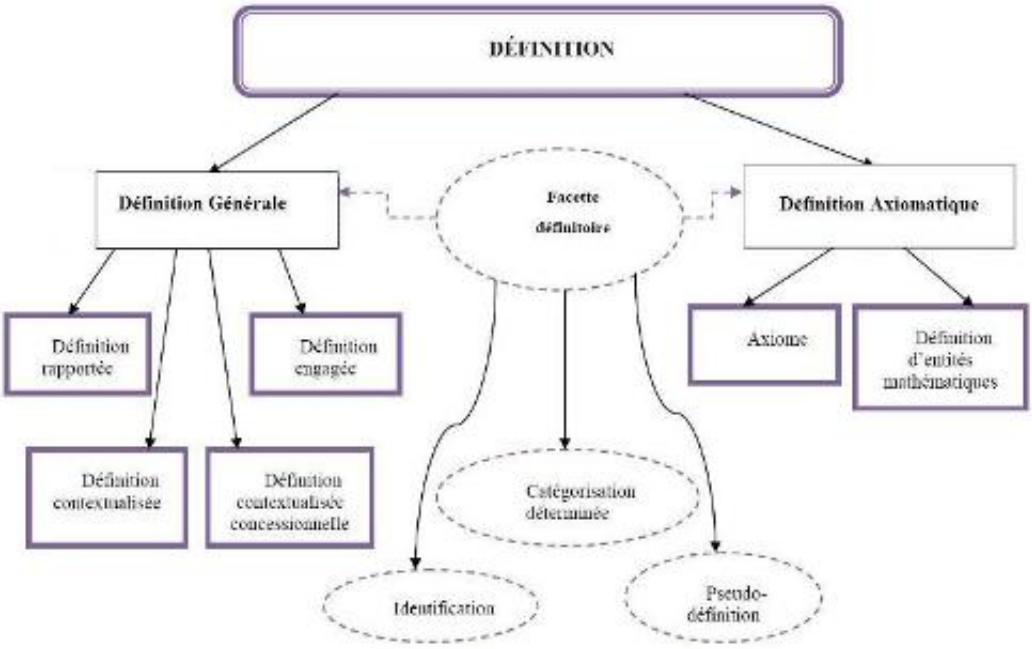

Schéma 1 :Carte sémantique de la définition

L'implémentation de cette analyse comme méthode d'exploration contextuelle fournira au moteur d'annotation l'ontologie linguistique qui autorisera une annotation automatique des segments textuels alors identifiés.

III. Méthodologie et implémentation

Exploration contextuelle

Nous venons de voir que l'étude du contexte permet de dégager des marqueurs afin d'annoter un segment textuel. En fait, la méthode d'Exploration Contextuelle (DESCLES, 2006) permet d'annoter automatiquement et sémantiquement des segments textuels pour une tâche donnée. Pour le cas présent, il s'agit de l'annotation selon le point de vue de la définition. La méthode consiste dans un premier temps à identifier des marqueurs linguistiques, que l'on appellera « *indicateur* », et qui portent la valeur sémantique de l'annotation. La présence d'un indicateur dans le contexte est une condition nécessaire pour l'exécution des règles. Cela se traduit par le fait que la méthode d'exploration contextuelle est avant tout une procédure applicative qui se présente sous forme d'un ensemble de règles d'exploration contextuelle. L'exploration contextuelle met donc en jeu des processus inférenciels qui sont déclenchés, dans un premier temps, par l'identification d'indicateurs linguistiques relatifs à un champ grammatical ou discursif précis. C'est en ce sens que ces indicateurs deviennent des marqueurs de valeurs sémantiques. La méthode d'exploration contextuelle s'appuie sur deux types de composants à savoir un ensemble de marqueurs linguistiques et un ensemble de règles d'explorations contextuelles. Étant donné un indicateur, on applique un ensemble de règles tel que pour chaque règle on recherche les indices associés à un espace de recherche. Si l'indice est positif, cela déclenche une annotation et si l'indice est négatif, la règle ne s'applique pas. L'indicateur est un déclencheur dont il faut vérifier par Exploration Contextuelle qu'il est bien associé à une annotation par la présence d'indice positif qui confirme la valeur de l'indicateur ou d'indice négatif qui inhibe la valeur associée à l'indicateur. Donc les règles, déclenchées par l'indicateur, identifient dans un contexte de recherche la présence d'indices qui autoriseront ou non l'annotation du segment textuel avec la catégorie correspondante. Les indicateurs, tout comme les indices peuvent être des listes de mots ou d'expressions régulières. Les catégories sémantiques pour les annotations sont organisées sous forme de treillis. En établissant une hiérarchie entre les indicateurs et les indices, nous définissons une catégorie sémantique. La construction des ressources linguistiques adéquates s'appuie uniquement sur une analyse de corpus permettant de trouver les catégories sémantiques d'une tâche donnée. Cela implique que la constitution de recherche est essentiellement basée sur le jugement humain. Cette méthode a été implémentée dans le système Excom2.
Nous faisons donc appel à des points de vue de fouille qui focalisent l'attention sur des segments et des organisations textuels. Ces points de vue sont des traces linguistiques d'opérations inscrites dans les documents textuels et les textes (relation entre les concepts, relation entre texte et image…). La hiérarchie étant : Point de vue - Notion - Catégorie - Instance. Cette approche qui repose sur des critères purement linguistiques est d'autant plus pertinente face à des indicateurs qui pourraient se révéler polysémiques. Nous serons donc face à plusieurs étiquettes sémantiques possibles pour un même segment textuel. L'application des règles d'Exploration Contextuelle permet de lever l'ambiguïté sémantique de ces indicateurs. Donnons un exemple en vue d'un repérage automatique d'expressions définitoires dans les textes, au travers de verbes comme « *définir* », « *signifier* » et «*désigner*». Ces verbes étant polysémiques, ils pourraient aller à l'encontre de notre objectif principal c'est-à-dire relever des énoncés non-définitoires. C'est pourquoi il est nécessaire d'analyser ces verbes et montrer qu'avec telle ou telle préposition on pourrait annoter ou non le segment comme « *définition* ». Regardons de plus près un de ces verbes afin de montrer un peu plus en détail la méthodologie. Le verbe « *définir* » peut avoir une signification autre que celle que l'on recherche. Prenons l'exemple, « *L'Etat définit l'impôt* ». Ici, le verbe «*définir* » a le sens de «*fixer une valeur* ». Pourtant, on peut remarquer que lorsqu'il est suivi d'une préposition telle que « comme » ou «*par* », il porte directement la relation définitoire : « *Joseph Nye définit la puissance comme la capacité de faire faire aux autres ce qu'ils ne voudraient pas faire dans un premier temps.* ». S'il est suivi de formules du type : «*en ces termes*», «*par ces mots* », «*de cette manière* », «*ainsi* », il sert d'introducteur à un énoncé définitoire. Remarquons également que la qualité de vie ou le bien-être qui est évidemment une notion subjective, peut-être défini ainsi : « Être *en bonne santé physique et morale* ». Mais, quelques contre-exemples peuvent être relevés notamment lorsque la préposition « *comme* » est de suite suivie d'une autre proposition telle que «*si* », «*dans* » ou « *par* »: «*Les instruments semblent errer dans le vide sans qu'aucun thème défini comme dans les autres mouvements ne prenne forme.* » De même, la préposition «*par* » à elle seule peut signaler une relation définitoire ambigüe : «*Il comprend différentes espèces, dont Dryopithecus fontani, définie par É. Lartet en 1856 à partir d'un fossile découvert à Saint-Gaudens (Haute-Garonne).* » Ce petit exemple permet de dégager la nécessité d'utilisation d'indices qui sont destinés à lever l'ambiguïté des indicateurs. Ces indices permettent ainsi, soit de leur attribuer une valeur sémantique (la valeur définitoire), soit au contraire de ne pas la leur attribuer (soit parce qu'aucun indice n'est présent, soit qu'un indice discriminant la valeur sémantique est présent).

La plate-forme d'Exploration COntextuelle Multilingue

La plate-forme EXCOM 2 ou Exploration COntextuelle Multilingue: (DJIOUA et alii, 2006) puis (ALRAHABI, 2008) est un moteur d'annotation sémantique opérant à partir de ressources linguistiques préalablement constituées sous forme d'ontologie linguistique. D'un point de vue technique, cette plate-forme repose essentiellement sur les technologies XML. Par ailleurs, la plate-forme propose une ouverture vers le multilinguisme en prenant en compte d'autres langues que le français tels que l'arabe et le coréen. Nous ne rentrerons pas dans la partie technique de l'architecture fonctionnelle de la Machine EXCOM2 à annoter les textes selon des points de vue de fouille, les articles suivants en font déjà référence (DJIOUA, 2008). La plate-forme Excom 2 prend en entrée un corpus de textes (provenant de n'importe quelle source comme internet, des œuvres numérisées, des textes personnels) enregistré au préalable sous UTF-8. Un module contenu dans cette plate-forme s'occupe de le segmenter via le module SegaTex et permet de distinguer titres, phrases et paragraphes. À l'issu de cette segmentation,

l’annotation sémantique peut se réaliser. Il s’agit d’attribuer des étiquettes à certains segments textuels selon des catégorisations discursives et sémantiques ou point de vue (définition, citation, rencontre, causalité, …). L'interface utilisant Excom2 permet à l'utilisateur d'enregistrer ses marqueurs et de concevoir ses propres règles d’exploration contextuelle en fonction du point de vue qui a été retenu.

IV. Réalisation informatique

Annotation de définition

L'analyse menée précédemment nous conduit à une implémentation informatique afin d'identifier et annoter des segments textuels porteurs d'une information définitoire. Les résultats obtenus sont issus de trois corpus différents. Les règles d’Exploration Contextuelle précédemment citées sont testées sur un corpus « *mixte* » composé de cinq textes de l’Encyclopédie Universalis (Cancer, Cellule, Liberté, Phénomène, Temps) et de cinq autres issus de Wikipédia (Géodésie, Intelligence, Mort, Territoire, Virus) afin de juger de la pertinence de notre implémentation. Nous obtenons 110 segments annotés composés de 60 annotations avec facettes définitoires et de 50 annotations en tant que définitions. Ces résultats montrent que les facettes définitoires ne sont point négligeables puisqu'elles apportent une information à un concept sans pour autant le définir. Voici quelques exemples de segments annotés par Excom2 ainsi que les catégories auxquels ils appartiennent. Pour les facettes définitoires, nous obtenons pour l'identification : « *La mémoire est la faculté de conserver les informations.* ». Pour la catégorisation Déterminée, un segment annoté est : «*Les ribosomes sont les organites cytoplasmiques qui synthétisent les chaînes polypeptidiques des protéines.* » Enfin pour la Pseudo-définition, nous obtenons : «*Cette unité du monde vivant à l'échelle cellulaire est l'héritage d'une longue évolution commencée il y a 4,5 milliards d'années avec la formation de la Terre, évolution qui a conduit à l'apparition des premiers organismes unicellulaires il y a 3,5 milliards d'années et qui s'est poursuivie jusqu'à maintenant.* »
En ce qui concerne les définitions, le segment textuel suivant a été annoté comme « *définition Contextualisée* ». Considérons l'exemple suivant qui met en évidence un collectif diffus : « *On définit ainsi le géoïde comme étant une surface équipotentielle du champ de pesanteur, choisie arbitrairement, mais très proche du niveau des océans que, par la pensée, nous pouvons prolonger sous les continents.* » Dans un cadre spatial, nous obtenons des segments du type suivant : «*Par exemple, en Europe le système EUREF ainsi obtenu est la base des systèmes de références géodésiques de tous les pays européens, dont la France qui, à son tour, a appuyé dessus sa référence nationale officielle, le RGF 93, sous la responsabilité de l'IGN.* » Dans un cadre temporel : « *Au XVIIe siècle, au moment du renouveau de la philosophie moderne, le terme de phénomène désigne les faits empiriques* ». Dans un domaine : «*En écologie, un territoire désigne la zone de peuplement d'une espèce végétale ou animale* ». Enfin, à replacer dans le contexte (déictique) : «*Ce cycle est appelé samsâra* ».
Pour la définition rapportée, le système propose les segments textuels suivants d'après un auteur : « *enfin, il n'est pas sans intérêt de rappeler que, pour Aristote, la science architecturale ou, pour mieux dire, architectonique, qui enveloppe toutes ces considérations sur le bonheur, la vertu et les vertus, sur le rapport de la préférence à l'excellence, sur le règne de la prudence, s'appelle le « politique ».* ». D'après une oeuvre : «*Dans le dictionnaire de géographie de Pierre George et Fernand Verger le territoire est défini comme un espace géographique qualifié par une appartenance juridique (on parle ainsi de «territoire national»);»*
Les règles n’ont pas été modifiées en fonction des résultats obtenus. C’est pourquoi, nous pourrons émettre quelques objections à certains énoncés relevés. Par exemple, un même segment aura deux annotations différentes. C’est une erreur que l’on retrouve uniquement dans les facettes définitoires : un même segment sera annoté catégorisation déterminée (a) et pseudo-définition (b).
(a) ces deux thèmes ont d'ailleurs des points communs puisque *la membrane plasmique, les chromosomes, les ribosomes* **sont des** organites *qui* existent dans toutes les cellules et que les thylakoïdes sont caractéristiques des cellules photosynthétiques possédant de la chlorophylle, qu'elles soient procaryotes ou eucaryotes.
(b) ces deux thèmes ont d'ailleurs des points communs puisque *la membrane plasmique, les chromosomes, les ribosomes* **sont des** organites *qui existent dans toutes les cellules et que* les thylakoïdes sont caractéristiques des cellules photosynthétiques possédant de la chlorophylle, qu'elles soient procaryotes ou eucaryotes.
(c) Comme l'a bien *vu Russell lui-même, l'instant* **est une** classe d'équivalence *qui* renferme les événements reliés par une relation symétrique transitive.
Il ne s’agit pas là d’un problème à considérer puisque l’annotation est tout de même convenablement effectuée et dans cet exemple, il s’agirait plutôt d’une catégorisation déterminée que d’une pseudo-définition.
De même, on pourrait se demander si l’énoncé (c) appartient à une catégorisation déterminée, tel qu’il a été annoté ici, ou à une définition rapportée. Ainsi, si l’on considère que l’expression « comme l’a bien vu » signale plutôt une définition rapportée, il nous suffira de l’ajouter dans la liste d’indices positifs avant en ce qui concerne la définition rapportée et de l’insérer dans la liste d’indices négatifs de la catégorisation déterminée.
Malgré ces quelques critiques, dans l’ensemble, ces règles ont permis l’extraction de segments textuels définitoires appartenant aux catégories implémentées de par les règles. Cette annotation sémantique et automatique nous permet, par exemple, d’obtenir deux définitions de la notion « *temps* » :

(d) c'est la « spatialisation du temps » analysée*par Bergson [cf. BERGSON (H.)], et la mise en oeuvre de la vieille définition d'Aristote [cf. ARISTOTE] : « Le temps* **est le** nombre du mouvement ».
(e) La formule (4) de la dynamique du point de *Galilée-Newton* **définit** donc le temps *comme* une grandeur mesurable en tant que référée universellement aux longueurs, aux forces et aux masses (considérées comme directement mesurables).
Le « *temps* » est d'une part défini par Bergson et d'autre part par «*La formule (4) de la dynamique du point de Galilée-Newton* ».
En revenant sur les scénarios précédemment énoncés, on peut émettre ici la remarque du gain de temps établi ici puisque sans avoir eu besoin de lire l'article « *Temps* » de l'Encyclopédie Universalis, nous avons deux positions de la définition du terme « *temps* ». Il ne reste qu'à l'utilisateur d'analyser ces deux définitions et d'observer si elles divergent ou convergent.

Application Bibliosémantique

Cette première application permet d'identifier les définitions dans le cadre de la fouille de texte et ainsi proposer aux utilisateurs un outil innovant. Mais il existe un autre type de besoin en relation avec la définition. Il est parfois nécessaire d'identifier les relations entre auteurs. Dans le discours savant, et à travers les publications scientifiques, il est pertinent de pouvoir identifier si un auteur est cité en tant que définition ou hypothèse. Pour ces résultats ou bien négativement. Ces travaux de recherche rentrent dans le cadre de ce qui se nomme la bibliosémantique (BERTIN, 2008). Cette application permet l'identification et la classification des jugements d'auteurs citant d'autres auteurs en qualité de définition. Pour cela, nous nous référons à une ontologie linguistique qui permet d'identifier des marqueurs discursifs au sein des corpus. D'un point de vue de la bibliométrie, la définition à un rôle fort à jouer. La citation en tant que définition, permet d'une part de faciliter la lecture en évitant les redites, mais surtout d'intégrer dans son raisonnement ce qui a déjà été considéré comme un acquis à la connaissance. Citer des travaux en s'y référant en tant que citation est important au sein du réseau. Un terme ou un concept défini dans un discours savant peut être repris ou non par la communauté, discuté, disséqué, commenté, et à travers le jeu des citations dresser un réseau de relations entre le point de vue des auteurs et la définition proprement dite. L'identification et l'annotation de segments textuels sur le jugement d'un auteur sont des points clés pour mieux appréhender la bibliométrie, la nature de certains réseaux et comprendre la structure cumulative de la science. L'une des contributions que cette approche peut apporter à la théorie de la définition relève de la possibilité d'étudier les définitions de termes ou de concepts à l'éclairage de ces citations qui peuvent être positives ou négatives.
Price a proposé comme définition de l'article que ce dernier fût un quantum d'information scientifique. Cela sous-entend que le produit final du travail d'un scientifique, d'une équipe ou d'un laboratoire est l'article qu'il publie, et que l'article est l'expression des travaux menés. Mais selon Estivals, la statistique bibliographique est inconcevable sans une méthode qualitative. Pour lui et son équipe, la méthode qualitative constitue, pour reprendre ses termes, le soubassement et l'environnement de toute étude statistique bibliographique. La méthode qualitative est plus détaillée alors que la méthode quantitative est plus abstraite et schématique. En fait, la complexité de l'interaction entre les deux méthodes est liée à la nature de l'objet étudié. Comme nous l'avons vu précédemment, la méthode qualitative est le fondement de la méthode quantitative, mais lors de l'exploitation, il y a une interdépendance en fonction de l'objet et de la conception de la recherche, notamment en sciences humaines.
L'identification des articles proposant des définitions et la catégorisation que l'on peut apporter aux jugements d'auteurs qui citent leurs pairs en tant que définition prennent alors à la lumière de cette nouvelle compréhension un sens plus complet.
La première étape du traitement repose sur l'identification des différents types de renvois bibliographiques permettant d'identifier le segment textuel sur lequel s'appliquera le traitement informatique. L'annotation sémantique résultante dégagera un ensemble de relations permettant alors une catégorisation de l'utilisation de ces renvois. Nous proposons d'utiliser les renvois bibliographiques d'un article afin de déterminer des segments textuels sur lesquels nous pourrons appliquer la méthode d'exploration contextuelle. L'appel de citation dans un texte peut prendre différentes formes. Il peut s'agir principalement d'un renvoi numérique ou d'un renvoi par nom d'auteur. Afin de traiter automatiquement cette tâche d'identification et d'extraction, nous avons défini un alphabet adéquat permettant d'appliquer au corpus un automate fini déterministe. Cette extraction nous permet dans un premier temps d'étiqueter le corpus, puis de dresser des listes d'auteurs, de renvois ainsi qu'une bibliographie complète de l'auteur et de ces co-auteurs. Pour ce travail, nous avons utilisé les normes, mais également les « *coutumes* ». En effet, les renvois bibliographiques dans le texte sont plus ou moins normalisés selon les normes [17] ISO 690-1 (Z 44-005) et ISO 690-2, tout en tenant compte des pratiques dépassant le simple renvoi numérique ou alphanumérique afin de pouvoir traiter exhaustivement l'ensemble des renvois bibliographiques. Cette approche s'appuie sur l'hypothèse que la prise de position d'un auteur vis-à-vis de ces confrères se trouve dans un espace proche d'un renvoi bibliographique. Dans le cadre de cette approche, nous considérons les renvois bibliographiques comme des indicateurs. Tout comme nous n'allons pas définir ici le rôle et la nature des renvois bibliographiques, nous ne détaillerons pas non plus en détail l'ontologie linguistique sous-jacente à cette problématique mais plutôt nous focaliser sur la catégorie traitant de la définition. La catégorie de la définition est importante. Les indices peuvent être: « ils caractérisent | la notion ... introduite dans |... ».
Le corpus est constitué de textes issus d'Intellectica, année 2001 et de l'ALSIC, année 2006 et 2007. Voici quelques exemples extraits de forme que nous trouvons généralement dans la littérature scientifique. La première ligne donne le nom du fichier ainsi que le corpus à partir duquel il est extrait. Entre parenthèses se trouve l'annotation donnée au segment textuel par la plateforme EXCOM.

« *2. fr.utf8.alsic.2006.v9.ADEN.txt.xml (ALSIC2006)*
Nous faisons l'hypothèse que la gestion didactique de ces supports multimédia nécessite une stratégie de développement de compétences culturelles, en particulier d'une compétence de repérage des stéréotypes et lieux communs dans le discours multimédiatique, et, de façon plus large, nous souhaitons explorer la compétence doxale ([Sarfati05]: 109-110), définie en partie comme " l'aptitude des apprenants à sélectionner et à identifier – à la production comme à la réception – les lieux communs...(citation, definition). »
Pour cet exemple, il est intéressant de souligner le caractère partiel de cette définition.

« *3. fr.utf8.alsic.2006.v9.ADEN.txt.xml (ALSIC2006)*
La doxa est définie par Bouthoul (1966, cité par [Sarfati05]: 96) comme "l'ensemble des jugements qui font l'objet de croyance pour des individus composant une société" et dont les apprenants ne sont pas conscients. (citation, definition) »
Le marqueur linguistique est facilement identifiable pour ce segment: « *est définie* ».

« *4. fr.utf8.alsic.2006.v9.ADEN.txt.xml (ALSIC2006)*
Il est également important de prendre en considération le fait que le processus de construction des stéréotypes est lié au phénomène de projection tel que défini par Giddens (cité par [MorganGrimshaw05]) comme le fait d'attribuer à l'autre, de façon inconsciente, ses propres désirs et caractéristiques (1993: 257). (definition) »

Les deux exemples suivants montrent que la recherche par mots-clés est insuffisante. Les systèmes reposant sur ces types d'approches, à savoir occurrence de terme et recherche par mots-clés ne peuvent pas identifier les phrases suivantes en tant que définition.

« *2. fr.utf8.intellectica.2002.Vol1.Num34_11_Lecuyer.txt.xml (Intellectica)*
Gelman & Gallistel (1978) ont proposé le concept de numéron: représentation interne de la numérosité (ce qui est représenté), idée reprise par (Wynn, 1995). (definition) »

« *4. fr.utf8.alsic.2007.v10.n2.Narcy-Combes.txt.xml (ALSIC2007)*
[1] Ce concept désigne pour de Rosnay [deRosnay06], la " nouvelle classe d'usagers des réseaux numériques" autrement dit les internautes qui s'emparent des "nouveaux outils d'empowerment" (d'autonomisation) qui sont à leur disposition sur le réseau. (definition) »

Cette application permet d'identifier les auteurs et les travaux qui sont cités en tant que définition. Elle met également en évidence les insuffisances des indicateurs bibliométriques actuels en démontrant l'existence de catégories discursives. L'ontologie linguistique résultante a donc un rôle important dans le cadre de l'évaluation scientifique et de la connaissance qu'elle peut apporter à la structure du raisonnement scientifique. La terminologie des textes savants a, en fonction du courant ou de la théorie prônée, des sens quelque peu différents et des interprétations, propres à l'auteur. Il est donc important de dresser une carte de ce tissu cognitif de l'activité scientifique et de savoir si un terme, une notion, une théorie sont cités dans le cadre en tant que vérité, c'est-à-dire en tant que connaissance et reconnaissance. L'identification et l'annotation des segments textuels définitoires jouent un rôle important dans le cadre de ce travail.

Schéma 2:Application BiblioExcom : Recherche automatique d'annotation liée à la définition.

Étude comparative de Google Define et Excom2

La dernière application que nous décrirons ici est une étude comparative entre le « *define* » de Google et notre plateforme d'annotation automatique. Afin de mettre en valeur la spécificité de notre approche et les annotations de la plate-forme EXCOM-MOCXE, nous avons effectué une comparaison avec le service « *Google Define* » afin de montrer la pertinence d'une annotation sémantique plein texte. Pour cela, nous avons recherché des informations autour de trois concepts que nous avons essayé de définir : Développement durable, ontologie et grippe aviaire. Le protocole expérimental est le suivant : recueillement des 100 premières réponses obtenues sur chacun de ses termes. Les trois cents documents obtenus ont été analysés par EXCOM avec le point de vue de la définition. L'indexation a été faite par MOXCE afin de pouvoir effectuer des requêtes sur la base de données indexée. Pour un ensemble identique de requêtes, « *Google Define* » a retenu 28 énonciations, 24 d'entre elle étant approprié (transmettant aux définitions) : 1 définition pour « *grippe aviaire* », 10 pour « *ontologie* », 13 pour « *développement durable* » et 4 réponses qui sont sans rapport. Le moteur d'annotation sémantique EXCOM-MOCXE a identifié 25 définitions sur 26 énonciations indexées : 7 définitions pour « *grippe aviaire* », 13 pour « *ontologie* », 5 pour « *développement durable* ». Ces résultats montrent que l'approche sémantique et discursive permet d'extraire des connaissances que Google Define n'identifie pas. Cela est lié au fait que google Define n'effectue aucun traitement textuel mais repose sur des glossaires alors que les énonciations extraites par EXCOM sont catégorisées et différenciées. Le détail complet de cette étude a été publiée dans (TEISSEDRE, 2008).

Conclusion

Il est important de disposer d'outils informatiques pertinents et dédiés à une tâche donnée comme l'étude de la définition. À travers les deux scénarios décrits, nous avons identifié un besoin et les faiblesses d'un outil grand public. Pour appréhender un peu mieux la définition à travers la littérature et l'activité scientifique, il est nécessaire de disposer d'applications informatiques qui prennent en charge les annotations sémantiques et automatiques de corpus. Cela est rendu possible grâce à la méthode mise en œuvre et applicable pour des problématiques qui peuvent sembler éloignées d'un premier abord. L'ontologie linguistique qui est utilisée par la plate-forme permet une approche qualitative et se révèle pertinente pour des problématiques comme l'étude de la définition. L'implémentation informatique qui en résulte, en autorisant la segmentation, l'identification et l'annotation automatique et sémantique de segments textuels, permet de mettre à disposition des outils qui n'ont pas encore leurs équivalents. Ces applications dédiées peuvent donc jouer un rôle important dans l'étude de la définition. Nous avons vu l'exemple de la définition au sein de l'activité scientifique dont la reconnaissance passe à travers le jugement de ses pairs. En effet, le mot ou le concept, une fois défini par un auteur, doit être soumis à l'étude, la critique, l'acceptation ou l'approbation de ses pairs afin d'être entièrement mis à nu et d'en obtenir toute la quintessence. Nous pensons que l'étude de la définition peut être grandement facilitée par l'utilisation d'outils dédiés.

Bibliographie

M. ALRAHABI, J-P. DESCLÉS, «*Automatic annotation of direct reported speech in Arabic and French, according to semantic map of enunciative modalities*», 6th International Conference on Natural Language Processing, *GoTAL 2008*, Gothenburg, Sweden, August 25-27, 2008
ARISTOTE, (Topiques, I, 4, 101b37-38)
M. BERTIN, «Categorizations and annotations of Citation in Research Evaluation»,*FLAIRS-21 (21th International Florida Artificial Intelligence, Research Society Conference)*, Miami, Florida, 15-17 may, p. 456-46, 2008
J-P. DESCLÉS, «Contextual Exploration Processing for Discourse Automatic Annotations of Texts»,*FLAIRS 2006*, Melbourne, Floride, 11-13 mai, Invited Speakers, p. 281-284, 2006
J-P. DESCLÉS, «De la définition chez Pascal aux définitions en Logique Combinatoire»,*Travaux de logique*, N°19, à paraître, Université de Neuchâtel, p. 73-113, 2008
J-P. DESCLÉS, B. DJIOUA, «La recherche d'informations par accès aux contenus sémantiques: vers une nouvelle classe de Systèmes de Recherches d'Informations... »,*Linguistique Informatique*, Roumanie, 2008
B. DJIOUA, J. GARCIA FLORES, A. BLAIS , J-P. DESCLÉS , G.GUIBERT , A. JACKIEWICZ , F. LE PRIOL, L. NAIT-BAHA, B. SAUZAY, «EXCOM: an automatic annotation engine for semantic information», FLAIRS 2006, Melbourne, Floride, 11-13 mai, 2006
B. PASCAL, «De l'esprit géométrique et de l'art de persuader»,*Oeuvres complètes, Bibliothèque de la Pléiade*, texte établi et annoté par Jacques Chevalier, p. 575-604, , 1954
C. TEISSEDRE, B. DJIOUA, J-P. DESCLÉS, «Automatic Retrieval of Definitions in Texts, in Accordance with a General Linguistic Ontology», *FLAIRS-21 (21th International Florida Artificial Intelligence, Research Society Conference)*, Miami, Florida, 15-17 may, p. 518-523, 2008

Pour citer cet article :

Marc BERTIN, Jean-Pierre DESCLÉS, Taouise HACÈNE, *Comment extraire des définitions des textes ?*, Autour de la définition, Publifarum, n. 11, pubblicato il 2010, consultato il 13/12/2018, url: http://publifarum.farum.it/ezine_pdf.php?id=133