

Le lexique électronique de la terminologie de Ferdinand de Saussure : une première

1. Introduction

La terminologie de Ferdinand de Saussure a depuis toujours fait l'objet de nombreuses études. Parmi les plus remarquables, il convient de citer le petit lexique contenu dans l'ouvrage exégétique de Godel (1957) « Les sources manuscrites du cours de linguistique générale », ainsi que l'étude la plus complète à ce jour, le « Lexique de la terminologie saussurienne » de Engler (1968) dans lequel sont définis, et souvent commentés, 441 termes clés de la pensée saussurienne. Concernant la terminologie, seules des questions ponctuelles ont été traitées depuis lors. C'est pourquoi, à la lumière des acquisitions théoriques liées à la découverte des manuscrits de l'Orangerie, il a semblé opportun d'actualiser ces études. Ainsi, en mettant à profit les potentialités offertes par l'informatique dans le domaine des humanités et l'état de l'art de la Lexicographie Computationnelle, le premier lexique électronique de la terminologie linguistique saussurienne a vu le jour.

2. Un modèle ontologique

L'utilisation des nouvelles technologies de l'information dans le domaine des sciences humaines a permis de stocker d'énormes quantités de données linguistiques qui ne se prêtent cependant à être exploitées et partagées qu'au prix d'une organisation préalable et d'une représentation rigoureusement structurée et formalisée. Ceci explique l'intérêt croissant que suscitent, depuis près de trente ans, la définition et la formalisation d'ontologies lexicales.

Pour le terme 'ontologie'¹, nous adopterons ici la définition de Studer (1998) qui englobe celles de Gruber (1993) et de Borst (1997) « An ontology is a formal, explicit specification of a shared conceptualization »². Les ontologies offrent en effet une modélisation des connaissances : représentation formalisée et structurée des concepts caractérisant un domaine, de leurs propriétés et restrictions, de leurs relations, et des instances de ces concepts. À partir de ces données, elles permettent la formulation

¹ Le terme 'ontologie', né dans le domaine de la philosophie, y est défini comme l'étude des propriétés les plus générales de l'être.

² « Une ontologie est une spécification formelle et explicite d'une conceptualisation partagée ».

de requêtes, la réalisation d'inférences, ainsi qu'une navigation intelligente dans les bases de données textuelles annotées par leurs concepts.

Cet aspect de support à la recherche d'information conceptuelle a motivé notre choix d'un modèle basé sur une ontologie lexicale pour la construction de ce lexique. L'adaptation d'un modèle existant nous est d'autre part apparue comme un choix judicieux, tant en termes d'économie de temps et de coût qu'en terme de qualité des résultats. Après un examen attentif des principaux modèles lexicaux à structure ontologique, conçus pour le développement de grandes ressources lexicales informatisées, nous avons opté pour le modèle SIMPLE (Lenci, 2000).

SIMPLE, inspiré du modèle WordNet et plus tard inspirateur du standard ISO LMF, est un modèle sémantique basé sur les principes fondamentaux de la théorie du Lexique Génératif (Pustejovsky, 1995). Il a permis la création de 12 lexiques sémantiques pour autant de langues européennes (Ruimy, 2003). Ce modèle offre un réseau conceptuel articulé en 157 classes sémantiques, reliées par des liens hiérarchiques et non hiérarchiques. Chaque sens d'un terme y est décrit au moyen de relations et de traits sémantiques ; l'information relationnelle, très riche et en grande partie fondée sur une relecture et extension de la structure des *qualia*³, incorpore également les liens synonymiques, dérivationnels et de polysémie logique. La richesse expressive de ce modèle, la finesse de ses descriptions, son niveau de granularité descriptive et sa flexibilité nous ont ainsi incités à entreprendre sa customisation en vue de construire la ressource lexicale que nous présentons ici.

3. Le modèle Simple_FdS⁴

3.1. L'ontologie de domaine

La connaissance spécifique de notre domaine d'intérêt a été modélisée dans l'ontologie de domaine Simple_FdS, en fonction de son objectif, de la typologie des utilisateurs, des types de recherche qui seront effectuées et des réponses attendues ; ceci, conformément aux principes de conception de l'ontologie générale SIMPLE.

Il existe différentes méthodologies de construction manuelle d'une ontologie⁵. L'approche 'middle-out' ou centrifuge, qui a présidé à la création de cette ontologie, nécessite l'identification des concepts essentiels à la connaissance du domaine, suivie

³ La structure des *qualia* est composée de quatre rôles permettant respectivement de situer l'entité dénotée parmi d'autres (rôle formel), d'en indiquer la fonction (rôle téléique), l'origine (rôle agentif) et la composition (rôle constitutif).

⁴ Dans les différentes sous-sections, nous ne fournirons que quelques éléments concernant l'ontologie, les relations sémantiques et les propriétés des termes du lexique. Une description complète est consultable à l'adresse : <www.ilc.cnr.it/saussure_prg/papers/Specifications_Linguistiques_Lexique.pdf>.

⁵ L'approche 'top-down' ou descendante qui opère par spécialisation des concepts les plus généraux ; l'approche 'bottom-up' ou ascendante où l'on procède à une généralisation des concepts les plus spécifiques ; l'approche 'middle-out' ou centrifuge.

de la recherche de concepts plus généraux et plus spécifiques en vue de la création du schéma conceptuel global. Notre point de départ a consisté en un noyau de termes significatifs, extraits des vocabulaires définis par Godel et Engler, d'où ont émergé les concepts les plus représentatifs de la pensée de Saussure. Une fois structurés en liens hiérarchiques et non hiérarchiques, ces concepts, situés à un niveau intermédiaire du schéma ontologique, ont été généralisés et spécialisés. Parallèlement, la structure interne des classes conceptuelles ainsi que les propriétés et restrictions des relations les unissant ont été définies : propriétés obligatoires ou optionnelles ; valeurs autorisées et cardinalité.

L'ontologie Simple_FdS est composée de 43 concepts hiérarchisés, répartis sur 4 niveaux de profondeur (cf. fig. 1).

À son sommet, et regroupés sous une classe nommée 'Thing', racine de la hiérarchie, figurent 4 classes sémantiques qui reflètent les 4 rôles de la structure des qualia, conformément au fondement théorique du modèle lexical. Il s'agit des classes 'Entity' (rôle formel), pour les entités dont la nature concrète ou abstraite n'est pas déterminée (*entité*), 'Telic', 'Agentive' et 'Constitutive'. Ces trois dernières classes permettent de regrouper des entités dont l'hyperonyme ne peut être déterminé et quiinstancient principalement une dimension de sens téléique (*objet_linguistique*), agentive (*origine_du_langage*) ou constitutive (*système*).

Seules quelques-unes des classes conceptuelles de l'ontologie seront à présent brièvement décrites.

Dans la classe 'Convention' sont regroupées les lois, les conventions en général, mais aussi celles qui régissent les phénomènes linguistiques (*Lautverschiebung*) ; 'Domain_of_Knowledge' héberge les différents sous-domaines (*signologie*). La classe 'Property', où sont regroupées les propriétés générales (*faculté*) ainsi que les propriétés plus spécifiques des éléments linguistiques (*arbitraire*), a pour sous-hiérarchie un ensemble de classes à usage exclusif des adjectifs, permettant d'en représenter les propriétés physiques, psychologiques, relationnelles et temporelles. Les activités mentales sont rassemblées dans trois classes : 'Cognitive_Event' (*abstraction*) et ses sous-classes : 'Cause_Change_Cognitive_Event', pour les activités mentales impliquant un changement et un résultat (*analyse_subjective*) et 'Mental_Creation' pour les processus mentaux de création (*paraplasme*). 'Speech_Act' encode les termes liés à l'acte de la parole (*circuit_de_la_parole*). 'State' rassemble les états (*synchronie*) et subsume les classes 'Relational_State', pour les termes dénotant une relation stative (*ablaut*) et 'Identificational_State', pour une relation d'identité ou d'opposition (*homorganie*) ; tandis que les instances de 'Change_of_State' sont des termes exprimant des changements d'état (*diachronie*). 'Representation', classe multidimensionnelle qui incorpore une dimension agentive et une dimension téléique, regroupe des entités ayant une fonction évocatrice (*carré_linguistique*). 'Linguistic_System' rassemble des termes dénotant la langue en tant que code et en tant qu'idiome, ainsi que les différentes langues (*sanscrit*). Les instances de 'Relational_Entity' sont, quant à elles, des termes clés de la pensée saussurienne qui dénotent toujours des entités relationnelles (*terme*).

La classe ‘Linguistic_Entity’ réunit des termes généraux concernant les unités linguistiques soit composites soit irréductibles résultant d’un processus de délimitation ou d’articulation (*unité*). Elle subsume six classes conceptuelles caractéristiques du domaine, parmi lesquelles ‘Basic_Element’, qui regroupe des termes dénotant les éléments irréductibles des différents niveaux de description linguistique (*consonne*); ‘Complex_Element’, pour les éléments composites des différents niveaux de description linguistique (*syntagme*); enfin deux classes particulièrement représentatives de la pensée de Saussure: ‘Concrete_Element’, dont les instances sont des termes dénotant la partie perceptible, concrète du signe linguistique (*figure_vocale*) et ‘Mental_Element’, qui regroupe les termes dénotant la partie conceptuelle, psychique, mentale du signe linguistique (*valeur_significative*).

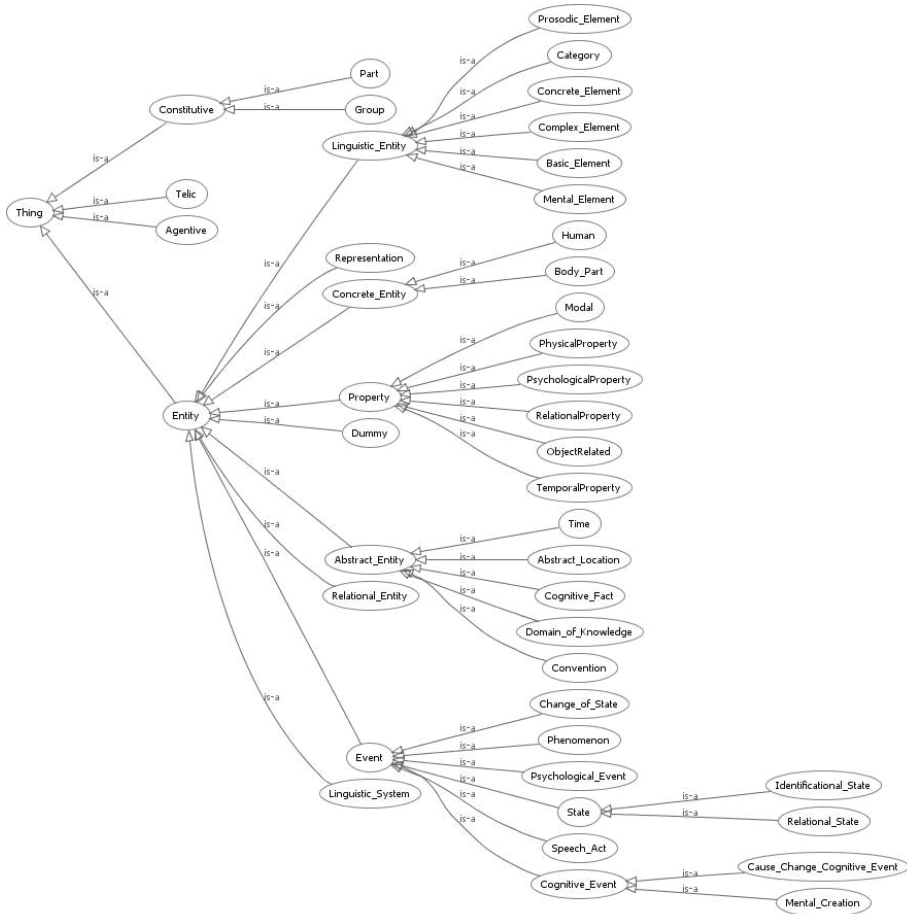


Fig.1. Représentation graphique de l'ontologie Simple_FdS

3.2. Les relations sémantiques

Les relations sémantiques constituent le cœur du modèle SIMPLE. La plupart d'entre elles sont le fruit d'une relecture et d'une extension de la structure des *qualia* du Lexique Génératif, la '*Extended Qualia Structure*'. Sous forme de triplets <unité sémantique source, relation, unité sémantique cible>, elles permettent de représenter de manière structurée la multi-dimensionnalité du sens d'un mot et de décrire finement le réseau de liens que les instances des différentes classes ontologiques entretiennent, tant au plan paradigmatique qu'au plan syntagmatique. Les relations classiques de synonymie, antonymie et de dérivation morphologique complètent ce large éventail de liens.

La flexibilité du modèle de référence a permis de conserver uniquement les relations jugées pertinentes à notre domaine de la connaissance et d'en créer de plus spécifiques afin d'exprimer des liens qui caractérisent l'organisation conceptuelle propre à ce domaine. Voyons ici quelques-unes des relations sémantiques les plus représentatives de la pensée saussurienne.

La relation 'duality' relie certains termes clés des réflexions saussuriennes, comme *langue* et *parole*, *synchronie* et *diachronie*. Pour le lecteur expert, le nom de cette relation évoque le célèbre manuscrit de l'Orangerie intitulé "De l'essence double du langage", et en particulier la note pour le second cours de linguistique générale⁶. Cette relation implique la forte interaction entre les deux concepts en jeu, contrairement à la traditionnelle interprétation dichotomique de ces couples, héritage de la conception structuraliste des niveaux hiérarchiques.

Liée à la précédente, la relation 'directlyImplies' relie, par exemple, *signifié* à *signifiant*. Il s'agit là d'une dualité non antinomique ; les deux entités entretiennent un rapport d'implication réciproque en ceci qu'elles constituent les deux faces indissociables du signe linguistique⁷.

La relation 'belongsTo' permet de situer un concept, soit dans une perspective synchronique ou diachronique, soit dans la sphère d'appartenance à la *langue* ou à la *parole*. Ce sont là deux oppositions capitales. D'une part, la distinction entre « la linguistique de la langue et la linguistique de la parole »⁸. D'autre part, celle entre deux différents ordres de faits : la linguistique synchronique, qui a pour centre l'état d'une langue et les rapports entre des « termes coexistants », et est concentrée sur la

⁶ « le langage est réductible à cinq ou six DUALITÉS ou *paires de choses* » (ELG, p. 298).

⁷ « Il est aussi vain de vouloir considérer l'idée hors du signe que le signe hors de l'idée » (ELG, p. 44).

⁸ Distinction que Hjelmslev définit la « thèse primordiale » du CLG et qui semble constituer la « première vérité » sur laquelle se fonde toute la linguistique générale, (Riedlinger, 1911 (SM 30)). La langue, vue comme la partie passive du circuit de la parole, le code social qui organise le langage et représente l'instrument fondamental pour l'exercice de la faculté du langage, se distingue de la parole, l'acte individuel, momentané du sujet parlant utilisant le code.

perspective du sujet parlant et la linguistique diachronique, centrée sur les rapports entre des « termes successifs » et fortement liée au facteur temps⁹.

Les relations ‘hasPreviousDenomination’ et ‘hasSubsequentDenomination’ permettent de saisir et de formaliser les changements terminologiques typiques de Saussure. S’élevant contre l’inexactitude de certains termes traditionnels dans le domaine de la linguistique, il tente d’en définir de manière rigoureuse des concepts centraux afin que celle-ci accède véritablement au statut de science. À cet effet, il confère des sens nouveaux à des mots existants, change la dénotation de certains concepts au fil du temps, forge des néologismes dont certains, jugés inadéquats, sont abandonnés par la suite.

Dans cette optique, deux autres relations ont été créées. La première, ‘hasNear-Synonym’ relie des termes caractérisés par une sémantique lexicale très semblable (*diachronie* et *évolution*)¹⁰. La seconde, ‘hasOtherDenomination’, permet de signaler l’introduction d’un terme nouveau, souvent créé *ad hoc*, pour indiquer un concept précédemment exprimé de façon différente. Ainsi, *morphologie* rebaptisée *théorie des signes*¹¹.

3.3. Propriétés exprimées par des couples attribut-valeur

Dans sa version actuelle, le modèle *Simple_FdS* est caractérisé par 30 propriétés appartenant à trois catégories différentes : les propriétés sémantiques, caractérisant soit une classe conceptuelle tout entière soit une instance spécifique ; les propriétés relatives à l’usage des termes ; les propriétés morphosyntaxiques (catégorie grammaticale).

Les propriétés sémantiques apportent un complément d’information concernant la portée sémantique des termes. Représentant des caractéristiques transversales à la hiérarchie des classes, elles favorisent le regroupement d’unités lexicales, indépendamment de leur classification ontologique. De plus, elles permettent de représenter des dimensions de sens, clairement perçues dans le contenu sémantique d’un mot, mais cependant trop imprécises pour être exprimées par un terme, au sein d’une relation sémantique.

Voici quelques-uns des traits sémantiques spécifiques des classes nominale et verbale :

‘Mental’ désigne la nature typiquement mentale d’une entité ou d’un événement (*image*, *Mental* *yes*). ‘Intentionality’ indique la présence ou l’absence d’intentionnalité

⁹ « Tous les objets à étudier forment deux domaines : le ou les champs synchroniques, le champ diachronique, à quoi correspondent des perspectives et des méthodes différentes. Dans le champ diachronique, la méthode se dédouble ; dans le champ synchronique, l’unique méthode est l’observation de ce qui est ressenti par les sujets parlants. » (R 84-85) (Godel, 72).

¹⁰ « Il ya un certain nombre de termes à peu près synonymes... : *histoire*, *évolution* *altération* – et on peut proposer aussi le terme de *diachronie*. » (D232).

¹¹ « Le vrai nom de la morphologie serait : la théorie des signes, et non des formes » (3293 §2).

dans un événement, selon la distinction chère à Saussure entre les activités conscientes opérées sur la langue par le linguiste et les activités inconscientes typiques de la masse parlante (*analyse_objective*, Intentionality yes; *analyse_subjective*, Intentionality no). ‘AgentiveF’ et ‘TelicF’ indiquent respectivement une dimension de sens agentive ou téléique inexprimable au niveau lexical dans une relation sémantique (*innovation*, Agentive yes; *symbole*, Telic yes). Il en est de même pour ‘ResultingState’ qui signale la présence d’un résultat imprécisable, conséquence d’une action (*altération*, ResultingState yes).

À ces propriétés sémantiques s’ajoute un ensemble de traits spécifiques des adjectifs. Six ‘Meaning Component’ précisent l’information fournie par la classe sémantique (*cosystème*: RelationalProperty, MeaningComponent : SetMembership). D’autres traits marquent la distinction entre adjectifs extensionnels et intensionnels; indiquent leur graduabilité; la durée temporelle d’une propriété physique; et pour les adjectifs modaux, le contrôle humain sur un événement ainsi qu’une nécessité, une prédiction, ou une possibilité d’actualisation.

Les propriétés concernant l’usage des termes fournissent des indications relatives à la période d’attestation ou d’apparition d’un terme (*temps_homogène*, Attestation-Period 1910-1911); sa fréquence, au moyen de valeurs non numériques (*circuit_de_la_parole*, Frequency hapax) ainsi que la référence à l’ouvrage dans lequel il apparaît (*terme*, Source N 20).

4. La base de connaissance lexicale

Les moyens descriptifs qui ont été présentés ont permis la création, à ce jour, de 500 entrées lexicales¹² ayant un contenu informatif particulièrement riche. Chaque entrée décrit une seule acception d’un lexème simple ou complexe et fournit une représentation structurée de la définition que Saussure lui-même donne de ce terme ou, à défaut, de celles qu’en proposent Godel et Engler. Dans les exemples d’entrées illustrés (cf. tableau 1), l’on peut observer, outre les susdites définitions en langage naturel, la présence d’informations concernant la classe ontologique d’appartenance du terme, le domaine d’usage, la période d’attestation, la fréquence d’occurrence, la source, les collocations, ainsi qu’un ensemble de relations qui établissent un lien entre le sens décrit et un certain nombre d’événements ou d’entités qui lui sont strictement liés.

¹² Dont 379 noms, 112 adjectifs et 9 verbes.

<p>id = "uni-spatialité"</p> <p>Saussure definition = " « Principe de l'uni-spatialité (si l'on considère le sôme) ayant pour conséquence dans le sême la divisibilité par tranches (toujours dans le même sens et par coupures identiques) l'opposition entre deux sêmes se règle au moyen de tranches allant dans le même sens et n'arrivant qu'une à une] au lieu de la divisibilité pluriforme qu'on aurait par exemple dans un système visuel direct (= écriture idéographique). » (3317.2); « Il faut s'astreindre à dire uni-spatialité du signe linguistique chaque fois, afin de faire sentir que ce n'est pas un caractère général du sême. » (3318.1) "</p> <p>Definition = " «Linéarité du signifiant.» (Engler, 51) "</p> <p>Features= "SemanticType Property SuperType Entity Domain Linguistics PoS Noun"</p> <p>SemRelations:</p> <p>Isa: propriété (Property)</p> <p>isPropertyOf: signifiant2 (Concrete Element)</p> <p>isPropertyOf: signé4 (Concrete Element)</p> <p>isPropertyOf: sôme (Concrete Element)</p> <p>isPropertyOf: discours (Speech_Act)</p> <p>isPropertyOf: chaîne_de_la_parole (Speech_Act)</p> <p>isPropertyOf: syntagme (Complex Element)</p> <p>isPropertyOf: parole (Speech_Act)</p> <p>Concerns: divisibilité (Property)</p> <p>hasSynonym: linéarité (Property)</p> <p>hasOtherDenomination: succession_temporelle (Property)</p> <p>dejectivalNoun: spatial (Physical Property)</p> <p>belongsTo: parole (Speech Act)</p>	<p>id = "parole"</p> <p>Saussure definition = " « Dans le langage, la langue a été dégagée de la Parole, elle réside dans [] l'âme d'une masse parlante, ce qui n'est pas le cas pour la parole. » (ELG, 333); « Quand on défalque du Langage tout ce qui n'est que Parole, le reste peut s'appeler proprement la Langue et se trouve ne comprendre que des termes psychiques, le nœud psychique entre idée et signe, ce qui ne serait pas vrai de la parole. » (ELG, p. 334) "</p> <p>Definition = " « langage moins la langue » (CLG, Engler, 38) « Tout ce qui est amené sur les lèvres par le besoin du discours et par une opération particulière. » (IR2.24); « Acte de l'individu réalisant sa faculté du langage au moyen de la convention sociale qui est la langue. » (IR6-7); « Partie active et individuelle [du langage], mais où il faut distinguer : 1°) l'usage des facultés en général en vue du langage (phonation, etc.); 2°) l'usage individuel du code de langue selon la pensée individuelle. » (D178; cf. 184 J) (Godel, 271) "</p> <p>Collocations = "parole effective; parole potentielle; circuit de la parole ; linguistique de la parole"</p> <p>Features = "SemanticType Speech_Act SuperType Act Domain Linguistics EventType Process Intentionality Yes Sound Yes AttestationPeriod 1811-1911 PoS Noun"</p> <p>SemRelations:</p> <p>Isa: exécution (Speech_Act)</p> <p>hasInstrument: langue2 (Linguistic_System)</p> <p>isTypicalOf: sujet_parlant (Human)</p> <p>hasProperty: linéarité (Property)</p> <p>hasProperty: uni-spatialité (Property)</p> <p>duality: langue2 (Linguistic_System)</p> <p>hasSynonym: langue_discursive (Speech_Act)</p> <p>hasSynonym: discours (Speech_Act)</p> <p>Constitutive: langage (Property)</p>
--	--

Tab. 1 Deux entrées lexicales : *uni-spatialité* et *parole*

5. Migration d'Access à Protégé-OWL

Il existe actuellement deux versions de la base de données lexicales. Dans la première version, le lexique est mémorisé dans une base de données relationnelle Microsoft Access. Celle-ci est gérée par une interface implémentée dans le langage de programmation VBA et articulée en une série de masques pour la création, la consultation et la modification des entrées lexicales. Le système permet en outre d'imprimer des rapports, d'exporter des parties spécifiques du lexique et d'effectuer des contrôles de cohérence.

Dans le but de satisfaire aux impératifs de standardisation et d'interopérabilité, la ressource lexicale a été successivement traduite de manière semi-automatique en langage OWL (Web Ontology Language), standard du W3C (World Wide Web Consortium) pour la représentation et le partage d'ontologies sur le Web. La version 4.1 de la plateforme Protégé¹³ a été utilisée pour la consultation et la révision du lexique OWL. Ce système intègre des plugins pour l'interrogation des données, tels que DL Query et OWL2Query, qui basent leur fonctionnement sur des raisonneurs sémantiques. D'autres plugins facilitent la phase de consultation et de révision de l'ontologie. Protégé permet en outre d'effectuer, toujours grâce à des moteurs de raisonnement, nombre de tests de cohérence de l'ontologie.

¹³ <protege.stanford.edu/>

Bien que les deux versions du lexique électronique (Access et OWL) soient encore actives à ce jour, la désactivation de la version Access est prévue dès que seront développés et intégrés dans Protégé des plugins open source dédiés aux fonctionnalités manquantes de répertoriage, d'exportation et de vérification. Ces nouveaux plugins permettront de personnaliser l'environnement Protégé pour la création et la consultation de lexiques électroniques.

Cette activité prévoit par ailleurs la production de composants supplémentaires pour la version en ligne de Protégé (Web Protégé¹⁴), qui ne requiert aucune installation sur ordinateur et est utilisable à travers un navigateur commun.

6. Interrogation du lexique

Sur le lexique, un grand nombre de requêtes peuvent être effectuées très simplement en partant d'une donnée mémorisée : une relation, un couple attribut-valeur ou une unité sémantique ; données interrogeables tant individuellement que simultanément. Il est possible d'extraire des groupes de termes partageant une ou plusieurs propriétés, selon des critères établis par le spécialiste pour satisfaire aux exigences de sa recherche. Effectuées dans un environnement formalisé garantissant la complétude des données extraites, le résultat de ces requêtes offre une vision exhaustive et approfondie de la nature componentielle et relationnelle du sens des mots.

Voici, à titre d'exemple, quelques-unes des nombreuses requêtes pouvant être effectuées :

- Quels sont les termes définis comme étant une faculté, un état ou un acte cognitif ?
- Quels sont les termes définis comme étant des créations résultant d'un processus d'association ?
- Quels sont les termes appartenant au domaine de la diachronie qui ont changé de dénomination au fil du temps ?
- Quels sont les termes dont la période d'attestation se situe entre 1893 et 1894 ?

Outre l'utilisation des plugins de l'éditeur Protégé, et afin de guider l'utilisateur peu familier avec les outils informatiques dans la formulation des requêtes, une interface conviviale de questions-réponses a été récemment développée à l'ILC. Elle se présente comme un ensemble de modèles de requêtes ('templates') en langage naturel contrôlé (Schwitter, 2010). Plusieurs types de requêtes peuvent être formulées à partir d'un seul 'template', grâce à des menus déroulants permettant de sélectionner, parmi les éléments proposés, les critères de la requête.

Chaque 'template' est composé de trois parties :

- une partie fixe qui caractérise un modèle d'interrogation (ex. : 'quels sont') ;
- une partie variable permettant de choisir dans un menu déroulant le type d'élément du lexique sur lequel effectuer la requête, ex. : 'relation sémantique' ou 'propriété' (cf. fig. 2,

¹⁴ <protegewiki.stanford.edu/wiki/WebProtege>

requête 1) ; ‘propriété sémantique’, ‘propriété morphosyntaxique’ ou ‘propriété d’usage du terme’ (cf. requête 2) ;

- une seconde partie variable permettant le choix d’une certaine propriété et de sa valeur (ex. : ‘Intentionality’, ‘no’) ; d’une certaine relation avec ou sans indication de son terme objet (ex. : ‘belongsTo’, ‘diachronie’) ou d’un certain terme du lexique (ex. : ‘diachronie’).

L’utilisateur, guidé par le modèle choisi, peut donc ainsi composer une requête en utilisant exclusivement le langage naturel.

Supposons qu’il souhaite savoir quels sont les faits intentionnels appartenant au domaine de la diachronie. Le modèle de requête sera le suivant, les choix qu’il devra opérer étant entre crochets angulaires :

Quels sont les termes caractérisés par la propriété sémantique <attribut> <valeur> et par la relation sémantique <nom de la relation> dont l’objet est <second terme de la relation> ? (cf. fig. 2, requête 4)

L’attribut sélectionné sera la propriété ‘Intentionality’, sa valeur ‘yes’ ; la relation, ‘belongsTo’ et le second terme, ‘diachronie’.

La composition des requêtes est un processus incrémental. En sélectionnant la valeur désirée dans chaque menu, celles contenues dans les menus déroulants successifs sont recalculées en conséquence. Reprenant l’exemple précédent, si l’utilisateur sélectionne dans <attribut> la propriété sémantique ‘Intentionality’, seules les valeurs associées à cette propriété apparaîtront dans le menu déroulant <valeur>. Le choix de la valeur ‘yes’ restreindra, à son tour, la liste des relations sémantiques : seules seront affichées dans le menu <nom de la relation> les relations liées aux termes ayant la propriété ‘Intentionality yes’. Enfin, le choix d’une des relations affichées, ici ‘belongsTo’, restreindra plus encore la liste des termes sur lesquels opérer un choix dans <second terme de la relation> puisque seules seront proposées les unités lexicales associées aux deux caractéristiques précédemment sélectionnées.

Les différentes typologies de requêtes sont illustrées dans la figure 2.

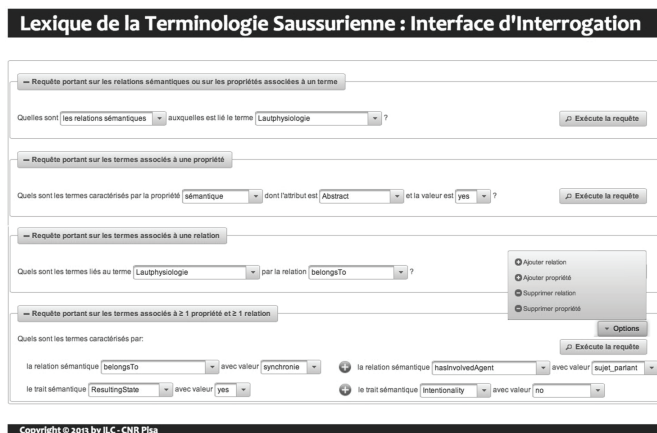


Fig. 2 Interface conviviale pour l’interrogation du lexique

Observons plus en détail le point 4 de la figure 2. Il s'agit là de la représentation formelle de la requête suivante :

Quels sont les termes dénotant des activités inconscientes (<Intentionality> <no>), typiques du sujet parlant (<hasInvolvedAgent> < sujet_parlant>), qui ont un résultat (<ResultingState> <yes>) et qui appartiennent au domaine de la synchronie (<belongsTo> <synchronie>) ?

La figure 3 illustre la visualisation sous forme de tableau du résultat de cette requête. L'utilisateur peut classer par ordre alphabétique le contenu de chaque colonne, afin de mettre tour à tour en relief différents aspects intéressants des données extraites et conserver le résultat de sa requête en l'exportant au format pdf.

TERME SOURCE	TYPE ONTOLOGIQUE
groupement	Cause_Change_Cognitive_Event
comparaison1	Cognitive_Event
analyse_subjective	Cause_Change_Cognitive_Event
classement2	Cause_Change_Cognitive_Event
articulation3	Cause_Change_Cognitive_Event
décomposition	Cause_Change_Cognitive_Event
délimitation2	Cause_Change_Cognitive_Event

Fig. 3 Visualisation des résultats d'une requête

Il apparaît donc clairement que même un utilisateur non expert en recherche d'informations peut ainsi formuler tout à fait aisément des requêtes précises et complexes en restreignant progressivement le champ de recherche par la sélection de propriétés et relations liées aux termes du lexique.

6.1. Technologies utilisées pour l'interface d'interrogation

L'interface décrite a été développée comme application Web et est accessible sur le site de l'ILC¹⁵. Le dorsal de l'application a été développé avec la technologie Java EE. À partir des données contenues dans le lexique, JENA¹⁶ construit le modèle ontologique adéquat en tenant compte de la connaissance inférée par son algorithme de raisonnement. Les modèles de requêtes sont traduits par le logiciel en requêtes SPARQL¹⁷ qui interrogent ce graphe et fournissent à l'utilisateur les réponses désirées, grâce également aux informations inférées par le raisonneur. Le frontal de l'application, qui implémente les modèles en langage naturel assisté pour les requêtes a

¹⁵ <www.ilc.cnr.it/>.

¹⁶ Framework Java pour les applications web sémantique. <jena.apache.org>.

¹⁷ Langage de requêtes pour des données représentées sous forme de triples RDF <www.w3.org/TR/rdf-sparql-query/>.

été, quant à lui, développé avec des technologies Java Server Faces (JSF)¹⁸ et Primefaces¹⁹.

7. Conclusion

Ce lexique a été créé dans le but d'offrir un instrument de recherche lexicale performant pour les études saussuriennes. Il a pour ambition de permettre une étude rigoureuse de la terminologie du grand linguiste genevois en fournissant une structuration multidimensionnelle des concepts du domaine de référence et une représentation formelle et hautement structurée de la sémantique lexicale de chacun des termes caractéristiques de sa pensée, ainsi qu'une vision d'ensemble du réseau de liens qui les unissent. Cet outil pourrait permettre une nouvelle approche de l'œuvre saussurienne et contribuer de manière déterminante à en éclairer certains aspects originaux.

CNR-Pise, Istituto di Linguistica
 Computazionale *Antonio Zampolli*
 Computazionale *Antonio Zampolli*
 Computazionale *Antonio Zampolli*

Silvia PICCINI
 Emiliano GIOVANNETTI
 Andrea BELLANDI
 Nilda RUIMY

Références bibliographiques

- Engler, Rudolf, 1968. *Lexique de la terminologie saussurienne*, Utrecht-Anvers, Spectrum, Comité international permanent des linguistes. Publication de la commission de terminologie.
- Francopoulo, Gil et al., 2007. «Lexical Markup Framework: ISO standard for semantic information in NLP lexicons», *Proceedings of Lexical-semantic and Ontological Resources*, Workshop at Gesellschaft für linguistische Datenverarbeitung, Biennial Spring Conference, Tübingen.
- Godel, Robert, 1957. *Les sources manuscrites du Cours de linguistique générale de Ferdinand de Saussure*, Genève, Droz.
- Gruber, Tom Robert, 1993. «A translation approach to portable ontologies», *Knowledge Acquisition* 5 (2), 199-220.
- Guarino, Nicola et al., 2009. «What Is an Ontology?», in: Staab, Steffen et Studer, Rudi, (ed.), *Handbook on Ontologies*, second edition, Berlin-Heidelberg, Springer Verlag, 1-17.
- Lenci, Alessandro et al., 2000. «SIMPLE: A General Framework for the development of Multilingual Lexicons», *International Journal of Lexicography* 13 (4), Special issue 'Dictionaries, Thesauri and Lexical-Semantic Relations', 249-263.
- Pustejovsky, James, 1995. *The Generative Lexicon*, Cambridge MA, The MIT Press.

¹⁸ <www.oracle.com/technetwork/java/javaee/javaserverfaces-139869.html>.

¹⁹ <primefaces.org/>.

- Ruimy, Nilda et al., 2003. « A computational semantic lexicon of Italian: SIMPLE », in : Zampolli, A. et al. (ed.), *Computational Linguistics in Pisa, Special Issue*, Pisa-Roma, IEPI, Vol. XVIII-XIX, II, 821-864.
- Saussure, Ferdinand de, 1916. *Cours de linguistique générale*, Bally, Charles et Sechehaye, Albert (ed.), Lausanne-Paris, Payot.
- Saussure, Ferdinand de, 1968-1974. *Cours de linguistique générale*, Engler, Rudolf (ed.), Vol. 1-2, Wiesbaden, Otto Harrassowitz.
- Saussure, Ferdinand de, 1922. *Recueil des publications scientifiques de Ferdinand de Saussure*, Bally, Charles et Gautier, Léopold (ed.), Lausanne, Payot.
- Saussure, Ferdinand de, 2002. *Écrits de linguistique générale*, Bouquet, Simon / Engler, Rudolf (ed.), Paris, Gallimard.
- Schwitler, Rolf, 2010. « Controlled Natural Languages for Knowledge Representation », *Proceedings of COLING 2010*, Beijing, China, 1113-1121.
- Sowa, John F., 2002. « Building, Sharing, and Merging Ontologies ». <www.jfsowa.com/ontology/ontoshar.htm>.
- Studer, Rudi, Benjamins, V. Richard, Fensel, Dieter, 1998. « Knowledge engineering: principles and methods », *Data & Knowledge Engineering*, Elsevier Ltd, Volume 25, Issues 1-2, 161-197.
- Vossen, Piek, 1999. *Eurowordnet. A multilingual database with lexical semantic networks*, Dordrecht, Kluwer.
- Apache Jena Project! Apache Jena™. <jena.apache.org>.
- RDF Vocabulary Description Language 1.0: RDF Schema. <www.w3.org/TR/rdf-schema/>.
- SPARQL Query Language for RDF W3C Recommendation 15 January 2008. <www.w3.org/TR/rdf-sparql-query/>.