

Terminologie & Ontologie : Théories et applications



Actes de la conférence

TOTh 2010

Annecy – 3 & 4 juin 2010

avec le soutien de :

- Ministère de la Culture et de la Communication, Délégation Générale à la Langue Française et aux Langues de France
- Association Européenne de Terminologie
- Société française de terminologie
- Ecole d'ingénieurs Polytech'Savoie – Université de Savoie
- Université de Sorbonne nouvelle
- Association EGC (Extraction et Gestion des Connaissances)
- ISKO (International Society for Knowledge Organization) France



Institut Porphyre
Savoir et Connaissance

<http://www.porphyre.org>

Comité scientifique

Président du Comité Scientifique : Christophe Roche

Comité de pilotage

Loïc Depecker	Professeur, Université de Sorbonne nouvelle
André Manificat	Directeur, GRETh
Christophe Roche	Professeur, Université de Savoie
Philippe Thoiron	Professeur émérite, Université de Lyon II

Comité de programme

Bruno de Bessé	Professeur, Université de Genève
Franco Bertaccini	Professeur, Université de Bologne
Gerhard Budin	Professeur, Université de Vienne
Marc van Campenhoudt	Professeur, Termisti, ISTI, Bruxelles
Danielle Candel	CNRS, Université Paris Diderot
Stéphane Chaudiron	Professeur, Université de Lille 3
Rute Costa	Professeur, Universidade Nova de Lisboa
Luc Damas	MCF, Université de Savoie
Sylvie Desprès	Professeur, Université Paris 13
François Gaudin	Professeur, Université de Rouen
Anne-Marie Gendron	Chancellerie fédérale suisse, Section terminologie
Jean-Yves Gresser	Ancien Directeur à la Banque de France
Olivier Haemmerlé	Professeur, Université de Toulouse
Michèle Hudon	Professeur, Université de Montréal
John Humbley	Professeur, Université Paris 7
Michel Ida	Directeur MINATEC, CEA
Hendrik Kockaert	Professeur, Lessius Hogeschool (Anvers)
Michel Léonard	Professeur, Université de Genève
Pierre Lerat	Professeur honoraire, Equipe Condillac
Widad Mustafa	Professeur, Université de Lille 3
Fidelma Ní Ghallchobhair	Foras na Gaeilge (The Irish-Language Body)
Henrik Nilsson	Terminologocentrum TNC, Suède
Jean Quirion	Professeur, Université d'Ottawa
Renato Reinau	Suva, Lucerne
François Rousselot	MCF, Université de Strasbourg
Gérard Sabah	CNRS, Orsay
Michel Simonet	CNRS, Grenoble
Marcus Spies	Professeur, Université de Munich
Dardo de Vecchi	Professeur associé, Euromed-Management

Comité d'organisation :

Responsable : Luc Damas
Samia Chouder, Joëlle Pellet

Avant propos



Cette année la conférence a été précédée d'une journée de formation consacrée à la terminologie et l'ontologie, à leurs liens et leurs apports mutuels. L'intérêt qu'a suscité cette journée nous amènera certainement à réitérer l'opération les années suivantes.

Le succès de la conférence d'ouverture de notre collègue Frédéric Nef, portant sur l'ontologie prise dans sa dimension philosophique, a montré, s'il en était encore besoin, la richesse d'une approche pluridisciplinaire.

Animées par différents présidents, les sessions ont alterné présentations théoriques et démonstrations de systèmes, offrant ainsi l'opportunité à plusieurs industriels de nous parler de leurs projets. L'éventail des sujets abordés, à travers les quatorze présentations retenues (incluant la conférence d'ouverture) réparties sur deux jours, illustre la richesse mais aussi la vitalité de notre communauté : aide à la traduction, thésaurus multilingue, phraséologie, entité nommée, recherche d'information, etc. L'« actualité » n'a pas été oubliée à travers une ontologie des risques financiers.

Enfin, les Conférences TOTh sont devenues internationales à partir de cette année avec le français et l'anglais comme langues officielles. Le comité de programme s'est ouvert à de nouveaux membres portant à dix le nombre de pays représentés et à plus de 40% le nombre de personnalités étrangères. Gageons que cette ouverture sera prometteuse.

Christophe Roche
Président du Comité Scientifique

Table des matières

CONFERENCE INVITEE

<i>L'Ontologie au miroir de la Terminologie</i>	9
Frédéric Nef	

ARTICLES

<i>Le travail sur la représentation (visuelle) des connaissances en terminologie : un retour d'expérience</i>	31
Dardo de Vecchi	
<i>Une « ontoterminologie » pour les interprètes de conférence</i>	53
Elisa Veronesi, Franco Bertaccini	
<i>Semiotic Triangle Revisited for the Purposes of Ontology-based Terminology Management</i>	83
Igor Kudashev, Irina Kudasheva	
<i>L'ontoterminologie pour la recherche d'information sémantique</i>	101
Luc Damas, Christophe Tricot	
<i>Modélisation des dénominations ontologiques</i>	117
Benjamin Diemert, Marie-Hélène Abel, Claude Moulin	
<i>Filtrage des Entités Nommées par des méthodes de Fouille de Textes</i>	141
Mathieu Roche	
<i>Ontologies des risques financiers – Continuité d'activité, gestion de crise, protection des infrastructures critiques financières</i>	155
Jean-Yves Gresser	
<i>Vers une ontologie pour le domaine de l'analyse de sécurité des systèmes de transport automatisés</i>	177
Lassaâd Mejri, Habib Hadj-mabrouk, Patrice Caulier	

DEMONSTRATIONS

<i>Une « ontoterminologie » pour les interprètes de conférence – Un outil développé au sein de l’environnement académique</i>	203
Elisa Veronesi, Franco Bertaccini	
<i>ITM, une infrastructure sémantique pour la maintenance du thésaurus multilingue Eurovoc</i>	207
Thomas Francart, Charles Teissède	
<i>Approche onomasiologique de la phraséologie transdisciplinaire des écrits scientifiques : la recherche sémantique dans les textes dans le cadre du projet Scientext</i>	211
Falaise Achille, Tutin Agnès	
<i>Ontoterminologie : méthode et mises en œuvre</i>	217
Marie Calberg-Challot, Christophe Tricot	
<i>Libellex, plateforme de travail multilingue et référentiel terminologique d’entreprise</i>	225
François Brown de Colstoun, Estelle Delpech	
<i>Pages blanches</i>	230

Une « ontoterminologie » pour les interprètes de conférence

Elisa Veronesi, Franco Bertaccini

Résumé : La terminologie et la création d'ontologies s'avèrent de plus en plus utiles pour tous ceux qui travaillent dans les domaines de la traduction et de l'interprétation. Ce projet a donc pour but d'étudier l'usage d'arbres ontologiques en tant que support à l'interprétation simultanée. Nous avons essayé de mettre la terminologie au service des interprètes en concevant une ontologie dynamique et modulaire pouvant être utilisée par les interprètes directement en cabine. Pour ce faire, notre projet s'est basé sur la création d'une conceptualisation formelle d'un domaine spécialisé, développée à partir de ses termes et concepts fondamentaux. La méthodologie de sa structuration repose donc sur la façon dont les termes (et les concepts auxquels ils font référence) sont reliés entre eux d'un point de vue sémantique et logique. Pour cette raison, ce support pourrait être défini comme une « ontoterminologie ». Le domaine conceptualisé est celui du SIDA. En particulier, notre analyse a porté sur l'étude des manifestations cutanéomuqueuses se présentant au cours de l'infection par le VIH. Nous avons créé deux corpus de langue spécialisée, un en français et un en italien, desquels nous avons extrait les concepts et les termes fondamentaux du domaine et du sous-domaine, ainsi que les relations sémantiques les reliant. Ensuite, nous avons développé une ontoterminologie bilingue italien-français très spécialisée et strictement liée au domaine. Le champ de la biomédecine étant très complexe et articulé, les instances, les relations, ainsi que la structure des diagrammes ont été définies manuellement, en utilisant le logiciel graphique Visio 2007. On a essayé, d'une part, de produire une ontologie flexible, synthétique mais exhaustive ; de l'autre, de faire en sorte qu'elle permette de prévoir (ou de suivre) l'évolution des discours de la conférence, en affichant non seulement les termes et les équivalents qui relèvent du sujet, mais aussi les relations sémantiques les reliant. Enfin, nous avons mené un test empirique pour vérifier la « lisibilité » et l'utilité effective du support avec des interprètes professionnels. L'analyse de leurs performances et opinions nous a permis d'évaluer les avantages concrets apportés par l'outil réalisé.

Mots-clés : Terminologie, ontologie, ontoterminologie, interprétation de conférence, modulaire, multi-niveaux, relations sémantiques, corpus, ontologie régionale, ontologie de domaine, SIDA, VIH, pathologies cutanéomuqueuses, terminologie médicale, médecine.

1. Introduction

Dans un monde de plus en plus globalisé et multiculturel, la terminologie et la structuration du savoir par les ontologies sont deux disciplines essentielles pour le développement de nombreux domaines scientifiques et industriels, ainsi que pour les recherches relatives aux langues et à la linguistique. En parallèle, la traduction et l'interprétation revêtent une importance majeure pour ce qui est, d'une part, de la préservation du multilinguisme en lui-même, et, de l'autre, de la croissance et de la continuation de toute activité industrielle, commerciale, sociale et de recherche. Interprètes et traducteurs ont donc largement recours à la terminologie et aux ontologies pour développer des bases de données exhaustives, modulaires et interactives, représentant tout domaine de connaissance par des critères sémantiques, afin de relier chaque concept au terme correspondant en toute langue ou bien langage de spécialité. Cependant, le foisonnement multilingue et multiculturel international a atteint un stade où une progression qualitative devient urgente. Étant donné que le catalogage informatique des données multilingues évolue assez rapidement et que l'on peut désormais compter sur des logiciels et des langages de programmation de pointe, l'urgent est de se pencher sur les outils de structuration et référencement des contenus multilingues. Bref, il devient primordial de pouvoir créer des réseaux sémantiques multilingues qui soient mieux articulés et plus flexibles que ceux que l'on peut développer aujourd'hui, afin de produire des structurations du savoir qui soient concrètement utiles pour l'apprentissage et le travail.

Ce projet se situe précisément à ce point de la recherche en interprétation, et a pour but d'étudier l'usage d'arbres ontologiques complexes en tant que support à l'interprétation simultanée. Nous avons essayé de mettre la terminologie au service des interprètes en concevant une ontologie dynamique et modulaire pouvant être utilisée par les interprètes en cabine, durant leur travail. Pour ce faire, notre projet s'est basé sur la création d'une conceptualisation formelle d'un domaine spécialisé (une ontologie), développée à partir de ses termes et concepts fondamentaux. La méthodologie de sa structuration repose par conséquent sur la façon dont les termes (et les concepts auxquels ils font référence) sont reliés entre eux d'un point de vue sémantique et logique. Pour cette raison, ce support pourrait être défini comme une « ontoterminologie ». Nous allons maintenant décrire plus en détail comment notre projet a été structuré et quelles approches ont été suivies pour son développement.

2. Terminologie et ontologies

Aujourd'hui, la terminologie et la création d'ontologies sont devenues deux facteurs clés pour le repérage, la catégorisation et la structuration des savoirs et des connaissances, un sujet très important pour les linguistes, les traducteurs et les interprètes. En particulier, la terminologie joue un rôle primordial, puisqu'elle permet à la fois d'encadrer les connaissances dans des bases de données structurées et plus ou moins « standardisées » et de partager ces connaissances avec d'autres utilisateurs/experts du domaine. Elle se situe donc au carrefour de plusieurs disciplines. Le terme lui-même est donc un élément « clé », parce qu'il véhicule non seulement les concepts et les idées fondamentales que l'émetteur souhaite transmettre, mais aussi la structure du système notionnel du domaine à l'intérieur duquel il est utilisé. Le système notionnel d'une terminologie est donc défini par un ensemble de concepts qui constitue, pour une communauté donnée, une « description du réel ». Cela nous amène à la notion d'ontologie, pratique strictement liée à celle de la terminologie.

La notion d'ontologie est devenue un élément clé dans toute une gamme d'applications faisant appel à des connaissances. Ces connaissances sont modélisées de manière structurée, et « exprimées à l'aide de langages ayant une sémantique bien définie » (Charlet, 2002 : 3). Plus en détail : « Les ontologies sont des systèmes conceptuels destinés à fournir les notions élémentaires à la formulation des connaissances dont on dispose sur un sujet donné » (Bachimont, 2004 : 128), et exprimés à l'aide de langages opérationnels standards. Une ontologie organise donc les concepts et les relations pertinents et consensuels dans un domaine pour une application donnée. Par conséquent, une ontologie est, en d'autres mots, la conceptualisation des objets reconnus comme existant dans un domaine, de leurs propriétés et des relations les reliant. Comme en terminologie, la signification des termes repose sur une conceptualisation du monde et un vocabulaire de termes pour en parler. L'ontologie vise, *ipso facto*, les mêmes propriétés de consensus, de cohérence, de partage et de possibilité de réutilisation que la terminologie.

2.1. Un support pour l'interprétation simultanée

Dans ce contexte, la terminologie et la construction d'ontologies peuvent être d'une certaine manière très utiles pour la traduction et l'interprétation, aussi. L'énorme contribution que la terminologie apporte au domaine de la traduction spécialisée (en termes de création et consultation de bases de données spécifiques uni- ou multilingues et d'aide au choix des mots/termes les plus adéquats pour un certain type de texte ou registre) est bien connue,

tandis que l'utilisation d'ontologies de la part des traducteurs ou de terminologues est en train de se frayer un chemin. Au contraire, l'usage de banques terminologiques et d'ontologies en interprétation de conférence est encore assez négligé ou méconnu.

Pourtant, lors de la préparation en vue d'une conférence, les bases de données terminologiques seraient bien entendu très utiles pour les interprètes, surtout s'ils ne doivent pas les produire eux-mêmes, puisque la création de fiches terminologiques est un procédé lent et complexe et les interprètes n'ont souvent que quelques jours pour étudier le sujet de la conférence. Mais, une fois en cabine, il est encore très difficile d'utiliser un ordinateur portable au moment de l'interprétation simultanée et de consulter des fiches, puisque, en principe, l'orateur parle trop vite pour que l'interprète puisse rechercher le terme souhaité. Des recherches sont en cours pour comprendre comment fournir aux interprètes de simultanée un support informatique qui puisse les aider dans l'accomplissement de leur tâche. Les glossaires traditionnels sont difficiles à consulter lors de l'interprétation, puisque l'interprète devrait les feuilleter au moment même où il parle et, à cause de l'effort produit pour cette recherche, il risquerait de commettre des erreurs ou de perdre des informations durant la phase d'écoute ou de reformulation. Par contre, l'ordinateur portable en cabine pourrait résoudre le problème, dans la mesure où il permettrait une consultation rapide d'un dictionnaire ou d'une base de données terminologique grâce à un logiciel spécifique. La voie dans la direction de l'utilisation d'outils numériques en tant que support aux interprètes de conférence semble donc ouverte.

2.2. Une « ontoterminologie » pour interprètes

Ce travail a donc eu pour but l'étude d'un outil linguistique gérable avec un ordinateur portable comme support à l'interprétation simultanée. Nous avons analysé la possibilité de mettre la terminologie au service des interprètes à l'aide d'ontologies, qui – tout comme les fiches terminologiques – représentent la conceptualisation d'un domaine, mais en soulignant les relations sémantiques et logiques qui lient les termes à travers des arbres et/ou des graphes. Le type d'ontologie nécessaire étant destiné spécifiquement à des interprètes de conférence et, par conséquent, la création d'arbres modulaires et très spécialisés s'est imposée. Nous avons besoin en fait de terminologies et d'ontologies bilingues – ou même trilingues – liées spécifiquement au sujet de la conférence en préparation. Nous avons donc décidé de modéliser un domaine de connaissance particulier et restreint par le truchement d'une ontologie régionale et structurée *ad hoc* répondant aux besoins des interprètes

de conférence, puisque un support linguistique de ce type pourrait constituer une aide précieuse pour le travail en cabine.

Pour la structuration et le développement de notre projet nous nous sommes basés en particulier sur les travaux menés en France par le professeur Christophe Roche et l'Équipe Condillac du Laboratoire Listic, ainsi que sur les projets développés en Italie par le professeur Franco Bertaccini au sein du laboratoire de Terminologie de l'Université pour Interprètes et Traducteurs de Forlì (SSLMIT). Leurs études portent en fait sur la nécessité de fusionner les disciplines de la terminologie et de l'ontologie pour pouvoir développer une approche que l'on peut qualifier de « ontoterminologique »¹. Cette approche repose sur l'idée qu'à l'heure actuelle il est désormais nécessaire de replacer le concept et sa dénomination au centre de la terminologie, tout en préservant sa dimension sociolinguistique par la prise en compte des termes d'usage à travers la langue de spécialité. Si employée en tant que critère de construction pour la réalisation d'ontologies (schémas conceptuels) relatives à un domaine donné, cette méthodologie permet de conceptualiser de façon explicite ce champ de connaissance et de rendre formelles les informations collectées et schématisées – qui sont donc manipulables, modifiables et partageables par un ordinateur. Pour cela, nous avons « emprunté » le nom créé par Christophe Roche, et appelé notre prototype une « ontoterminologie ».

De surcroît, on a également pris en compte les principes de la « terminologie conceptuelle »², une approche qui préconise l'élaboration d'une terminologie spécifique à un champ de connaissance sur la base de l'analyse des concepts auxquels les termes font référence et qui aboutit à une systématisation de ceux-ci en structures cohérentes telles que des arbres et des graphes conceptuels. De telles structures permettent de représenter graphiquement les domaines et les sous-domaines d'application de certaines connaissances, ainsi que les relations sémantiques qui relient les concepts qui en sont à la base. Ces schémas sont modélisés à partir des requêtes spécifiques des commanditaires et/ou des usagers finaux du catalogue terminologique réalisé. Par conséquent, si la conceptualisation ainsi créée est bien en langue

¹ En employant la dénomination créée par C. Roche dans son article *Le terme et le concept : fondements d'une ontoterminologie*, TOTh, 2007.

² Cette approche a été en effet récemment suivie également par l'ISO, qui a employé les principes de la terminologie conceptuelle pour les règles, les définitions et la structure de la nouvelle norme ISO 704: 2000. Plus de détails concernant les canons de travail relatifs à la terminologie ainsi dite "conceptuelle" au sein de la norme ISO 704: 2000 peuvent être trouvés dans l'article de Wright "The Once and Future ISO 704", publié dans le magazine de secteur *eDITion* n. 1/2007.

naturelle, elle est également définie dans un langage formel selon des principes épistémologiques. Une conceptualisation terminologique ainsi définie donne, de ce fait, la structure de base pour une ontologie qualifiée de « régionale » et, donc de « domaine », qui ne cherche pas à modéliser le monde en général, comme le font les ontologies de type « *top level* », mais un contexte d'usage bien spécifique dans un domaine ciblé.

En effet, les interprètes requièrent un support qui leur permette non seulement de retrouver des termes et leur équivalent, mais aussi de les placer dans le contexte spécifique de la conférence grâce à la représentation graphique des relations entre les termes et les concepts. Cela signifie que la sélection et la structuration de la terminologie à la base de ce type d'ontologie devraient être étudiées attentivement et préparées *ad hoc* sur la base du sujet de la conférence et de la combinaison linguistique demandée à l'interprète. Et, question qui ne peut être négligée, l'ontologie elle-même devrait permettre à l'interprète de prévoir (ou de suivre) le(s) discours(s) en affichant non seulement tous les termes et les équivalents qui relèvent du sujet abordé, mais aussi les relations sémantiques et logiques qui les relient.

2.3. Les objectifs à atteindre par la création du prototype

Nous nous sommes donc attachés à modéliser une ontologie régionale schématisant la terminologie relative à un domaine scientifique spécifique qui se base sur le primat des concepts et des relations sémantico-logiques les reliant. Nos objectifs étaient les suivants :

- Produire une ontoterminologie bilingue italien–français interactive, dynamique, modulaire, taillée sur le domaine pour lequel elle a été conçue, contextuelle, répondant aux besoins spécifiques des interprètes de conférence.
- Développer le prototype d'un outil concis et exhaustif en même temps, qui soit intuitif et facile à utiliser non seulement lors de la préparation au travail, mais aussi au cours de la conférence-même, en simultanée.
- Produire des schémas ontologiques présentant une hiérarchie des concepts du domaine médical analysé basée sur :
 - les relations hiérarchiques verticales d'identité (*is_a*) et d'appartenance (*part_of*) ;
 - les relations non hiérarchiques causales et temporelles, qui sont d'importance majeure pour la conceptualisation des domaines de médecine et de biomédecine.

- Faire en sorte que les diagrammes ontologiques permettent à l'interprète de « prévoir » ou, au moins, suivre l'évolution du discours, pour que le contexte d'usage des termes fournisse à l'interprète des points de repère.
- En parallèle, permettre à l'utilisateur de créer différents parcours conceptuels de consultation des fiches basés sur différents « points de départ » (selon le concept ou le terme duquel on commence à « lire » les arbres ontologiques).

3. Les phases concrètes de réalisation du projet

Pour réaliser notre projet, nous avons mené une étude empirique portant sur la création d'une ontologie spécifique à un domaine scientifique. Les conférences scientifiques, en effet, se prêtent davantage à être « prévues » (ou analysées, si l'interprète a la chance d'avoir les textes à interpréter en avance) et représentées de façon graphique et schématique. La raison en est que ces types de discours ou de relations utilisent en principe un langage très spécialisé (et qui, donc, ne devrait pas poser de problème d'interprétation au niveau des métaphores ou de la polysémie) et sont d'ordinaire assez rigoureusement structurés. Ainsi, il devrait être possible de représenter la démarche logique et sémantique d'un texte par une ontologie modulaire basée sur un corpus terminologique créé *ad hoc* sur la base de textes scientifiques accrédités (le support d'un expert du domaine s'impose donc).

Le domaine choisi est le champ médical, qui est particulièrement approprié pour notre étude aussi bien pour ses caractéristiques intrinsèques (l'existence d'une langue très spécialisée dont les termes ont, dans la majorité des cas, un équivalent dans les autres langues étrangères ; la facilité de repérer les documents à partir desquels on peut extraire une terminologie bilingue) que pour ses caractéristiques extrinsèques (beaucoup de conférences internationales consacrées à ce secteur sont organisées chaque année ; par conséquent, comme nombre d'interprètes choisissent de se spécialiser en médecine, ce travail pourrait s'adresser à un groupe de professionnels assez vaste). En particulier, l'étude a porté sur le VIH et, plus spécifiquement encore, sur les symptômes dermatologiques liés à l'apparition de cette pathologie. Ce choix a été fait de sorte que notre « simulation » se rapproche le plus possible de la réalité des conférences, où, généralement, les experts et les professionnels se focalisent sur un sujet très spécifique.

Les langues des corpus terminologiques ont été le français et l'italien, soit deux de nos langues de travail, pour lesquelles il est assez facile de repérer du

matériel dans les bibliothèques ou sur la toile. L'ontologie a donc été produite pour une interprétation du français en italien.

3.1. L'étude du domaine et la création des corpus de référence

Tout d'abord, nous avons donc décortiqué le sujet du SIDA et du VIH pour mieux nous orienter dans la recherche et comprendre les textes lors de la phase d'extraction terminologique. Ci-après, un exemple des graphes réalisés en italien pour schématiser le domaine à représenter dans notre ontoterminologie.

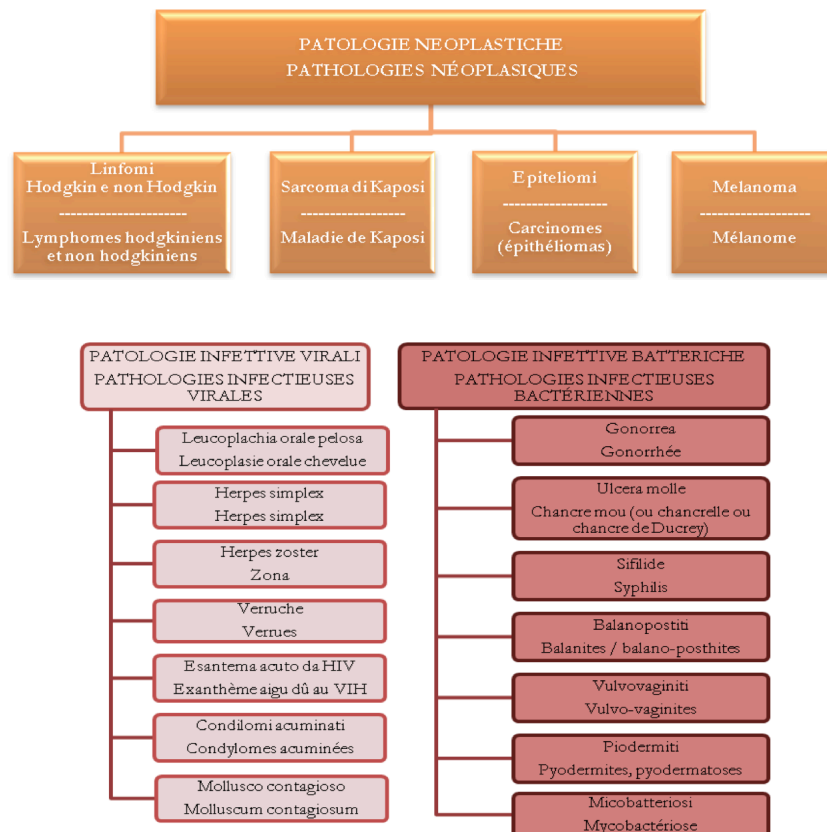


FIG. 1 – *pathologies néoplasiques, pathologies infectieuses virales et bactériennes dues à l'infection par le VIH.*

Ensuite, il a été nécessaire de repérer un ou plusieurs experts du secteur. Leur expertise a été essentielle pour choisir les documents pour la création des corpus terminologiques et la rédaction d'un discours *ad hoc* pour tester

L'ontologie produite durant l'interprétation simultanée. Les deux corpus créés ont été composés comme suit :

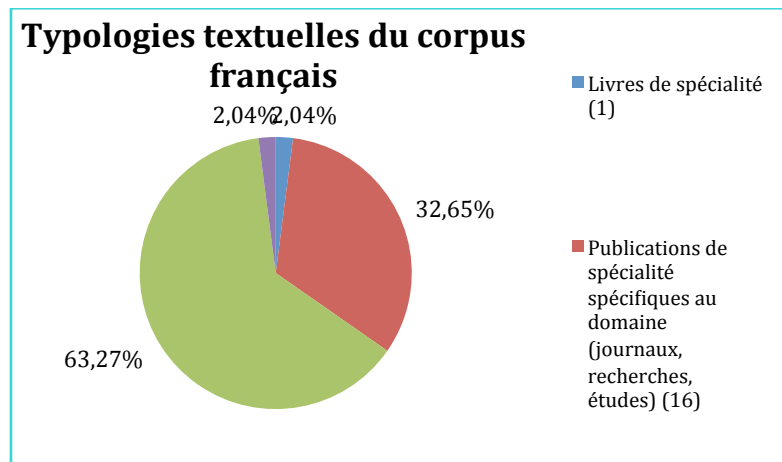
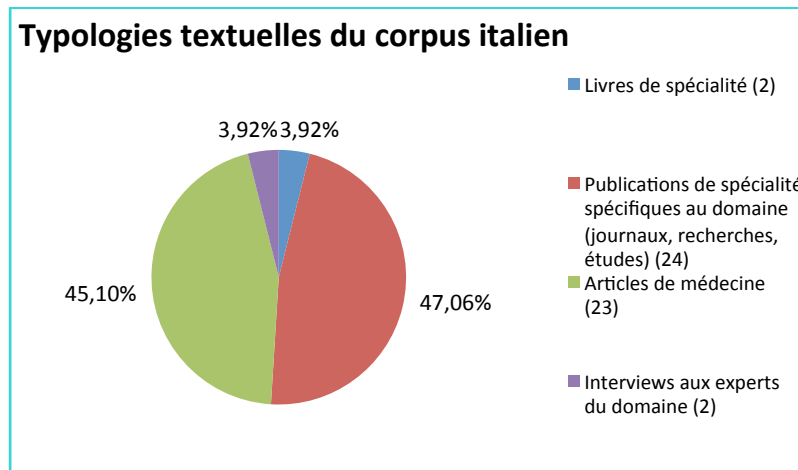


FIG. 2 – Diagrammes présentant les typologies textuelles des documents composant les corpus italien et français

Les corpus sont homogènes en termes de fiabilité, niveau de spécificité et destinataire (experts ou semi-experts) ; mais ils ne le sont pas complètement en termes de typologie textuelle. Ils incluent en effet des livres, des articles et des publications spécialisés sur papier, ainsi que des recherches, des articles et des études fouillés spécifiques publiés sur des sites Internet de médecine.

a) L'extraction des candidats termes et l'identification des termes et des relations pour l'ontoterminologie

Nous sommes ensuite passés à la phase d'extraction terminologique bilingue, pour laquelle nous avons utilisé Wordsmith, un logiciel qui permet d'identifier les candidats termes d'un domaine de spécialité à travers une comparaison statistique de la fréquence d'apparition des mêmes termes à l'intérieur d'un corpus de spécialité et d'un corpus de type générique. Les deux corpus ont les dimensions suivantes :

- Corpus de langue de spécialité en italien: 551.257 mots
- Corpus de référence en italien: autour de 2 millions de mots
- Corpus de langue de spécialité en français: 552.079 mots
- Corpus de référence en français: autour de 2 millions de mots

L'analyse à travers des paramètres configurés par nous-mêmes au sein de Wordsmith nous a permis d'extraire une longue liste de candidats termes simples et composés (syntagmes et relations sémantiques) dans les deux langues. Nous avons extraits des listes de un à cinq termes regroupés (*clusters*), formant des termes composés ou des syntagmes. L'extraction a affiché une correspondance intéressante entre les positions des candidats dans les listes italiennes et françaises: les termes définissant le même concept sont très souvent placés à des positions similaires dans le classement. Cet aspect nous a aidés lors de la recherche des équivalents interlinguistiques et démontre qu'il est tout à fait possible de schématiser un domaine scientifique et suffisamment spécifique par une ontologie bilingue, puisqu'en principe les mêmes concepts sont présents dans les deux langues. Comme on peut le voir dans le tableau qui suit, très souvent les candidats termes ont une position et une fréquence très similaires.

N	WORD	FREQ.		N	WORD	FREQ.
1	HIV	4.659		1	VIH	4.036
2	INFEZIONE	2.537		2	TRAITEMENT	3.613
3	PAZIENTI	2.370		3	CHEZ	2.598
4	TERAPIA	2.482		4	ÊTRE	2.277
5	VIRUS	1.805		5	PATIENTS	1.992
6	AIDS	1.570		6	INFECTION	1.888
7	FARMACI	1.605		7	PEUT	1.552
8	ANTIRETROVIRALE	1.293		8	SIDA	1.352
9	INFEZIONI	1.225		9	VIRUS	1.441
10	CELLULE	1.071		10	PERSONNES	1.300

11	LESIONI	940	11	ÉTÉ	1.232
12	RISCHIO	1.330	12	LÉSIONS	1.035
13	VIRALE	865	13	PATIENT	1.249
14	SOGGETTI	1.202	14	VIRALE	898
15	HAART	821	15	INFECTIONS	875
16	PAZIENTE	756	16	CHARGE	1.343
17	LINFOCITI	764	17	RISQUE	1.279
18	CASI	1.391	18	CAUSE	1.856
19	PATOLOGIA	1.067	19	PEUVENT	720
20	TRATTAMENTO	1.199	20	CELLULES	642
21	DIAGNOSI	712	21	PEAU	639
22	CUTE	575	22	IMMUNITAIRE	589
23	CUTANEE	2.010	23	THÉRAPEUTIQUE	577
24	KAPOSI	430	24	TRAITEMENTS	734
25	MALATTIA	415	25	PORTE	7.670
26	OPPORTUNISTICHE	411	26	CUTANÉES	638
27	SARCOMA	410	27	AYANT	544
28	TOSSICITÀ	439	28	LYMPHOCYTES	535
29	CLINICA	515	29	PRÉVENTION	511
30	CAUSA	419	30	CAS	1.815

FIG. 3 – Les 30 premiers résultats de l'extraction des candidats termes simples (unigrammes) par Wordsmith, en italien et en français. Le montant total de candidats extraits s'élève à 500 termes (simples, complexes et composés)

b) L'identification des relations sémantiques fondamentales

Les relations essentielles pour la conceptualisation ont été repérées lors de nos entretiens avec les experts, ainsi que par l'extraction des termes des corpus de langue de spécialité. L'ontoterminologie inclut:

- Des relations hiérarchiques, et en particulier des relations génériques (surtout la relation de subsumption) et des relations partitives;
- Des relations non hiérarchiques, et donc associatives pragmatiques, et en particulier des relations séquentielles (temporelles) et causales.

Dans notre cas, nous avons dû "dilater" et adapter sensiblement les standards d'identification et d'emploi des relations sémantiques. Tout d'abord, parce que le sous-domaine est plutôt complexe. Ensuite, puisque, vu la nécessité de produire un outil, même si encore prototypique, qui réponde aux

besoins des interprètes, l'exigence s'est imposée d'adapter de façon pertinente la structure ontologique tout entière. La biomédecine est en effet un domaine tellement articulé et complexe que la théorie formelle de la classification médico-biologique est encore à l'état embryonnaire³. De suite, les relations principales employées pour l'ontoterminologie:

Italien	Français
Causa	Cause
Colpisce, interessa	Atteint, touche (à)
È	Est
È caratterizzato/a da	Est caractérisé/e par
Si manifesta, si esprime con / tramite	Se manifesta, s'exprime par
Si tratta, si cura con	Est traitée à / par Est soigné(e) à / par Le traitement repose sur
Dimostra, prova	Démontre, prouve Fait porter le diagnostic de
Origina, scatena, porta a	Cause, engendre, déclenche, porte à
Si differenzia / distingue in Può differenziarsi / distinguersi in	Se divise, se différencie en Peut se diviser, se différencier en
Sviluppa, porta allo sviluppo di	Développe, Porte à, conduit à développer

TAB. 1 – Tableau des principales relations sémantiques employées pour l'ontoterminologie

3.2. La création d'un glossaire thématique bilingue

Une fois les candidats termes extraits, nous avons procédé à les analyser avec l'aide des experts du secteur, afin de les valider et détecter les relations sémantiques et logiques les plus importantes. En mettant en relation les termes

³ Smith (2003: 90) affirme en fait: "Biological classes are marked always by an opposition between standard or prototypical instances and a surrounding penumbra of non-standard instances (not all the instances of the class human being are marked by the presence of amputation stumps or pituitary tumors)".

simples et composés identifiés, nous avons construit un glossaire très spécifique, à partir duquel nous avons ensuite modélisé notre ontologie *ad hoc*. La méthode de sélection terminologique et de réalisation du glossaire a suivi les étapes suivantes:

- identification des termes de domaine présents dans les textes analysés;
- analyse du signifié et de la compréhension des concepts véhiculés par les termes, ainsi que de leur intension et extension;
- analyse contrastive entre les systèmes conceptuels des deux langues et recherche des équivalents dans la langue d'arrivée de l'interprétation simultanée (le français).

Manifestazioni cutaneo-mucose in presenza di HIV	Manifestations cutanéomuqueuses en présence de VIH
Accesso	Poussée
Affezione cutanea	Affection cutanée
Aftosi, stomatite aftose afta - forma <i>minor</i> - forma <i>maior</i> - forma erpetiforme	Stomatite aphteuse, aphtose aphte - forme mineure - forme majeure - forme herpétiforme
Alone ecchimotico	Halo ecchymotique
Angiolipoma	Angiolipome
Angioma	Angiome
Angiomatosi epitelioide	Angiomatose épithélioïde
Anorettale	Anorectal
Balanopostite manifestazioni eritemato-essudative	Balanite manifestations érythémateuses exsudatives
Benignità	Bénignité
Benigno	Bénin
Bolla rilievo cutaneo circoscritto, costituito da una cavità a contenuto liquido, di dimensioni superiori a quelle della vescicola	Bulle dermatologique phlyctène rempli d'une sérosité contenant ou non du sang, ou l'ampoule due aux frottements et aux brûlures, plus grand qu'une vésicule

TAB. 2 – Extrait du glossaire réalisé à partir des termes extraits

4. La réalisation et la formalisation de l'ontologie

Aujourd'hui, les langages artificiels et les logiciels utilisés pour la création automatique d'ontologies (tels que Protégé, OWL, RDF) portent principalement sur deux relations sémantiques : la relation de subsomption (*is-a*) et la relation mérologique (partitive, *part-of*). La raison en est qu'il s'agit des relations principales qui lient les termes insérés dans les bases de données traditionnelles. Ces types de bases de données cataloguent les termes d'un domaine donné de façon hiérarchique suivant la direction hyperonyme \Rightarrow hyponyme (de « vin blanc » à « Chardonnay ») ou bien classe supérieure \Rightarrow classe inférieure (de « mammifère » à « chien »). Toutefois, ces relations ne suffisent pas à conceptualiser un vrai discours, lors duquel un orateur utilise des liens sémantiques et logiques beaucoup plus nombreux pour expliquer ou décrire tel ou tel sujet.

En d'autres mots, ce dont nous avons besoin est de pouvoir formaliser « informatiquement » les données ontologiques elles-mêmes. Il existe pour cela différents langages et modèles de représentation qui, toutefois, ne permettent pas de représenter ces informations en gardant une différence graphique « rationnelle » entre les différentes relations sémantiques à la base d'un domaine de connaissance. De surcroît, un logiciel adapté aux besoins des interprètes devrait permettre de garder et formaliser ce qu'on pourrait appeler « la valeur ajoutée » conférée à toute conceptualisation par son concepteur, c'est-à-dire la façon dont l'« ontoterminologue » a représenté visuellement les éléments de l'ontologie, en les approchant, les éloignant, les mettant sur une feuille plutôt qu'une autre, etc.

La modélisation manuelle des arbres ontologiques. La représentation visuelle d'une ontoterminologie pour interprètes de conférence est donc un élément déterminant de la qualité du résultat. C'est justement là, à notre avis, la plus grosse différence avec ce qui existe déjà dans l'état de l'art : les différents outils et modèles que nous avons testés ne s'intéressent qu'à la modélisation des relations et des concepts, avec plus ou moins de richesse de reproduction ; mais la représentation graphique qui peut en découler est, *a priori*, calculée par l'ordinateur selon des algorithmes assez basiques, qui n'envisagent pas la complexité de visualisation. Ces logiciels sont configurés selon des algorithmes très efficaces pour créer des arbres géométriquement bien structurés, qui calculent la largeur des branches inférieures pour positionner les parents et les instances supérieures de façon équilibrée et correcte. Toutefois, peu de contrôle est exercé sur le rendu visuel effectif qui, très probablement, sera moins efficient que celui que l'« ontoterminologue » a déterminé manuellement pour l'usage de l'interprète. Et la moindre altération du graphe

lui-même sera susceptible de déséquilibrer l'algorithme, et donc de produire un résultat sensiblement différent, pouvant rendre incorrecte ou invalide la conceptualisation. Pour cette raison, pour produire notre prototype nous avons créé manuellement les ontologies nécessaires par le logiciel graphique Visio 2007 de Microsoft, qui permet de réaliser des arbres et des graphes très variés.

4.1. La construction de l'ontoterminologie

Pour la construction et la structuration des diagrammes nous avons suivi l'approche suivante:

- 1) Construction d'un lexique « orthonormé », composé par les termes identifiés comme fondant le domaine (et dénotant donc un concept spécifique) et, éventuellement, par leurs synonymes (selon les lexiques employés par les utilisateurs) à partir de la documentation fournie par les spécialistes du secteur.
- 2) Identification des typologies épistémologiques des termes simples et composés repérés (en prenant en compte l'essence des termes et la définition de leur signification par l'expression de leurs différences spécifiques, des termes étant par exemple les véritables instances conceptuelles du domaine, d'autres étant leurs attributs ou des relations).
- 3) Modélisation et conséquente construction des diagrammes conceptuels schématisant le domaine sur l'empreinte de l'« Arbre de Porphyre », qui est à la base des ontologies modernes.

Nous nous sommes donc basés sur une combinaison et un usage concerté des méthodes suivantes:

- 1) Application du modèle épistémologique.
- 2) Application du modèle computationnel.
- 3) Formalisation logique *a posteriori*, sur la base de nos connaissances du domaine et surtout des indications fournies par les médecins spécialistes.

Ainsi, nous avons été à même de confectionner des ontologies très spécifiques et pertinentes en mettant en évidence les relations sémantiques les plus utiles et appropriées selon le discours que l'on prévoit d'interpréter.

4.2. Critères de développement des diagrammes

Représentation des relations. Les instances sont liées par des relations sémantiques de différent type (hiérarchiques et non hiérarchiques, verticales et

horizontalles), explicitées par des « satellites » pour faciliter la lecture. Nous avons repéré ces relations lors du choix des candidats termes pour l'ontologie, puisqu'elles définissent et caractérisent le domaine de la biomédecine autant que les relations d'identité/appartenance.

Structuration graphique. Les formes, les couleurs et l'aménagement des diagrammes et des instances se basent sur des critères logiques ainsi qu'épistémologiques. Toute instance est caractérisée par une forme, une couleur et une position qui dépendent de son niveau, ses attributs, son essence et sa fonction. Nous présentons de suite une feuille de la conceptualisation en guise d'exemple. Les correspondants en français apparaissent tous, pour mieux montrer ce que l'interprète pourrait voir en choisissant de visualiser le correspondant d'un terme simple ou composé.

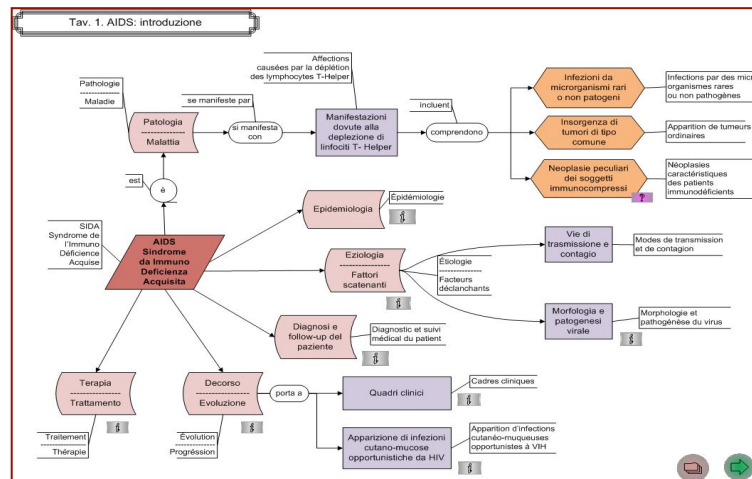


FIG. 4 – Page d'accueil de l'ontoterminologie, de type « top level » par rapport à la spécificité des concepts inclus.

Visualisation des équivalents interlinguistiques des instances. Nous avons inséré les équivalents français à l'intérieur de fenêtres de type *tooltip*, pour faire en sorte qu'ils s'affichent seulement au passage de la flèche de la souris. Ceci permet à l'interprète de visualiser les équivalents des termes le plus rapidement possible. Ces équivalents sont accompagnés, le cas échéant, d'un hyperonyme, un synonyme et/ou un hyponyme (en français et/ou en italien) pour que l'interprète ait une compréhension globale de la phrase qu'il est en train de reformuler. Nous avons choisi de faire apparaître les termes en français seulement si l'interprète en a besoin pour éviter que des graphiques

trop compliqués ou étendus ne le fassent s'égarer, l'empêchant ainsi de suivre aisément la démarche conceptuelle du discours à travers les arbres.

Aménagement et organisation des diagrammes et des feuilles. La structuration des arbres a été basée sur l'évolution la plus vraisemblable d'un discours prononcé au sein d'une conférence spécialisée de médecine pour experts et/ou semi-experts (déterminée grâce à l'aide des experts du domaine). Selon le sous-domaine et la partie du discours à conceptualiser, donc, la structure suit les parcours suivants (l'un après l'autre ou bien en même temps):

- introduction \Rightarrow corps du discours \Rightarrow conclusion (pour l'ordre d'apparition des feuilles et des instances au sein des arbres);
- concepts supérieurs - à bas ou moyen niveau de granularité \Rightarrow sous-concepts spécifiques - à haut niveau de granularité (pour la représentation des pathologies, de leur étiologie et des traitements) ;
- origine \Rightarrow évolution \Rightarrow fin (pour la progression chronologique des pathologies).

L'ontologie se compose de plusieurs feuilles de travail et de différentes sections, à cause des limites de représentation que Visio nous a imposées. Chaque page contient donc une partie de l'arbre ontologique du domaine, qui se développe en suivant la démarche typique d'un discours scientifique. Les instances et les concepts relatifs sont présentés dans l'ordre de parution le plus probable d'un point de vue logique et conceptuel – ordre proposé aux experts et validé par ceux-ci. D'un point de vue graphique, l'ordre logique et chronologique attribué aux concepts est mis en exergue par l'emploi d'une forme et une couleur spécifique pour chaque niveau de la conceptualisation.

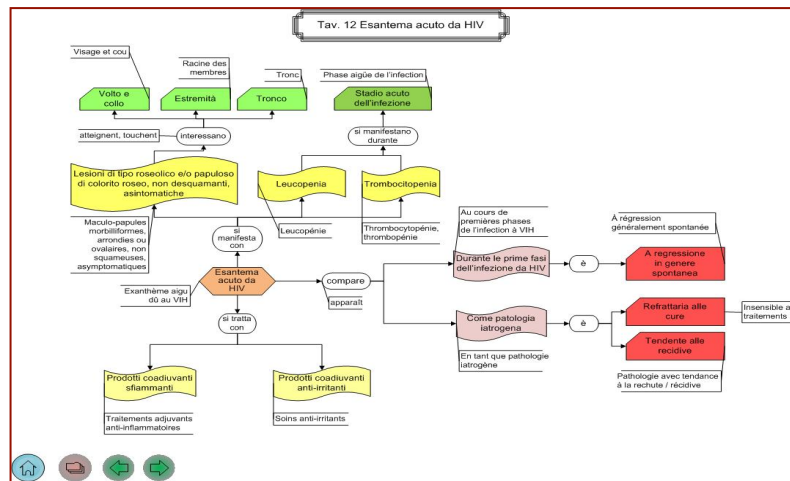


FIG. 5 – Fiche relative à l'exanthème aigu dû au VIH. Les instances faisant partie d'une même classe sont insérées à l'intérieur de cases de la même forme et couleur. La structure de l'arbre suit l'ordre logique et chronologique qui sera sans doute suivi par l'orateur.

Navigation de l'ontoterminologie. Pour rendre notre ontoterminologie prototypique bien navigable et facile à consulter, nous avons inséré des liens intertextuels qui permettent à l'utilisateur de se déplacer à l'intérieur des modules de la conceptualisation. Les liens hypertextuels ont été employés pour:

- faciliter le passage d'une feuille à l'autre et la consultation des diagrammes;
- produire un support souple et adaptable à plusieurs type d'usages /recherches (même si encore prototypique) ;
- permettre aux usagers de suivre différents parcours discursifs et logiques lors de la conférence ou de la préparation au travail.

Plus précisément, les hyperliens relient:

- les instances des arbres de type *top-level* aux feuilles de conceptualisation spécifiques;
- les termes spécifiques aux fiches terminologiques relatives;
- tous les diagrammes à l'arbre introductif de type *top-level* (*homepage*) et au tableau présentant le sommaire des pathologies cutanéomuqueuses conceptualisées en détail.

Pour la mise en exergue et l'activation des liens intertextuels, nous avons utilisé différents boutons. Le bouton « Home » (bleu, avec le symbole d'une maison) permet de revenir à la page d'accueil, pour pouvoir changer de volet quand l'orateur passe à la phase suivante de son discours. Le bouton « Informations » (gris, avec le symbole « i » ou celui d'une loupe) permet d'accéder directement à la conceptualisation de l'instance à laquelle il se réfère (par exemple, la schématisation de l'épidémiologie du Sida). Le bouton « Fiche » (rose, avec un point d'interrogation) conduit directement à la fiche terminologique créée pour les termes les plus intéressants, pour lesquels il était souhaitable de mener une analyse linguistique plus fouillée. Enfin, sur toutes les feuilles sont présentes les flèches « Suivant » et « Précédent », pour en faciliter la consultation.

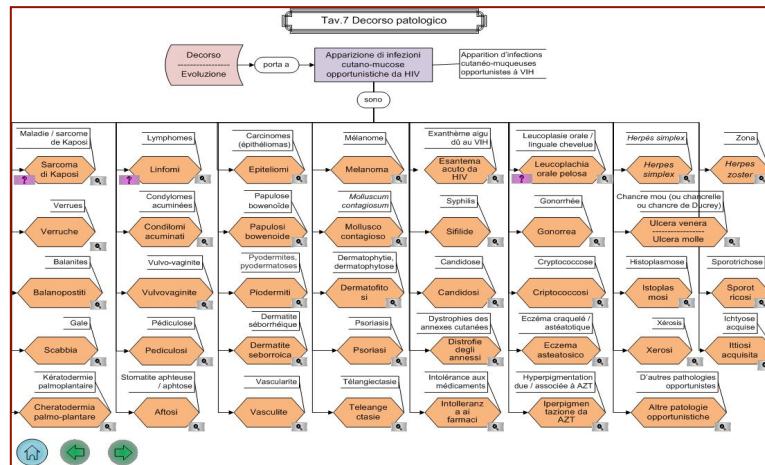


FIG. 6 – L'arbre des maladies cutané-muqueuses dues au VIH. Le bouton gris avec une loupe permet d'accéder directement à la fiche relative à chaque pathologie recensée et analysée.

Comme on peut le voir, l'ontologie a été structurée pour qu'elle soit non seulement une représentation conceptuelle graphique d'un domaine spécifique, mais pour qu'elle suive aussi de près la démarche de la conférence à interpréter, pour faire en sorte que l'ontoterminologie soit un véritable support pour le travail de l'interprète.

4.3. Quelques chiffres...

Notre prototype d'ontoterminologie bilingue se compose de 46 feuilles, chacune contenant un différent diagramme ontologique. Pour chaque langue, l'ontoterminologie inclut approximativement:

- 800 termes simples;
- 1000 syntagmes ou termes composés;
- 500 phrases définitives / descriptives (descriptions ou définitions strictement liées à un terme ou concept);
- 15 relations sémantiques (de quatre types: de subsomption, mérologiques, causales, temporelles).

5. Essai empirique de l'outil ontoterminologique

Au terme de ce travail, nous avons testé l'ontologie ainsi créée de façon empirique. Nous avons demandé à deux types d'utilisateurs différents d'interpréter un discours créé *ad hoc* sur le VIH et ses symptômes dermatologiques en cabine, pour tester l'utilité concrète de notre ontologie. Un premier groupe qui a testé l'outil se composait d'interprètes de conférence professionnels, alors qu'un deuxième rassemblait des étudiants expérimentés en interprétation. Cette démarche nous a permis de vérifier également si l'expérience joue un rôle dans l'usage d'outils numériques de support à l'interprétation en cabine.

De surcroît, les deux groupes ont été ultérieurement divisés en un groupe A et un groupe B. Une semaine avant le test, les interprètes/étudiants du groupe A ont reçu un glossaire traditionnel sur papier contenant, entre autres, les termes que l'orateur utiliserait dans le discours d'épreuve. Par contre, les interprètes/étudiants du groupe B ont appris à consulter et utiliser l'ontoterminologie, qu'ils ont employée pour se préparer au test, ainsi que pendant l'interprétation en cabine. Les deux groupes ont effectué le test de la façon suivante:

- 1) **Groupe A:** composé de 7 personnes, les membres ont interprété le discours tous seuls, sans l'aide d'un collègue passif;
- 2) **Groupe B:** composé de 8 personnes, les membres ont été divisés en 4 couples. Dans chaque couple, un interprète a interprété activement le texte en français en simultanée, tandis que l'autre a fait fonction de collègue passif et a aidé le collègue à consulter l'ontoterminologie. En outre, le collègue passif a précédemment interprété le texte préparé *ad hoc* à l'aide du seul glossaire thématique.

Lors du test, les participants ont enregistré leur performance, que nous avons collectée et analysée. Enfin, nous leur avons demandé de remplir un formulaire spécifiquement rédigé pour vérifier si l'ontologie a été utile ou pas, et quels changements pourraient améliorer l'outil ainsi conçu. L'analyse des

performances des deux groupes et la lecture des formulaires nous a permis d'évaluer l'utilité effective de l'ontologie et le niveau de difficulté maximal qu'un support pour l'interprétation simultanée devrait afficher pour qu'il soit une aide et non une contrainte en cabine.

5.1. Analyse des opinions des interprètes

En général, le test empirique de l'ontoterminologie a donné des résultats satisfaisants. Les interprètes ont déclaré à l'unanimité qu'il s'agit d'un excellent outil de préparation au travail. En outre, les participants ont été d'accord, même si à différents niveaux, pour affirmer que l'ontoterminologie peut représenter aussi un bon support pour l'interprétation simultanée, en particulier en présence d'un collègue passif qui aide la consultation. Les réponses les plus significatives :

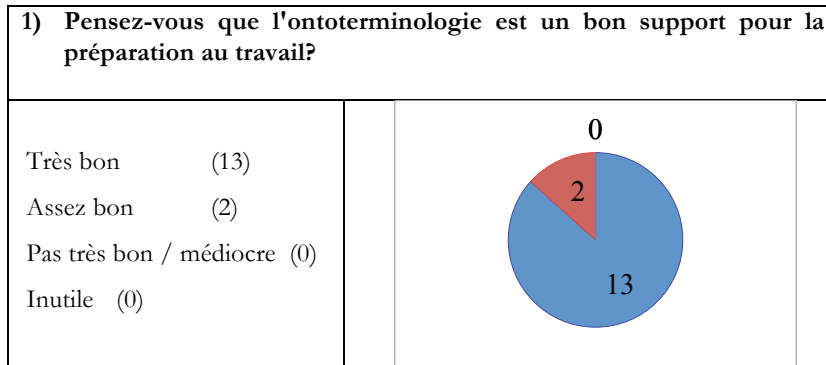


FIG. 7 – L'ontoterminologie en tant que support pour la préparation au travail

2) L'ontoterminologie est-elle représentative du domaine analysé?

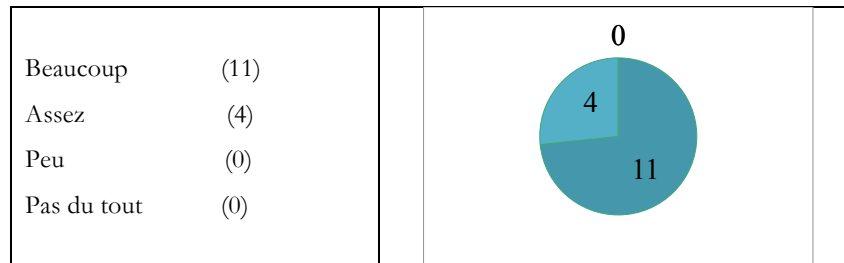


FIG. 8 – La représentation et l'adhérence au domaine analysé

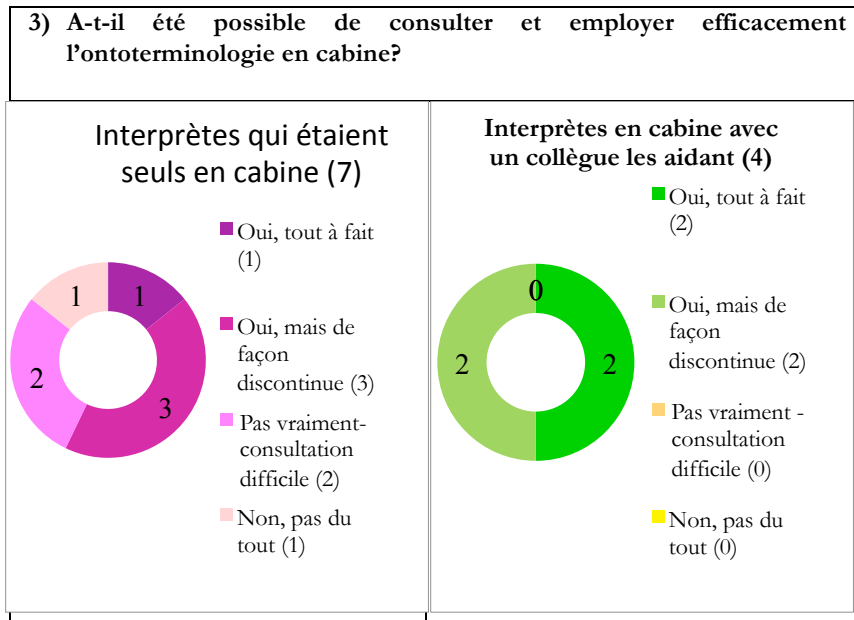


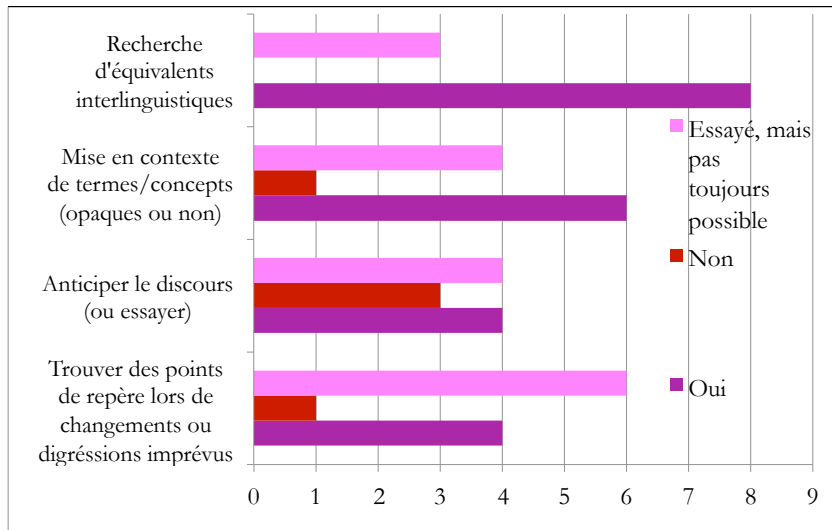
FIG. 9 – L'emploi de l'ontoterminologie en cabine (facilité d'usage).

Page suivante :

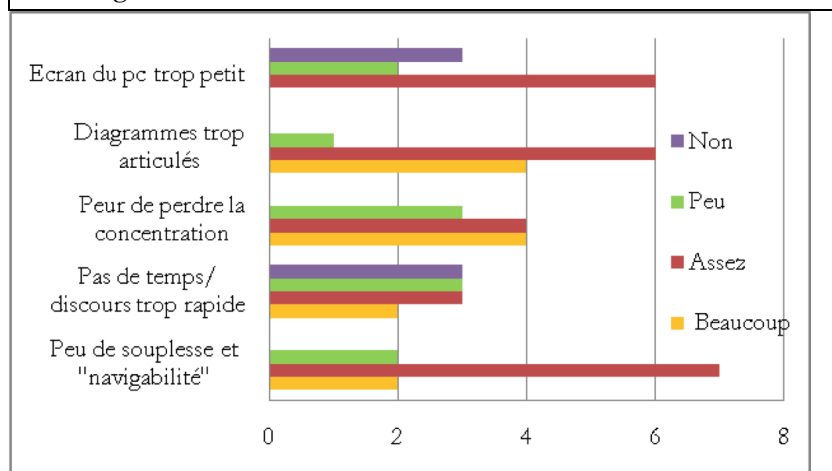
FIG. 10 – L'emploi de l'ontoterminologie en cabine (opérations)

FIG. 11 – L'emploi de l'ontoterminologie en cabine (difficultés)

4) Pour quel type d'opération avez-vous utilisé l'ontoterminologie en cabine?



5) Quelles sont les difficultés que vous avez rencontrées lors de l'usage en cabine?



NB. Les participants qui ont rencontré plus de problèmes ont été les interprètes qui ont travaillé seuls en cabine. Ils ont eu beaucoup ou assez de difficultés à cause de diagrammes trop articulés, d'un manque de temps ou de la crainte de perdre la concentration en consultant l'outil de recherche.

6) Vous estimez que le prototype que vous avez testé est:

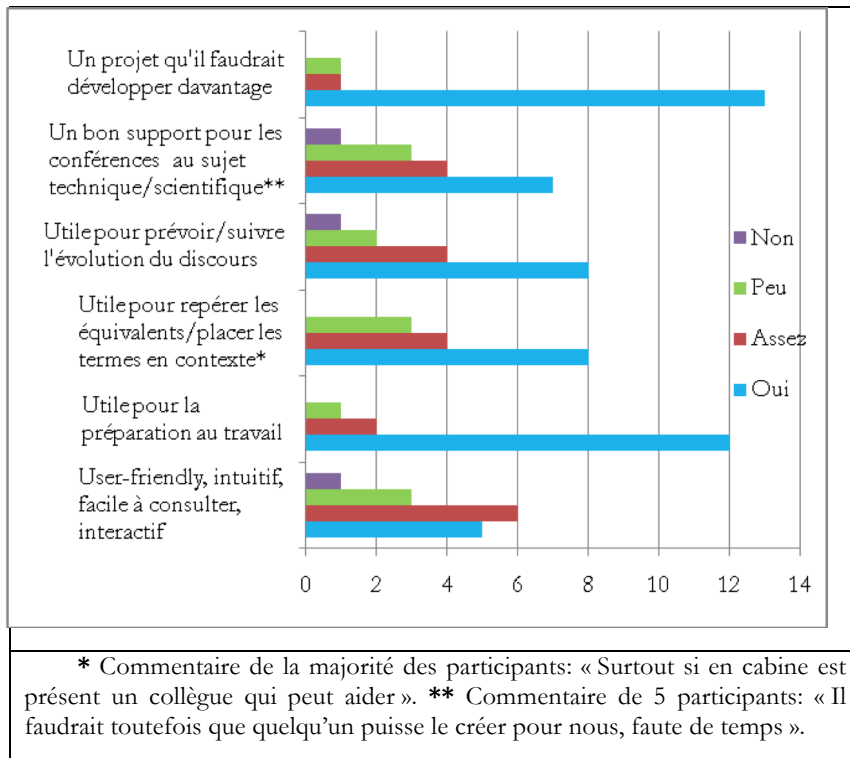


FIG. 12 – Évaluation générale de l'ontoterminologie

Considérations générales. Le niveau de complexité de certains diagrammes devrait être diminué par des logiciels permettant d'en cacher/réduire certaines portions et afficher seulement la partie souhaitée de l'arborescence. Ces problématiques relèvent toutefois du fait que le support ontoterminologique a été développé avec un logiciel de création graphique non approprié pour ce type d'usage et de projet.

De toute façon, le prototype a déjà apporté des bénéfices en cabine, notamment grâce aux liens intertextuels insérés, à l'application aux instances d'une fenêtre *tooltip* pour faire apparaître l'équivalent au passage de la souris, et à la fonction *Find* (« Trouver ») de Visio, qui permet de rechercher un terme spécifique à l'intérieur de toutes les tables. Nous avons donc pu déjà atteindre plusieurs objectifs, puisque notre support permet, si les circonstances contingentes le consentent, de :

- anticiper ou, du moins, suivre l'évolution d'un discours technique/scientifique;
- mettre en contexte un terme/concept opaque;
- rechercher de façon ponctuelle un terme et/ou son équivalent interlinguistique, ainsi qu'un concept, à l'intérieur des diagrammes.

6. Conclusions

L'utilité d'un support ontoterminologique pour interprètes de conférence. L'évaluation de l'ontoterminologie de la part des utilisateurs a été généralement très ou assez positive, surtout de la part des interprètes qui étaient en cabine avec un collègue et pouvaient donc compter sur son aide pour la consultation et la recherche d'un terme spécifique à l'intérieur des arbres. En effet, le discours livré par les interprètes qui ont utilisé l'ontoterminologie pour la préparation et lors du test a été généralement plus précis et assuré d'un point de vue de registre et d'exactitude terminologique, affichant moins de fautes conceptuelles ou de grammaire.

Toutefois, des limites extrinsèques demeurent malheureusement, puisque le logiciel utilisé pour la création des diagrammes (Microsoft Visio 2007) n'est pas conçu pour la création automatique et formelle d'arbres ontologiques. Puisque l'outil utilisé n'est pas spécifiquement créé pour ce type de réalisation graphique, la conceptualisation peut paraître peu lisible, et l'apparition des équivalents linguistiques est parfois trop lente.

Pourtant, la majorité des interprètes, et surtout les professionnels, se sont dits très favorables au développement d'un tel support à l'interprétation de conférence, puisqu'il serait très utile non seulement pour la préparation au travail, mais aussi pendant l'interprétation même. Un tel outil permettrait en fait de récupérer les équivalents linguistiques des termes plus opaques et techniques, ainsi que de suivre l'orateur au fur et à mesure qu'il présente son discours. Ainsi, l'interprète serait à même de placer un terme qu'il ne connaît pas ou dont il ne se souvient pas dans le correct contexte conceptuel et d'usage, ce qui lui permettrait de trouver des points de repère même si le discours affiche un haut degré de spécificité.

6.1. Petit cahier des charges pour la création d'un logiciel adéquat pour ontoterminologies

Création des corpus linguistiques et extraction des candidats termes. Pour ce qui est de la création des corpus et de l'extraction automatique des candidats termes, il serait souhaitable de pouvoir compter sur

un logiciel qui fournisse une meilleure interopérabilité entre un système d'analyse statistique et les résultats d'une analyse distributionnelle. On pourrait par exemple intégrer une extraction à deux niveaux: une extraction des formes du texte pour la validation humaine des données, et une extraction de la forme lemmatisée des textes et des unités lexicales sélectionnées, qui seraient stockées dans la base de données.

Un logiciel pour la création d'ontotermologies pour interprètes.

Pour ce qui est de l'outil « idéal » de structuration ontologique du savoir, les énoncés définitoires des classes et des instances devraient pouvoir être réalisés sur la base de différentes relations sémantiques, qui soient aussi transversales et, donc, plus complexes par rapport à l'hyperonymie ou la synonymie. Il serait donc souhaitable d'avoir des logiciels qui permettent la réalisation de ce qu'on appelle « heavyweight ontologies », c'est-à-dire des ontologies fondationnelles ou régionales basées sur l'emploi d'un langage formel très expressif et à même d'aboutir à des structurations mieux articulées des arbres ontologiques, pour mieux représenter le domaine conceptualisé. Le logiciel dont nous aurions besoin pour créer un outil conçu pour les interprètes de conférence devrait donc:

- fournir un ensemble riche de distinctions conceptuelles (au moins, par rapport au domaine de l'application) ;
- être à même de calculer et modéliser clairement et correctement les relations sémantiques temporelles et de cause-effet par rapport aux relations d'identité et d'appartenance (ou faire en sorte qu'elles soient « remodelisables » manuellement, après la création automatique de l'arbre).

De surcroît, un outil adéquat pour la représentation de réseaux ontologiques pour interprètes devrait permettre de:

- appliquer aux instances l'option de fenêtre *tooltip*, pour insérer plus d'informations et les afficher seulement si nécessaire;
- créer automatiquement un index avec aperçu des tables, auxquelles pouvoir accéder par des liens intratextuels ou par une *overview* ;
- élargir et réduire des parties des diagrammes par une option de zoom, pour que les diagrammes se présentent aérés et bien lisibles. Il serait souhaitable que la navigation soit réalisable via une sorte de « loupe » qui « zoome/dé-zoome » des portions de l'arborescence au fil des clics de l'utilisateur (suivant des hyperliens), ou suivant une action explicite (comme l'emploi de la roulette de la souris) ;

- effectuer une recherche directe des termes, en tenant compte de toutes les tables de l'ontoterminologie ;
- modéliser de façon automatique ainsi que semi-automatique les relations sémantiques, pour que le concepteur ou l'utilisateur puisse décider si les rendre ou non explicites.

Nous estimons que la production d'un logiciel basé sur notre cahier de charge permettrait de réaliser des ontoterminologies adaptées aux besoins spécifiques non seulement des interprètes de conférence et des traducteurs, mais également et plus génériquement des linguistes et des terminologues. Cela, parce que les difficultés auxquelles il faut faire face pour conceptualiser des domaines de spécialité très spécifiques sont encore considérables, vu qu'à présent les logiciels de structuration du savoir permettent de créer automatiquement et formellement des ontologies ne représentant que des relations sémantiques verticales et de simple inclusion/exclusion.

Il serait donc souhaitable de développer des outils à même de traiter des données plus complexes et, surtout, basés sur des algorithmes qui prennent en considération des paramètres de structuration « logiques », afin de créer une représentation graphique vraiment représentative du domaine analysé et dont les éléments soient agencés de façon rationnelle. Ainsi, la position, la forme et les couleurs des instances, des arbres et des ramifications pourraient, elles-mêmes, fournir aux utilisateurs des informations importantes sur les connaissances conceptualisées.

6.2. Considérations finales

La nécessité de pouvoir compter sur des outils plus appropriés pour un classement conceptuel de la terminologie de domaine est ressentie non seulement par traducteurs et interprètes, mais aussi par les spécialistes-mêmes de médecine et de biomédecine (Yapi J.H. *et al.*, 2003). Il serait donc souhaitable de poursuivre ce projet (ou la recherche en ce domaine) afin d'améliorer et développer davantage le prototype que nous avons produit et pouvoir construire des ontoterminologies appropriés pour le travail des terminologues, des linguistes et des interprètes / traducteurs.

Dans cette perspective, s'il était possible de développer des logiciels adéquats, il serait même licite de postuler la naissance de la profession de l'"ontoterminologue": un professionnel spécialisé dans la construction d'ontoterminologies spécialisées complexes, de type fondationnel, modulaires et dynamiques, applicables à plusieurs domaines et très souples et expressives, pour les tailler sur les besoins spécifiques des utilisateurs. Les

ontoterminologies ainsi créées pourraient être réunies à l'intérieur de véritables bases de données ontoterminologiques, qui apporteraient des bénéfices à plusieurs professionnels (interprètes, traducteurs, linguistes, terminologues, spécialistes des domaines conceptualisés, etc.). Étant produits de façon formelle, ces outils pourraient être créés de façon automatique ou semi-automatique, mis à jour, partagés et réutilisés. S'il en sera ainsi, il sera possible dans un futur proche de réaliser des ontoterminologies « fondationnelles » complètes et bien représentatives de notre réalité (ou, au moins, d'une partie de celle-ci). Nous pourrions donc dire d'être enfin arrivés à réunir le moderne sens du mot « ontologie » avec sa signification la plus ancienne: l'étude de « l'être en tant qu'être » d'Aristote.

Bibliographie

Bachimont, B., (2004) : *Arts et sciences du numérique : ingénierie des connaissances et critique de la raison computationnelle*, habilitation à diriger des recherches, Université de Technologie de Compiègne, in Malaisé V., (2005) : *Méthodologie linguistique et terminologique pour la structuration d'ontologies différentielles à partir de corpus textuels*, thèse de doctorat en linguistique, université Paris 7 – Denis Diderot, version en ligne.

Charlet, J., (2002) : *L'ingénierie des connaissances : développements, résultats et perspectives pour la gestion des connaissances médicales*, habilitation à diriger des recherches, CHU Pitié-Salpêtrière, 10 décembre 2002.

Golbreich, C., Dameron, O., Gibaud, B., Burgun, A., *Web Ontology Language Requirements w.r.t Expressiveness of Taxonomy and Axioms in Medicine.*, in "International Semantic Web Conference", 2003, pp. 180-194.

Roche, C., *Le terme et le concept : fondements d'une ontoterminologie*, TOTb 2007 : «Terminologie & Ontologie : Théories et Applications », Annecy, 2007 .

Roche, C., Marty, J-C., Lacroix_S., *Ontologie et terminologie: le modèle OK*, in Rint - Réseau international de néologie et de terminologie , «Terminologies nouvelles », 19, 6-1999, pp. 101-110.

Wright, S.E., (2007) : *The Once and Future ISO 704*, in eDITion 2007/1, pp. 1-4.

Yapi, J.H., Lasquellec, A., Gibaud, B., *Evolution de Schéma et Migration d'Instances. Prise en compte des Besoins d'une Application Médicale.*, BDA, 1996, pp. 353-372.

A propos des auteurs

Veronesi Elisa

Elisa Veronesi est interprète de conférence et traductrice sectorielle. Elle travaille en collaboration avec le professeur Bertaccini et mène plusieurs recherches dans le domaine de la terminologie, de l'interprétation et de la traduction.

Via Nello Buzzi 4, 48124 Piangipane / Ravenna (Italy)

everonesi@gmail.com

Bertaccini Franco

Franco Bertaccini est professeur de terminologie à l'université de Bologne, responsable du département de traduction et directeur du laboratoire de recherche terminologique. Parmi ses succès il compte la publication de plusieurs dizaines d'articles en différents domaines.

Campus Universitario di Forlì - Via G. della Torre, 5 – 47100 Forlì (FC)
(Italy)

bertacci@sslmit.unibo.it

<http://terminologia.sslmit.unibo.it/>



TOTh 2010. *Actes de la quatrième conférence TOTh - Annecy – 3 & 4 juin 2010*

Editeur : Institut Porphyre, *Savoir et Connaissance*

<http://www.porphyre.org>

Annecy, 2010

ISBN 978-2-9536168-1-1

EAN 9782953616811

© Institut Porphyre, *Savoir et Connaissance*