

La Terminologie Universelle

CHAPITRE 14: DÉFINITIONS

Ce chapitre porte sur les définitions des entités du domaine de l'anatomie. La référence historique TA98 a été publiée sans définitions, partant du principe que tous les utilisateurs partagent une connaissance commune de l'anatomie et sont capables de définir eux-mêmes l'entité qu'ils manipulent. C'est une situation assez courante : la plupart des terminologies sont publiées sans définitions. L'une des raisons principale de cette absence de définitions est l'insuffisance de ressources humaines. Mais on peut aussi croire que les auteurs des terminologies précédentes ne sont pas convaincus de la nécessité de définitions de qualité et préfèrent consacrer leurs ressources à d'autres tâches.

Dans ce chapitre, nous tenterons de démontrer le rôle majeur de bonnes définitions dans une terminologie. Nous développerons ensuite une stratégie pour développer ces définitions. Enfin, nous ouvrirons une fenêtre sur l'avenir où un modèle formel de définitions pourrait générer automatiquement les définitions attendues dans plusieurs langues. Le but final est d'obtenir des définitions qui soient lisibles ou compréhensibles à la fois par les ordinateurs et les êtres humains.

Ce document constitue le chapitre 14 de l'ouvrage *La Terminologie Universelle*, qui présente une documentation globale sur la **T**_{logie}.

Contents

14.1 Le rôle des définitions	3
14.2 Définitions taxonomiques	5
14.3 Exemples de définitions	7
14.3.1 Vertèbre cervicale	7
14.3.2 Vertèbre	7
14.3.3 Décussation des pyramide	8
14.3.4 Thalamus	8
14.3.5 Artère basilaire	8
14.4 Définitions semi-formelles	9
14.5 Définitions formelles	10
14.5.1 Un réservoir de propriétés	10
14.5.2 Les propriétés définissantes	12
14.5.3 Taxonomie des structures anatomiques	12
14.5.4 Taxonomie des relations	12
14.5.5 Taxonomie des substances	14
14.5.6 Génération automatique des différentias	14
14.5.7 Un autre exemple	16

14.5.8 Et encore un exemple	18
14.6 Discussion à-propos des définitions	20
14.6.1 Acteurs humains et machines	20
14.6.2 Que comprend une machine?	21
14.6.3 Génération de textes	22
14.6.4 Implémentation des définitions	23
14.7 Journal des mises à jour	24
14.8 Crédits	24

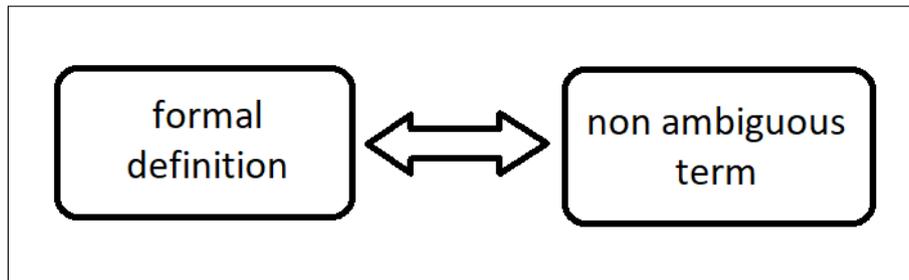


Figure 14.1: La dualité des termes et des définitions. Toute faiblesse d'un côté se propage rapidement à l'autre. Une définition formelle signifie une définition lisible par machine.

14.1 Le rôle des définitions

Considérons d'abord une approche simpliste du problème des définitions. On pourrait dire qu'il n'est pas nécessaire de définir le cœur, car cet organe est connu de tous. Il en va de même pour certaines parties du corps connues du grand public. Les professionnels de santé peuvent imaginer qu'une définition consensuelle existe, probablement pour 75 % des entités anatomiques. Même si ce chiffre était porté à 90 %, qu'en serait-il des autres entités ?

Le rôle de la terminologie est précisément de se préoccuper des entités problématiques, car peu connues, ambiguës ou faisant l'objet d'interprétations divergentes, et ce, au sein d'une même population. Et dès que ces critères sont examinés en fonction de différentes populations utilisant des langues différentes, le problème s'aggrave.

Il existe en effet une dualité parfaite entre le terme désignant une entité et sa définition, voir la figure 14.1. À quoi servirait un terme donné si l'on ne savait pas précisément de quelle entité il s'agissait en réalité ? Quel serait l'intérêt d'une définition exacte si l'on ne disposait pas d'un terme pertinent permettant de communiquer sur cette entité ? La précision finale est proportionnelle à la qualité du terme et à sa définition précise. Le but ultime de la \mathbf{T}_{logie} est de satisfaire aux besoins de clarté pour les entités problématiques, car les entités les plus courantes peuvent être suffisamment maîtrisées sans terminologie !

Plusieurs auteurs ont souligné la nécessité de définitions.

[Cimino, 1998]

La liste des desiderata contient, dans sa sixième partie, la mention de définitions formelles. Cet article examine brièvement la possibilité de formaliser les définitions par le développement d'un modèle et souligne la nécessité de ressources humaines suffisantes pour tout développement de définitions. L'intérêt de cet article réside dans sa publication précoce et dans le fait qu'il mentionne déjà la nécessité de définitions parmi ses desiderata fondamentaux.

[Michael et al., 2001]

Cet article est similaire à celui de Rosse (2003) et le complète. Il présente également un ensemble de dix desiderata pertinents pour la rédaction de définitions.

[Rosse and Mejino, 2003]

Citations de cet article : 1) L'approche de modélisation rigoureuse employée pour le développement de l'AMF repose sur un ensemble de principes déclarés... les définitions aristotéliennes. 2) L'objectif des définitions est d'aligner tous les concepts du domaine au sein d'une hiérarchie ou taxonomie cohérente des types d'héritage. 3) Pour paraphraser Aristote, l'essence d'une entité est constituée de deux ensembles d'attributs définissants : l'un, le genre, nécessaire pour assigner une entité à une classe, et l'autre, les différenciations, nécessaires pour distinguer l'entité des autres entités également assignées à cette classe. Cet article étaye les principaux arguments que nous avons adoptés pour la tlogie.

[Rosse and Mejino, 2007]

Cet article, dans sa section 4.4.5, résume la position de FMA concernant les définitions taxonomiques, à laquelle nous adhérons pleinement. Cependant, la mise en œuvre des définitions par FMA est limitée à quelques centaines d'entités après plus de 20 ans de développement, ce qui limite considérablement la validité de leur proposition. De plus, aucune génération automatique de définitions n'est prévue. Voir l'annexe de ce chapitre pour une liste des définitions des entités de premier niveau de la taxonomie.

La diversité des définitions est aussi grande qu'il y a d'auteurs dans un domaine donné. Les définitions les plus courantes, auxquelles nous sommes tous habitués, sont les *définitions encyclopédiques*, constituées exclusivement de texte libre. En réalité, il s'agit de plus qu'une définition : il s'agit d'un ensemble de propriétés relatives à l'objet à définir. Les dictionnaires et les encyclopédies ne distinguent généralement pas la définition des propriétés : dire que le rein est un organe de 120 à 160 grammes filtrant le sang et produisant l'urine est correct, mais le poids du rein est une propriété et non une définition, et les fonctions énumérées ne sont pas exhaustives. En effet, le rôle d'une définition est de distinguer une entité d'une autre, autrement dit de la différencier.

Lors de la construction d'une ontologie, notre tâche doit se limiter aux stricts besoins de cette discipline. Les ressources humaines étant limitées (tant lors de la création que lors de l'entretien), nous devons éviter tout effort superflu : nous ne voulons pas réécrire une autre encyclopédie, même partiellement. Cela nous conduit immédiatement à conditionner le corpus de définitions par des principes directeurs forts. Ce contexte contraignant est notamment obtenu avec les *définitions taxinomiques* décrites ci-après. De plus, il repose sur le cœur de l'ontologie, la taxonomie du domaine.

14.2 Définitions taxonomiques

Comme indiqué dans [Rosse and Mejino, 2003], l'essence d'une entité anatomique peut être saisie sous la forme de deux attributs : le genre et la différenciation, selon le principe d'Aristote.

Nous extrayons une citation de la référence ci-dessus : *Le but des définitions est d'aligner tous les concepts du domaine de l'ontologie au sein d'une hiérarchie ou d'une taxonomie cohérente des types d'héritage.* C'est précisément notre orientation concernant les définitions.

Le genre est par définition l'entité taxonomique mère. Mais l'entité mère elle-même possède un genre, et ainsi de suite, de sorte que le genre complet est l'ensemble des ancêtres taxonomiques. En pratique, nous limitons le genre au père taxonomique, mais l'essence de la définition peut être recherchée dans d'autres ancêtres. Voir les exemples ci-dessous concernant cet aspect de la définition taxonomique.

Cela signifie que le genre est totalement contraint et peut être généré par ordinateur à volonté. Par exemple, la première partie d'une définition taxonomique est : *Le rein est un organe corticomédullaire qui...* . Ce texte est facilement généré par ordinateur. Si l'on examine les ancêtres supérieurs, on obtient les informations suivantes : *organe parenchymateux* et *organe solide*. Ces dernières informations ne seront pas affichées dans la définition, mais elles sont accessibles au lecteur ou à une application informatique.

La **T***logie* a implémenté la génération de la partie genre des définitions. Notre implémentation étant disponible en quatre langues modernes, le texte généré est multilingue.

Le différentia d'une définition doivent être trouvées dans le corpus de propriétés disponible dans le domaine. Cependant, elles doivent se limiter à la documentation de l'essence de l'entité définie. À proprement parler, les différenciations doivent distinguer l'entité définie de ses homologues dans la taxonomie. Rien de plus n'est nécessaire. Autrement dit, cela signifie que certaines propriétés peuvent être qualifiées de *propriétés définissantes*.

Certaines règles générales s'appliquent à la création de nouvelles définitions dans le contexte actuel des définitions taxonomiques. L'une d'elles consiste à éviter les arguments fonctionnels, sachant que cela n'est pas toujours possible, ni du moins souhaitable. Nous sommes dans le domaine de l'anatomie et les arguments fonctionnels sont empruntés à la physiologie. Les arguments issus de la morphologie, de l'histologie, voire de l'embryologie, sont à privilégier.

Un autre principe directeur est de maintenir une définition simple et facile à comprendre. Chaque définition, autant que possible, doit être parfaitement comprise dès la première lecture par tout utilisateur occasionnel. Mais bien sûr, en cas de conflit entre la précision de la définition et l'exigence ci-dessus, la précision prime.

Certaines définitions sont triviales et peuvent paraître absurdes. Par exemple : *un gyrus orbitalis antérieur est un gyrus orbitalis qui est en position antérieure*. Inutile d'en ajouter davantage, cette définition est bonne, précise et non ambiguë. Quoi d'autre ? En fait, nous sommes dans une situation où le terme lui-même contient la définition. C'est loin d'être vrai dans 90 % des entités. Mais cela existe bel et bien.

Une critique possible des définitions taxinomiques tient au fait que la discipline imposée et le formalisme ne favorisent pas une présentation didactique con-

A kidney is a corticomedullary organ [organum corticomedullare] which has as its part renal pelvis [pelvis renalis] shared with and connected to ureter [ureter].
(Foundational Model of Anatomy)

Un rein est un organe corticomédullaire [organum corticomedullare] qui a pour partie le bassin rénal [pelvis renalis] partagé avec et connecté à l'uretère [ureter].

Uno reno es uno órgano corticomedular [organum corticomedullare] que tiene como parte la pelvis renal [pelvis renalis] compartida y conectada al uréter [ureter].

почка - это корково-медуллярный орган [organum corticomedullare] который имеет в своей части почечную лоханку [pelvis renalis], общую и связанную с мочеточником [ureter].

Figure 14.2: La définition du rein est disponible en quatre langues. L'outil de génération semi-automatique garantit l'équivalence des textes dans toutes les langues : il est essentiel de fournir une définition identique à tous les utilisateurs de la **T**logie. Cette définition est construite à partir de deux références à d'autres entités : le *bassin* et l'*uretère*. Sur le site web de la **T**logie, ces références sont des hyperliens vers ces entités, qui possèdent elles-mêmes une définition.

viviale. Le risque est que nous affichions la définition et que l'utilisateur la rejette parce qu'elle est trop compliquée, pour ensuite se référer à une définition encyclopédique. C'est une réalité incontournable. Quoi qu'il en soit, les définitions encyclopédiques existent en grand nombre et ne doivent pas être considérées comme une alternative exclusive. Au contraire, pourquoi ne pas mélanger les sources, en y ajoutant des atlas d'anatomie en ligne ? Il faut comprendre que la discipline des définitions taxinomiques n'est qu'une autre facette de la terminologie anatomique.

14.3 Exemples de définitions

L'objectif de cette section est d'illustrer les règles de base pour la création de nouvelles définitions par des exemples significatifs tirés du domaine de l'anatomie. Chaque exemple est présenté sous forme de sous-section et est accompagné de commentaires.

Il est à noter que le corpus de définitions de la \mathbf{T}_{logie} est actuellement en cours d'évaluation et qu'une validation formelle par des experts du domaine sera organisée ultérieurement.

14.3.1 Vertèbre cervicale

Vertèbre cervicale

Une vertèbre cervicale est une vertèbre [vertebra] qui forme la partie supérieure de la colonne vertébrale [columna vertebralis] et soutient la tête [caput].

Cette définition classique repose sur la localisation. La deuxième assertion concernant la tête n'est pas nécessairement présente et relève du choix de son auteur.

14.3.2 Vertèbre

Vertèbre

Une vertèbre est un os de la colonne vertébrale [os columnae vertebralis], qui constitue l'unité de construction de sa partie articulée au-dessus du sacrum [os sacrum].

La définition découle ici du fait que l'entité définie est une partie constitutive d'un tout. En remontant un peu la taxonomie, on apprend que la vertèbre est un os irrégulier.

14.3.3 Décussation des pyramide

Décussation des pyramides

Une décussation de pyramides est une décussation de névraxe [decussatio neuraxis] qui se situe dans le rhombencéphale caudal [rhombencephalon caudale], là où les fibres du faisceau corticospinal latéral [tractus corticospinalis lateralis] croisent du côté ipsilatéral au côté controlatéral de leur origine.

Cette définition repose sur deux assertions : premièrement, l'entité est localisée dans le rhombencéphale caudal ; deuxièmement, le faisceau impliqué dans la décussation est déterminé.

14.3.4 Thalamus

Thalamus

Un thalamus est un composant de l'organe du neuraxis [componens organi neuraxis] qui agit principalement comme relais des fibres sensibles afférentes et des fibres motrices efférentes entre le cortex cérébral [cortex cerebri] et les autres segments du neuraxis [neuraxis].

Le rôle de la taxonomie est ici important : le thalamus n'est pas un organe, mais un composant du névraxe qui est lui-même un organe. Le modèle FMA est clair sur ce point. La définition est axée sur les connexions de la substance blanche et insiste sur le rôle principal du thalamus comme relais entre le cortex cérébral et les autres composants du système nerveux central.

Un complément sur la localisation du thalamus pourrait être ajouté à cette définition, par exemple sous le télencéphale. Ce n'est pas strictement nécessaire, car le cortex cérébral est mentionné et toute indication supplémentaire est facultative. Ceci est sujet à discussion.

Cet exemple illustre la grande liberté dont disposent les auteurs pour sélectionner un aspect de cette entité et ignorer d'autres propriétés. D'autres auteurs insistent sur la localisation entre le télencéphale et le mésencéphale ou la proximité du troisième ventricule.

14.3.5 Artère basilaire

Artère basilaire

L'artère basilaire est une artère systémique [arteria systemica] formée par la confluence de l'artère vertébrale gauche [arteria vertebralis sinistra] et de l'artère vertébrale droite [arteria vertebralis dextra].

Il s'agit d'une artère particulière, formée par l'anastomose de deux branches de l'artère sous-clavière. En effet, il ne s'agit pas d'une branche d'une autre artère. La définition est explicite dans ce cas.

14.4 Définitions semi-formelles

La différentia est plus complexe à générer automatiquement et reste aujourd'hui dominée par le texte libre. Elle commence par un verbe à la troisième personne et peut être de longueur quelconque. Cependant, nous essayons de la limiter à la différenciation de l'entité spécifiée par rapport à tous ses frères. Afin de ne pas être trop formel et de faciliter la lecture des définitions résultantes, nous pouvons accepter du texte supplémentaire basé sur des propriétés ou des fonctions, mais cela doit rester une exception. Ainsi, nous obtenons, pour l'exemple ci-dessus : *Organe corticomédullaire dont le bassinot rénal est partagé et relié à l'uretère* (source : FMA).

Nous observons cependant que la différentia possède deux liens avec d'autres entités anatomiques : le bassinot rénal et l'uretère. Il est relativement facile de localiser ces deux entités dans le texte libre et de les remplacer par le lien réel vers ces entités, grâce à leur identifiant. De plus, comme le lien donne accès à l'entité entière, nous pouvons insérer automatiquement le terme latin. Le résultat en 4 langues est visible sur la figure 14.2.

Cette approche est semi-formelle, car une partie des différenciations est automatisée et le reste reste en texte libre. C'est l'état actuel de l'enveloppement. Nous présenterons les projets futurs de création de définitions entièrement formalisées dans une section ci-dessous.

Nous avons vu que la génération du genre est automatisée et que la référence à d'autres entités dans les différenciations de la définition l'est également. Ces étapes sont importantes, mais les tâches restantes sont complexes. L'examen de la forme des différenciations dans plus de 1 000 exemples révèle qu'elle est constituée d'un ensemble d'assertions de la forme Relation + Entité, avec un maximum de 3 assertions.

La prochaine étape du processus d'automatisation serait la modélisation des différentielles. Notre évaluation actuelle constate que la diversité des textes est importante et qu'un modèle trop simple ne serait pas adéquat. Les objets référencés dans les différentielles peuvent se situer hors du domaine strict de l'anatomie et le nombre de propriétés définissantes à créer est important. Une extension du modèle du domaine est nécessaire, fournissant de nouveaux objets tels que des structures, des relations ou des fonctions, tous présents dans une nouvelle branche de la taxonomie.

Il est intéressant de citer ici une initiative similaire, aboutissant à la même conclusion : seule une couverture partielle de toutes les définitions a été modélisée. Des définitions formelles complètes n'ont pas encore été atteintes.

[Mungall, 2004]

Cet article passe en revue une expérience menée par l'auteur pour saisir la partie définitionnelle des termes eux-mêmes et utiliser ces informations à des fins de validation de la taxonomie. Comme prévu, le succès a été réel, mais partiel, car tous les termes ne sont pas directement significatifs. Cependant, cette expérience montre que la modélisation des textes anatomiques est une source de connaissances à envisager pour l'avenir.

14.5 Définitions formelles

Il n'y a aucun doute : *le but ultime d'une ontologie moderne est de nommer précisément tous les objets de son domaine et de les identifier de manière lisible par machine.*

Cela signifie que l'aspect terminologique doit être complété par un aspect définitionnel strict. De plus, les définitions doivent être partagées entre les ontologies : elles doivent être compréhensibles par les ordinateurs. Tant que cet objectif n'est pas atteint, notre ontologie n'est qu'une œuvre incomplète.

À l'heure où l'information est partagée en permanence entre des applications informatiques d'origines très diverses, il est impératif de disposer d'un moyen d'identifier ce dont nous parlons et ce que nous communiquons. Sans un tel principe d'identification, aucune communication ni aucun partage d'informations sérieux ne sont possibles. Aujourd'hui, lorsque deux applications communiquent, elles supposent simplement partager une compréhension commune de leurs objets constitutifs, mais cette hypothèse n'est pas prouvée. En réalité, elle est toujours partielle et certainement éloignée de la situation finale attendue. C'est la limite fondamentale de la communication informatique. La communication humaine est également limitée, mais contrairement aux ordinateurs, les humains peuvent savoir quand les limites sont atteintes et interrompre leurs échanges, avant que des erreurs majeures ne surviennent.

Comme nous l'avons vu précédemment, les définitions semi-formelles sont déjà possibles grâce à la disponibilité d'une taxonomie du domaine : le genre des définitions taxonomiques est généré automatiquement. Le problème réside alors dans l'automatisation de la génération des différentiels. Où devons-nous chercher des informations complémentaires pour servir de base de connaissances à la génération de définitions lisibles par ordinateur ?

14.5.1 Un réservoir de propriétés

La première version d'une terminologie est généralement vide et seules les hiérarchies verticales sont présentes : la taxonomie et la paronomie. Les autres liens horizontaux entre entités sont peu développés, voire absents. Passons en revue les liens horizontaux déjà développés dans la \mathbf{T}_{logie} .

Dans un premier temps, le vocabulaire a été entièrement formalisé et les entités utilisant des mots communs dans leurs termes peuvent être facilement retrouvées. Chaque entrée du vocabulaire est un ensemble de mots partageant une étymologie et un sens communs. Cet ensemble comporte deux extensions : 1) par type syntaxique des mots, noms, adjectifs et préfixes ; 2) par langue utilisée dans la présentation de la terminologie. En général, chaque terme de la \mathbf{T}_{logie} est généré à partir d'une formule universelle, indépendante de la langue. Cette formule est constituée de pointeurs vers l'entrée de vocabulaire adéquate. Un mot, quelle que soit la langue, même s'il apparaît plusieurs centaines de fois, n'est déclaré qu'une seule fois dans la terminologie source. Cette approche permet d'afficher facilement toutes les occurrences d'un mot donné dans la \mathbf{T}_{logie} comme *testis*.

Deuxièmement, la \mathbf{T}_{logie} a développé un mécanisme d'expansion pour la formulation de ses termes : chaque terme construit à partir d'un autre terme est explicitement déclaré et le lien est rendu accessible à une application informatique.

Considérons maintenant le vaste domaine des propriétés des entités anatomiques. En effet, l'état actuel de l'art des sciences biologiques a accumulé une somme de connaissances sur ces entités : la plupart d'entre elles possèdent plusieurs dizaines de propriétés liées à leur position, leur forme, leurs fonctions, leur aspect physique, leur typologie, leurs dépendances, etc. C'est ce que l'on appelle un réservoir de propriétés. En général, ces connaissances ne sont pas formellement représentées, mais sont accessibles grâce à la littérature scientifique et à tout autre média associé. Tout anatomiste amateur est censé maîtriser cette source d'information, qui représente sa compréhension globale de l'anatomie. Cependant, toutes les propriétés n'apportent pas la même importance à notre compréhension du domaine. Certaines propriétés sont immédiatement pertinentes, d'autres présentent peu d'intérêt général, d'autres encore sont fondamentales.

Une stratégie recommandée est la suivante : sélectionner quelques propriétés pertinentes comme propriétés de définition. Cela signifie qu'une, deux ou plusieurs propriétés suffisent pour construire des définitions sans équivoque. Cet exercice manuel a déjà été tenté avec des centaines de propriétés et a démontré que, dans la majorité des cas, deux propriétés suffisent. Prenons un exemple allant dans ce sens.

entité sélectionnée :	CECI est <small>LA:</small> <i>facies lateralis testis</i> .
ancêtre taxonomique :	CECI <small>LA:</small> <i>isa</i> <small>LA:</small> <i>regio superficialis testis</i> .
première propriété :	CECI <small>LA:</small> <i>est_face_de</i> <small>LA:</small> <i>testis</i> .
Motif anglais :	"est une face orientée de".
seconde propriété :	CECI <small>LA:</small> <i>a_direction</i> <small>LA:</small> <i>latéral</i> .
Motif anglais :	"dans * direction".
définitions formelles :	3276 1053 ; 3276 16482 3273 ; 3276 16483 10.
Définition anglaise :	une face latérale du testicule est une région superficielle du testicule qui est une face du testicule orientée dans une direction latérale.
Définition française :	une face latérale de testicule est une région superficielle de testicule qui est une face orientée du testicule en direction latérale.
Définition espagnole :	une face latérale du testicule est une région superficielle du testicule qui est une face latérale du testicule orientée vers la direction latérale.

Table 14.1: Exemple de définition. Le * dans un modèle est un espace réservé pour la valeur générée correspondant à l'entité de droite dans la propriété (par défaut, c'est la position la plus à droite).

Cet exemple est une définition assez triviale : son terme à lui seul contient la définition, du moins pour un lecteur humain, mais certainement pas pour un lecteur informatique ! Le cœur de cet exemple est la relation `is_face_of` qui doit être formellement définie. Pour ce faire, nous définissons d'abord la relation générale `is_surface_of` qui s'applique entre une surface et une entité matérielle. Ensuite, notre relation spécifique est définie comme un descendant taxonomique possédant une propriété `is_oriented` avec la valeur latérale. La définition complète est disponible soit sous forme d'énoncés formels, soit sous forme de phrases textuelles dans tous les langages reconnus. Les définitions

formelles doivent être spécifiées à l'aide d'un formalisme mathématique adéquat afin de pouvoir être transférées à des ordinateurs externes.

Si nous adoptons cette stratégie de collecte de propriétés spécifiques pertinentes comme base de définition, nous sommes naturellement confrontés à un problème essentiel : comment passer des propriétés exprimées en texte libre à un corpus de représentations formelles de ces mêmes propriétés ? C'est le principal problème de la représentation des connaissances.

14.5.2 Les propriétés définissantes

Comme hypothèse de travail, nous avons décidé que les relations binaires de type $A \text{ rel } B$ suffisent à représenter les connaissances nécessaires à la construction des différentielles de définitions. Des relations unaires pourraient éventuellement être envisagées. Cette hypothèse pourrait ne pas être suffisante à long terme et risque d'être reconsidérée à un moment donné du développement avec l'inclusion de relations plus complexes. Mais nous avons de bonnes raisons de penser qu'elle est suffisante et, deuxième argument, nous ne sommes pas prêts à gérer des relations plus complexes ! Ce deuxième argument n'est pas scientifique, mais pragmatique : seuls des résultats positifs dans cette approche pourraient la justifier.

Dans la relation de la forme $A \text{ rel } B$, A représente la classe de l'entité à définir. B représente soit une autre entité anatomique, soit une entité non physique décrivant une structure présente dans le monde de l'anatomie, comme un canal, un tractus, un aspect morphologique, une forme, etc. Ces entités sont explicitement créées pour faciliter la préparation des définitions. Elles représentent un ensemble fini et récurrent de structures, chacune étant réutilisée plusieurs fois. Le *rel* lui-même peut être tout lien pertinent entre A et B , mais il est restreint par $B \text{ isa } C$, C étant toute entité de la taxonomie spécifiée dans la définition de *rel*. Cela permet de limiter l'utilisation de ce lien à un sous-ensemble spécifique d'entités (ce sous-ensemble étant dû au caractère transitif de *isa*). Un nombre illimité de liens peut être créé, mais si une réutilisation importante des liens n'est pas observée, notre hypothèse de travail serait un échec.

14.5.3 Taxonomie des structures anatomiques

Une structure anatomique est une entité non physique servant de modèle à la représentation des structures des entités anatomiques. Ces structures sont hiérarchiquement organisées dans la taxonomie sous l'appellation LA: *structura corporea* ou *structure corporelle*. Cette entité s'étend à toutes les structures matérielles nécessaires à la description des entités matérielles du corps humain. Elles ne sont pas des parties du corps, mais servent de modèle aux structures observables des entités anatomiques matérielles.

Les enfants de la structure matérielle sont les suivants :

14.5.4 Taxonomie des relations

Les propriétés à considérer dans le corpus de propriétés du corps humain et de ses entités constitutives sont nombreuses, certainement quelques centaines. Elles doivent être identifiées et formalisées. Cette identification s'effectue via

1	structure de soutien : Structures servant de support, de protection ou de consolidation aux entités du corps humain.
2	structure vectorielle : Structures assurant le transport des substances dans le corps humain, qu'il s'agisse de substances matérielles ou de signaux.
3	structure opératoire : Structures produisant une substance ou exerçant des actions motrices.
4	structure sensorielle : Structures détectant différents signaux.

Table 14.2: Les différentes classes de structures.

une branche d'entités non physiques de la taxonomie. Cette branche était déjà présente pour l'identification des sous-parties de la relation *partie.de*. Elle doit être étendue à toutes les relations de propriété.

Les enfants de l'entité supérieure de la relation sont les suivants :

1	relation taxonomique : Toutes les relations pour la spécification de la taxonomie.
2	relation structurelle : Toutes les relations spécifiant les propriétés d'une entité anatomique.
2a	relation partonomique : Toutes les relations qui constituent une spécialisation de la relation <i>partie.de</i> .
2b	relation d'association spatiale : Toutes les relations pour la spécification d'une association spatiale.
2c	relation de structure anatomique : Toutes les relations spécifiant un modèle représentant une structure anatomique.
3	relation physique : Toutes les relations spécifiant une propriété physique d'une entité anatomique.

Table 14.3: Les différentes classes de relations. Toute relation utilisée dans la \mathbf{T}_{logie} est nécessairement identifiée comme une entité non physique de la taxonomie. Elle est également formellement définie ailleurs.

Les relations taxonomiques sont bien connues dans la conception de la taxonomie. La relation principale est la relation *isa*, dont la caractéristique essentielle est la conservation des propriétés. Un chapitre de ce livre est consacré aux relations taxonomiques.

Les relations structurales sont consacrées à la description des propriétés anatomiques et constituent le récepteur représenté ci-dessus comme le réservoir des propriétés. Elles se divisent en trois sous-groupes : les relations partonomiques, les relations d'association spatiale et les relations de structure anatomique.

Les relations partonomiques sont toutes des relations qui subsument la relation *partie/de*. Elles sont nombreuses, selon le type d'entités auxquelles elles sont liées. Un chapitre de ce livre est consacré aux relations partonomiques.

Les relations d'association spatiale traitent de la représentation spatiale jusqu'à trois dimension des entités anatomiques.

Les relations de structure anatomique constituent le sous-groupe principal, avec des centaines d'éléments. Elles constituent le cœur d'un modèle d'entités anatomiques, jusqu'à un niveau de granularité détaillé. Elles comprennent les

structures de soutien, comme un os ou une membrane, les structures de transport, comme un tractus ou un canal, et les structures opératoires, comme un muscle ou un ligament. Une relation de structure anatomique est conçue pour représenter les différents modèles de structure, appliqués à un grand nombre de parties du corps.

Par exemple, considérons une structure de soutien, généralement un tubercule osseux, qui est une projection osseuse, une éminence ou une excroissance présente sur un os. Ce modèle de structure osseuse peut être référencé par une relation du type *A est_os_tubercule_de B*, déclarant que A est une réalisation du modèle du tubercule sur un os B, par exemple, *LA:tuberculum orbitale ossis zygomatici*, ou explicitement *tuberculum orbitale est_os_tubercule_de l'os zygomaticus*.

Une variante de la relation structure anatomique concerne la modélisation des voies : une voie dans le corps humain pour le transport des substances corporelles. Voici un exemple de structure de transport. Il s'agit typiquement du tractus gastro-intestinal (*LA:16385*). En général, une telle structure est composée de plusieurs *étages* orientés dans une direction définie : on en compte quatre : l'*LA:oesophage**LA:2509*, le *LA:gaster**LA:2526*, l'*LA:intestinum tenue**LA:2559* et l'*LA:intestinum crassum**LA:intestinum crassum*. Cette voie transporte le *bol alimentaire* de la bouche au rectum. Il semble naturel de définir les organes du système digestif en relation avec de telles voies fonctionnelles, avec une propriété telle que *intestinum tenue is_distal_of gaster*, qui ne se comprend que dans le contexte du tractus gastro-intestinal.

Les relations physiques constituent un groupe de relations distinct, car elles sont largement définies en dehors de la terminologie actuelle et sont valables dans plusieurs domaines.

14.5.5 Taxonomie des substances

Les substances corporelles sont présentes partout dans le corps humain, mais n'en font pas partie. De ce fait, elles sont généralement absentes de la terminologie anatomique traditionnelle. En réalité, elles sont étroitement liées au corps humain ou à l'une de ses entités anatomiques par la relation *contenue_dans*, à ne pas confondre avec *partie_de*. Les substances corporelles sont nécessaires à de nombreuses définitions. Elles doivent être formellement définies.

Les enfants de la substance corporelle sont les suivants :

- | | |
|---|--|
| 1 | substances matérielles : Les substances, solides, liquides ou gazeuses, qui entrent, sont produites, transportées ou sortent du corps humain. |
| 2 | signaux immatériels : Les signaux ou l'énergie transportés entre deux points, soit l'extérieur du corps, soit tout point interne. |

Table 14.4: Les différentes classes de substances corporelles.

14.5.6 Génération automatique des différentias

L'étape suivante, étant donnée une entité pour laquelle une définition taxonomique doit être créée après avoir collecté le lien *isa* vers son ancêtre tax-

onomique direct pour le genre de la définition, et l'ensemble des propriétés de définition sous la forme $A \text{ rel } B$, éventuellement une, deux ou plusieurs propriétés, consiste à générer automatiquement les différentiels en faveur des lecteurs humains. Pour ce faire, il est nécessaire d'associer à chaque lien un modèle de texte prédéfini pour chaque langue de la \mathbf{T}_{logie} . Ce modèle fait partie intégrante de la spécification du lien.

Pour le transfert vers un autre ordinateur, les liens mentionnés ci-dessus suffisent. Bien entendu, l'ordinateur récepteur a déjà collecté l'ontologie complète, y compris les définitions formelles de chaque *rel*. Ces définitions formelles ont été préparées avec soin, conformément aux directives de [?].

Concernant le dernier exemple de table 14.1 ci-dessus, les propriétés de définition sont les suivantes :

1		faciès latéral du testicule est une région superficielle du testicule
2		faciès latéral du testicule est_face_du testicule
3		faciès latéral du testicule a_direction latérale
4		3276 10545 16413; 3276 16482 3273; 3276 16483 10

Table 14.5: L'ensemble des propriétés définissantes. Dans cet exemple, nous affichons les trois propriétés collectées, toutes récupérées grâce à leur entité gauche commune *3276 faciès latéral du testicule*. La forme codée 4 de cet ensemble de propriétés est ensuite affichée sur une quatrième ligne : il s'agit de la forme lisible par machine de la définition. Elle est simplement constituée des identifiants des entités et des relations ci-dessus.

La génération du genre s'effectue grâce à 1 de la forme $A \text{ isa } B$. Le langage naturel repose sur le schéma suivant : $a A \text{ est un } B \text{ [] } \text{ qui... } ,$ où [] représente le terme latin principal de B.

La génération des différentielles est une séquence de contributions, une pour chaque propriété définissante présente. On suppose que la plupart des définitions peuvent être construites de manière adéquate par concaténation de fragments successifs de textes, ce qui est vrai pour toutes les langues. Il s'agit là encore d'une hypothèse de travail qui doit être confirmée par une mise en œuvre réussie : si elle ne fonctionne pas dans la majorité des cas, une nouvelle stratégie sera nécessaire. Il est possible d'envisager deux stratégies de génération, la seconde étant activée en cas d'échec de la première. Si cela est vrai, nous pouvons poursuivre notre hypothèse de travail, même si nous n'avons actuellement aucune idée de la seconde stratégie.

Sous cette hypothèse de travail, notre tâche consiste maintenant à générer une phrase partielle pour chaque propriété disponible concernant la définition. Cette tâche peut être réalisée indépendamment des autres propriétés possibles. Considérons maintenant tour à tour les deux propriétés de notre exemple.

La première propriété de la forme $A \text{ rel } B$ relie la propriété à définir *LA:faciès lateralis testis* à une autre entité, ici le *LA:testis*. La relation *rel* est *_face_of* et constitue notre argument principal. Elle doit être construite avec deux objectifs : 1) la génération du bloc de texte attendu ; 2) la définition formelle de sa signification sous une forme lisible par ordinateur.

La génération du texte repose sur un schéma textuel représentant cette relation. Ce schéma est actuellement le suivant : *est une face orientée de*. On dit que l'on représente une face et que cette face est orientée, autrement dit,

qu'elle a une direction principale, qui sera précisée par la deuxième propriété. Un tel schéma doit toujours commencer par un verbe à la troisième personne du singulier. En regroupant le schéma textuel avec le membre approprié de la propriété, nous obtenons le texte attendu : *est une face orientée de testis []*. La paire de crochets représente un espace réservé pour le terme latin *testis*. Cet espace réservé, sur un site web, peut servir d'hyperlien vers l'entité mentionnée. Il est à noter que notre exemple est en anglais, mais qu'en réalité, le travail peut être réalisé dans n'importe quelle langue de la \mathbf{T}_{logie} .

Pour la deuxième propriété, la relation *a.direction*, reliant notre entité à définir à une valeur de direction. Nous obtenons ici la valeur *latéral*. Le schéma textuel étant *en direction **, nous devons insérer la valeur à la place de l'astérisque, ce qui donne le bloc de texte final *en direction latérale*.

La différentielle attendue est maintenant entièrement construite, donnant *est une face de testicule orientée [] en direction latérale* et peut être ajoutée au genre disponible. Ceci démontre la génération d'une définition lisible par l'homme, garantie conforme aux informations formelles sur l'entité.

Enfin, nous envisageons la création d'une représentation formelle de chaque relation impliquée dans les propriétés. Cette tâche doit être effectuée à chaque création de relation. Il s'agit d'un processus assez complexe, en partie basé sur des formalismes mathématiques, qui ne sont pas nécessairement maîtrisés par tous. Nous en présentons ici les principales étapes sans entrer dans les détails. Une présentation complète de ces formalismes sera présentée ailleurs.

La deuxième propriété utilise la relation *est_face_de*. Elle s'applique uniquement à une entité matérielle, ici le testicule. Toute entité matérielle peut être liée au volume qu'elle occupe par la relation *a.volume*, ce qui donne : *testis a.volume spatium testis*. De même, chaque volume est lié à sa surface externe par la relation *a.surface*, ce qui donne *spatium testis a.surface surface testis*. Il suffit maintenant de définir la relation suivante : *facies lateralis testis est_face_de surface testis*. Il apparaît immédiatement que l'argument de gauche est une sous-surface de l'argument de droite, sous une condition restrictive. Cette condition stipule que toutes les sous-parties de l'argument de gauche sont orientées dans la même direction. Cette formulation peut être transformée en formule mathématique. Ceci finalise notre processus de génération automatique des différentielles de nos définitions.

14.5.7 Un autre exemple

Considérons maintenant une définition plus complexe, impliquant quatre propriétés déterminantes, du testicule. Ici, quatre propriétés déterminantes sont nécessaires. Premièrement, le testicule est le point de départ du tractus spermatique masculin, qui comprend tous les organes définis sous les organes génitaux internes masculins. Deuxièmement, le testicule est le générateur des spermatozoïdes. Troisièmement, le testicule est également le générateur de testostérone. Quatrièmement, la localisation du testicule dans le scrotum doit être précisée dans la définition. Cela donne :

À l'examen de cette définition, il apparaît qu'aucune propriété déterminante ne peut être négligée : chacune des quatre propriétés est nécessaire. On pourrait penser que les deux propriétés associées au *is_generator_of* ne le sont pas, mais en réalité, le modèle global doit pouvoir déterminer la provenance de toute substance corporelle. La dernière propriété, concernant la localisation

entité sélectionnée :	CECI est <i>LA: testis</i> .
ancêtre taxonomique :	CECI <i>LA: isa LA: organum lobulare</i> .
première propriété :	CECI <i>LA: est_premier_stade_de LA: tractus spermaticus masculinus</i> .
modèle anglais :	"est le stade initial de".
deuxième propriété :	CECI <i>LA: est_générateur_de LA: substantia spermatozoida</i> .
modèle anglais :	"produisant".
troisième propriété :	CECI <i>LA: est_générateur_de LA: substantia testosteron</i> .
Modèle anglais :	"produisant".
quatrième propriété :	CE <i>LA: est_contenu_dans LA: scrotum</i> .
Modèle anglais :	"produisant".
définitions formelles :	3273 10545 10404 ; 3273 16482 16459 ; 3273 16487 16485 ; 3273 16487 16486 ; 3273 10566 3414.
Définition anglaise :	Un testicule est un organe lobulaire [organum lobulare] qui constitue l'étape initiale du voie spermatique [tractus spermaticus masculinus] produisant les spermatozoïdes [substantia spermatozoidi] et la testostérone [substantia testosteron] et contenues dans le scrotum [scrotum].
Définition française :	Un testicule est un organe lobulaire qui est l'étape initiale de la voie spermatique [tractus spermaticus masculinus] produisant les spermatozoïdes [substantia spermatozoidi] et la testostérone [substantia testosteron] et localisée dans le scrotum [scrotum].
Définition espagnole :	Un testículo es un órgano lobular [organum lobulare] qui est l'étape initiale de la via espermática [tractus spermaticus masculinus] qui produit des espermatozoïdes [substantia spermatozoidi] et testosterona [sustancia testosteron] et ubicada en el escroto [escroto].

Table 14.6: Exemple de définition. Dans ce cas, quatre propriétés déterminantes sont nécessaires, afin de définir précisément quel est le testicule.

dans le scrotum, est nécessaire, car sinon, aucune localisation ne serait ni présente ni déductible. Quant à la première propriété, elle contribue significativement à l'approche semi-fonctionnelle activement développée pour les définitions. L'existence de plusieurs passages décrit le circuit logique en place dans le corps humain et fournit des informations sur les connexions internes naturelles, où circulent les substances corporelles.

Dans cette définition, le tractus spermatique masculin (LA16459) est évoqué. Ce conduit décrit le circuit des spermatozoïdes dans le corps masculin. Les spermatozoïdes sont créés dans le testicule, stockés et maturés dans l'épididyme, puis acheminés par le canal déférent jusqu'à la jonction avec la glande séminale où ils reçoivent une substance protectrice pour leur sortie du corps. Ils continuent ensuite par le canal éjaculateur interne à la prostate où ils reçoivent une substance alcaline, puis continuent dans l'urètre masculin avant d'être finalement éjectés.

14.5.8 Et encore un exemple

Considérons maintenant une définition plus complexe, impliquant quatre propriétés définissantes, au sujet du FR: *trigone fémoral*. Ici, quatre propriétés définissantes sont nécessaires. Premièrement, le trigone est un triangle, ce qui signifie qu'il est visible comme tel. Secondement, ce triangle est délimité par trois organes définissant trois lignes droites lesquelles se trouvent en première approximation dans un même plan. Chaque côté du triangle nécessite une propriété. Toutes ces informations sont assemblées dans la figure 14.7.

selected entity:	THIS is <small>LA</small> : <i>trigonum femoris</i> .
taxonomic ancestor:	THIS <small>LA</small> : <i>isa</i> <small>LA</small> : <i>divisio regionis femoris</i> .
first property:	THIS <small>LA</small> : <i>has_shape_of</i> <small>LA</small> : <i>triangle</i> .
English pattern:	"is un triangle".
second property:	Triangle <small>LA</small> : <i>is_delimited_by</i> <small>LA</small> : <i>musculus sartorius</i> .
English pattern:	"bounded by".
third property:	Triangle <small>LA</small> : <i>is_delimited_by</i> <small>LA</small> : <i>musculus adductor longus</i> .
English pattern:	"bounded by".
fourth property:	Triangle <small>LA</small> : <i>is_delimited_by</i> <small>LA</small> : <i>ligamentum inguinale</i> .
English pattern:	"bounded by".
formal definitions:	224 10545 10600; 224 xxx yyy; yyy zzz 2086; yyy zzz 2096; yyy zzz 1882.
English definition:	A femoral trigone is a subdivision of anterior region of thigh [divisio regionis femoris] which is a triangle bounded by the sartorius muscle [musculus sartorius], the long adductor muscle [musculus adductor longus] and the inguinal ligament [ligamentum inguinale].
French definition:	Le trigone fémoral est une division de la région antérieure de la cuisse [divisio regionis femoris] qui est un triangle délimité par le muscle sartorius [musculus sartorius], le muscle adducteur long [musculus adductor longus] et le ligament inguinal [ligamentum inguinale].
Spanish definition:	es un triángulo delimitado por el músculo sartorio [musculus sartorius], el músculo aductor largo [musculus adductor longus] y el ligamento inguinal [ligamentum inguinale].

Table 14.7: Exemple de définition. Dans ce cas quatre propriétés définissantes sont nécessaires.

La première propriété définissante consiste à affirmer que le FR: *trigone fémoral* est un triangle. Cette déclaration appartient à la géométrie plane qui n'est pas à proprement parler ce que voit un anatomiste qui dissecte le haut de la cuisse d'un cadavre. Entre cette affirmation et la réalité du monde biologique, il y a une grande différence. Il en résulte le besoin de définir une fonction de passage entre l'abstrait et le concret: comment interpréter un triangle dans une collection de tissus biologiques? Une telle situation est typique et permanente dans ce

processus de définition.

Les trois autres propriétés définissantes sont dans la même problématique et consistent à faire apparaître une ligne droite à partir d'un organe ou autre partie du corps. Ces lignes droites seront destinées à former les côtés du triangle. Mais il est évident que les trois lignes ne seront jamais dans un même plan et qu'il y a peu de chance qu'elles se rejoignent pour former les sommets du triangle. De fait dans le meilleur des cas elle seront à peu près droite et pas trop éloignées d'un plan commun. On va donc rechercher ce plan commun et projeter les lignes dessus, ce qui va finalement donner un triangle.

Cet exemple explique la difficulté de définir les fonctions présentes dans les propriétés. Elles doivent non seulement être concrètes avec la réalité du terrain; elles doivent en plus être comprises par un ordinateur, autrement dit faire l'objet d'une définition formelle. La bonne réalisation des fonctions des propriétés sont le principal obstacle sur la route des définitions.

14.6 Discussion à-propos des définitions

Cette section est une discussion sur les définitions, leur rôle, leur mise en œuvre, leur complexité, etc.

14.6.1 Acteurs humains et machines

Il a été affirmé que les définitions doivent être compréhensibles par les humains et par les ordinateurs. C'est une condition essentielle à une terminologie pertinente. Bien entendu, les informations transmises aux humains et aux ordinateurs ne seront pas les mêmes. Dans les deux cas, les messages transmis doivent être adaptés. Cependant, le sens profond du domaine de l'anatomie véhiculé par les définitions, quelle que soit leur forme, doit impérativement être le même. Autrement dit, la forme peut être différente, mais pas le contenu. Ce desiderata doit évidemment être compris également dans un environnement multilingue.

Cette dépendance à la cible soulève plusieurs problèmes, et il est certainement difficile de les résoudre. L'un des principaux problèmes réside dans la difficulté pour les humains, par rapport aux ordinateurs, d'adapter le contenu d'une définition aux concepts existants dans leur esprit, grâce à des mécanismes mal compris. De plus, un tel processus humain n'est pas sujet à erreur, et il incombe aux auteurs de définitions d'être suffisamment explicites et d'éviter toute ambiguïté. À l'inverse, les ordinateurs ont besoin de détails très détaillés et précis, que les humains peuvent considérer comme inutiles, excessifs ou embarrassants. Le dilemme est là, et la solution doit être suffisamment approfondie pour combler ce fossé.

Prenons un exemple simple de définition destinée aux humains : le raphé du scrotum est une décussation située sur la ligne médiane du scrotum. Ici, la décussation est l'ancêtre taxonomique du raphé du scrotum et, à ce titre, est définie ailleurs. Cette définition permet de savoir précisément ce qu'est une décussation. Cela signifie que la définition elle-même ne doit pas contenir d'informations disponibles à un autre niveau. Il en va de même pour le scrotum, défini séparément. Un autre commentaire concerne la présence du terme *ligne médiane* : à proprement parler, il n'est pas nécessaire, car il n'y a qu'un seul raphé dans le scrotum, ce qui rend cette information superflue. Cependant, les auteurs des définitions ont pris l'initiative d'apporter plus d'informations que nécessaire. Nous considérons cette définition comme un bon exemple satisfaisant à la discipline des définitions taxonomiques : elle est précise, concise et répond à une demande humaine.

Que devient cette définition lorsque la cible est un ordinateur ? Comme d'habitude avec les définitions taxinomiques, le genre n'est pas sujet à discussion et, pour l'ordinateur, il se présente sous la forme d'une relation : *raphe scrotum isa decussation* ou sous sa forme codée : *3415 10545 9649*. L'ordinateur sait que toutes les propriétés de la décussation s'appliquent à raphe scrotum. Pour la différentia, il faut une propriété définissante exprimant le fait que le raphé est médian sur le scrotum. Cela peut être fait avec une propriété comme : *raphe scrotum is_midline_of scrotum* ou sous sa forme codée : *3415 xxx 3414*. En effet, avec une telle propriété, on transfère simplement la difficulté à la spécification de cette nouvelle relation *is_midline_of*. Cette relation doit être définie comme partie intégrante de la terminologie : elle doit être compréhensible par ordinateur et traduisible dans n'importe quelle langue humaine.

La relation *A est_médiane_de B* est une relation spatiale : elle spécifie que A est situé dans le plan sagittal médian, à la surface de B. Plus précisément, le plan scrotal médian définit, sur la surface externe du scrotum, une ligne suivie par le raphé matériel du scrotum. Il ne fait aucun doute que l'intersection du plan sagittal médian et du scrotum existe également sur une surface interne du scrotum, mais cela ne nous intéresse pas ! C'est trivial pour un humain, mais pas pour un ordinateur. La surface externe doit être spécifiée dans cette relation, mais cette information doit de préférence rester cachée à un lecteur humain.

La formulation de la relation peut être présentée plus en détail :

A est_médiane_de B si :

- il existe un sous-plan sagittal C défini comme B est_intersection_de S, où S représente le plan sagittal médian ;
- il existe une droite L définie comme L limite_extérieure_de C ;
- il existe une propriété de A définie comme A a_droite L.

Dans cette présentation, les relations utilisées sont toutes des relations spatiales primitives définies ailleurs.

Quelles conclusions pouvons-nous tirer de cet exemple ? Il apparaît clairement que le niveau de détail requis pour une représentation informatique pertinente est bien plus élaboré que notre vision naturelle. L'ordinateur ne peut pas fournir une vision approximative de la réalité : il doit prendre en compte toutes les exceptions ou irrégularités possibles, car elles surviendront tôt ou tard, et si elles ne sont pas anticipées, une erreur perturbera le processus.

Cependant, au prix de la préparation d'un vaste ensemble de relations spécifiques, probablement quelques centaines pour l'ensemble de la terminologie, la construction d'une base solide pour des définitions taxonomiques générées automatiquement est réalisable. Cela représente un effort acceptable en termes de ressources humaines.

14.6.2 Que comprend une machine?

L'ordinateur dispose de trois sources d'information :

- les propriétés définissantes ;
- la connaissance de la terminologie ;
- toute connaissance générale externe, codée dans le code informatique ou sous forme de données.

À partir de ces connaissances, et exclusivement de celles-ci, l'ordinateur doit comprendre les définitions et être capable de faire des déductions logiques sur les entités anatomiques. La troisième branche ci-dessus est une source de connaissances ouverte : nous n'envisageons pas un tel développement et nous nous limitons donc aux deux premières sources. Les propriétés définissantes ont été définies ci-dessus et représentent un volume limité de connaissances. Le deuxième axe, quant à lui, est la partie adaptable de la connaissance anatomique. Ce deuxième axe peut être soit limité aux informations strictement nécessaires, soit étendu à volonté à tout ensemble de propriétés générales.

Les connaissances minimales à transférer de la terminologie à un ordinateur étranger pour comprendre les définitions concernent les relations utilisées dans les propriétés de définition, ainsi que toute autre entité ou relation utilisée pour spécifier ces relations initiales. Cette masse de connaissances constitue le cœur des définitions. Même si un utilisateur humain occasionnel n'en a pas

nécessairement conscience, on suppose qu'il les connaît implicitement. Cette hypothèse émane des auteurs des définitions. Bien entendu, si elle est fautive, le transfert des connaissances de la terminologie à ses utilisateurs est compromis. Ces connaissances issues de la terminologie sont disponibles globalement sous forme d'entrées dans la taxonomie.

L'approfondissement du développement des connaissances terminologiques est assurément un objectif de toute terminologie spécifique. Qu'elle se limite au strict nécessaire, avec ou sans connaissance des définitions, relève du choix des auteurs. Il convient de préciser ici que la plupart des terminologies sont minimales et que même les connaissances nécessaires aux définitions ne sont pas disponibles. De toute façon, l'ordinateur ne comprendra que ce qu'on lui a fourni : pas de miracle ! À cela s'ajoute le fait que la plupart des terminologies dépendent d'une ou deux langues. Autrement dit, d'importants domaines de développement restent ouverts et partiellement explorés (décembre 2023).

14.6.3 Génération de textes

Cette sous-section examine comment générer des définitions en texte libre à partir des propriétés définissantes retenues. Le cas typique est un ensemble de deux ou trois propriétés définissantes, à partir desquelles la définition taxonomique doit être construite.

La différenciation est toujours censée être une phrase unique commençant par un verbe à la troisième personne du singulier, à ajouter après la phrase du genre. L'exigence d'une phrase unique est une contrainte arbitraire, elle a été retenue car elle simplifie la présentation de toutes les définitions et limite de facto le nombre de propriétés définissantes, ce qui poserait de sérieuses difficultés si le nombre était trop élevé. Cependant, cette contrainte est tout à fait acceptable et ne devrait pas nuire à la qualité ou à la précision des définitions. La plupart des définitions doivent être réalisées avec trois propriétés définissantes ou moins, et certaines peuvent en comporter quatre.

Le schéma générateur commun consistera en une séquence de textes générés, ajoutés les uns après les autres, chacun issu d'une seule propriété de définition. L'ordre des textes générés n'est certainement pas arbitraire. Un facteur de pondération issu des relations utilisées pour les propriétés de définition devrait suffire à gérer cet aspect, mais d'autres stratégies pourraient être adoptées si nécessaire. Les textes générés issus d'une paire de propriétés ne sont pas pris en compte. Chaque texte généré doit être préparé comme un texte prêt à être ajouté directement à la partie déjà générée de la définition.

Le texte d'une propriété dépend essentiellement de la relation utilisée dans la propriété de définition. Chaque relation définie dans la terminologie est associée à un modèle de texte exprimant sa signification. Ce lien entre une relation et un modèle de texte doit être reformulé dans toutes les langues de la terminologie. Le texte généré doit être suivi du nom de l'entité correspondante dans la propriété de définition. Cependant, ce nom peut être placé à l'intérieur du texte du modèle : un astérisque dans le texte du modèle servirait d'espace réservé lorsque la position par défaut la plus à droite est modifiée.

Lorsqu'un texte généré ne correspond pas à la première propriété, il doit être coordonné avec le texte généré précédemment. Cela peut se faire de différentes manières. La situation la plus courante est l'utilisation de la virgule ou de la conjonction *et*. Cependant, d'autres solutions sont possibles.

Enfin, des contraintes syntaxiques liées au genre et au nombre dans les différentielles peuvent être présentes et doivent être résolues. L'utilisation d'articles doit également être prise en compte dans certaines langues.

14.6.4 Implémentation des définitions

L'implémentation des définitions pour la \mathbf{T}_{logie} est une tâche colossale. Plus de 12 000 entités attendent une définition et au moins quatre langues sont utilisées comme langues cibles. En 2023, personne n'a jamais accompli une telle tâche, même en partie. Bien entendu, ces affirmations ne concernent pas les éditeurs de dictionnaires et d'encyclopédies, qui ne considèrent jamais ces documents comme lisibles par ordinateur.

De plus, nous ne pouvons pas prétendre que notre approche de l'échange de définitions entre différentes applications informatiques constitue une solution définitive : nous avons peut-être omis certaines difficultés importantes de ce processus de représentation des connaissances. Mais, même si nous sommes encore loin d'une solution viable, nous sommes convaincus que nous réalisons des progrès significatifs.

Heureusement, un grand nombre de définitions (plus de 80 %) sont soit triviales, soit reproduisent une autre définition avec une légère modification. De plus, le nombre de langues cibles ne pose pas de problème de main-d'œuvre important, car le traitement des langues est largement automatisé de manière similaire dans toutes les langues.

Nous proposons aujourd'hui, en 2024, une mise en œuvre en trois étapes :

Court terme Réalisation des définitions formelles pour un échantillon de 20 entités sélectionnées dans différents contextes.

Moyen terme Extension à un chapitre entier de la \mathbf{T}_{logie} avec 200 entités, avec moins de 10 % d'échecs. Validation interne par des experts sélectionnés.

Long terme Planification et réalisation des définitions pour l'ensemble de la terminologie, impliquant un comité exécutif d'anatomistes pour un processus de validation.

Bibliography

- [Cimino, 1998] Cimino, J. (1998). Desiderata for controlled medical vocabularies in the twenty-first century. *Methods Inf Med.* 1998 November ; 37(4-5): 394-403. [PDF](#).
- [Michael et al., 2001] Michael, J., Mejino, J., and Rosse, C. (2001). The role of definitions in biomedical concept representation. *In proceedings of AMIA 2001*. [PDF](#).
- [Mungall, 2004] Mungall, C. (2004). Obol: integrating language and meaning in bio-ontologies. *Comp Funct Genom* 2004; 5: 509-520. [PDF](#).
- [Rosse and Mejino, 2003] Rosse, C. and Mejino, J. (2003). A reference ontology for biomedical informatics: the foundational model of anatomy. *Journal of Biomedical Informatics* 36 (2003) 478-500. [PDF](#).
- [Rosse and Mejino, 2007] Rosse, C. and Mejino, J. (2007). The foundational model of anatomy ontology. *Anatomy Ontologies for Bioinformatics: Principles and Practice*, Springer, New York. [PDF](#).

14.7 Journal des mises à jour

21 Sep 2025 Création du chapitre à partir de la version anglaise. La version française devient la version de référence.

14.8 Crédits

Ce document fait partie de l'ouvrage *La Terminologie Universelle* qui accompagne le site web Terminologia Anatomica Humana. Il exprime la vision des auteurs de la **T**_{logie} sur les fondements de la science de l'ontologie, étayant la terminologie présentée ici. Bien qu'il soit aussi exact que possible, proche de la réalité de la base de données terminologique et du logiciel qui l'entoure, des approximations, des erreurs et des ambiguïtés sont possibles et doivent être considérées comme indépendantes de leur volonté et de leurs intentions.

Tout commentaire concernant le contenu du site web et sa présentation est le bienvenu. Une réponse appropriée sera apportée le cas échéant.

URL authentique de ce fichier :

<https://ifaa.unifr.ch/Public/TNAEntryPage/help/Chap14FR.pdf>