

Exploitation de l'expertise humaine dans un processus de constitution de terminologie

Thierry Hamon et Adeline Nazarenko
LIPN – UPRES-A 7030
Université Paris-Nord & CNRS
avenue J.B. Clément
93430 Villetaneuse, FRANCE
{*prenom.nom*}@lipn.univ-paris13.fr

Résumé - Abstract

Le processus de construction de terminologie ne peut être entièrement automatisé. Les méthodes et des outils de la terminologie computationnelle permettent de prendre en charge une partie de la tâche, mais l'expertise humaine garde une place prépondérante. Le défi pour les outils terminologiques est de dégrossir les tâches qui sont soit trop longues soit trop complexes pour l'utilisateur tout en permettant à ce dernier d'intégrer ses propres connaissances spécialisées et en lui laissant le contrôle sur la terminologie à construire. Nous montrons ici comment le rôle de cette expertise est pris en compte dans SynoTerm, l'outil d'acquisition de relation de synonymie entre termes que nous avons développé.

The terminology building process cannot be fully automatized. Various terminological tools have been designed but human expert knowledge is still required. This paper shows how SynoTerm, our system for the discovery of synonymy relations between terms, deals with human knowledge. The challenge for terminological tools is to complete the time-consuming tasks under human control and to combine the algorithmic results with human specific information.

Mots clés : Structuration de terminologie, aide à la validation, synonymie

1 Introduction

Devant la masse croissante des documents produits dans les domaines de spécialité, des applications de plus en plus variées (recherche d'informations, outil de navigation dans des textes, aide à la rédaction, contrôle de spécification, etc.) ont besoin de disposer de terminologies structurées, où les termes sont en relation les uns avec les autres, plutôt que de simples listes de termes. Par exemple, pour consulter des documents techniques, il est utile de savoir que les termes *matériel électrique* et *équipement électrique* sont synonymes. Cela permet d'identifier les différentes sections du document se référant aux équipements électriques.

La structuration d'une terminologie est une tâche complexe qui doit tenir compte de différents paramètres comme la pratique du terminologue (Srinivasan, 1992; Gouadec, 1990), le contexte

applicatif et le corpus (Bourigault & Slodzian, 1999), ce qui met l'utilisateur au cœur du processus. D'un autre côté, cet utilisateur n'est pas forcément un terminologue expérimenté et il a besoin d'être guidé dans son travail terminologique¹.

Plusieurs approches ont été développées pour acquérir des informations sémantiques : relations d'hyponymie (Morin, 1999), relations de méronymie (Séguéla, 1999), relations de synonymie (Hamon & Nazarenko, 2001), classes de termes (Assadi & Bourigault, 1995), etc. Toutes intègrent le fait que les résultats obtenus doivent être proposés à l'utilisateur qui les valide.

Nous nous sommes intéressés à la relation de synonymie et nous avons développé un outil d'aide à l'acquisition de ces relations, SynoTerm. Nous avons montré l'intérêt d'utiliser des ressources lexicales, fussent-elles générales, pour amorcer l'acquisition de relations de synonymie. La propagation par compositionnalité des informations lexicales qu'elles contiennent permettent de proposer de nouveaux liens sémantiques.

Les résultats sur lesquels nous nous appuyons ici sont issus de deux séries d'expériences : la première sur un corpus de 160 000 mots, constitué de documents décrivant le fonctionnement d'une centrale nucléaire (DSE) ; la seconde sur le corpus Menelas composé de 85 000 mots, portant sur la médecine coronarienne (Zweigenbaum, 1994). Nous avons travaillé sur les candidats termes issus de l'analyse du corpus par LEXTER (Bourigault, 1994).

Au cours de nos expériences, nous avons constaté que le processus d'acquisition de relations proposé est complémentaire du travail manuel de l'utilisateur et que les interactions entre le système et ce dernier sont complexes. Non seulement, l'utilisateur valide les résultats obtenus automatiquement mais il ajoute ses propres connaissances parfois suggérées par les résultats du système lui-même. Dans certains cas, l'utilisateur juge intéressants des liens qu'il ne retient pas pour la terminologie parce qu'ils suggèrent d'autres liens auxquels il n'aurait pas pensé spontanément (Hamon & Nazarenko, 2001). Par exemple, lors de la validation des résultats du corpus DSE, l'utilisateur a rejeté la synonymie de *échelle logarithmique / gamme logarithmique* et *échelle logarithmique / mesure logarithmique* mais ajouté de son propre chef la relation de synonymie *gamme logarithmique / mesure logarithmique*. Dans d'autres cas, l'utilisateur conserve un lien mais en modifiant son type : *vanne de purge* a ainsi été considéré comme un hyponyme de *vanne de vidange* plutôt que comme un synonyme (DSE).

Ces observations nous ont conduits à voir l'acquisition de relations sémantiques spécialisées comme un tout, intégrant le processus d'inférence mais aussi l'étape de validation. Il s'agit d'aider l'utilisateur en lui proposant des relations inférées à partir de différentes ressources (Hamon *et al.*, 1999), mais aussi de lui permettre d'intégrer ses propres connaissances et de guider son intervention sur les résultats lors de la validation. Pour faciliter la coopération entre le système et l'utilisateur, nous avons conçu une interface de validation qui permet à l'utilisateur de revenir sur le processus d'inférence, d'assurer la cohérence de la validation, et d'intégrer des connaissances supplémentaires au fur et à mesure de son travail.

Dans un premier temps, nous revenons à la section 2 sur la place des utilisateurs dans les outils terminologiques. Après avoir décrit notre méthode d'acquisition de relation à la section 3, nous montrons la complémentarité de l'approche automatique par rapport au travail de l'utilisateur (section 4). Dans la section 5, nous montrons comment les fonctionnalités de l'interface de

1. Dans la suite du texte, nous employerons le terme neutre de l'utilisateur pour désigner la personne en charge de construire la terminologie qui utilise donc les outils d'aide. Sauf mention expresse, nous ne faisons aucune hypothèse sur la compétence en terminologie de cet utilisateur. Nous supposons seulement qu'il a une bonne connaissance du domaine de l'application visée.

validation de SynoTerm permettent à l'utilisateur d'intégrer ses propres connaissances dans ce processus de constitution d'une terminologie structurée.

2 Quelle place pour l'utilisateur dans les outils terminologiques ?

L'utilisation d'outils effectuant une analyse automatique des données demande toujours une étape d'interprétation pour l'utilisateur (Bourigault & Slodzian, 1999). La présentation des résultats joue donc un rôle important dans la validation et la structuration d'une terminologie. Dans la mesure du possible, l'utilisateur doit en avoir une vision globale. Il faut également qu'il puisse intervenir directement sur les données non seulement pour effectuer des corrections ou des modifications mais aussi pour intégrer des connaissances supplémentaires en fonction des observations réalisées. Habert et Nazarenko (1996) montrent ainsi que les résultats de l'analyse distributionnelle des termes extraits d'un corpus doivent nécessairement être interprétés avant la constitution de classes sémantiques.

La plupart des outils proposent une visualisation très classique des résultats focalisée sur un terme à la fois, alors que la présentation globale de tous les termes reliés les uns aux autres est souvent nécessaire. Par ailleurs, la visualisation et la gestion des résultats sont souvent conçues indépendamment de l'outil qui produit ces résultats, ce qui rend d'autant plus difficile l'ajout d'informations par l'utilisateur au cours de la validation. Ainsi, le réseau terminologique de LEXTER est représenté dans une interface de consultation, l'Hypertexte Terminologique, basée sur un système de gestion de bases de données indépendant du processus d'analyse de LEXTER (Bourigault, 1994). L'utilisateur effectue dans cette interface le travail de dépouillement et de fouille qui lui permet de construire la terminologie, mais il ne peut pas contrôler le processus d'extraction effectué par LEXTER qui engendre des cas d'erreurs récurrentes.

Il est donc important de présenter graphiquement le réseau terminologique à l'utilisateur. Le système CODE (Skuce, 1993) développé pour faciliter la construction de bases de connaissances, propose ainsi une présentation des données sous forme de graphes combinés à des listes fournissant des informations extraites pour chaque nœud. Nous avons nous-même constaté, au cours de nos expériences, qu'une vision graphique partielle ou globale du réseau terminologique facilite la validation. Il est tout aussi important concevoir le travail de validation comme un travail interactif et pour le système de prendre en compte dynamiquement le résultat des validations précédentes. La validation interactive des résultats réalisée dans le système d'apprentissage de classes sémantiques ASIUM (Faure & Nédellec, 1998) se rapproche en partie de cet objectif. À chaque étape de l'algorithme d'apprentissage, l'utilisateur affine les classes proposées en tenant compte des itérations précédentes.

3 Acquisition de relations sémantiques spécialisées

3.1 Approche générale

Nous nous intéressons à l'acquisition de relations de synonymie entre des candidats termes complexes. Celle-ci repose sur l'exploitation de relations sémantiques extraites de ressources de nature différente : dictionnaire de langue générale *Le Robert*, thesaurus et classes de synonymes constituées manuellement par un terminologue.

Nous concevons la synonymie comme une relation contextuelle, avec une définition proche de celle de WordNet (Fellbaum, 1998) et de (Ploux & Victorri, 1998) : X est un synonyme de Y dans un contexte C si les deux termes sont syntaxiquement identiques et substituables dans le contexte C .

Dans le cadre applicatif qui est le nôtre, nous n'interprétons pas cette relation de synonymie comme une relation d'équivalence car les ressources lexicales que nous exploitons fournissent des relations hors-contexte qui ne permettent pas d'appliquer la transitivité². Par exemple, un mot tel que *traitement* peut être synonyme de *thérapeutique*, dans un contexte médical, et de *salaire*, dans un autre contexte, sans pour autant qu'il existe un contexte commun où *salaires* soit synonyme de *thérapeutique*.

3.2 Propagation compositionnelle de relations de synonymie entre candidats termes

Notre approche repose sur l'hypothèse que la compositionnalité des candidats termes complexes préserve la synonymie. Cette hypothèse est évidemment simplificatrice : ce n'est pas parce que *Le Robert* donne *épuisement* et *appauvrissement* comme synonymes dans certaines acceptions que le terme complexe *épuisement de la bêche ASG*³ doit s'entendre au sens de *appauvrissement de la bêche ASG* (l'épuisement de la bêche signifie qu'elle est vide). En pratique, l'inférence d'un lien de synonymie nécessite que deux termes complexes construits selon le même schéma syntaxique et comportant des éléments synonymes, soient attestés en corpus. Nous considérons donc que deux termes sont synonymes si leurs composants sont identiques ou synonymes. Dans l'exemple ci-dessus, *appauvrissement de la bêche ASG* n'étant pas attesté dans le corpus, aucun lien de synonymie n'est inféré. En revanche, dès lors que *épuisement du combustible* et *appauvrissement du combustible* sont tous les deux des termes attestés, nous faisons l'hypothèse qu'ils sont synonymes.

Nous considérons que deux candidats termes complexes sont synonymes si leurs composants sont identiques ou synonymes. Nous proposons trois règles pour inférer des liens de synonymie entre des candidats termes complexes. Ainsi, un lien de synonymie est ajouté entre deux candidats termes du réseau terminologique si l'une des trois conditions suivantes est vérifiée :

Variation sur l'expansion : les têtes sont identiques et les expansions sont synonymes (*action de protection / action de sauvegarde*) ;

Variation sur la tête : les têtes sont synonymes et les expansions sont identiques (*capacité faible / puissance faible*) ;

Variation sur la tête et l'expansion : les têtes sont synonymes et les expansions sont synonymes (*classement d'équipement / classification de matériel*).

Les liens extraits d'une ressource lexicale sont d'abord utilisés pour amorcer la détection des candidats termes complexes. Puis, nous réitérons le processus en propageant les liens précédemment détectés sur des termes de plus en plus complexes.

2. Pour appliquer la transitivité entre deux relations de synonymie, il faudrait en effet être sûr qu'elles sont toutes les deux valides dans le même contexte, une information que nous n'avons pas.

3. ASG signifie *alimentation de secours des générateurs de vapeur* et une bêche est synonyme de réservoir dans le domaine des centrales nucléaires.

4 Complémentarité des approches algorithmique et humaine

Nous avons effectué des expériences sur différents corpus et en exploitant différentes ressources lexicales (dictionnaire de langue générale, thesaurus, classes de synonymes) (Hamon *et al.*, 1999). Les résultats montrent l'intérêt de notre approche. L'utilisateur retient comme pertinents près de 40% des relations qui lui sont proposées. Cette aide qui peut paraître numériquement modeste est jugée précieuse par l'utilisateur avec qui nous avons travaillé. En dégrossissant le travail de structuration, SynoTerm le rend possible. En pratique nous constatons que les résultats de SynoTerm sont qualitativement et numériquement complémentaires des connaissances que l'utilisateur peut lui-même apporter.

4.1 Analyse de la couverture des règles

Nous avons calculé le rappel relatif moyen⁴ de chaque règle d'inférence, quel que soit le corpus de travail (cf. tableau 1).

Variation sur l'exp.	Variation sur la tête	Variation sur la tête et l'exp.
29,11%	65,70%	5,19%

TAB. 1 – *Rappel relatif moyen des règles d'inférence.*

L'analyse de ces résultats met tout d'abord en évidence le faible rappel de la règle de variation sur la tête et l'expansion (par ex. *altération important / trouble majeur, fonctionnement continu / marche permanente* ou *calibre acceptable / qualité correcte*). Cette règle qui est la moins productive (à peine 1/20ème des liens validés sont inférés par cette règle) est néanmoins précieuse car les liens doublement synonymes qu'elle produit sont rarement détectés par le terminologue, l'association d'idée jouant plus difficilement dans ce cas.

Les résultats montrent également que la règle de variation sur la tête est la plus productive. Près de deux tiers des liens validés sont inférés à partir de cette règle, alors que le rappel de la règle de variation sur l'expansion est inférieure à 1/3.

La différence de productivité entre ces deux dernières règles s'explique par le fait que les termes en position d'expansion entrent dans plus de relations de dépendance que les termes en position de tête. Nous observons qu'en moyenne un mot en position tête est en relation de dépendance avec trois mots en expansion, alors qu'un mot en expansion se combine avec quatre têtes différentes. Le champ d'application de la règle de variation sur la tête est donc plus grand⁵.

4.2 Un complément indispensable au travail du terminologue

Pour apprécier la complémentarité de ce processus inférentiel par rapport aux liens de synonymie auquel l'utilisateur peut penser spontanément, nous avons confronté les résultats produits par SynoTerm avec les relations données directement par l'utilisateur à partir de sa connaissance du domaine et de l'application visée. Dans le cadre de l'expérience sur le corpus DSE

4. C'est-à-dire la proportion de liens inférés par cette règle et validés, par rapport au nombre total de liens validés.

5. En moyenne, on peut donc potentiellement détecter 6 relations de synonymie différentes entre les têtes liées à une expansion donnée et seulement 3 relations entre les expansion d'une même tête.

(Hamon *et al.*, 1999), en effet, des informations nous ont été fournies par l'utilisateur (qui est, dans ce cas, à la fois terminologue et expert du domaine).

Ces informations se présentaient sous la forme de 500 classes de synonymes construites manuellement dont nous avons extrait 3 456 relations de synonymie. Nous avons considéré ces relations comme le reflet de son travail de structuration terminologique.

Nous avons évalué la proportion de relations inférées à partir du dictionnaire de langue et présentes dans les classes de synonymes. Parmi les 590 liens obtenus automatiquement⁶, peu de liens (18) ont été par ailleurs fournis par l'utilisateur.

On note par ailleurs, au vu des résultats que l'utilisateur privilégie une variation sur l'expansion. L'invariant syntaxique se situe en tête des candidats termes. Cette observation confrontée à l'analyse des règles présentée à la section 4.1 montre la complémentarité de notre approche. En effet, celle-ci favorise l'acquisition de relation entre des termes possédant des têtes synonymes, ce type de variation semblant plus difficile à identifier manuellement. La règle de variation sur la tête, mais aussi la règle de variation sur la tête et l'expansion, sont d'autant plus précieuses qu'elles infèrent des liens que généralement l'utilisateur ne recherche pas manuellement.

5 Intervention de l'utilisateur sur les résultats

En dehors des informations qu'il peut fournir, l'utilisateur intervient également lors de la validation des résultats : selon les cas, il doit éliminer certains liens, modifier le type de certaines relations ou revenir sur le processus d'inférence pour en corriger un biais. Il s'agit de l'aider dans ce travail de validation en lui donnant une vue globale et cohérente sur l'ensemble des résultats et en lui permettant de revenir au texte aussi souvent qu'il le souhaite.

5.1 Modifier le typage des relations acquises

Lors de la validation des résultats, nous avons laissé à l'utilisateur la possibilité de renommer les relations exprimées par les liens inférés lorsque celles-ci ne correspondent pas à la relation de synonymie. L'interface de validation propose ainsi différents type de liens : outre les relations de synonymie (valeur par défaut), on trouve les liens associatifs "Voir aussi", les relations d'hyponymie, de méronymie, mais aussi d'antonymie. L'analyse des résultats sur les expériences que nous avons faites permet d'évaluer l'importance de cette fonctionnalité.

Nous avons calculé la répartition moyenne des types de relations dans les résultats obtenus sur les corpus DSE et Menelas. Parmi l'ensemble des liens validés, inférés à partir de ressources de différente nature, près de deux tiers expriment des relations de synonymie : *prescription de sûreté / règle de sûreté*. On constate à l'inverse qu'un tiers des liens sont retypés. La proportion de liens validés comme des relations Voir-Aussi représente 1/5 des liens : *état coronaire / bilan coronarien*. Au total, les relations exprimant une forte proximité sémantique entre des candidats termes, représentent près de 85% des liens validés.

Alors que les relations de méronymie sont quasiment inexistantes (moins de 0,5%), 1/8 des liens validés sont des relations d'hyponymie : *phénomènes naturels / phénomènes physiques*.

6. Parmi les 590 liens proposés par SynoTerm, l'utilisateur en a reconnu 199 comme valides.

Type de relation	Répartition Moyenne
Synonymie	62,73%
Voir-Aussi	21,31%
Hyperonymie	12,32%
Méronymie	0,36%
Antonymie	3,28%

TAB. 2 – Répartition moyenne des types de liens inférés.

Deux phénomènes peuvent expliquer la présence de relations d'hyperonymie dans nos résultats. D'une part, les relations proposées par les ressources lexicales sont contextuelles ou plus lâches que la synonymie. Par exemple, le dictionnaire de langue générale propose un lien de synonymie entre les mots *triple* et *multiple* ou entre *appoint* et *alimentation*. D'autre part, la propagation compositionnelle de la synonymie est parfois mise en défaut. En effet, dans un couple de synonymes, le sens d'un des deux termes peut se spécialiser lorsqu'il devient le composant d'un terme complexe. Le lien inféré exprime alors plutôt une relation d'hyperonymie. Par exemple, alors que *purge* et *vidange* sont considérés comme des synonymes dans le dictionnaire de langue générale, l'utilisateur considère que les liens inférés à partir de ce lien initial⁷ sur le DSE expriment des relations d'hyperonymie : une *ligne de purge* est une *ligne de vidange*.

La présence de relations d'antonymie est due à des raisons techniques. La lemmatisation des textes, réalisée en amont de l'extraction des candidats termes et indépendante de SynoTerm, peut poser problème. Ainsi, elle modifie certains termes portant des marques de négation. Par exemple, le candidat terme *matériels non-statiques* devient sous sa forme lemmatisée *matériel statique*. Étant donné que nous travaillons sur les formes lemmatisées en nous approchant le plus possible des conditions réelles de la structuration d'une terminologie, nous inférons le lien *matériel statique / équipement statique*. Cependant, lors de la validation, les termes sont présentés sous leur forme fléchée afin d'assurer une bonne lisibilité des résultats. Ainsi, le lien présenté sera *matériels non statiques / équipement statique* que l'utilisateur validera donc comme lien d'antonymie.

La possibilité de modifier le typage des liens inférés lors de la validation est importante pour l'utilisateur. Elle lui permet de corriger certains biais dus aux ressources lexicales ou aux outils utilisés. De plus, il peut plus facilement ajuster le type des relations proposées en fonction des spécificités du corpus et conserver des relations qui auraient été rejetées dans une validation trop stricte.

5.2 Organiser la présentation des résultats

Le travail de validation est un travail long et coûteux. Une première expérience de validation (Hamon *et al.*, 1998) a montré la nécessité d'un filtrage et d'une structuration préalable des résultats. Il s'agit d'assurer la cohérence de la validation mais aussi de faciliter l'élimination de certaines erreurs répétitives. Nous proposons donc deux axes d'organisation des liens inférés, l'un centré sur le processus inférentiel (regroupement par familles), l'autre sur les termes (regroupement par classes). Nous considérons que ces axes apportent deux vues distinctes et complémentaires des relations inférées.

7. Nous définissons un lien initial comme une relation entre deux mots ou deux termes, proposée par une ressource et pouvant être exploité par les règles pour inférer des liens entre des termes complexes.

5.2.1 Regroupement de liens par familles

L'étude des résultats par les règles d'inférence montre que certaines erreurs sont apparentées et peuvent donc être éliminées rapidement si elles sont présentées ensemble lors de la validation. Ainsi lorsque le lien initial n'est pas pertinent pour un domaine donné, toutes les relations inférées à partir de ce lien sont à rejeter. C'est le cas, par exemple, pour le lien initial *capacité / puissance*, tous les liens inférés sur le corpus DSE sont rejetés :

<i>capacité maximale</i>	<i>puissance maximale</i>
<i>capacité faible</i>	<i>puissance faible</i>
<i>capacité nulle</i>	<i>puissance nulle</i>
<i>capacité petite</i>	<i>puissance faible</i>

Afin de réduire le coût de la validation par l'utilisateur, nous avons donc choisi de regrouper au sein d'une même famille l'ensemble des liens inférés à partir d'un même lien initial⁸. Par exemple, le lien *calibre correct / calibre acceptable* figure dans la famille dont le lien initial est *correct / acceptable*. Le lien *calibre acceptable / qualité correcte* relève des familles dont les liens initiaux sont *calibre / qualité* et *correct / acceptable*.

La plupart des familles sont validées de manière homogène : tous les liens sont corrects ou aucun ne l'est. La répartition des familles suivant la précision montre qu'environ deux tiers des familles sont rejetés dans leur totalité (resp. 70 % et 46 %, sur les corpus DSE et Menelas) alors qu'environ un tiers est complètement validé (resp. 22 % et 44 %). Très peu de familles (moins de 4 %) possèdent un nombre égal de liens corrects et de liens rejetés.

Nous avons calculé l'homogénéité de chaque famille i , ainsi définie :

$$\text{Homogénéité}_i = \frac{|NbVal_i - NBerreurs_i|}{NbLiens_i}$$

où $NbLiens_i$ est la taille de la famille, $NbVal_i$ est le nombre de liens validés et $NBerreurs_i$ est le nombre de liens rejetés. L'homogénéité d'une famille est égale à 1 si tous les liens sont, soit validés, soit rejetés. Elle est nulle lorsque la moitié des liens est validée. L'analyse des résultats sur les corpus montre que la validation est homogène pour plus de 90 % des familles.

Cette organisation des résultats est donc conçue pour faciliter le rejet des liens d'une famille si l'utilisateur considère que le lien fourni par la ressource de départ n'est pas pertinent pour le corpus.

5.2.2 Regroupement des liens dans des classes

Afin d'assurer une validation cohérente, il est également important de présenter les liens inférés avec leur voisinage. Quand l'utilisateur valide le lien de synonymie entre deux termes, il doit pouvoir consulter tous des autres liens dans lesquels entrent t_1 et t_2 . Par exemple, lors de la validation du lien *indicateur local / signalisation locale*, les liens *indicateur local / mesure locale* et *signalisation locale / signalisation lumineuse locale* sont également présentés à l'utilisateur. Il peut ainsi effectuer une validation cohérente par rapport aux décisions déjà prises et au besoin, modifier ces dernières. Cette organisation permet ainsi de visualiser le voisinage des

8. Les liens inférés à partir des règles de variation sur l'expansion et de variation sur la tête appartiennent à une seule famille alors que les liens inférés à partir de la troisième règle relèvent de deux familles à la fois.

termes d'un lien. Pour en avoir une vision plus globale, nous regroupons les liens possédant un terme en commun dans des classes⁹. Par exemple, les liens de la classe présentée à la figure 1 sont visualisés ensemble lors de la validation.

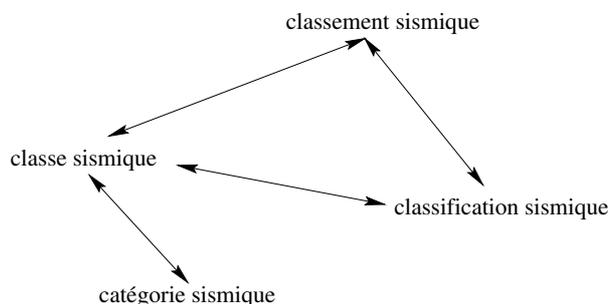


FIG. 1 – Représentation graphique d'une classe.

La taille de ces classes est compatible avec le travail de validation : elle est suffisamment réduite pour que les classes puissent être visualisées et appréhendées globalement¹⁰.

6 Conclusion

L'analyse globale des résultats obtenus par l'application des règles en regard des classes de synonymes construites manuellement montre la complémentarité de notre approche par rapport à la démarche d'un utilisateur. Cependant, dans le cadre d'outil assistant le travail d'un utilisateur, la structuration des termes en vue de constituer une terminologie ne doit pas se limiter au processus d'acquisition de relations sémantiques. Elle nécessite d'organiser les résultats pour faciliter le travail de validation de l'utilisateur. Nous avons ainsi proposé deux modes de regroupement des liens inférés par SynoTerm. Le regroupement des liens par familles permet de prendre en compte les biais de l'approche et des données utilisées pour inférer les liens. Le regroupement par classes permet d'assurer une validation cohérente de liens sémantiques entre des candidats termes en présentant une vision globale des liens inférés. De plus, l'utilisation de ces modes de structuration des résultats dépasse le seul cadre de SynoTerm. La structuration par classes sont exploitables par d'autres méthodes d'acquisition de relations et, le regroupement des résultats par famille peut être adapté à d'autres types d'approches. Il s'agit de regrouper les résultats en fonction d'une origine commune (patrons lexico-syntaxiques, règles, relations).

Les présentations proposées ne sont toutefois pas suffisantes. Nous avons constaté que la présentation par classes a un pouvoir suggestif qui doit être pris en compte : l'utilisateur est tenté de proposer des liens qui n'ont pas été inférés. Ces observations montrent la nécessité de permettre à l'utilisateur d'interagir de manière plus importante sur le processus d'acquisition des relations lors de la validation.

L'intervention de l'utilisateur a, par ailleurs, une incidence sur la manière d'évaluer de tels outils. L'évaluation des outils terminologiques ne doit plus se cantonner à un calcul des taux de précision et de rappel des résultats. L'élaboration d'une grille prenant en compte diverses paramètres est nécessaire.

9. Les classes sont alors équivalentes à des composantes connexes.

10. Sur les données des corpus DSE et Menelas, moins de 2% des classes possèdent plus de 10 liens.

Références

- ASSADI H. & BOURIGAULT D. (1995). Classification d'adjectifs extraits d'un corpus pour l'aide à la modélisation de connaissances. In *Proceedings of the 3rd International Conference on Statistical Analysis of Textual Data*, p. 313–320, Rome, Italy.
- BOURIGAULT D. (1994). *LEXTER un Logiciel d'EXtraction de TERminologie. Application à l'extraction des connaissances à partir de textes*. Thèse en mathématiques, informatique appliquée aux sciences de l'homme, École des Hautes Études en Sciences Sociales, Paris, France.
- BOURIGAULT D. & SLODZIAN M. (1999). Pour une terminologie textuelle. *Terminologies nouvelles*, (19), 29–32.
- FAURE D. & NÉDELLEC C. (1998). A corpus-based conceptual clustering method for verb frames and ontology acquisition. In *Proceedings of the First International Conference on Language Resources and Evaluation – Workshop on Adapting Lexical and Corpus Resources to Sublanguages and Applications*, p. 5–12, Granada, Spain.
- C. FELLBAUM, Ed. (1998). *WordNet: an electronic lexical database*. Language, Speech and Communication. Cambridge, Massachusetts: The MIT Press.
- GOUADEC D. (1990). *Constitution des données*. France: Afnor Gestion.
- HABERT B. & NAZARENKO A. (1996). La syntaxe comme marche-pied de l'acquisition des connaissances: bilan critique d'une expérience. In *Actes des Journées Acquisitions des Connaissances*, p. 137 – 142, Sète, France.
- HAMON T., GARCIA D. & NAZARENKO A. (1999). Détection de liens de synonymie : complémentarité des ressources générales et spécialisées. *Terminologies Nouvelles*, (19), 61–69.
- HAMON T. & NAZARENKO A. (2001). Detection of synonymy links between terms: experiment and results. In *Recent Advances in Computational Terminology*. John Benjamins. À paraître.
- HAMON T., NAZARENKO A. & GROS C. (1998). A step towards the detection of semantic variants of terms in technical documents. In *Proceedings of 17th International Conference on Computational Linguistics (COLING-ACL'98)*, p. 498–504, Université de Montréal, Montréal, Quebec, Canada.
- MORIN E. (1999). Acquisition de patrons lexico-syntaxiques caractéristiques d'une relation sémantique. *Traitement Automatique des Langues*, **40**(1).
- POUX S. & VICTORRI B. (1998). Constructions d'espaces sémantiques à l'aide de dictionnaires de synonymes. *Traitement Automatique des Langues*, **39**(1), 161–182.
- SÉGUÉLA P. (1999). Adaptation semi-automatique d'une base de marqueurs de relations sémantiques sur des corpus spécialisés. *Terminologies Nouvelles*, (19), 52–60. Acte du colloque Terminologie et Intelligence Artificielle, 10-11 mai 1999.
- SKUCE D. (1993). A system for managing knowledge and terminology for technical documentation. In INDEKS-VERLAG, Ed., *Proceedings of TKE'93: Terminology and Knowledge Engineering*, p. 428–441, Frankfurt.
- SRINIVASAN P. (1992). Thesaurus construction. In W. B. FRANKES & R. BAEZA-YATES, Eds., *Information Retrieval: Data Structures and Algorithms*, chapter 9. New Jersey: Prentice Hall.
- ZWEIGENBAUM P. (1994). MENELAS: an access system for medical records using natural language. *Computer Methods and Programs in Biomedicine*, **45**, 117–120.