

Construction et maintenance d'une ressource lexicale basées sur l'usage

Laurie Planes¹,

(1) Inbenta France, 164 route de Revel, 31400 TOULOUSE
lplanes@inbenta.com

Résumé. Notre société développe un moteur de recherche (MR) sémantique basé sur la reformulation de requête. Notre MR s'appuie sur un lexique que nous avons construit en nous inspirant de la Théorie Sens-Texte (TST). Nous présentons ici notre ressource lexicale et indiquons comment nous l'enrichissons et la maintenons en fonction des besoins détectés à l'usage. Nous abordons également la question de l'adaptation de la TST à nos besoins.

Abstract.

Lexical resource building and maintenance based on the use

Our company develops a semantic search engine based on queries rephrasing. Our search engine relies on a lexicon we built on the basis of the Meaning-Text theory. We introduce our lexical resource and explain how we enrich and update it according to the needs we detect. We also mention the customization of the Meaning Text Theory to our needs.

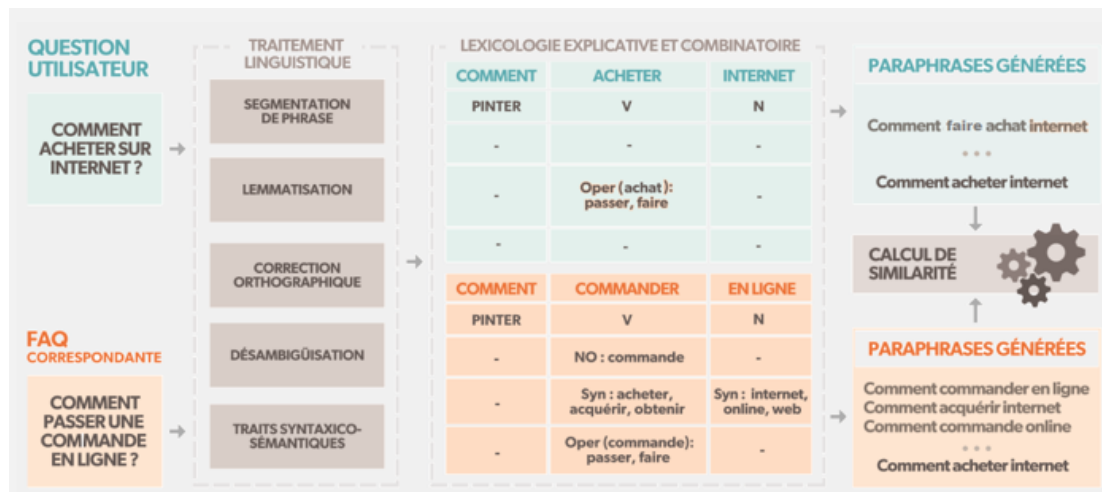
Mots-clés : Ressources lexicales, Théorie Sens-Texte, Recherche d'Information.

Keywords: Lexical Ressources, Meaning-Text Theory, Information Retrieval.

1 Contexte

Notre société a développé son propre Moteur de Recherche Sémantique. Nous utilisons notamment ce moteur dans une Foire Aux Questions (FAQ) dynamique. Une FAQ dynamique est un outil permettant d'accéder à une liste de questions-réponses, notamment en posant une question en langage naturel dans un moteur de recherche.

Pour cela, le moteur procède à l'analyse de la question utilisateur afin d'en produire une représentation sémantique. Dans l'exemple ci-dessous la question « Comment acheter sur internet ? » est lemmatisée et la préposition est supprimée. Ensuite une paraphrase de la question est générée grâce à la fonction lexicale *Oper* (présente dans notre ressource linguistique) qui indique que « faire un achat » a le même sens qu'« acheter ». Parallèlement, nous générons un ensemble de reformulations des titres des FAQs (questions-réponses) à l'aide des relations stockées dans notre ressource : le lexique. Ci-dessous, en orange, des paraphrases du titre de la FAQ sont générées à partir des Fonctions Lexicales associées aux lemmes (formes canoniques) de la question. Enfin, nous comparons ces différentes reformulations à la représentation sémantique de la question utilisateur, ce qui nous permet de retrouver les FAQs dont les titres sont les plus similaires à la question utilisateur. Il s'agit du calcul de similarité.



Comme on le voit dans ce schéma, le cœur du fonctionnement du moteur de recherche repose sur le lexique. Il est donc essentiel de disposer de ressources de « qualité ». Comme le souligne (Saint-Dizier, 2002) la qualité d'une ressource lexicale dépend de son adéquation aux besoins qui sous-tendent son utilisation. « *A défaut d'une analyse fine et précise des besoins, il revient donc au concepteur de ressources de faire lui-même les choix, peut-être de façon arbitraire, et ensuite de tenter de les faire correspondre aux intuitions et aux besoins des utilisateurs. Une attitude prudente consiste à développer des ressources cohérentes, par rapport à sa propre vision théorique et technique, puis de mettre en œuvre des ponts, dits d'insertion lexicale (Strigin, 1998), qui sont autant de canaux qui vont filtrer et reformater ces ressources selon les besoins des applications.* ».

Notre besoin est donc de disposer d'une ressource nous permettant de formaliser le sens d'un énoncé et de générer des reformulations de ce sens. C'est pourquoi nous avons choisi de construire un lexique inspiré de la Théorie Sens-Texte (TST) puisque cette théorie fournit un modèle de traduction du sens en énoncés (textes).

Nous aborderons ici les questions de construction et maintenance de nos ressources lexicales avant de présenter les principes généraux de la TST. Nous verrons ensuite dans quelle mesure nous avons adapté cette dernière.

2 Les méthodes de constitution du lexique

Afin de disposer d'une base de départ nous avons importé un lexique existant pour le français que nous avons retravaillé par la suite. Nous avons limité notre lexique aux domaines qui nous concernent, à savoir la banque et l'assurance. Nous ne visons pas la description exhaustive du français, c'est pourquoi nous avons éliminé certains termes non pertinents pour nos domaines afin de limiter les ambiguïtés.

2.1 Enrichissement du lexique

Pour chaque nouveau site sur lequel nous intégrons une FAQ dynamique nous créons un « projet ». Chaque projet a sa propre base de questions-réponses et son lexique. Ce dernier est constitué d'une base commune à l'ensemble des projets ainsi que d'éléments spécifiques au projet. En général, l'enrichissement du lexique coïncide avec l'arrivée de nouveaux projets et de nouvelles thématiques. Ajouter les termes au fur et à mesure nécessite une réactivité certaine pour maintenir les performances de notre moteur mais nous permet en contrepartie de démarrer rapidement un projet et de constituer un lexique adapté, répondant exactement à nos besoins.

Lorsque de nouvelles questions font ressortir des termes absents du lexique, nous nous appuyons sur l'expertise de nos clients pour intégrer ces termes et définir les relations qu'ils entretiennent avec les termes déjà présents. En effet, nos clients sont capables de nous indiquer si la question X fait référence à la FAQ Y. Les linguistes enrichissent donc le lexique sur la base de cette expertise ainsi que de leurs propres intuitions, suivant une démarche introspective qui sera par la suite validée et complétée par une analyse de corpus.

De plus, nous restons informés des actualités du domaine banque-assurance pour ajouter les termes et relations en rapport avec les nouveautés de ce domaine. Nous avons par exemple enrichi notre lexique de termes relatifs à l'ANI (L'Accord National Interprofessionnel) du 11 janvier 2013, qui généralise la complémentaire santé pour tous les salariés.

Nous repérons les failles du lexique à travers des tâches de maintenance, comme nous le verrons dans le 3.2, ou bien ponctuellement aux cours de nos analyses des performances du moteur basées sur les notions de bruit et de silence.

Pour les questions restées sans réponse (silence), soit il s'agit d'un manque dans nos FAQs (aucune FAQ ne traite de cette thématique), soit le moteur n'a pas été capable de ramener la FAQ pertinente. Dans le second cas nous pouvons être amenés à modifier ou ajouter des relations lexicales. Par exemple, lorsque nous repérons une formulation utilisant des synonymes des mots de la FAQ, nous devons alors ajouter ces synonymes ou bien seulement ajouter une relation de synonymie entre ces mots s'ils sont déjà présents dans le lexique.

En l'absence de gold standard, nous définissons la notion de bruit comme suit : lorsque des FAQs ramenées pour une question ne sont pas cliquées (et donc pas consultées), nous en déduisons qu'elles n'étaient probablement pas pertinentes. Il faudra alors effectuer des ajustements dans le lexique, par exemple en réduisant l'importance (« poids sémantique ») d'un mot qui génère du bruit. Dans ce sens, nous avons minoré l'importance du verbe « vouloir » qui a, par conséquent, un poids faible dans le calcul de similarité. Nous évitons ainsi qu'une FAQ comme « Je *souhaite* changer de formule. Comment faire ? » remonte pour les questions du type « je *veux* résilier mon assurance » (souhaiter et vouloir étant synonymes).

L'enrichissement du lexique peut également se faire à travers des analyses de corpus. Par exemple, nous avons procédé à l'analyse des cooccurrences au sein des questions utilisateurs. Cette analyse, visant à repérer les termes ou groupes de termes qui apparaissent au sein des mêmes contextes, nous a permis d'identifier qu'une des relations présente dans la TST serait pertinente au sein de notre lexique. Il s'agit de la relation entre « faire un paiement » et « payer », intitulée « *Oper* ». Ainsi, nous faisons appel à des scripts automatiques pour enrichir notre lexique. Cependant, une étape de vérification manuelle est ajoutée afin de garantir la qualité de la ressource.

2.2 Maintenance

Nous avons mis en place un système de maintenance reposant sur des tâches effectuées régulièrement par les linguistes de la société. Nous utilisons des outils spécifiques pour repérer automatiquement les anomalies au sein du lexique. Le lexique est stocké sous forme de base de données SQL. Nous pouvons extraire tous les lemmes qui n'ont pas de valeur associée pour un type de relation. Cela nous permet de repérer notamment les verbes qui n'auraient pas de participe passé associé via la relation *PartVerb*.

Dans le même esprit, un autre outil contrôle que le nombre de formes associées aux lemmes des verbes est supérieur à minimum défini et repère ainsi d'éventuels manques.

Dans le but de travailler sur les ambiguïtés catégorielles, nous extrayons tous les mots reconnus comme identiques après normalisation des caractères. Ainsi, les formes identiques associées à des lemmes différents sont listées avec les catégories grammaticales de leurs différents lemmes. Nous disposons de ce fait de la liste de tous les mots grammaticalement ambigus. Cette réserve de cas spécifiques nous sert à rechercher des exemples pour tester et améliorer nos règles de désambiguïsation. Nous pouvons, par exemple, extraire tous les mots ayant l'ambiguïté Verbe vs Nom vs Participe passé. En lançant une recherche sur ces mots parmi les questions utilisateurs, nous obtenons une liste des contextes d'apparitions de l'ambiguïté.

D'autres outils utilisés à l'échelle du projet permettent de repérer les mots de la base de questions-réponses du projet qui sont absents du lexique. Nous détectons ainsi de nouveaux mots à ajouter à notre lexique ainsi que des relations à ajouter avec les lemmes existants.

A l'inverse, nous pouvons voir parmi les questions utilisateurs quels sont les mots absents du lexique. Dans ce cas, cela signifie, soit qu'il faut ajouter un mot, soit qu'un mot a été mal écrit et n'a pas été géré par le correcteur orthographique. Nous améliorons alors notre système de correction orthographique.

3 La lexicologie explicative et combinatoire en théorie vs. en pratique

3.1 Introduction à la TST

Comme nous l'avons expliqué précédemment, le moteur de recherche génère des reformulations sur la base de relations entre les mots. La théorie Sens-Texte (TST) de I. Mel'čuk (Mel'čuk, 1995) propose un modèle dit « traductif » qui permet de mettre en correspondance des représentations sémantiques avec toutes les représentations phoniques qui peuvent les exprimer dans une langue donnée. Il s'agit donc d'un modèle reposant sur le paraphrasage.

Pour permettre ces mises en correspondance, Mel'čuk a produit un dictionnaire explicatif et combinatoire du français contemporain (DEC) dans lequel il décrit les lexies (lemmes) à différents niveaux. Ces niveaux, appelés « zones » permettent la description exhaustive des unités : zone phonologique, syntaxique, sémantique, de combinatoire lexicale, de combinatoire syntaxique, d'exemple et de phraséologie.

La zone de combinatoire lexicale permet le choix du mot juste, notamment à travers l'usage des Fonctions Lexicales (FL) « *La vocation des fonctions lexicales est de fournir au locuteur la totalité des moyens lexicaux nécessaires à l'expression la plus riche, la plus variée et la plus complète de la pensée et, en même temps, de garantir le choix le plus précis de la formulation appropriée. En d'autres termes, les FL [...] alimentent un système puissant de paraphrasage, qui est à la fois une raison d'être des FL et un outil fondamental de leur vérification.* » (Mel'čuk, 1995)

Exemple : 'synonyme'(accuser) = inculper

Les FL syntagmatiques relient les lexies apparaissant en relation de cooccurrence alors que les FL paradigmatiques relient les lexies qui sont liées par des relations sémantiques communes. En d'autres termes, les FL syntagmatiques visent la

combinatoire lexicale (un élément de la valeur de la FL est utilisé **à côté de** son mot-clé) tandis que les FL paradigmatiques visent la sélection lexicale (un élément de la valeur de la FL est utilisé **à la place** de son mot-clé). (Mel'čuk, 2003)

3.2 Notre utilisation de la TST

Notre société a fait le choix de s'intéresser uniquement à la zone de combinatoire lexicale, notre but n'étant pas de capitaliser un nombre maximal d'informations sur les lexies contrairement à l'objectif de Mel'čuk lors de la création de son D.E.C. Cette zone permet le choix du mot juste et la production de la combinaison lexicale adéquate dans un paradigme sémantique donné.

Notre lexique s'inspire de la TST puisqu'il reprend le concept des Fonctions Lexicales (FL) pour le paraphrasage. Il réutilise 5 FL telles qu'elles sont décrites dans le DEC (*Phrase*, *Loc*, *Neg*, *Oper*, *Real*) et en adapte d'autres telles que *Syn 0*, *Syn 1* et *Syn 2*. Nous utilisons également *OperInv*, adaptation de la FL *Oper* pour les formes passives. Les relations *Nominalisation* et *Verbalisation* ont été regroupées en une seule relation réciproque « *NounVerb* ».

5 FL supplémentaires ont été créées pour nos besoins :

- *AddSic* permet de contracter les expressions multitermes : retard + paiement = impayé.
- *ClearSic* permet de réduire une expression en ignorant les mots non porteurs de sens : merci + d'avance → merci.
- *Origen Final* permet de matérialiser un lien de cause à effet ; lorsqu'un utilisateur utilise le lemme « trompé » (ex : « je me suis trompé dans ma date de naissance ») la finalité de sa question est de savoir comment rectifier son erreur. Il faut donc lui répondre par une FAQ concernant la modification des informations personnelles. Par conséquent, nous associons « modifier » à « trompé » via la fonction *Origen Final*.
- *Local FL* est une FL « joker » permettant d'utiliser, au sein du projet seulement, un FL ne répondant pas aux propriétés des FL existantes.
- *PartVerb* : matérialise le lien réciproque entre un Verbe et son participe passé. En effet, il était important pour nous de pouvoir distinguer les verbes de leurs participes passés : la question « comment bloquer ma carte » est relative au fait de faire opposition sur sa carte, alors que « ma carte est bloquée » fait référence à un problème de fonctionnement de la carte. Nous avons donc créé deux catégories distinctes. Nous souhaitons tout de même conserver un lien entre le verbe et le participe car dans certaines situations ils sont utilisés indifféremment. C'est pourquoi nous avons créé la relation *PartVerb*.

Comme évoqué précédemment, notre lexique vise la description limitée des domaines qui nous intéressent (principalement la banque et l'assurance) dans la limite des mots effectivement employés dans nos projets afin d'éviter de générer des ambiguïtés qui n'ont pas lieu d'être au sein d'un domaine spécifique. Par exemple, lorsque nous rencontrons le mot « franchise », nous savons que dans le cadre des assurances il ne s'agit pas de « sincérité » mais bien de la somme restant à la charge de l'assuré dans le cas où survient un sinistre. Inutile donc d'ajouter deux versions du lemme « franchise », qu'il faudrait ensuite désambigüiser.

Cependant, certaines ambiguïtés persistent même en se limitant au domaine banque-assurance. Elles sont alors gérées par des règles de désambigüisation par mot. Ces règles, sous forme de patrons, utilisent les lemmes du contexte précédant ou suivant le mot ambigu. Par exemple, la forme « suis » existe pour les verbes « être » et « suivre ». Il ne s'agit pas d'une ambiguïté catégorielle, il faut donc créer une règle spécifique pour ce mot.

A l'issue d'une analyse de corpus effectuée dans les questions utilisateurs (requêtes) de nos projets, nous avons constaté que « suis » est presque toujours utilisé pour le verbe « être ». Peu d'occurrences de « suivre » ont été constatés pour la forme « suis » : « suivre une procédure » et « suivre des instructions ». Notre règle de désambigüisation par mot indique que « suis » devant un déterminant puis le lemme « procédure » sera identifié comme « suivre », idem pour « suis » devant un déterminant puis le lemme « instruction ». Par défaut, « suis » sera identifié comme « être ».

1. [Ambigüité] + [*|DET] + [procédure|N] = [suivre|V]

2. [Ambigüité] + [*|DET] + [instruction|N] = [suivre|V]

3. Conditionnel par défaut = [etre|V]

Cette règle, bien trop simpliste pour un corpus de textes tout-venant, fonctionne bien dans le cadre limité de nos projets. Une fois de plus, notre lexique est adapté aux besoins spécifiques de nos projets et non à l'ensemble des textes du français.

De plus, nous adaptions certaines FL à nos besoins. Par exemple, la FL « synonyme » proposée par Mel'čuk distingue les synonymes absolus des synonymes plus spécifiques (notion d'hyponymie), des synonymes moins spécifiques (hyperonymie) et des synonymes à intersection. Alors que nous distinguons nos synonymes par degré de proximité (niveaux 0, 1 et 2). Les *syn 0* sont exactement substituables, les *syn 1* ont un sens proche, les *syn 2* ont des sèmes communs mais entretiennent une relation plus lointaine que les *syn 1*. Ces degrés de proximité subissent un traitement distinct lors du calcul de similarité entre un titre de FAQ et une question utilisateur. En effet, le score de similarité sera plus élevé pour un *syn 0* que pour un *syn 1*, et plus élevé pour un *syn 1* que pour un *syn 2*. Cela traduit une confiance plus grande dans le score de similarité pour le *syn 0* que pour le *syn 2*.

Afin de déterminer ce degré de proximité, nous testons la pertinence de la substitution d'un lemme par son synonyme dans nos différents projets.

- Si la substitution est pertinente dans tous les cas nous utilisons la relation *syn 0*. Par ex. : « deuxième » est un *syn 0* de « second ».
- Si elle fonctionne dans la majorité des cas mais peut interférer avec d'autres FAQs, nous utilisons la relation *syn 1*. Par ex. : « annuler » est un *syn 1* de « résilier » car nous rencontrons des formulations du type « Comment annuler mon assurance ? ». Cependant, pour les questions contenant « annuler », il n'est pas souhaitable que les FAQs sur la résiliation obtiennent un score de similarité plus important que les FAQs du type « Comment annuler la modification de mon assurance ? ».
- Si le cas est plutôt incertain nous utilisons la relation *syn 2* qui permettra de réduire le silence sans trop risquer de produire du bruit, étant donné son score plus faible lors du calcul de similarité. C'est ce que nous faisons pour « devis » avec son *syn 2* « tarif ». Nous voulons en effet ramener la FAQ « Comment obtenir un devis d'assurance auto ? » pour les questions du type « Quels sont les tarifs de l'assurance auto ? ». Cette relation relève de la connaissance du monde mais ne se vérifie pas d'un point de vue strictement linguistique, d'où l'usage de la relation *syn 2*.

4 Structure de notre lexique

4.1 Le lexique en quelques chiffres

Nous distinguons les lemmes (formes canonique considérées comme des concepts), des formes qui leurs sont associées, que nous appelons « mots ». Au lemme « payer » sont associés les mots « payer, payons, payez, payent » etc.

Dans notre lexique français nous avons actuellement :

- 23013 lemmes et 163337 mots
- 52170 relations lexicales.
- Une moyenne de 30 mots locaux (mots spécifiques à un projet) par projet
- Une moyenne de 85 particularisations/projet (mots particularisés et relations locales, cf. 4.2)

Pour chaque mot (forme) du lexique nous disposons du lemme correspondant associé à sa catégorie grammaticale ainsi que de la catégorie sémantique du mot. La catégorie sémantique permet de graduer le poids du mot dans le projet, c'est-à-dire l'importance qui lui sera accordée lors du calcul de similarité entre question utilisateur et titre de FAQ.

Pour chaque lemme nous disposons de sa catégorie grammaticale ainsi que des valeurs qui lui sont associées pour l'ensemble des Fonctions Lexicales utilisées dans le lexique.

Notre société gère plus de 20 langues dont le catalan, le castillan, le galicien, le basque, l'anglais, le français, le portugais, l'italien, l'allemand, le néerlandais, le russe, le turque. Chacune de ces langues possède un lexique dédié. Une détection de la langue est faite en entrée, ce qui permet de déterminer le lexique à utiliser. A ce jour, les lexiques sont indépendants les uns des autres.

4.2 Spécificités

Comme évoqué précédemment, nous devons adapter notre lexique aux spécificités des domaines considérés ainsi qu'à celles du projet concerné. Ainsi, le lexique est organisé à différents niveaux : nous distinguons un dictionnaire général partagé par l'ensemble des projets, des dictionnaires locaux, qui sont propres à chaque projet. Un lemme peut être ajouté spécifiquement dans un projet et rester absent des autres projets (éventuellement pour ne pas produire d'ambiguïté inutile). Il s'agit alors d'un lemme « local ». C'est le cas notamment des noms de produits spécifiques à un projet.

Il est également possible d'ajouter des particularisations. Il s'agit d'attribuer des caractéristiques et des relations spécifiques à un mot au sein d'un projet alors que ce terme est utilisé dans l'ensemble des projets avec des caractéristiques différentes.

Par exemple, dans un projet dédié exclusivement aux assurances nous pouvons créer une relation de synonymie *syn I* entre « relevé d'information » (attestation de bonus/malus) et « relevé » car dans la majorité des cas, les utilisateurs emploient « relevé » pour « relevé d'information », alors que dans un projet banque ou banque/assurance, cette relation risque de poser problème car « relevé » peut également faire référence au « relevé de compte ». « Relevé d'information » aura donc un *syn I* « relevé » uniquement dans le projet d'assurance. Dans le lexique général, partagé par les autres projets, il n'aura pas ce synonyme.

Des particularisations sont également effectuées sur la notion de « poids sémantique » traduisant l'importance d'un lemme dans le calcul de similarité entre FAQ et Question utilisateur. Dans le projet du groupe « Banque Accord » la plupart des noms de produit contiennent le mot « accord », comme par exemple « garantie hospitalisation Accord », « compte carte Accord ». Le mot « accord » revient donc très souvent et n'apporte pas de sens aux questions. Nous avons donc diminué son importance (« poids sémantique ») au sein du projet. Ainsi, il est peu pris en compte dans le calcul de similarité de ce projet mais conserve une importance normale dans les autres projets.

Cette structure par niveaux nous permet une grande flexibilité. Nous sommes ainsi en mesure de modéliser le lexique précisément en fonction des besoins du projet considéré sans perturber d'autres projets, tout en construisant une base commune qui sert de socle à l'ensemble des projets.

En plus de ces niveaux (local vs général), nous avons choisi une organisation en conformité avec la structure du langage telle que décrite dans la linguistique structuraliste. Nous reprenons la distinction Saussurienne entre axe paradigmatique et syntagmatique pour qualifier nos différents types de relations entre lemmes. Nous sommes ainsi en mesure d'opérer un traitement différent au moment de la reformulation des questions.

Les fonctions paradigmatiques, qui concernent la sélection des mots, incluent la synonymie (divisée en degré de proximité), la relation *PartVerb* entre un Verbe et son participe passé, la relation *Nominalisation/Verbalisation* entre un nom et son déverbal. Au niveau du calcul de similarité entre questions utilisateurs et FAQ, ces FL permettent de générer une paraphrase par substitution d'un lemme par un autre.

Les relations syntagmatiques sont celles entretenues par les termes dans leur enchaînement au sein de la phrase. Lors de la reformulation des FAQs, ces FL permettent soit, de contracter une expression : remplacer plusieurs lemmes par un seul ou au contraire développer une forme contractée : remplacer un lemme par plusieurs lemmes. Nous avons, par exemple, la relation *Oper* : « faire un paiement » est substituable par « payer », ou bien la relation *Neg* : « rappelle plus » est substituable par « oublié » (ex : « j'ai oublié mon mot de passe » sera reformulé en « je ne me rappelle plus de mon mot de passe »). On remplace donc un lemme par plusieurs ou inversement.

5 Conclusions et perspectives

Nous venons d'évoquer un cas de construction de ressource lexicale adaptée à un usage spécifique. Ici, nous mettons en pratique une formalisation théorique du langage. Le passage à la pratique nécessite certaines adaptations et des ajustements permanents.

Nous avons notamment insisté sur la différence fondamentale entre une ressource visant la description exhaustive du langage et une ressource visant la description limitée des domaines de la banque et de l'assurance, dans la limite des mots effectivement employés dans nos projets. Ainsi, la maintenance et l'enrichissement de la ressource sont continus et nous permettent d'avoir une ressource de plus en plus performante.

Grâce à une structuration à différents niveaux, nous avons doté notre ressource d'une grande flexibilité. Elle s'adapte ainsi aux spécificités de chaque projet dans lequel elle intervient.

Notre ressource continue aujourd'hui de s'enrichir et nous sommes en recherche perpétuelle de moyens de l'affiner et de la compléter.

Références

MEL'CUK, I., POLGUERE, A. (1995) *Introduction à la lexicologie explicative et combinatoire*. Bruxelles : Duculot.

MEL'CUK, I. (2003) Collocations dans le dictionnaire, *Les écarts culturels dans les Dictionnaires bilingues*, Paris: Honoré Champion, p.19-64,

SAINT-DIZIER, P. (2002) Quelques défis et éléments de méthode pour la construction de ressources lexicales sémantiques, *Revue française de linguistique appliquée* I (Vol. VII), p 39-51

STRIGIN, A. (1998) : Lexical Rules as Hypothesis Generators. *Journal of Semantics* 15.