

PROJET EUREKA GENELEX

Rapport sur LE MULTILINGUISME

Consortium GENELEX

Version 2.0

Décembre 1994

Ce rapport présente le travail réalisé par l'ensemble des partenaires du consortium GENELEX France pour la définition d'un modèle décrivant la composante multilingue d'un ensemble de dictionnaires.

Il fait suite aux documents, de présentations comparables, produits par le même consortium, et intitulés "COUCHE MORPHOLOGIQUE Version 3.2", "COUCHE SYNTAXIQUE Version 4.0" et "COUCHE SEMANTIQUE Version 2.1". Les considérations générales et les déclarations préliminaires de ces rapports restent valables ici. On ne reprécisera pas dans ce rapport les notions et choix conceptuels détaillés dans les travaux précédents, mais il y sera fréquemment fait référence.

Au 28 décembre 1994, ce document n'a pas reçu sa validation définitive par le consortium.

A - LE MODELE 3

1. Introduction : situation du modèle GENELEX 4

1.1. Modèle monolingue du fran•ais 4

1.2. Généralisation - validation du modèle 5

1.2.1. Sur le fran•ais 5

1.2.2. Sur d'autres langues 5

1.3. flaboration d'un métamodèle indépendant des langues représentées 6

3. Problèmes posés par la mise en correspondance de dictionnaires monolingues 7

3.1 Divergences de contenu sémantique sans divergences structurelles 7

3.2 Divergences structurelles associées ou non à des divergences de contenu sémantique 9

4. Réponses offertes par le modèle GENELEX 11

4.1 L'approche pivot 12

4.2. Au niveau des unités sémantiques 13

4.3. Au niveau des prédicats lexicaux ou non 14

4.4. Au niveau conceptuel 15

4.5 Au niveau syntaxique 16

4.6 Mixte : couples Usyn-Usem 18

B - MANUEL DE L'UTILISATEUR 19

1. Généralités 20

1.1 Un modèle basé sur la sémantique et la syntaxe monolingues 20

1.2 Un modèle non-orienté 21

1.3 Un modèle ne déformant pas l'approche monolingue 21

1.4 éléments complémentaires 21

1.5 Un modèle offrant différents niveaux de mise en correspondance 22

2. Correspondance d'Usem 23

3. Correspondance de Prédicats 24

4. Correspondance d'Arguments 25

5. Correspondance de Concepts 26

6. Correspondances associant un Concept à un Predicat 27

7. Modifieur 28

8. Correspondance d'Usyn 29

9. Reecrit_Lexicalise, Ajoute_Position 31

10. Correspondance de couples Usem-Usyn 33

C - SCHEMAS ENTITES-RELATIONS 35

1. Correspondance d'Usem 36

2. Correspondance de Prédicats 37

3. Correspondance d'Arguments 38

4. Correspondance de Concepts 39

5. *Correspondances associant un Concept à un Prédicat* 40

6. *Modifieur* 42

7. *Correspondance d'Usyn* 43

8. *Réécrit_Lexicalise, Ajoute_Position* 44

9. *Correspondance de couples Usem-Usyn* 45

D- DTD SGML 47

I - Introduction - Traduction du Modèle Conceptuel 48

II - DTD GENELEX commentée 50

1. DTD lesgenelex.dtd 50

2. DTD genemono.dtd 51

3. DTD genemulti.dtd 53

4. CONTRAINTES Genelex Multilingue 61

A - Le Modèle

1. Introduction : situation du modèle GENELEX

L'approche de GENELEX concernant la modélisation de dictionnaires électroniques génériques a été de travailler d'abord en profondeur sur la modélisation monolingue d'une langue, en l'occurrence le français. L'adaptation / généralisation à d'autres langues du modèle établi pour le français a été entreprise par la suite. C'est ainsi que les travaux sur le portugais, l'italien et l'anglais se sont constamment appuyés sur les résultats obtenus pour le français, et ont permis de confirmer la validité du caractère générique que l'on a voulu donner à ce modèle. D'autres langues seront progressivement prises en compte, même si ces développements s'effectuent en dehors du cadre EUREKA.

Dans cette approche, le multilinguisme a été abordé non pas comme un préalable, mais comme la suite naturelle d'une modélisation monolingue parvenue à maturité. Le modèle GENELEX de mise en correspondance de dictionnaires monolingues qui est présenté dans ce rapport est donc l'étape finale de ce processus.

La philosophie globale de cette mise en correspondance a été de s'appuyer le plus systématiquement possible sur les couches monolingues, et en particulier d'utiliser les différents éléments descriptifs des couches syntaxique et sémantique quand c'est possible, l'unification du langage de description ne pouvant qu'augmenter la cohérence du tout, et la stratégie de modélisation fine monolingue prenant alors tout son sens.

1.1. Modèle monolingue du français

Le modèle GENELEX tel qu'il est défini pour le français dans les trois rapports concernant la morphologie, la syntaxe et la sémantique a été élaboré par les équipes de travail du consortium GENELEX pour la France en particulier sur la base de dictionnaires électroniques existants, et en s'appuyant sur des travaux théoriques concernant le français. Ce modèle est à l'heure actuelle un modèle générique monolingue, dans un état stable, rendant compte de cette langue.

Le modèle du français propose une certaine architecture du lexique (en trois couches : morphologique, syntaxique, et sémantique) et une certaine structuration des données, architecture et structuration étant par ailleurs des choix stratégiques de départ pour la généralisation aux autres langues.

1.2. Généralisation - validation du modèle

1.2.1. Sur le français

Les modèles monolingues du français ont été par ailleurs confrontés massivement à des données lexicales. Certains retours sur la modélisation ont eu lieu à la suite de ces confrontations aux données, et on peut considérer que les versions définitives des modèles produites à la fin du projet sont validées par les données qui ont pu être codées dans ces modèles.

Le modèle est donc validé en grandeur réelle. Les données issues des différents dictionnaires des membres du consortium ont été progressivement mises au modèle GENELEX.

Le travail de validation a porté dans un premier temps sur la morphologie, puis progressivement sur la syntaxe et sur la sémantique.

Au cours de la validation, certaines limites de la modélisation ont pu être mises en évidence et ont amené à des modifications ponctuelles et à certaines reformulations. Les modèles sont à présent stabilisés pour le français.

1.2.2. Sur d'autres langues

Le modèle générique proposé pour le français a été confronté à d'autres langues. L'équipe portugaise a en particulier travaillé sur celui-ci et déterminé en quoi il est adéquat ou non à la description de sa langue.

Le travail a porté dans un premier temps sur le modèle de la couche morphologique.

Il n'est pas surprenant que d'ores et déjà certains éléments du modèle de la couche morphologique demandent des aménagements :

- Un certain nombre de traits ont une liste figée de valeurs ; ces listes de valeurs ne conviennent pas nécessairement pour toutes les langues.
- Certains traits, qui n'existent pas dans le modèle du français parce que jugés inutiles, seront nécessaires pour d'autres langues.

La validation-généralisation pour le Portugais du modèle de la couche morphologique n'a pas remis en question la structuration globale ; elle a toutefois mené à une adaptation des listes de valeurs et des listes

de traits.

Il en est allé de même lorsque l'équipe française a travaillé à la modélisation de la couche morphologique de l'anglais. Il existe par ailleurs un premier ensemble de propositions pour adapter la couche morphologique du modèle à la langue allemande, mais ces travaux ne sont pas intégrés dans le cadre EUREKA - GENELEX.

Les couches syntaxique et sémantique du modèle GENELEX n'ont pas posé de problèmes particulier pour le portugais, et l'anglais n'a apporté qu'une ou deux modifications minimales sous forme de traits ou de valeurs inexistantes ou différents en français et en portugais. La structuration globale de ces deux couches et leur pouvoir d'expression et de description se sont avérés suffisamment puissants pour rendre compte des particularités de ces langues.

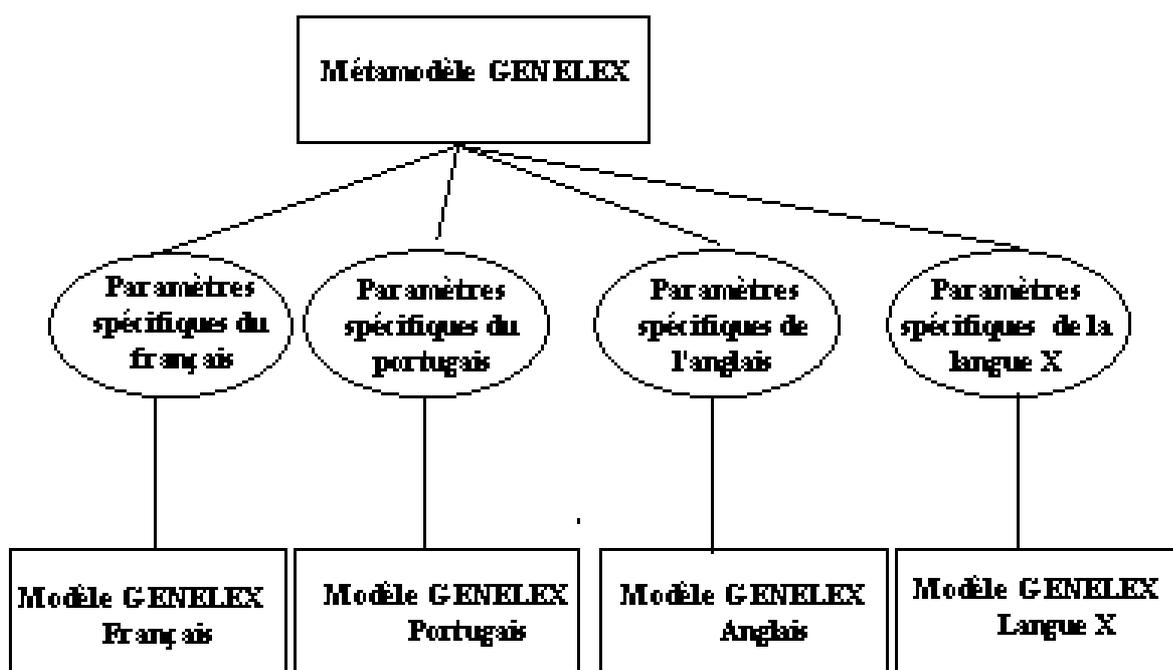
1.3. flaboration d'un métamodèle indépendant des langues représentées

La phase de validation et généralisation à d'autres langues que le français a abouti à la définition de DTD monolingues pratiquement semblables, et divergeant par des listes de valeurs différentes (pour les sous-catégories morpho-syntaxiques, les traits morphologiques et leur valeurs ...) Ces particularités de chaque langue sont exprimées dans des fichiers de valeurs propres à chaque langue.

Les divergences concernent la couche morphologique. Le modèle syntaxique comporte sans doute également des variations propres à une langue ou une autre qui s'expriment par des personnalisations différentes des listes de valeurs.

Le modèle de la couche sémantique est le plus ouvert et il n'a pas nécessité d'adaptation particulière pour traiter le portugais ou l'anglais. Le besoin de paramètres d'instanciation associés à une description par langue ne s'y est pas fait sentir.

Une façon de représenter les différents modèles monolingues exprimés en GENELEX serait la suivante:



Ce métamodèle est cependant, à l'heure actuelle, implicite et virtuel. Il se comprend comme l'ensemble du modèle du français, moins certaines listes fermées de traits et de valeurs.

3. Problèmes posés par la mise en correspondance de dictionnaires monolingues

Le modèle GENELEX *généralisé* est un modèle de dictionnaire monolingue capable de décrire des entrées lexicales de diverses langues européennes.

La mise en correspondance de deux dictionnaires GENELEX de deux langues différentes peut signifier la mise en correspondance, explicitement ou implicitement, de triplets "unité morphologique-unité syntaxique-unité sémantique". Cela signifie donc **associer non seulement des unités de sens, mais aussi des structures syntaxiques**, tout en accédant aux informations morphologiques. La mise en correspondance peut aussi être faite à un niveau de la couche sémantique plus abstrait. La couche "pivot" proposée dans le modèle sémantique est une réponse de type "INTERLINGUA" au problème du multilinguisme, et peut mener à l'absence de nécessité de mise en correspondance de deux dictionnaires monolingues.

Dans le modèle GENELEX, une entrée lexicale n'est pas décrite de façon isolée, mais dans un contexte. Cette façon d'aborder le problème est particulièrement intéressante pour résoudre les problèmes spécifiques que pose la mise en correspondance d'entrées lexicales de deux langues.

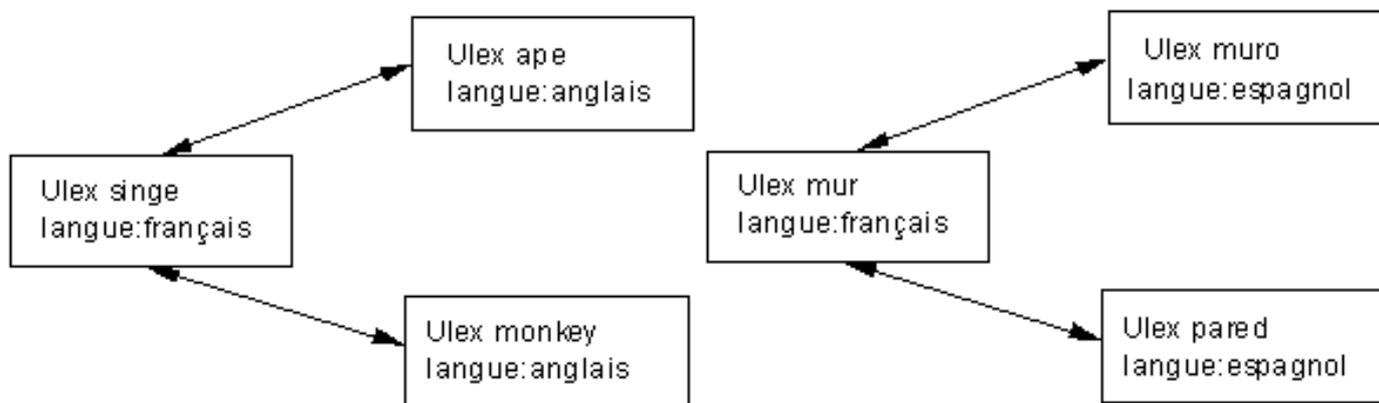
Cependant, il est bien connu que ce problème est plus compliqué que ce que laisse entendre la notion d'association "un pour un" entre unités lexicales. En effet, chaque langue a sa propre perception de la réalité décrite et son propre découpage de la réalité. Ces découpages différents font que la correspondance de langue à langue est généralement imparfaite, du point de vue du sens, et bien souvent divergente du point de vue de la structure de surface.

Différents types de divergences peuvent être identifiés. On en illustrera quelques-uns ci-dessous, sans prétendre rendre compte de façon détaillée et exhaustive de tous les problèmes posés par ces divergences. En particulier, on n'abordera pas les questions liées à des divergences où c'est la correspondance des objets dénotés plus que celle des signifiés qui fait problème (c'est alors en quelque sorte davantage le "multiculturalisme" que le multilinguisme qui est en jeu) : un exemple typique en est la désignation des entités juridiques, politiques ou administratives propre à un système institutionnel, et dont il est malaisé d'identifier les "correspondants" dans un autre système. Chaque situation particulière de traduction induit du reste, à ce sujet, des contraintes différentes (adaptation "transculturelle", conservation littérale, glose, etc.)

3.1 Divergences de contenu sémantique sans divergences structurelles

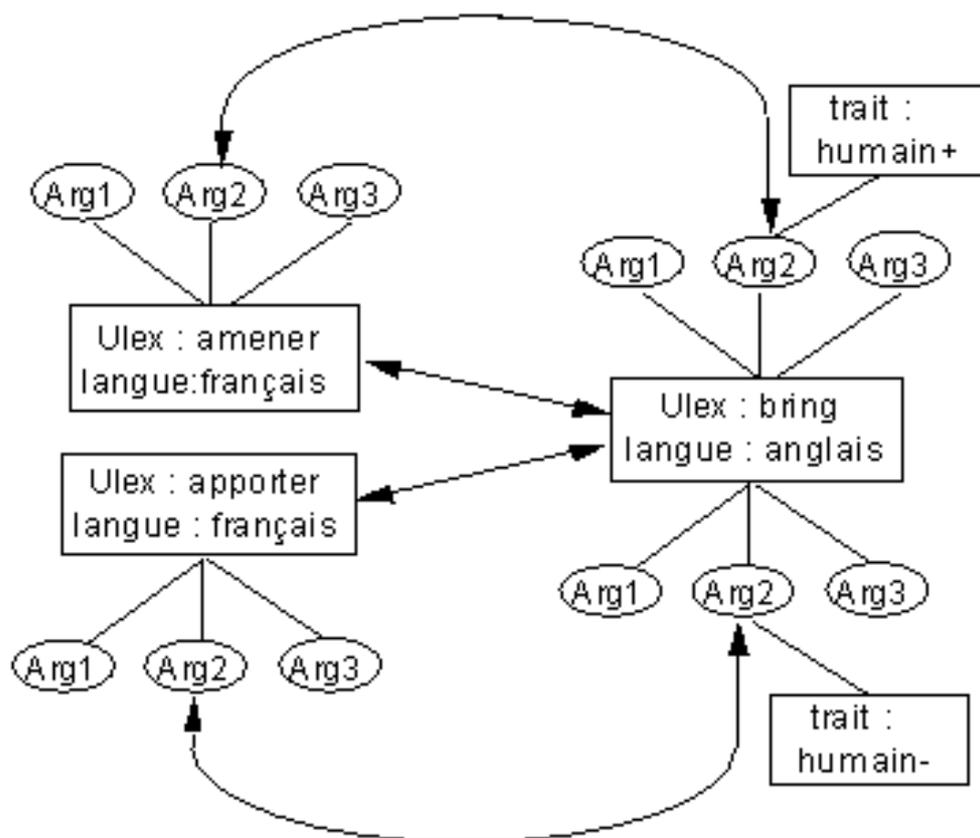
Il s'agit là de découpages différents de la réalité opérés par les langues, ou de choix de spécialisations lexicales divergents d'une langue à l'autre.

- Sens non entièrement recouvrant entre unités non-prédicatives. Cela pose, pour la correspondance, un problème de perte d'information dans un sens, et de choix d'unité lexicale dans l'autre sens.



(Cas simple de correspondance 1-2, en général les chevauchements d'unités sont plus complexes)

- Même problème de recouvrement partiel entre unités prédicatives, souvent associé à des structures prédicatives respectant l'ordre et le nombre des arguments, mais aux "signatures" argumentaires plus ou moins différentes, par exemple l'une plus contraignante que l'autre, ou pire, plus contraignante sur un argument et moins contraignante sur un autre argument.

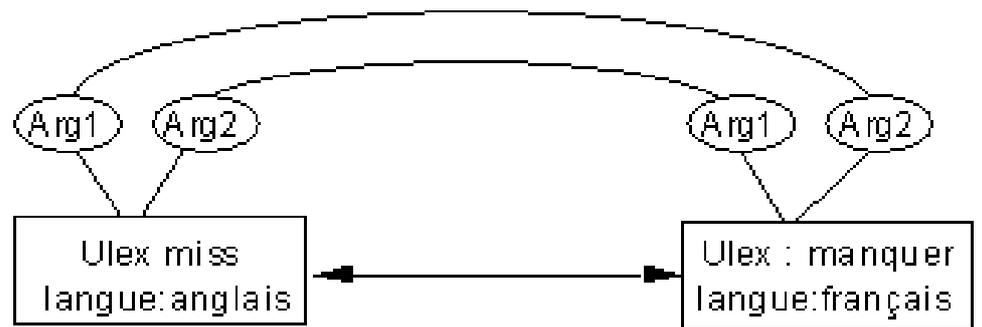


3.2 Divergences structurelles associées ou non à des divergences de contenu sémantique

- Thématization différente des arguments qui se manifeste par l'ordre différent des arguments dans des structures syntaxiques en correspondance. Cependant le nombre des arguments est conservé et prédicats et arguments se correspondent respectivement.

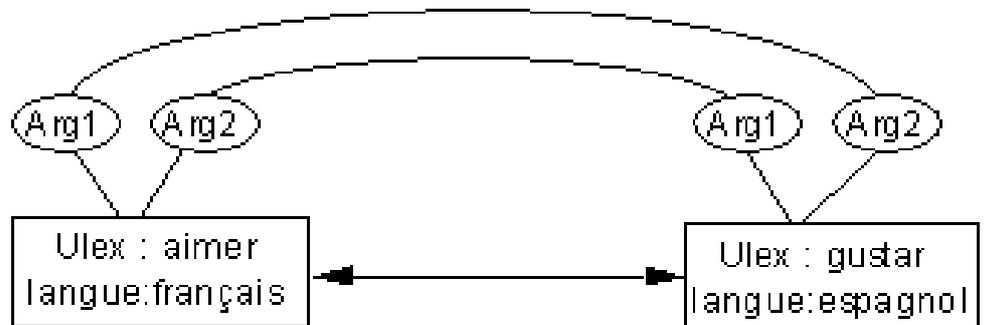
Ex: *I miss Max*

Max me manque



J'aime le football

El fútbol me gusta



- Divergences structurelles de surface: construction verbale/ verbe support, être + Adjectif/ verbe support + nom, etc

Ex: *J'ai faim*

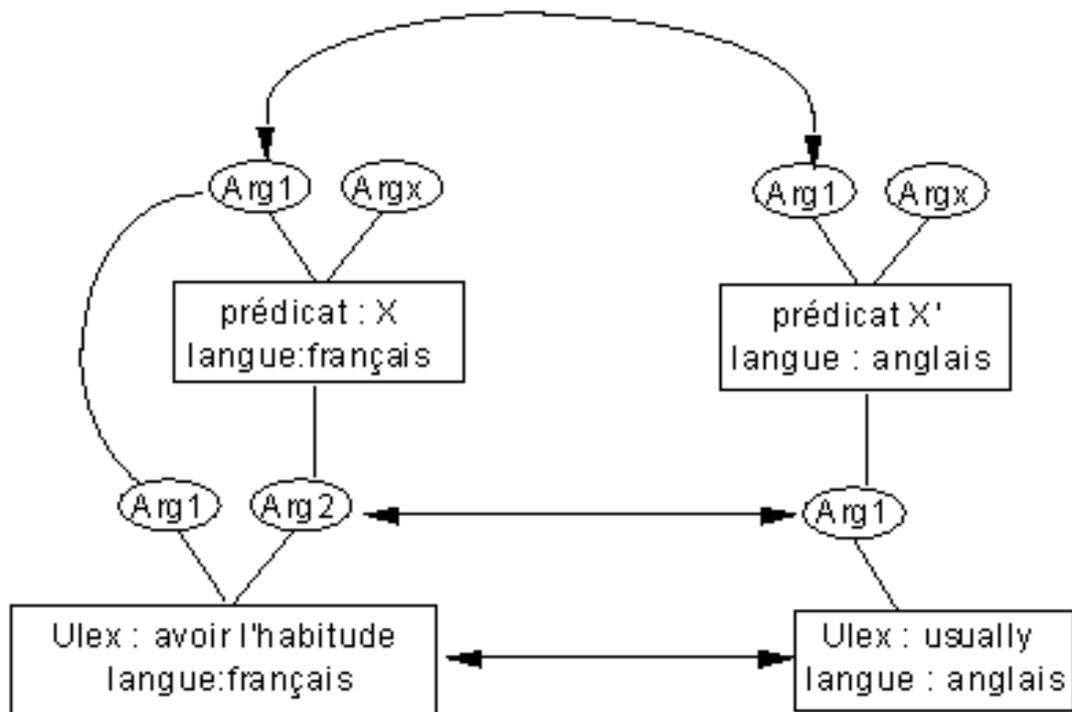
I am hungry

- Correspondance entre, dans une langue, un prédicat dont l'argument ou l'un des arguments est lui-même un prédicat, et un modifieur ou un verbe de modalité portant sur le correspondant du prédicat-argument (deux têtes syntaxiques et peut-être sémantiques suivant les choix de codage sont permutées). Cela mène à des structures syntaxiques fort différentes.

Ex: *J'aime manger*

Ich esse gern

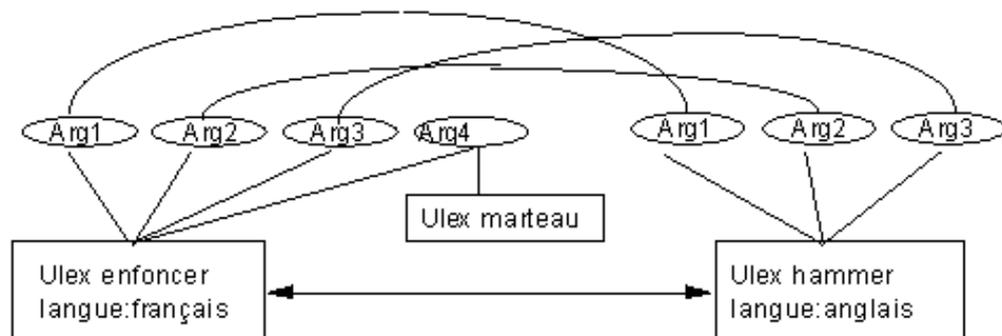
Max usually drinks at night

Max a l'habitude de boire le soir(expression composée *avoir l'habitude*)

- Un argument d'un prédicat est incorporé dans le prédicat équivalent: le nombre des arguments n'est pas le même, un argument est implicite dans une langue, explicite dans l'autre.

Ex Max hammers a nail in the wall

Max enfonce un clou dans le mur avec un marteau



4. Réponses offertes par le modèle GENELEX

L'hypothèse initiale de GENELEX au sujet du multilinguisme était que la plupart des questions épineuses concernant les informations de traduction, dès lors que c'est le plan du lexique qui est en jeu, dérivent d'une formalisation et d'une précision insuffisantes des informations lexicales monolingues. Il s'est avéré,

au cours des travaux sur le modèle multilingue, que la définition préalable du métalangage descriptif applicable aux dictionnaires monolingues facilitait considérablement l'établissement de propositions concernant la mise en correspondance multilingue. Celle-ci peut s'appuyer efficacement sur les mécanismes définis dans les modèles morphologique, mais surtout syntaxique et sémantique, même si des éléments de description contrastive ont bien évidemment dû être introduits. Au surplus, ces éléments eux-mêmes bénéficient largement des moyens d'expression utilisés dans les dictionnaires monolingues.

La conséquence immédiate de cette approche est que la composante multilingue, dans GENELEX, ne se conçoit pas indépendamment des dictionnaires monolingues qu'elle rapproche.

Comme nous l'avons vu plus haut, le problème est celui de la mise en correspondance à la fois d'unités de sens (Usem) et éventuellement de structures syntaxiques. Le modèle GENELEX offre un cadre où ces informations peuvent s'exprimer assez commodément, par le fait que les unités sémantiques sont, dans les dictionnaires monolingues GENELEX, associées à des unités syntaxiques, et que la mise en correspondance au niveau sémantique met implicitement en correspondance les unités syntaxiques associées aux unités sémantiques (on peut également préciser explicitement des correspondances mettant en jeu la syntaxe) .

Suivant les choix stratégiques globaux d'un dictionnaire de correspondance multilingue, suivant le type de descriptions existant dans les dictionnaires monolingues, on pourra placer les correspondances entre Usem, entre Prédicat, entre Concept, entre Prédicat et Concept, entre Usyn, et on pourra combiner différentes approches. Dans tous les cas, le besoin de compléter ces "connexions" par des informations précisant les éléments en correspondance se fait sentir. Cela s'exprime tout naturellement :

- dans la structure syntaxique par l'ajout de lexicalisations exprimées dans un langage commun aux descriptions syntaxiques du modèle ;
- dans la structure sémantique par l'ajout de modificateurs sémantiques et de traits valués, exprimés ici encore par des objets de description de la couche sémantique monolingue.

Avant d'aborder les différentes possibilités de mises en correspondance offertes par la modélisation fine monolingue, faisons le point sur l'approche "pivot" offerte par Genelex.

4.1 L'approche pivot

La couche sémantique propose un certain nombre d'éléments descriptifs, et parmi ceux-ci, certains éléments peuvent être marqués comme pivot. Différents choix stratégiques de gestion des pivots peuvent être faits, voyons de plus près ce que cela signifie.

Une couche descriptive "pivot" peut comprendre les objets descriptifs suivants : Prédicat, Concept, Trait sémantique valué, Relation entre Prédicat, entre Concept et entre Prédicat et Concept. Une couche descriptive pivot autonome suppose que ces éléments pivot se décrivent en ne s'appuyant eux-mêmes que sur des éléments de la couche pivot . Dans ce cas, nous avons une sorte de langage de description minimum, autonome, qui est déclaré pivot, et qui peut être partagé par les différentes couches monolingues.

La meilleure façon d'assurer un réel partage de cette couche serait de l'identifier comme telle et de l'inclure dans tout dictionnaire monolingue ayant choisi de se projeter (ne serait-ce que partiellement) sur le noyau pivot. Cette solution semble assez séduisante, car elle permet de fournir une approche

partiellement pivot par projection des objets propres à chaque langue sur ces objets pivot.

La difficulté qui demeure, bien sûr, est celle qui consiste à élaborer et mettre au point ce noyau pivot.

Les prédicats primitifs et certains concepts du modèle peuvent être considérés comme des primitives indépendantes des langues. Les unités sémantiques sont éventuellement projetées sur ce niveau dans le dictionnaire monolingue.

Le choix de primitives partagées par toutes les langues et la projection sur ce niveau des unités sémantiques permet donc d'avoir une représentation abstraite indépendante de la langue. Ils rendent donc inutiles les informations supplémentaires de correspondance entre langues, cette correspondance se faisant au niveau de la représentation intermédiaire unique. Évidemment, cette projection sur un niveau pivot se fait avec une perte de la finesse de l'information, et cette approche peut être complétée par un calcul sur certains traits portés par l'unité sémantique dont on cherche le ou les correspondants.

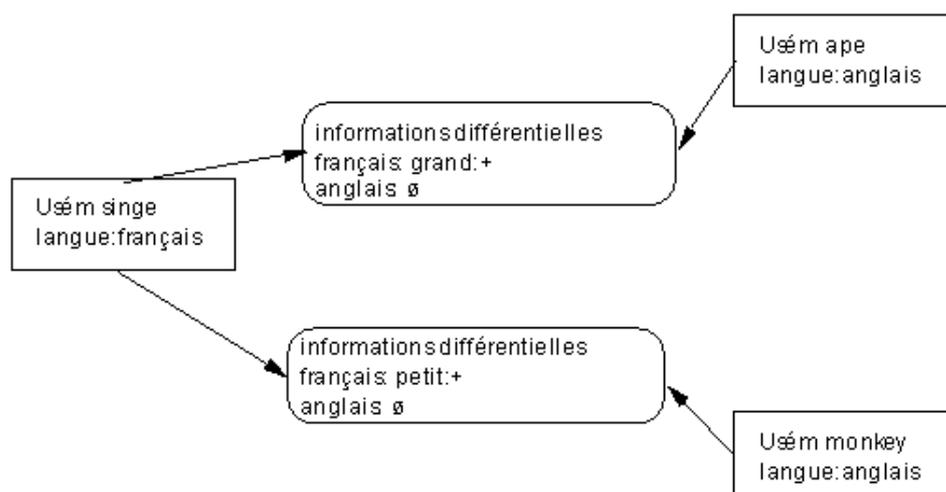
On peut imaginer que des traits dits *généraux* pourraient être *primitifs*, par exemple animé, humain, et certains traits de classes sémantiques, de niveau de langue, de domaine...

La plupart des relations sémantiques peuvent également être vues comme partagées par différentes langues, et donc *primitives*, même si certaines d'entre elles sont privilégiées par une langue et ignorées par une autre.

4.2. Au niveau des unités sémantiques

La correspondance semble pouvoir se faire assez bien entre unités sémantiques. Cependant, cette correspondance devra être complétée le cas échéant d'informations différentielles complémentaires.

- les divergences de sens peuvent s'exprimer par le fait que les unités de part et d'autre de la correspondance peuvent être porteuses d'informations complémentaires sous forme de traits sémantiques valués ou sous forme de "modifieurs" sémantiques, exprimés dans le langage de traits de la langue de l'unité concerné ou à l'aide de Prédicat de cette langue.

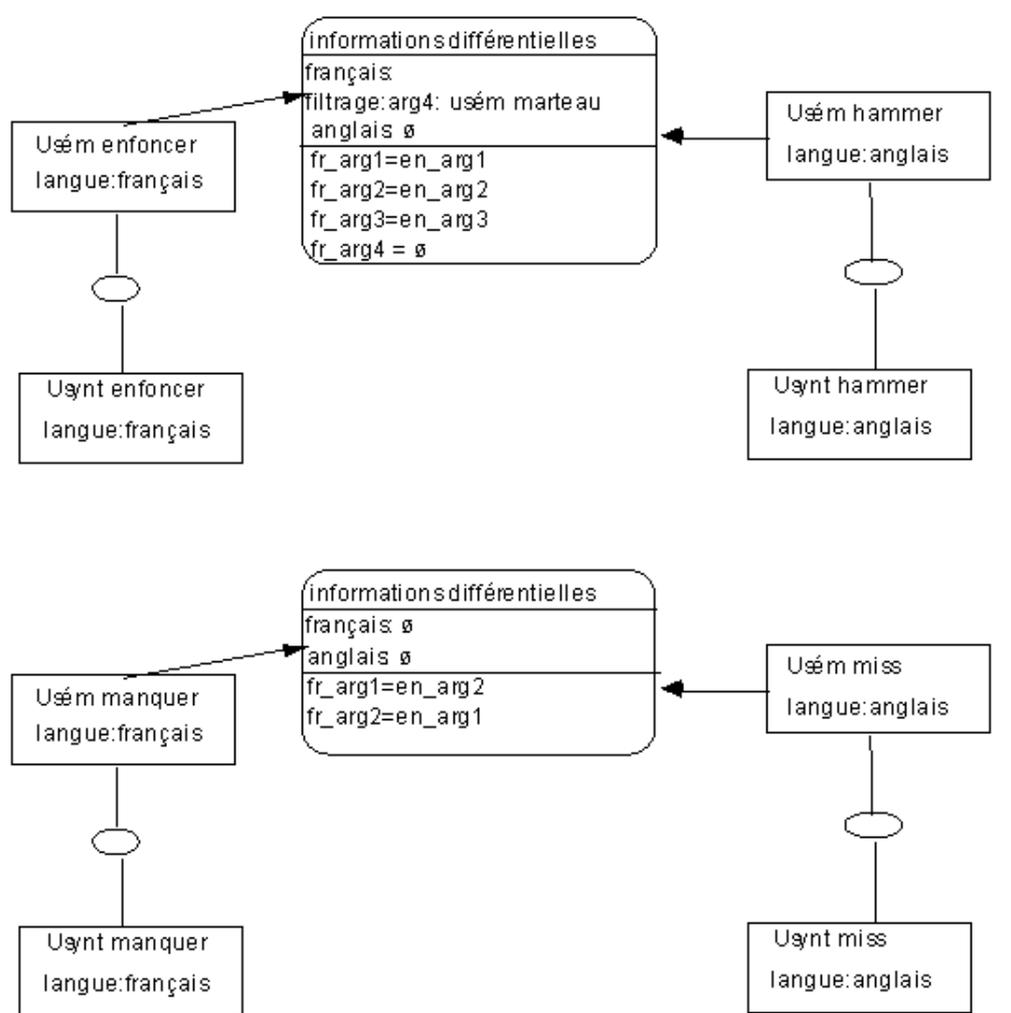


- le cas échéant, le filtrage des structures syntaxiques acceptées dans la correspondance peut se faire un peu comme il se fait dans le passage syntaxe-

sémantique: par des informations complémentaires de filtrage sur les unités syntaxiques mises en jeu par l'association de deux unités sémantiques de langues différentes.

La correspondance est alors une correspondance mettant en jeu Usem et Usyn, correspondance de couples.

- les valeurs des arguments incorporés, les correspondances d'arguments seront précisées pour les unités prédictives. Ces informations permettent à la fois de résoudre les problèmes de nombres d'arguments différents, de thématisations différentes. Les informations de structures syntaxiques différentes sont portées par les unités syntaxiques et n'ont pas besoin d'être précisées .

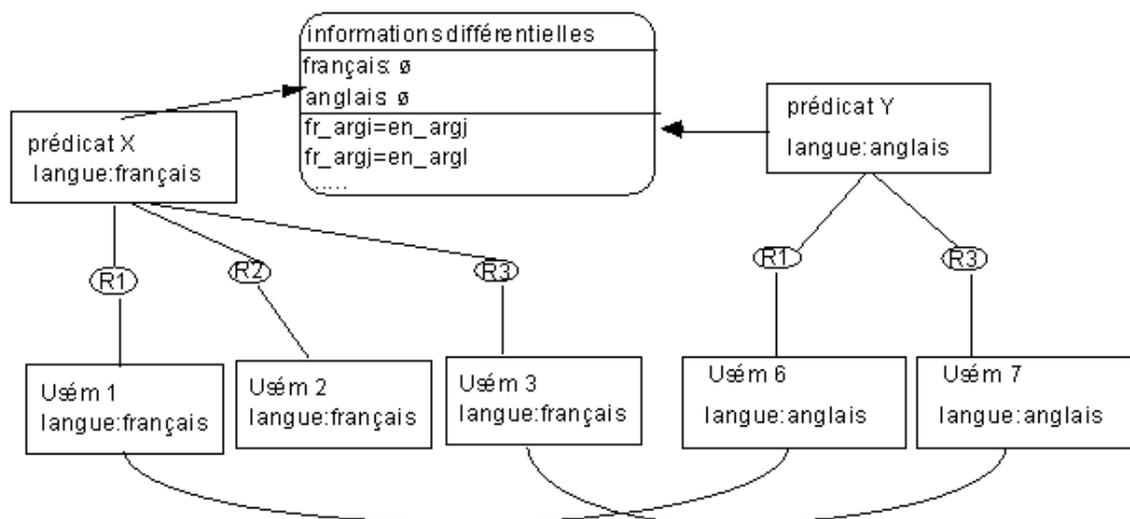


4.3. Au niveau des prédicats lexicaux ou non

Les informations concernant les prédicats permettent une certaine factorisation de l'information sémantique au niveau monolingue. On pourra donc factoriser parallèlement les correspondances entre langues en associant des prédicats lexicaux de chacune des langues quand ils se correspondent. Cette solution présente l'avantage de **mettre en relation un réseau dérivationnel avec un autre réseau**

dérivationnel, si l'on suppose que les relations sémantiques de dérivation permettent d'accéder aux différentes unités de ce réseau et que ces relations sont renseignées dans les deux langues mises en correspondance. Deux unités issues par une même relation de dérivation, de deux prédicats lexicaux en correspondance seraient implicitement en relation également.

Cela présente de plus l'intérêt de résoudre de façon économique le problème des dérivations sémantiques présentes dans une langue et absentes dans une autre, dans la mesure où la factorisation par la mise en correspondance des prédicats permet de s'orienter vers des traductions utilisant des paraphrases de l'unité source.



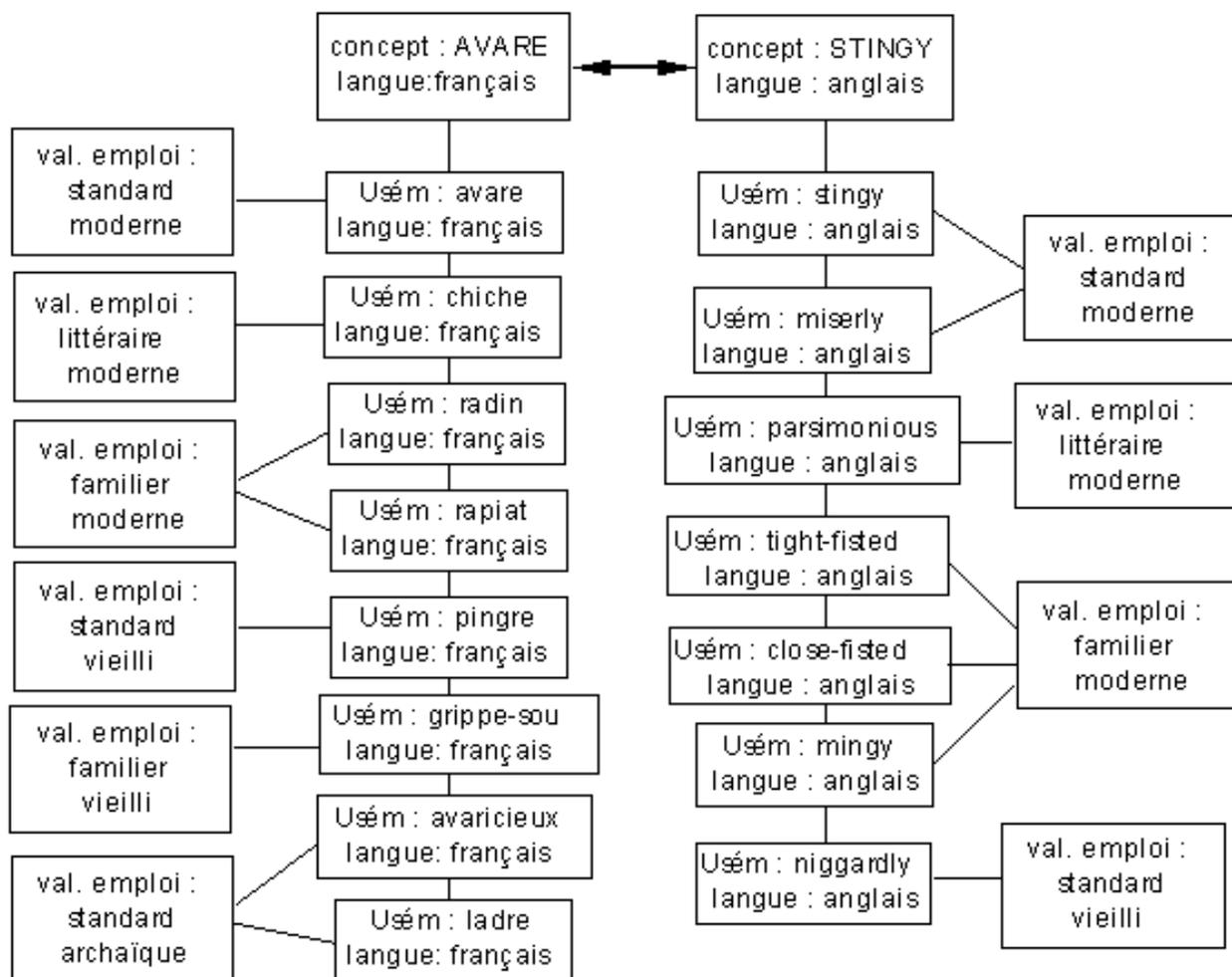
Les Prédicats ne se correspondant pas toujours exactement, l'information différentielle sera précisée de différentes façons suivant la nature des écarts sémantiques :

- en surchargeant l'information sémantique portée par les arguments, en les "instanciant"
- en surchargeant le Prédicat par des Traits sémantiques valués
- en ajoutant au Prédicat un Modifieur, lui-même un Prédicat mettant en relation le modifié et d'autres arguments dont on peut préciser l'instanciation.

4.4. Au niveau conceptuel

De la même façon, il sera possible de mettre en correspondance des Concepts représentant eux aussi une certaine abstraction sur le lexique, et potentiellement une factorisation d'unités lexicales. Ces Concepts ne sont pas forcément des unités "pivot", ce qui n'exclut pas toutefois qu'ils puissent avoir des représentations dans plusieurs langues différentes.

Ce type de mise en correspondance offre la possibilité de relier, par exemple, des paradigmes de synonymes plus ou moins complets dont chaque item n'a pas forcément de correspondant exact ni unique dans l'autre langue. Les informations, par exemple des valeurs d'emploi ou des traits pragmatiques, portées dans chacune des langues par les Usém lexicalisant un Concept peuvent alors prendre une valeur contrastive qui pourra être exploitée, dans une application, pour une sélection plus fine de l'unité cible pertinente.



L'exemple ci-dessus illustre l'intérêt qu'il y a parfois à adopter ce type de mise en correspondance, qui offre non seulement l'avantage d'une représentation "économique", mais aussi celui d'éviter un certain nombre de difficultés pour associer les unités sémantiques correspondantes deux à deux.

Comme pour les unités sémantiques, il sera possible de noter des différences de sens ou de portée entre les Concept ainsi associés, au moyen d'informations complémentaires sous forme de traits sémantiques valués ou sous forme de "modificateurs".

4.5 Au niveau syntaxique

Si le niveau sémantique, aussi bien par ses représentations proches du lexique que par les abstractions qu'il autorise, est particulièrement apte à servir de support aux informations multilingues, le modèle GENELEX prévoit néanmoins la possibilité d'établir des correspondances au niveau des unités syntaxiques.

Cette ouverture a semblé nécessaire notamment en raison de la mise en évidence, dans le travail du traducteur, d'éléments de rang supérieur aux unités lexicales (fussent-elles des unités syntaxiques composées) pour lesquels le dictionnaire multilingue peut avoir intérêt à enregistrer des équivalences. De tels éléments, que l'on peut appeler "tournures", relèvent d'une composante "phraséologique", intermédiaire entre des unités susceptibles de recevoir une description lexicale complète et des unités phrastiques telles que ce que peut enregistrer, par exemple, une mémoire de traduction. Elles présentent les caractéristiques suivantes :

- elles sont potentiellement récurrentes dans un sous-langage ;

- elles sont fortement contraintes structurellement et au moins partiellement contraintes lexicalement ;
- leur rang peut s'étager entre un syntagme nominal simple et une proposition, voire une phrase complexe, mais elles ne représentent pas forcément un constituant syntaxique unique ou complet ;
- elles ne nécessitent pas une interprétation sémantique fine de chacune des unités lexicales qu'elles impliquent pour pouvoir être traduites.

Des exemples typiques de ce type d'éléments se rencontrent dans les formules de politesse, certaines structures conventionnelles de la langue juridique ou administrative, etc.

C'est donc au niveau syntaxique que pourra être traitée la mise en correspondance de telles tournures, ou éléments de phraséologie. On s'appuiera alors sur les mécanismes descriptifs mis au point dans la couche syntaxique pour détailler les diverses contraintes structurelles et/ou lexicales nécessaires et suffisantes pour la détermination d'un contexte valide dans chacune des langues. On aura en pratique deux types de représentation de ces unités :

- une mise en correspondance impliquant des Usyn, simples ou composées, en surdéterminant structurellement le contexte défini par leur description, par l'expression de contraintes sur les syntagmes occupants de position ou par l'ajout de position supplémentaires ;
- l'intégration aux dictionnaires monolingues d'Usyn composées (dès lors que ces tournures peuvent avoir une valeur en langue indépendamment des besoins de traduction), sans que des unités sémantiques leurs soient nécessairement associées ; la mise en correspondance, au niveau multilingue, sera alors effectuée plus directement.

Il va de soi que ce mécanisme de mise en correspondance au niveau syntaxique peut être utilisé plus largement que pour le traitement des éléments phraséologiques, si l'on souhaite (et si l'on peut) éviter le passage par un niveau sémantique. Le système d'expression de contraintes structurelles dans la mise en correspondance d'Usyn permet notamment de rendre compte simplement de phénomènes de cooccurrences déterminant fortement le type de traduction à privilégier.

4.6 Mixte : couples Usyn-Usem

On aura également la possibilité de lier une Usem et une Usyn dans chacune des langues traitées, de manière à contraindre, dans la mise en correspondance, le contexte syntaxique à instancier dans une traduction. Ce mécanisme permet de privilégier, si on le souhaite, un certain isomorphisme syntaxique dans la traduction, ou d'indiquer une préférence parmi les structures syntaxiques possibles pour une unité.

B - MANUEL DE L'UTILISATEUR

1. Généralités

Cette version du modèle GENELEX multilingue, et donc la DTD qui en est l'expression de référence, ne prévoient pas de dispositions particulières pour la mise en place d'une approche fondée sur un noyau "pivot" d'informations lexicales. La correspondance entre plusieurs langues dans cette approche doit en effet découler de la formalisation adoptée dans la couche sémantique de chaque dictionnaire monolingue

concerné, et suppose que les informations déclarées dans ces dictionnaires comme "pivot" se retrouvent en effet dans chacun de ces dictionnaires. Aucune mise en correspondance explicite n'y serait donc nécessaire. En l'état, les différents éléments d'un noyau de type "pivot" ou "Interlingua" font partie de la couche sémantique d'un ou plusieurs dictionnaires GENELEX, o• ils sont marqués comme tels. Il serait néanmoins souhaitable que ces éléments, qu'on les ait définis par un travail de généralisation à partir de la description sémantique d'une langue, ou (à plus forte raison) qu'on les ait établis a priori, trouvent leur place dans la composante multilingue d'un ensemble de dictionnaires GENELEX, ne serait que pour faciliter la mise en cohérence de plusieurs dictionnaires monolingues de ce point de vue. Une version ultérieure du modèle pourra donc comprendre les aménagements nécessaires à cette remise à niveau.

Le modèle proposé est un modèle de mise en correspondance par couple de langues.

Dans les modèles Entités-Relations, afin de rendre les correspondances multilingues plus facile à interpréter et les différents éléments de la couche monolingue de chacune des langues bien différenciés, on a systématiquement distingué les deux langues (Langue1 et Langue2) et les objets de chacune de ces langues. On a donc une représentation du modèle qui fait intervenir les éléments des couches monolingues de façon symétrique et explicite. Cela ne doit pas faire penser que les modèles des deux langues divergent dans la couche sémantique ou syntaxique. Les éléments sur lesquels se basent les correspondances sont des éléments conceptuellement identiques et partagés par les modèles conceptuels des deux langues.

1.1 Un modèle basé sur la sémantique et la syntaxe monolingues

Le modèle du multilinguisme a pour caractéristique de mettre en relation deux dictionnaires décrits dans le modèle GENELEX. Les couches sémantique et syntaxique y sont modélisées de façon très fine, la description s'appuyant sur de nombreux éléments descriptifs combinés entre eux et visant à une explicitation maximum de l'information.

La mise en correspondance de deux dictionnaires monolingues se fait donc assez simplement en s'appuyant massivement sur les objets descriptifs des couches sémantique et syntaxique, la couche multilingue n'ayant besoin que d'assez peu de nouveaux objets descriptifs.

1.2 Un modèle non-orienté

La correspondance est définie comme "non orientée", et on n'aura pas à proprement parler de dictionnaire source et de dictionnaire cible. Cependant, dans la mesure o• les langues divergent nécessairement pour certaines unités, et suivant l'approche descriptive et méthodologique adoptée, certaines correspondances seront plus faciles à utiliser dans un sens que dans l'autre pour une application.

L'attribut *statut* permet de marquer le statut de la correspondance, en particulier, lorsque la correspondance se fait malgré certaines divergences, il est intéressant de savoir si l'un des éléments est plus générique que l'autre, plus spécifique, ou simplement s'il n'y a que recouvrement partiel. Les valeurs de statut ne sont pas contraintes, les utilisateurs pourront se définir la liste qu'ils souhaitent utiliser. Le statut permet de déduire l'orientation plus ou moins forte de la mise en correspondance multilingue.

Un attribut *point de vue* est porté par les correspondances, il sert à marquer les préférences de traduction ou d'association liées à différentes applications partageant les mêmes données lexicales, mais divergeant éventuellement pour certaines correspondances (par exemple pour tout ce qui concerne la terminologie et le langage spécialisé). Cet attribut permet à différents utilisateurs de "personnaliser" leurs correspondances tout en partageant un même dictionnaire multilingue.

1.3 Un modèle ne déformant pas l'approche monolingue

Le modèle multilingue a été défini afin de ne pas imposer une description monolingue qui ne serait pas justifiée dans la couche monolingue, mais par les besoins de mise en correspondance. En particulier, les distinctions de sens dans le dictionnaire monolingue sont faites sur des critères internes à la langue décrite, et non pas pour répondre aux besoins de mise en correspondance multilingue.

1.4 éléments complémentaires

Les écarts sémantiques entre des éléments des deux langues en correspondance sont comblés par des informations complémentaires apportées par la mise en correspondance.

Ces informations complémentaires s'expriment, pour les objets de la sémantique, par des "modificateurs sémantiques" qui viennent préciser par modification l'élément principal de la correspondance, et par des traits sémantiques valués qui viennent se surajouter à ceux que porte déjà l'élément en correspondance.

La modification pour le niveau syntaxique s'exprime par la réécriture et la lexicalisation partielle ou totale de syntagmes, et par l'ajout de syntagmes partiellement ou totalement lexicalisés jouant le rôle de modifieur.

1.5 Un modèle offrant différents niveaux de mise en correspondance

Le modèle monolingue offre beaucoup de possibilités de description, et la mise en correspondance peut se faire à différents niveaux : non seulement entre Usem, ce qui semble le plus immédiat, mais également, en tirant parti des généralisations offertes par la couche monolingue, entre Predicat et Concept. La correspondance au niveau syntaxique (entre Usyn) est également possible ainsi qu'une approche "mixte"

Dans un même dictionnaire de correspondance peuvent sans aucun problème cohabiter des mises en correspondance de différents niveaux, certaines divergences entre langues se compensant mieux à certains niveaux.

2. Correspondance d'Usem

La correspondance qui semble la plus évidente et la plus intuitive est celle qui met en relation deux Usem.

Cette correspondance multilingue permet d'associer des Usem de chacune des langues indépendamment des contextes syntaxiques qui leur sont associés, c'est à dire indépendamment de la ou des Usyn à laquelle est associée chacune des Usem dans la correspondance syntaxe-sémantique.

La correspondance d'Usem fait implicitement se correspondre toutes les Usyn compatibles avec l'Usem langue1 avec toutes les Usyn compatibles avec l'Usem langue2. étant donné que la correspondance syntaxe sémantique précise comment les arguments (des Usem prédicatives) et les positions sont associées (en monolingue), il suffit de préciser dans la correspondance multilingue comment se correspondent les arguments : les correspondances entre positions des différents couples d'Usyn en relation implicites sont calculables. Cette correspondance entre Usem ne privilégie aucune association d'Usyn, toutes les combinaisons sont également pertinentes du point de vue de la correspondance figurant dans le lexique. Si l'on souhaite filtrer sur la syntaxe ou associer préférentiellement certaines Usyn entre elles, on utilisera alors des Correspondances entre couples Usyn-Usem (voir ci-dessous).

La correspondance d'Usem fait se correspondre des structures argumentaires quand les deux Usem sont liées à des Predicat ; elles peuvent aussi lier deux Usem dont l'une est prédicative et l'autre non.

Les Corresp_Multilingue Arg-Arg précisent la correspondance entre les arguments des Prédicat associés aux Usem de chacune des langues. Lorsqu'il n'y a pas le même nombre d'arguments (ou lorsque l'une des

Usem est non-prédicative : pas d'argument du tout), on peut préciser d'un point de vue sémantique les arguments associés à l'une des Usem via son Predicat, on peut "instancier" les arguments comme cela se fait dans la couche sémantique monolingue dans le cas de relations entre prédicats à nombre d'arguments différents. Deux Usem peuvent donc se correspondre indépendamment du fait que l'une est "prédicative" et l'autre non.

Filtre_enrichit_par_trait (_1 et _2) permettent de "surcharger" les Usem d'un point de vue sémantique en ajoutant des traits sémantiques valués (de la couche monolingue)

Filtre_enrichit_par_modif (_1 et _2) complètent par un Modifieur qui précise par modification le sens des éléments en correspondance.

Ces deux types d'informations complémentaires ou différentielles permettent de combler (totalement ou partiellement) au niveau sémantique l'écart qui sépare le sens des deux entrées en correspondance.

Un attribut *autonomie* permet de préciser si la correspondance est autonome, c'est à dire valable et pertinente prise isolément, ou si elle n'est à prendre en considération qu'en tant qu'élément d'un couple de correspondance syntaxe-sémantique, auquel cas l'attribut marque sa non-autonomie.

3. Correspondance de Prédicats

La correspondance entre Predicat est presque identique à la correspondance entre Usem. Cependant, le Predicat représente un certain niveau d'abstraction (plus ou moins grand) par rapport à l'Usem directement lexicale, et il découle de cela que la correspondance entre Predicat se situe à un niveau plus abstrait, et donc plus loin des informations syntaxiques sous-jacentes.

En effet, un Predicat est, directement ou pas, lié à des lexicalisations, mais le cheminement vers ces lexicalisation peut être complexe, et la correspondance au niveau des Predicat permet de s'abstraire plus aisément de divergences structurelles du point de vue de la syntaxe sous-jacente. En effet, la correspondance de Predicat laisse une grande liberté d'utilisation des informations monolingues qui associent Usem et Prédicat, et en associant des Predicat, c'est en quelque sorte un ensemble d'Usem qu'on associe, indépendamment de contextes syntaxiques, de catégories syntaxiques qui peuvent être variées (un même Predicat peut être associé dans le lexique à un nom, un verbe, et même un adjectif ...).

4. Correspondance d'Arguments

La correspondance d'arguments se fait suivant le même principe que dans les Relations valuées entre Predicat ou Usem en ce sens qu'elle associe deux chemins vers des arguments :

- chemin_arg_langue1, chemin_arg_langue2 (suite de nombres qui permettent d'atteindre un nième argument, et si celui-ci est prédicatif, d'atteindre récursivement l'un de ses arguments...)

- informe_arg_langue1_1 , informe_arg_langue2_1 permettent de préciser par un ajout d'informations sémantiques ces arguments

les chemins vers les arguments sont optionnels, ce qui permet de décrire qu'un argument n'est en correspondance avec aucun dans l'autre langue, et de le préciser sémantiquement

il est donc possible de gérer les nombres différents d'arguments et d'"instancier" certains arguments.

Au moins l'un des chemin_arg_langue(1 ou 2) est présent, et aucun informe_arg_langue(1 ou 2)_1 ne peut être présent si le chemin_arg_langue(1 ou 2) de sa langue n'est pas présent.

5. Correspondance de Concepts

La correspondance entre Concept est analogue à la correspondance entre Prédicat, du point de vue de l'abstraction par rapport aux Usem. Il n'y a ici pas de correspondance d'arguments, et donc aucune référence, même par un cheminement implicite, à des correspondances de Position de la syntaxe.

La correspondance de Concept permet donc de se faire correspondre à un certain niveau d'abstraction des objets descriptifs sans tenir compte des écarts qui peuvent être importants en ce qui concerne les contextes de surface associés.

6. Correspondances associant un Concept à un Predicat

Afin de bien distinguer les objets de la Langue 1 de ceux de la Langue 2, nous avons deux Correspondances associant un Predicat (objet descriptif de la Langue1 -resp. Langue2) à un Concept (objet descriptif de la Langue2 -resp Langue1). Il s'agit de Correspondance utilisées lorsque les divergences entre les deux langues sont assez importantes : soit que des choix descriptifs opposés aient été pris pour des objets relativement proches, soit le plus souvent qu'une langue nomme ce qui correspond dans une autre à un Predicat aux arguments instanciés .

C'est la raison pour laquelle les Argument du Predicat seront le plus souvent précisés ou instanciés par les SelectEtPreciseArg (de la couche sémantique). En effet cette correspondance sera le plus souvent entre un Predicat ayant "instancié" ses arguments (tous ou non) et un Concept, associés ou non de part et d'autre à des modifieurs.

- SelectEtPreciseArg pointent vers les informations précisant les arguments du Predicat

7. Modifieur

Un Modifieur vient compléter un élément sémantique (Usem, Predicat ou Concept) en le modifiant.

La précision de sens se fait en s'appuyant sur un Predicat qui met en relation l'élément modifié et le ou les autres arguments modifiant. Le Prédicat permet de préciser la relation qui lie le modifié aux autres arguments qui ont bien sûr par ailleurs un ou des rTMles sémantiques. Il faut bien remarquer que les prépositions pleines du point de vue sémantique, peuvent être au niveau sémantique associées à des Predicat à au moins deux arguments : les deux éléments -minimalement, mais il peut y en avoir plus- qu'elles mettent en relation. Ainsi, toujours minimalement, le Predicat associé à l'un des sens d'une préposition peut venir modifier l'élément principal de la correspondance en intervenant dans un Modifieur mettant en relation le modifié et un autre argument qu'un ou des SelectEtPreciseArg viendra préciser.

Un Modifieur pointe donc vers le Predicat modifieur.

Les SelectEtPreciseArg précisent les arguments de ce Predicat.

L'un des arguments de ce Predicat n'est pas "instancié" autrement que par le fait qu'il est associé à

l'élément qu'il modifie. Le `Chemin_arg_modifie` précise le chemin vers l'argument du Predicat qui est le modifié et qui est donc associé à l'élément que le Modifieur vient préciser, élément principal de la correspondance qui peut être Concept, Predicat, ou Usem suivant la correspondance multilingue dans laquelle le modifieur intervient. Cet argument permet en quelque sorte de préciser la connexion entre le Modifieur et celui qu'il modifie.

8. Correspondance d'Usyn

La correspondance multilingue d'Usyn permet de se faire correspondre des Usyn indépendamment du ou des sens associés. Cette correspondance d'Usyn est utile lorsqu'on ne souhaite pas associer une représentation sémantique à certaines Usyn que cependant l'on souhaite pouvoir mettre en correspondance multilingue (c'est par exemple le cas très probablement de la phraséologie, qui pourra être décrite au moyen d'Usyn composées auxquelles on n'associera pas nécessairement d'Usem). C'est intéressant également comme raccourci quand les langues distinguent différents sens pour les Usyn, mais que les Usyn se correspondent indépendamment de ces différents sens (autrement dit, les deux langues identifient les mêmes sens pour les Usyn en correspondance). Un module de traduction pourra utiliser ces correspondances sans avoir à désambiguer et choisir les Usem concernées par l'énoncé à traduire.

Une approche syntaxiste de la traduction pourra également s'appuyer sur ce types de correspondances.

Les `CorrespMult_Pos_Pos` précisent les correspondances entre les positions des Descriptions associées aux Usyn de chacune des langues, que celles-ci soient présentes dans la couche monolingue ou surajoutée par la correspondance en réécrivant un syntagme non-terminal ou en insérant une Position dans une liste de Position.

- `contraint_syntaxe(_1 ou _2)` pointe vers un `ContraintDescription` afin de restreindre les réalisations des contextes syntaxiques associées aux descriptions (même mécanisme que lors de la correspondance syntaxe sémantique) : contraint ou interdit la réalisation d'une position optionnelle, force une position à se réaliser par un certain syntagme élément de la liste des syntagmes occupant cette position, rajoute des contraintes sous forme de traits à certain syntagmes... (voir le modèle et la DTD sémantique pour plus de détails) ;

- `Reecrit_Lexicalise_synt(_1 ou _2)` permet de contraindre et préciser structurellement par une réécriture avec lexicalisations ou non un syntagme non-terminal qui est présent dans la description sans être contraint structurellement. Les contraintes structurelles (réécriture sous forme de listes de Position dont l'ordre est pertinent) devront être compatibles avec les traits éventuellement portés par le syntagme.

Plusieurs `Reecrit_Lexicalise` peuvent être présents de part et d'autre.

Les `Ajoute_Position` permettent de rajouter des positions aux éléments de la description qui s'expriment par une liste de Position, liste dans laquelle s'insère la Position ajoutée.

Les Usyn sont ainsi "surchargées" par l'ajout de Position à fonction "modifieur", les réalisations de ces positions pouvant par ailleurs être précisées par des feuilles lexicalisées.

Les `Ajoute_Position` permettent donc d'insérer des modifieurs entièrement, partiellement ou pas lexicalisés dans la structure syntaxique de référence qui est associée à l'Usyn.

`Ajoute_Position` et `Reecrit_Lexicalise` jouent donc, au niveau syntaxique, le même rôle que les Modifieur et les `Trait_Sem_ValPond` ajoutés au niveau sémantique : ils permettent de combler explicitement l'écart qu'il y a entre deux unités en correspondance.

Un attribut *autonomie* permet de préciser si la correspondance est autonome, c'est à dire valable et pertinente prise isolément, ou si elle n'est à prendre en considération qu'en tant qu'élément d'un couple de correspondance syntaxe-sémantique, auquel cas l'attribut marque sa non-autonomie.

9. Reecrit_Lexicalise, Ajoute_Position

Reecrit_Lexicalise sert à contraindre structurellement (par une liste de Position) un syntagme occupant de position non réécrit de la syntaxe et pour éventuellement le lexicaliser entièrement ou partiellement par des traits lexicaux (de la couche syntaxique) portés par les syntagmes entrant dans sa réécriture à un niveau quelconque.

Dans le cas où une contrainte de structure est déjà présente mais où l'on veut ajouter une position "libre" (modifieur) à un certain niveau (le plus haut, construction, ou non), on peut le faire par un **Ajoute_position**, la position précisera alors son ou ses occupants, le plus souvent lexicalisés. (voir ci-dessous)

Les attributs portée, nième_composition, nième_composante ont le même sens que dans la couche sémantique : ils permettent d'identifier (en s'associant si nécessaire à CheminSyntagme) le Syntagme à réécrire qui peut dépendre de la Construction (externe), de la Structure interne, ou d'une description de composés syntaxique par Composition, seul cas où nième_composition, et nième_composante sont renseignés.

CheminSyntagme permet d'atteindre le Syntagme à réécrire. Il existe évidemment la contrainte que le Syntagme soit un Syntagme_NT dont la liste de Position est vide. Cette liste de Position (réécriture du Syntagme_NT) est donnée par position_1 qui pointe vers des Position.

AjoutePosition permet d'ajouter une Position (non essentielle du point de vue monolingue, sinon on peut supposer qu'elle serait présente dans la couche monolingue) à un niveau quelconque de la Description.

portée, nième_composition, nième_composante ont le même sens que dans la couche sémantique et que dans **Reecrit_Lexicalise** : ils permettent d'identifier (en s'associant si nécessaire à CheminSyntagme) où doit s'insérer la position.

CheminSyntagme permet d'atteindre le Syntagme réécrit dans la liste de Position duquel va s'insérer la Position ajoutée. Le syntagme doit donc bien être à la fois non terminal et réécrit.

L'attribut rang permet de préciser où cette Position s'insère. Rang est optionnel car la Position n'est pas nécessairement contrainte du point de vue de sa place dans la surface et donc du point de vue de son rang dans la liste. Rang indique l'ordre d'apparition dans la liste résultant de l'insertion de la Position et aussi de l'insertion de Self le cas échéant. Rang compte donc à partir de 1. Il ne faut surtout pas confondre le rang d'insertion avec le rang des Positions dans la liste, même si ce n'est pas sans rapport.

Ex portée = EXTERNE, pas de CheminSyntagme, rang = 2 signifie que la Position interviendra dans la liste de Position de la Construction, et qu'elle s'insère en deuxième dans la liste.

Afin d'éviter les problèmes que pose éventuellement l'insertion du SELF (inserterself), le rang comptera SELF inséré comme un élément virtuel de cette liste. Ainsi, dans l'exemple: si insereself = 1, et rang = 2, la Position insérée s'insère avant SELF, si rang = 3, la Position s'insère après SELF et avant la deuxième Position.

10. Correspondance de couples Usem-Usyn

Cette mise en correspondance multilingue permet de s'appuyer à la fois sur la couche sémantique et la couche syntaxique de chacune des langues mises en correspondance et permet d'associer des couples Usem-Usyn de chaque langue en correspondance

Ces couples comportent les Usem pointées par le CorrespMult_Usem référencé par `corresp_mult_usesm`, et les Usyn pointées par le CorrespMult_Usyn référencé par `corresp_mult_usyn`.

Ces couples doivent bien sur être compatibles, c'est-à-dire que dans chacune des langues, l'Usyn est associée à l'Usem par la correspondance syntaxe-sémantique monolingue.

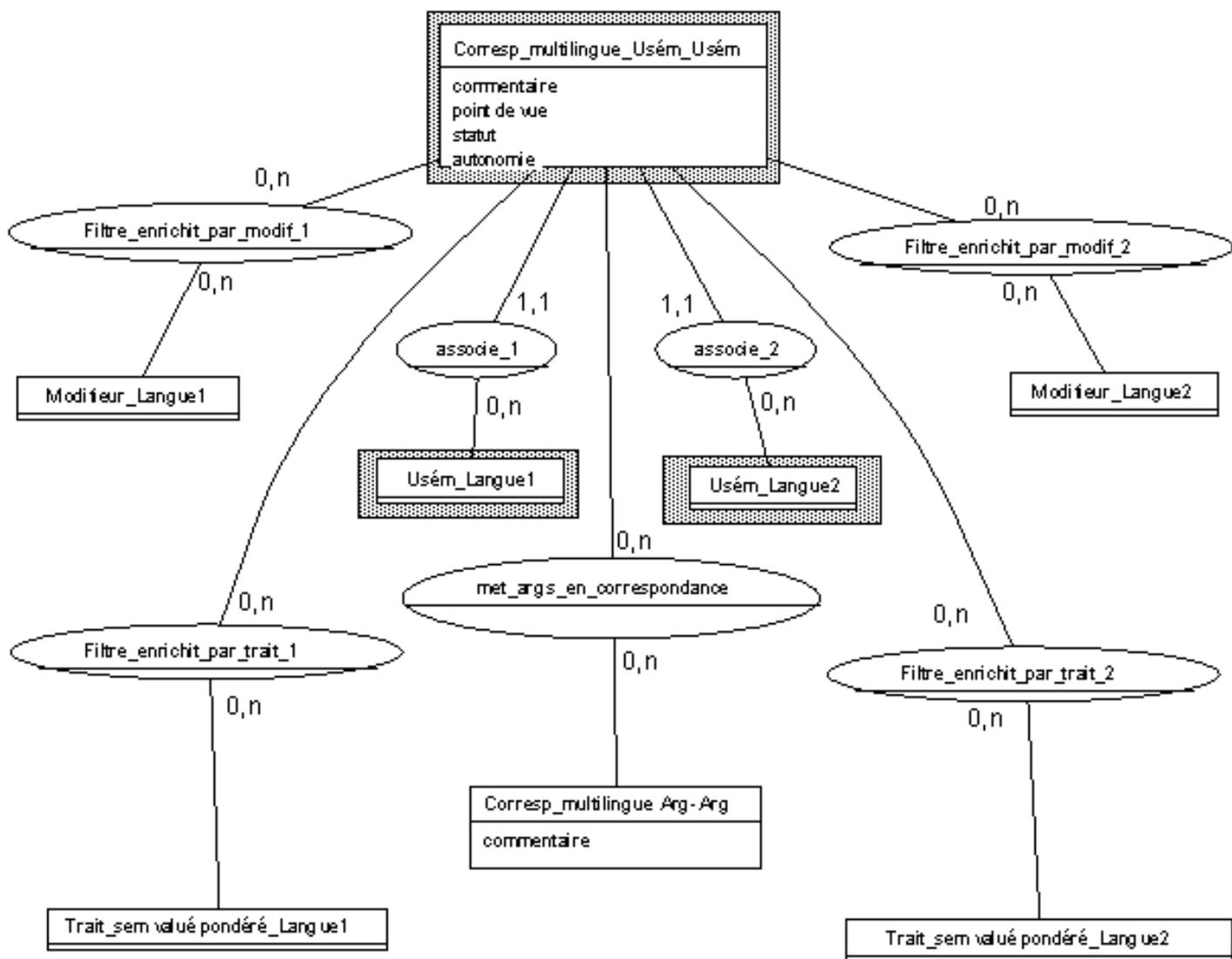
Aussi bien les Usyn que les Usem peuvent intervenir dans plus d'une CorrespMult_Usyn_Usem.

Les CorrespMult_Usyn et CorrespMult_Usem intervenant dans ces correspondances peuvent être autonomes (et donc être des correspondances pertinentes indépendamment de leur appartenance à des CorrespMult_Usyn_Usem et donc renseignées à un niveau suffisant de détails) ou être non-autonomes (auquel cas elle peuvent être renseignées précisément ou bien plus minimalement).

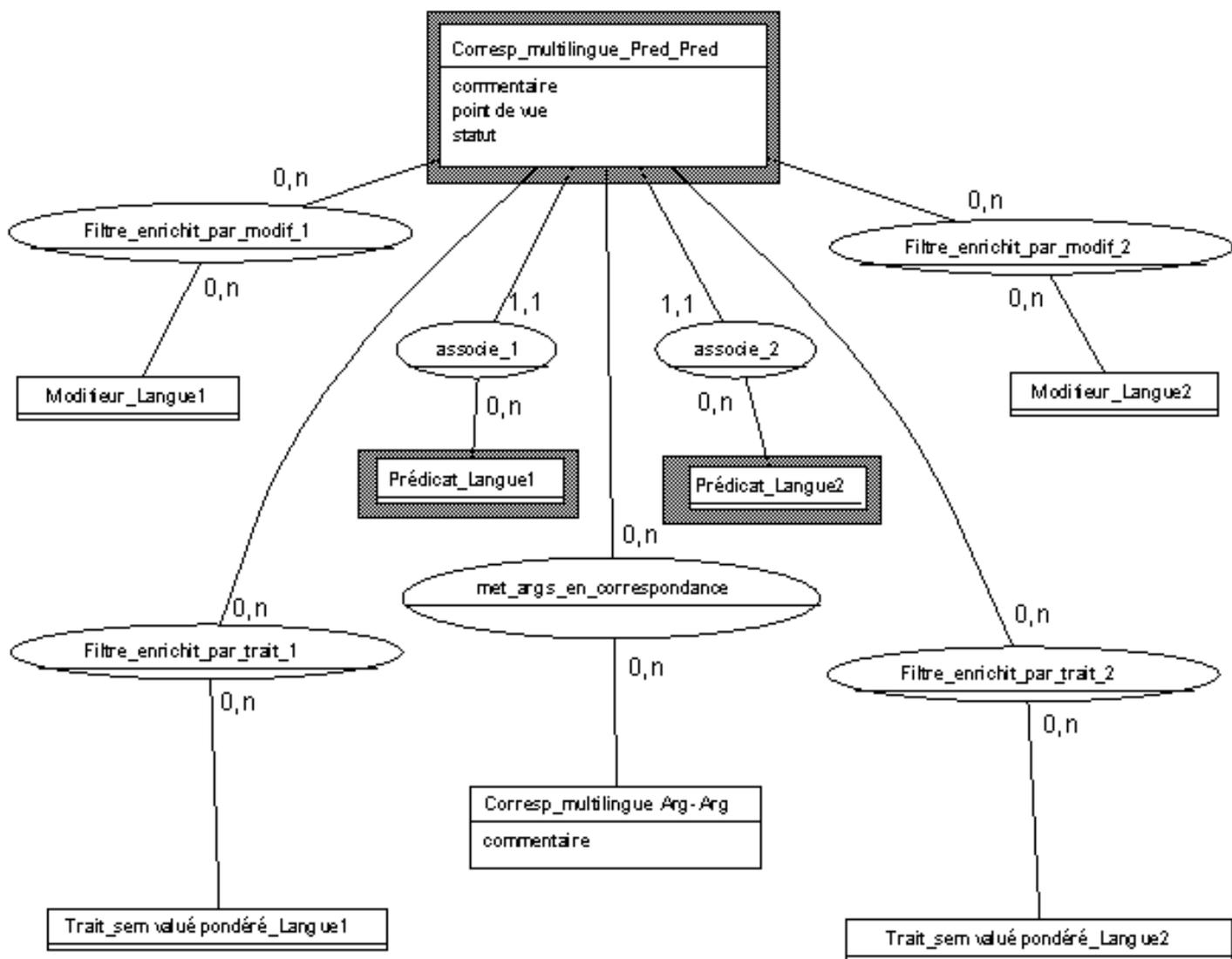
Si par exemple, on souhaite associer deux Usem et préciser simplement comment les différents contextes (associés aux différentes Usyn) s'associent, on pourra avoir une CorrespMult_Usem renseignée avec un bon niveau de précision (cette CorrespMult_Usem pourrait être autonome, par exemple) et une CorrespMult_Usyn qui se contente de pointer vers les Usyn à mettre en rapport. On peut imaginer aussi que l'on souhaite également filtrer sur les Description, auquel cas ConstraintDescription serait présent également dans la CorrespMult_Usyn, qui ne mettrait pas en correspondance les Position.

C - Schémas entités- relations

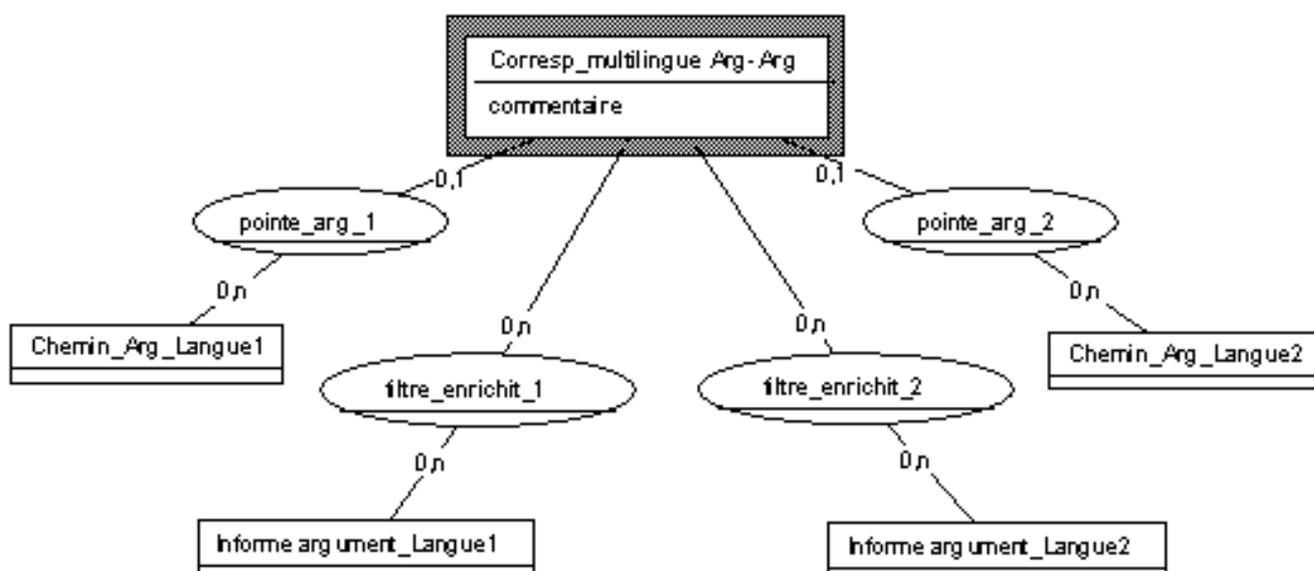
1. Correspondance d'Usem



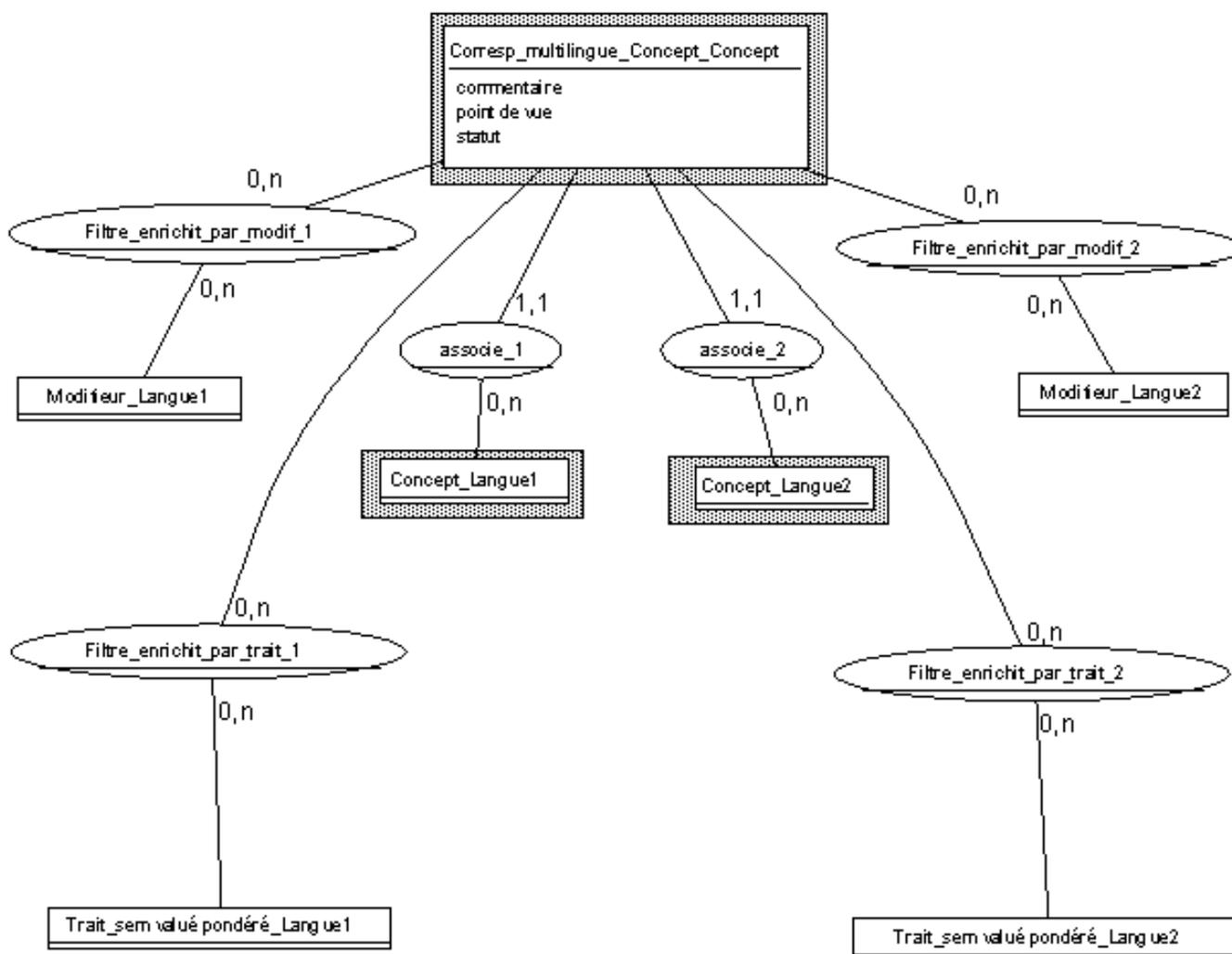
2. Correspondance de Prédicats



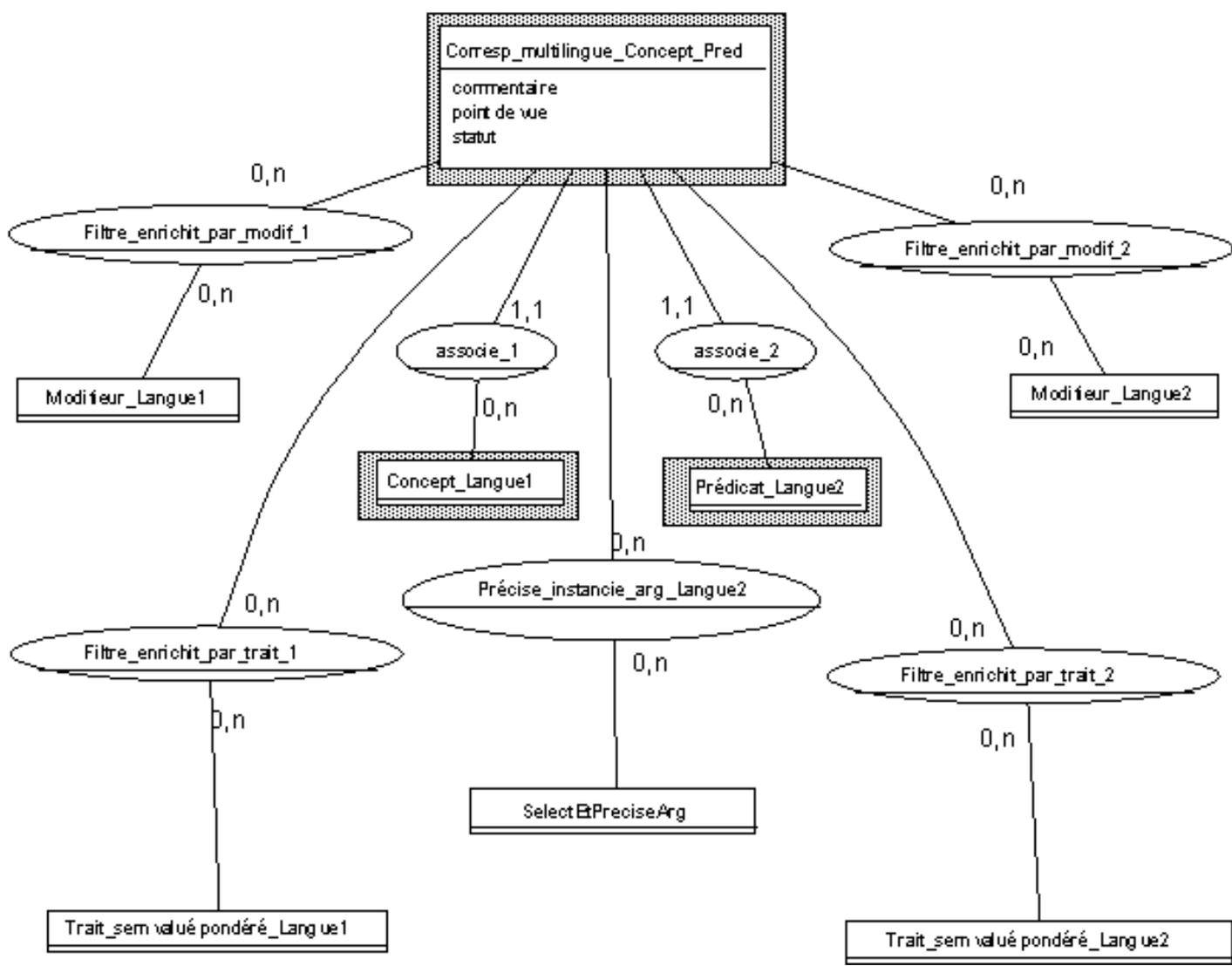
3. Correspondance d'Arguments

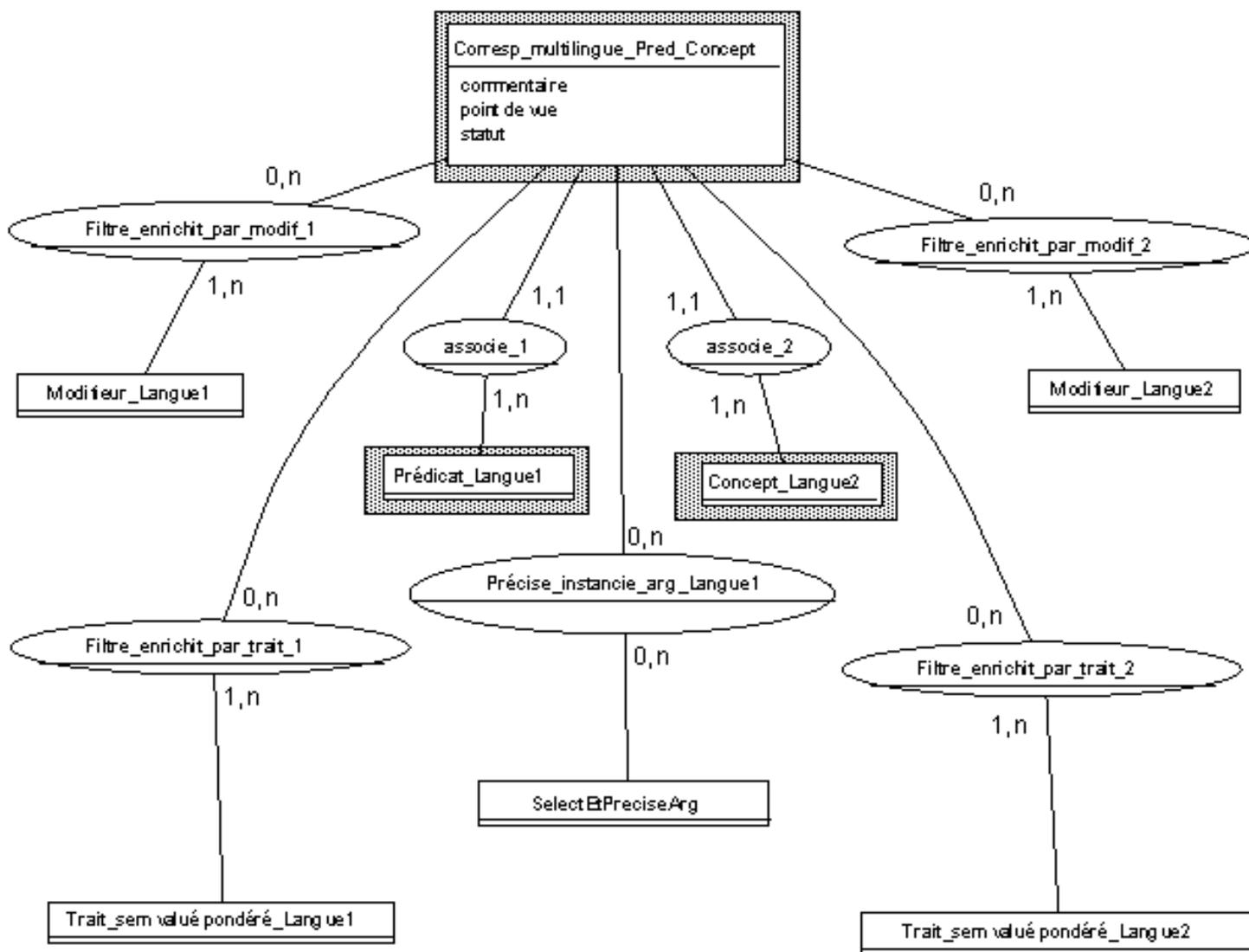


4. Correspondance de Concepts

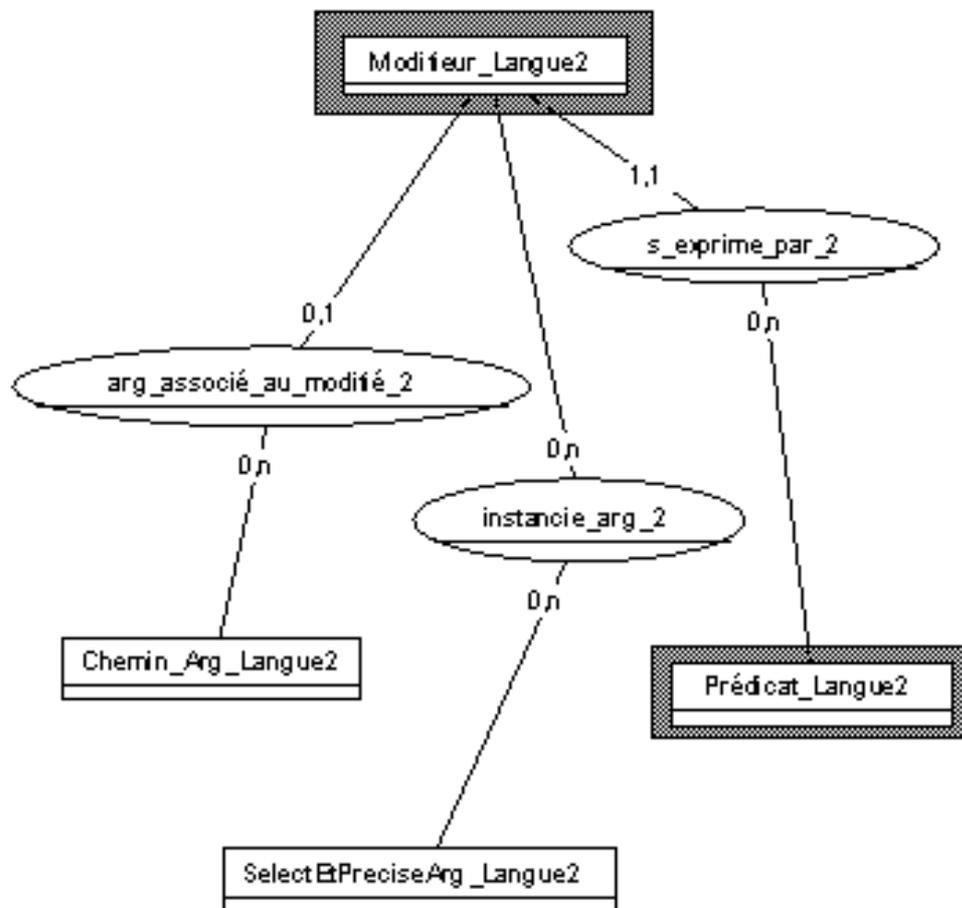
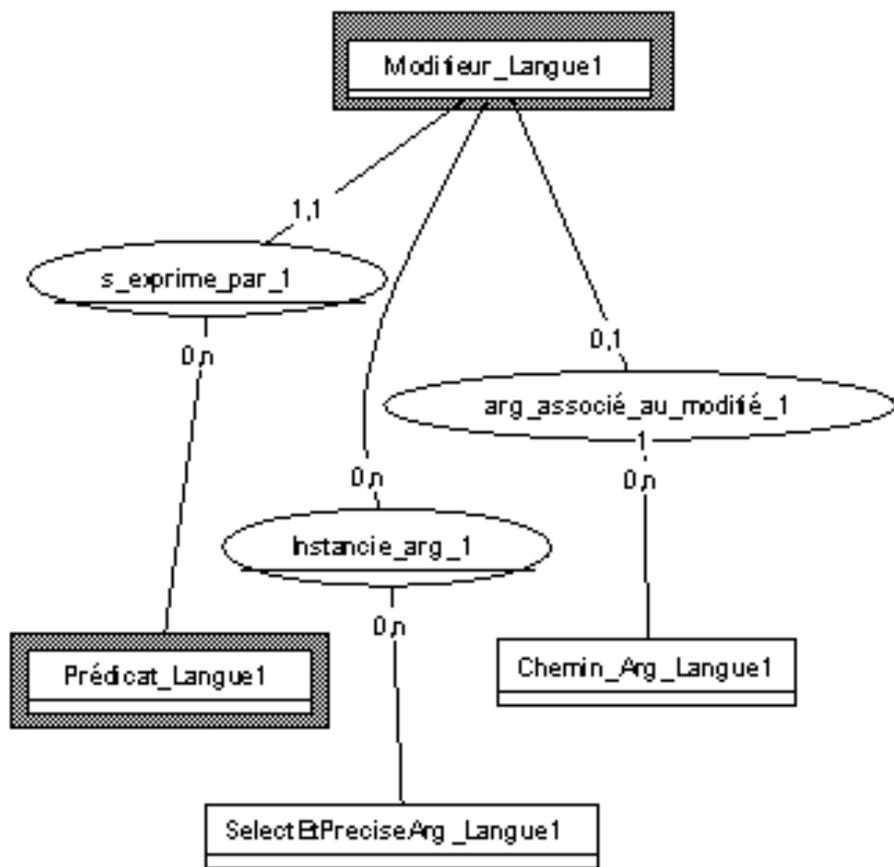


5. Correspondances associant un Concept à un Prédicat

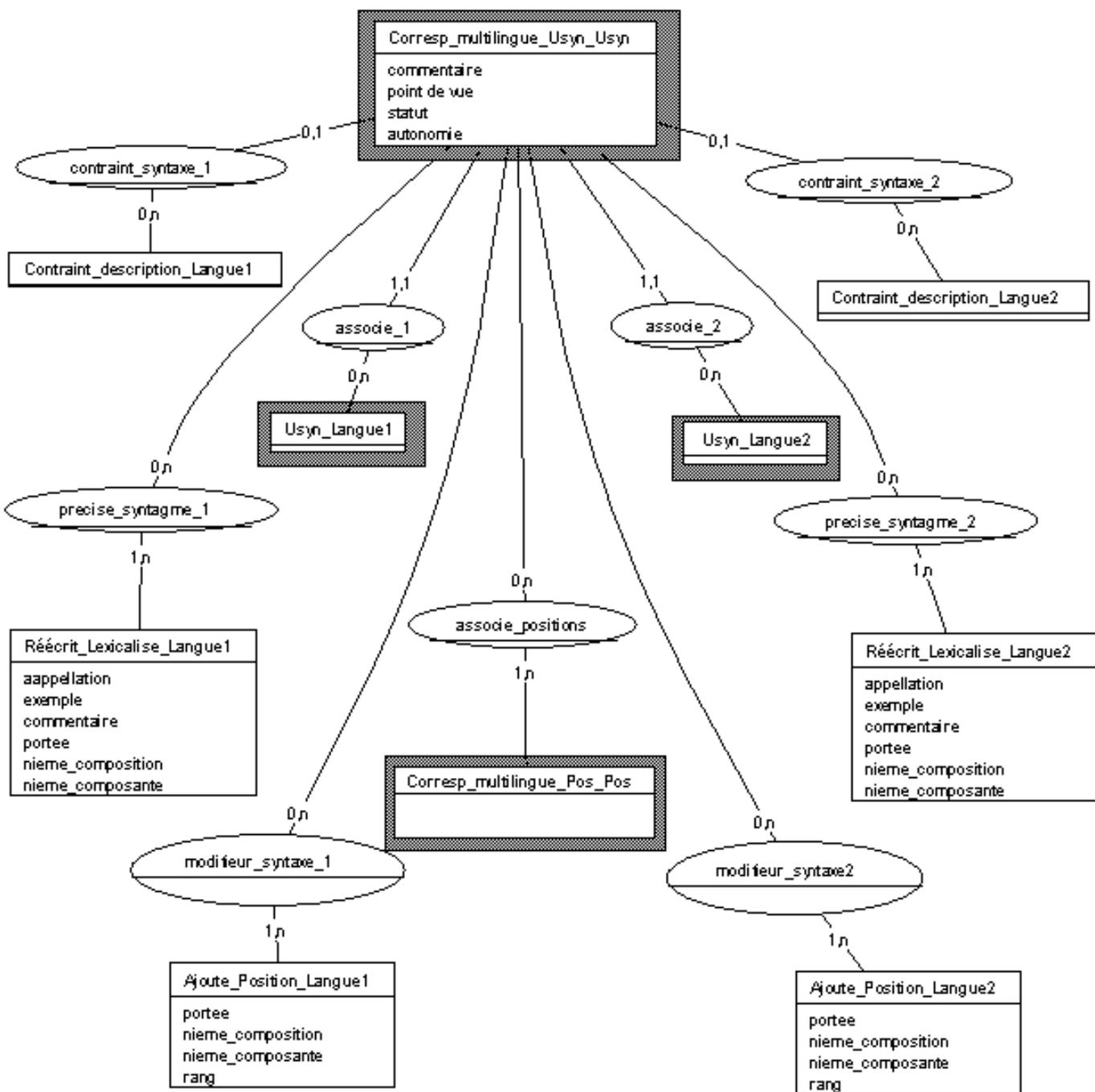




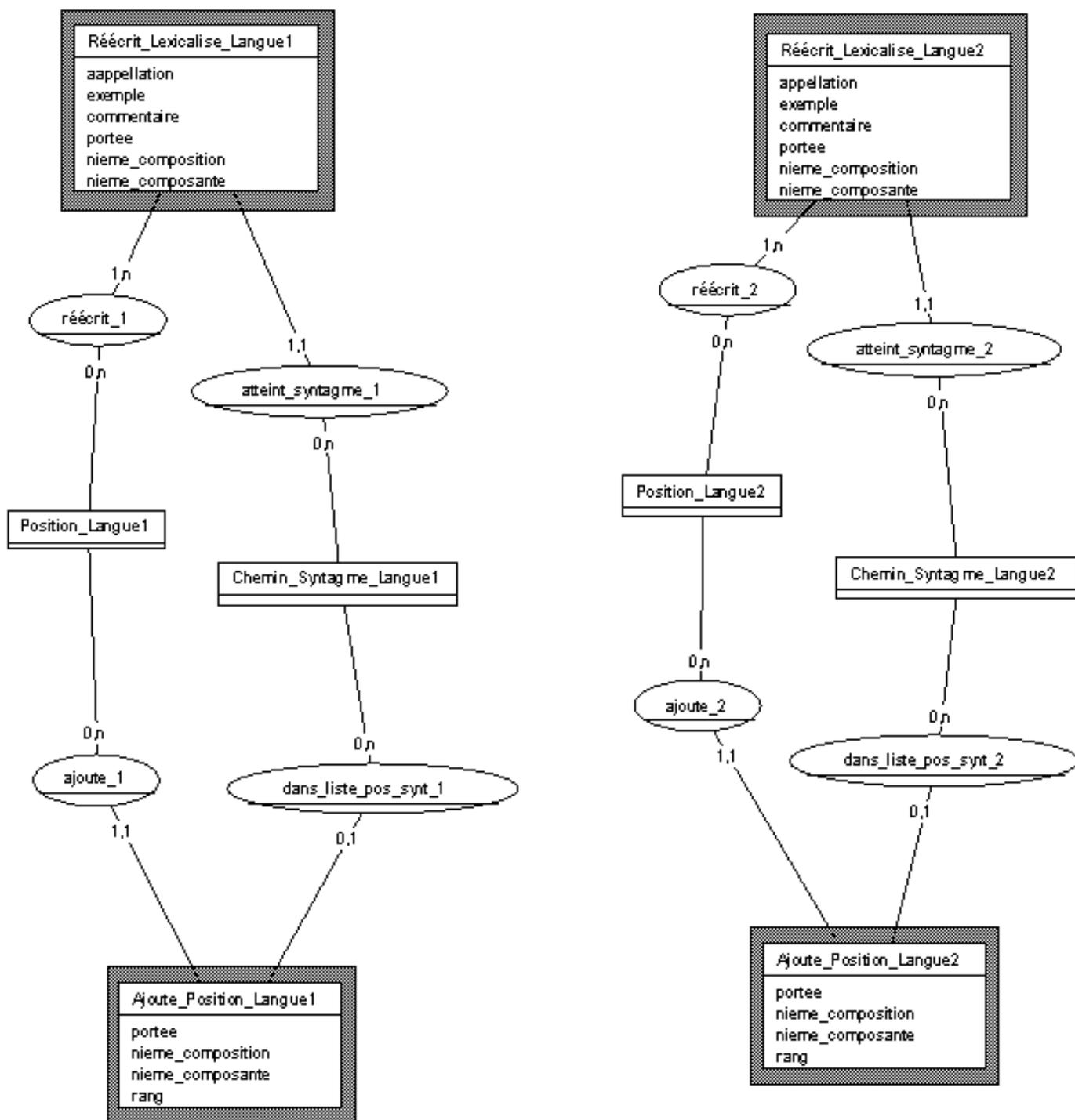
6. Modifieur



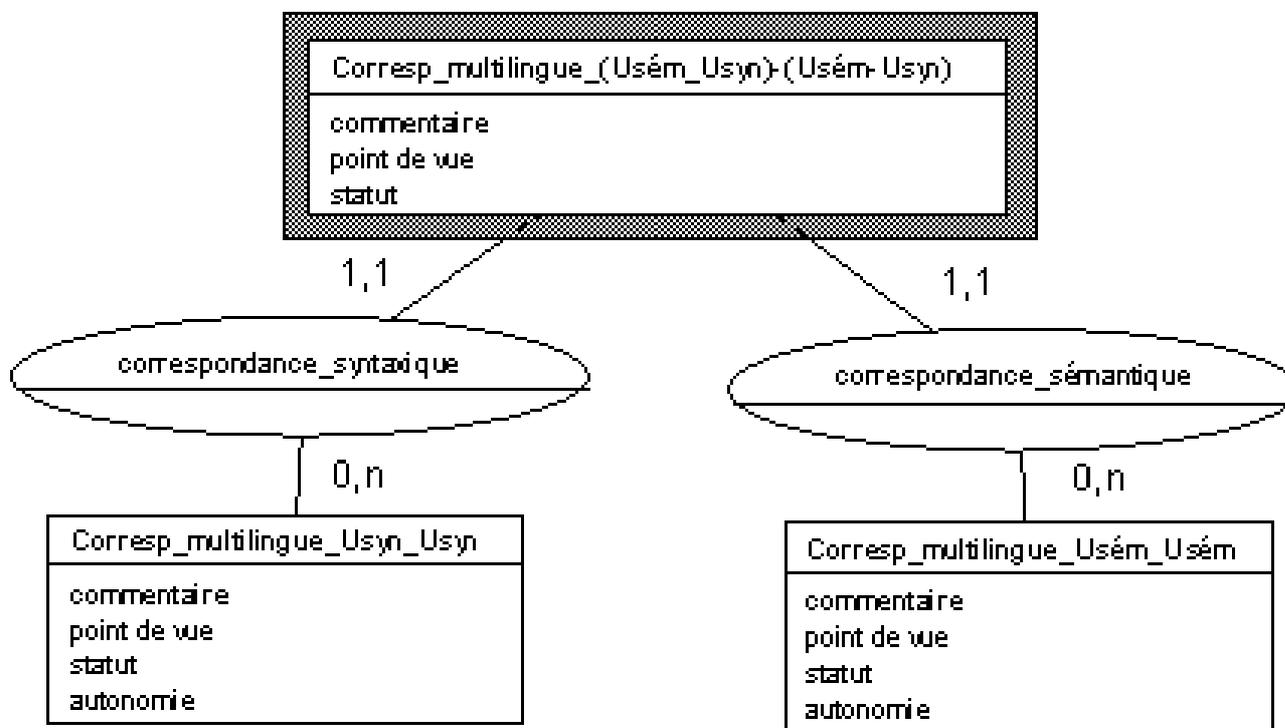
7. Correspondance d'Usyn



8. Réécrit_Lexicalise, Ajoute_Position



9. Correspondance de couples Usem-Usyn



D- DTD SGML

I - Introduction - Traduction du Modèle Conceptuel

Le modèle conceptuel GENELEX a été largement exprimé au travers de modèles Entité-Attribut-Relation (Merise).

Baucoup de contraintes d'intégrité sont exprimées dans ce formalisme : typage des objets, typage des relations, cardinalité des relations, etc. Cependant, ce modèle n'étant pas fait pour exprimer des règles - à moins de complications extrêmes-, certaines contraintes ont dû être exprimées dans le document d'accompagnement (restriction sur les combinaisons de valeurs). Il s'ensuit que le modèle Conceptuel de GENELEX combine l'utilisation du formalisme Entité-Attribut-Relation (EAR) et de commentaires en langage naturel.

Une DTD (Définition de Type de Document) SGML est un modèle physique du type grammair qui décrit le marquage des données.

Lors du passage du Modèle Conceptuel à la DTD GENELEX, nous nous sommes efforcés de traduire de manière systématique les modèles EAR et avons tenté d'exprimer formellement la plupart des contraintes d'intégrité décrites en langage naturel.

Certaines règles de traduction du formalisme EAR vers SGML ont été mises en œuvre :

- (i) Les Entités EAR deviennent des Eléments SGML.
- (ii) Les Attributs d'Entités EAR deviennent des Attributs d'Eléments.

Lorsque les valeurs d'un Attribut sont exclusives les unes des autres et qu'elles constituent un vocabulaire fermé, ces valeurs sont représentées sous forme d'Attributs SGML listés.

(iii) Les Relations non attribuées pointant sur une Entité EAR non partagée sont traduites par des liens de hiérarchie entre les Eléments de la DTD. Leur cardinalité est traduite par les indicateurs d'occurrences SGML : ? + *

(iv) Les Relations non Attribuées pointant sur une Entité EAR partagée sont exprimées par des liens de référence entre les Eléments.

(v) Les Relations Attribuées sont traduites par des Eléments SGML attribués mis en relation - hiérarchie ou référence - avec les Eléments traduisant les Entités EAR.

Un fichier de contraintes ("multiling.ctr") a été créé pour faciliter la lecture des références croisées. Ces contraintes apparaissent comme des commentaires et seront donc ignorées par un parser SGML ; elles expriment dans une syntaxe intuitive le typage des références.

D'autre part, la couche multilingue s'appuyant massivement sur les couches syntaxique et sémantique, on ne s'étonnera pas que la DTD qui suit s'appuie sur des "ELEMENT" des DTD de ces couches ; la compréhension fine de la DTD qui suit demande donc une certaine connaissance des DTD de la syntaxe et de la sémantique.

On remarquera en particulier, que de nombreux attributs pointent par des IDREF vers des ELEMENT décrits dans semant.dtd ou syntaxe.dtd, ainsi que certains ELEMENT enchâssés .

II - DTD GENELEX commentée

1. DTD *lesgenelex.dtd*

```
<!--Consortium GENELEX @(#) lesgenelex.dtd 1.1 @(#) 94/12/28 16:28:03-->
```

```
<!DOCTYPE LesGenelex [
```

```
<!ENTITY % ISOlat1 PUBLIC "ISO 8879-1986//ENTITIES Added Latin 1//EN"> %ISOlat1
```

```
<!ENTITY % CustEnt PUBLIC "-//GLX-TEAM//ENTITIES Custom Entity Set//FR"> <!ENTITY % MorpEnt PUBLIC "-//GLX-TEAM//ENTITIES Morphologie Entity Set//FR"> <!ENTITY % SyntEnt PUBLIC "-//GLX-TEAM//ENTITIES Syntaxe Entity Set//FR"> <!ENTITY % SemEnt PUBLIC "-//GLX-TEAM//ENTITIES Semantique Entity Set//FR">
```

```
%CustEnt
```

```
%MorpEnt
```

```
%SyntEnt
```

```
%SemEnt
```

```
<!ELEMENT LesGenelex - O (Genelex+ & GenelexMultilingue*)>
```

```
<!-- L'ensemble des dictionnaires Genelex est constitue de dictionnaires monolingues Genelex et de ponts multilingues entre ces dictionnaires. -->
```

<!ENTITY % GLXmono PUBLIC "-//GLX-TEAM//DTD Genelex Monolingue//FR"> %GLXmono

<!ENTITY % GLXmulti PUBLIC "-//GLX-TEAM//DTD Genelex Multilingue//FR"> <!ENTITY % MultiCtr PUBLIC "-//GLX-TEAM//DTD Contraintes Multilingue//FR"> %GLXmulti

%MultiCtr

2. DTD *genemono.dtd*

<!--Consortium GENELEX @(#) genemono.dtd 1.2@(#) 94/12/28 16:31:03-->

<!ELEMENT Genelex - O (GenelexMorpho? & GenelexSyntaxe? & GenelexSemant? & CombVE*)>

<!ATTLIST Genelex

id ID #REQUIRED

nom CDATA #REQUIRED

langue CDATA #REQUIRED

version CDATA #IMPLIED

date_creation1 CDATA #IMPLIED

date_creationglx CDATA #IMPLIED

date_modif CDATA #IMPLIED

propriete CDATA #IMPLIED

copyright CDATA #IMPLIED

integrite (SANS_B|%pBooleen) SANS_B

certification CDATA #IMPLIED>

<!-- ***** -->

<!ENTITY % pGlose

"appellation CDATA #IMPLIED

exemple CDATA #IMPLIED

commentaire CDATA #IMPLIED">

<!-- D'une maniere generale dans tout le fichier :

- appellation : permet de nommer de facon comprehensible

et si possible univoque l'objet

- exemple : permet d'illustrer l'emploi (citation d'auteur,

exemple de dictionnaire ou de linguiste)

- commentaire : champ libre pour l'utilisateur -->

<!-- ***** -->

<!ELEMENT CombVE - O EMPTY>

<!ATTLIST CombVE

id ID #REQUIRED

datation (SANS_D|%pDatation) SANS_D

niveaulgue (SANS_NL|%pNiveauLgue) SANS_NL

frequence (SANS_F|%pFrequence) SANS_F

vargeog CDATA #IMPLIED>

<!ENTITY % GLXmor PUBLIC "-//GLX-TEAM//DTD Description Morphologie//FR"> <!ENTITY % MorpCtr PUBLIC "-//GLX-TEAM//DTD Contraintes Morphologie//FR"> %GLXmor

%MorpCtr

<!ENTITY % GLXsyn PUBLIC "-//GLX-TEAM//DTD Description Syntaxe//FR"> <!ENTITY % SyntCtr PUBLIC "-//GLX-TEAM//DTD Contraintes Syntaxe//FR"> %GLXsyn

%SyntCtr

<!ENTITY % GLXsem PUBLIC "-//GLX-TEAM//DTD Description Semantique//FR"> <!ENTITY % SemaCtr PUBLIC "-//GLX-TEAM//DTD Contraintes Semantique//FR"> %GLXsem

%SemaCtr

3. DTD *genemulti.dtd*

<!--Consortium GENELEX @(#) genemulti.dtd 1.2@(#) 94/12/28 16:31:03-->

<!ELEMENT GenelexMultilingue - O (

CorrespMult_Usyn_Usen*

& CorrespMult_Pred*

& CorrespMult_Concept*

& CorrespMult_Pred_Concept*

& CorrespMult_Concept_Pred*

& CorrespMult_User*

& CorrespMult_Usyn*

& Modifieur*

& AjoutePosition*

& Reecrit_Lexicalise*

)>

<!ATTLIST GenelexMultilingue

commentaire CDATA #IMPLIED

genelex1 IDREF #REQUIRED

genelex2 IDREF #REQUIRED>

<!-- Un GenelexMultilingue est une mise en correspondance de deux dictionnaires Genelex monolingues. Les attributs genelex1 et genelex2 pointent vers ces deux dictionnaires monolingues. -->

<!ENTITY % pAutonomie_cor_m

"C_AUTONOME|C_NON_AUTONOME">

<!-- Les valeurs de l'attribut autonomie permettent de preciser si la correspondance est autonome, c'est a dire valable et pertinente prise isolement (valeur C_AUTONOME), ou si elle n'est à prendre en considération qu'en tant qu'element d'un couple de correspondance syntaxe-semantic, auquel cas l'attribut marque sa non-autonomie (valeur C_NON_AUTONOME). -->

<!ENTITY % pConstCorresp_Mult

"commentaire CDATA #IMPLIED

point_de_vue CDATA #IMPLIED

statut_corr CDATA #IMPLIED">

<!-- Differentes attributs communs aux differentes mises en correspondance multilingues.

- statut_corr : statut de la mise en correspondance - par exemple : equivalence, 1 plus generique que 2, 1 moins generique que 2, recouvrement partiel.

- point_de_vue : pour marquer que la mise en correspondance se fait suivant un certain point de vue, qui peut être par exemple celui d'un utilisateur;

différents utilisateurs peuvent ainsi avoir des préférences de traduction

différentes. -->

```
<!ENTITY % pConstFiltre_Ajoute
```

```
"ajoute_trait_sem_langue1 IDREFS #IMPLIED
```

```
ajoute_trait_sem_langue2 IDREFS #IMPLIED
```

```
modifieur_langue1_I IDREFS #IMPLIED
```

```
modifieur_langue2_I IDREFS #IMPLIED">
```

<!-- Différents attributs concernant le filtrage/ajout d'information en langue1 ou en langue2, permettant de combler l'écart sémantique qui sépare les éléments des deux langues intervenant dans la mise en correspondance multilingue.

- ajoute_trait_sem_langue1 complète l'information par l'ajout de Traits sémantiques values. Ces traits jouent un rôle de filtrage si langue1 est la langue source, et d'enrichissement si langue1 est la langue cible d'un processus de traduction.

- ajoute_trait_sem_langue2 joue un rôle analogue concernant la langue2

- modifieur_langue1 permet de préciser l'élément sémantique intervenant dans

la mise en correspondance en langue1 par des éléments modifieurs (Modifieur). Comme les traits sémantiques values, ces modifieurs expriment un filtrage

lorsque la langue qui les porte est langue source d'un processus de traduction s'appuyant sur ces mises en correspondance, et enrichissement de la représentation sémantique quand elle est langue cible de ce processus. -->

```
<!ELEMENT CorrespMult_Usyn_Usem - O EMPTY>
```

```
<!ATTLIST CorrespMult_Usyn_Usem
```

```
%pConstCorresp_Mult
```

```
corresp_mult_usyn IDREF #REQUIRED
```

```
corresp_mult_usem IDREF #REQUIRED>
```

<!-- Mise en correspondance multilingue qui permet de s'appuyer à la fois sur la couche sémantique et la couche syntaxique de chacune des langues mises en correspondance. et permet d'associer des couples Usem-Usyn de chaque langue en correspondance.

Ces couples comportent les Usem pointées par le CorrespMult_Usem référence par corresp_mult_usem, et les Usyn pointées par le CorrespMult_Usyn référence par corresp_mult_usyn.

Ces couples doivent bien sûr être compatibles, c'est à dire que dans chacune des langues, l'Usyn est associée à l'Usem par la correspondance syntaxe-sémantique monolingue.

Aussi bien les Usyn que les Usem peuvent intervenir dans plus d'une CorrespMult_Usyn_Usem. -->

<!ELEMENT CorrespMult_Pred - O (CorrespMult_Arg_Arg*)>

<!ATTLIST CorrespMult_Pred

id ID #REQUIRED

%pConstCorresp_Mult

%pConstFiltre_Ajoute

pred_langue1 IDREF #REQUIRED

pred_langue2 IDREF #REQUIRED>

<!-- Correspondance multilingue de Predicat permet d'associer des Predicat de chacune des langues independamment de la ou des Usem qui leur sont associees directement (s'il s'agit d'un Predicat LEXICAL)ou non, et a plus forte raison independamment des contextes syntaxiques qui leur sont associes par differents cheminements dans la couche monolingue. La correspondance de Predicat permet donc de se faire correspondre a un certain niveau d'abstraction des structures argumentaires liees par une relation que le predicat represente independamment d'ecarts qui peuvent etre importants en ce qui concerne l'expression de surface de ces Predicat.

On precise :

- CorrespMult_Arg_Arg : correspondances entre les arguments des predicats de chacune des langues

- pred_langue1, pred_langue2 pointent vers les Predicats en correspondance multilingue

- ajoute_trait_sem_langue1, ajoute_trait_sem_langue2 permet de "surcharger" les Predicat d'un point de vue semantique par des traits semantiques values (de

la couche monolingue) en pointant vers des Trait_Sem_ValPond

- modifieur_langue1, modifieur_langue2 pointe vers des elements Modifieur qui precisent par modification le sens des elements en relation. Cela permet de completer au niveau semantique l'ecart qui separe le sens des deux entrees en correspondance. -->

<!ELEMENT CorrespMult_Arg_Arg - O EMPTY>

<!ATTLIST CorrespMult_Arg_Arg

commentaire CDATA #IMPLIED

chemin_arg_langue1 NUMBERS #IMPLIED

chemin_arg_langue2 NUMBERS #IMPLIED

informe_arg_langue1_I IDREFS #IMPLIED

informe_arg_langue2_I IDREFS #IMPLIED>

<!-- La correspondance d'arguments se fait suivant le meme principe que dans

les Relations valuees entre Predicat ou Usem en ce sens qu'elle associe deux

chemins vers des arguments :

- chemin_arg_langue1, chemin_arg_langue2 (suite de nombres qui permettent de atteindre un nieme argument, et si celui-ci est predicatif, d'atteindre recursivement l'un de ses arguments...)

- informe_arg_langue1_l , informe_arg_langue2_l permettent de preciser par un ajout d'informations semantiques ces arguments.

Les chemins vers les arguments sont optionnels, ce qui permet de decrire qu'un argument n'est en correspondance avec aucun dans l'autre langue, et de le

preciser semantiquement.

Il est donc possible de gerer les nombres differents d'arguments et

d'"instancier" certains arguments.

Au moins l'un des chemin_arg_languex est present, et aucun

informe_arg_languex_l ne peut etre present si le chemin_arg_languex de sa langue n'est pas present. -->

<!ELEMENT CorrespMult_Concept - O EMPTY>

<!ATTLIST CorrespMult_Concept

id ID #REQUIRED

%pConstCorresp_Mult

%pConstFiltre_Ajoute

concept_langue1 IDREF #REQUIRED

concept_langue2 IDREF #REQUIRED>

<!-- Correspondance multilingue de Concept permet d'associer des Concept de chacune des langues independamment de la ou des Usem qui leur sont associees directement ou non, et a plus forte raison independamment des contextes syntaxiques qui leur sont associes par differents cheminements dans la couche monolingue. La correspondance de Concept permet donc de se faire correspondre a un certain niveau d'abstraction des objets descriptifs sans tenir compte des ecarts qui peuvent etre importants en ce qui concerne les contextes de surface associes.

On precise :

- concept_langue1, concept_langue2 pointent vers les Concept en correspondance multilingue

- ajoute_trait_sem_langue1, ajoute_trait_sem_langue2 permet de "surcharger" les Concept d'un point de vue semantique par des traits semantiques values (de la couche monolingue) en pointant vers des Trait_Sem_ValPond

- modifieur_langue1, modifieur_langue2 pointe vers des elements Modifieur qui precisent par modification le sens des Concept en correspondance Cela permet de completer au niveau semantique l'ecart qui separe le sens des deux entrees en correspondance. -->

<!ELEMENT CorrespMult_Pred_Concept - O (SelectEtPreciseArg*)>

<!ATTLIST CorrespMult_Pred_Concept

id ID #REQUIRED

%pConstCorresp_Mult

%pConstFiltre_Ajoute

pred_langue1 IDREF #REQUIRED

concept_langue2 IDREF #REQUIRED>

<!-- Correspondance associant un Predicat (objet descriptif de la langue1) a un Concept (objet descriptif de la langue2), les divergences de descriptions sont ici importantes.

Le Predicat sera le plus souvent precise ou instancie par les InformeArg pointes par informe_arg_langue1_I En effet cette correspondance sera le plus souvent entre un Predicat ayant "instancie" ses arguments (tous ou non) et un Concept, associe ou non a des modifieurs.

- pred_langue1 pointe vers le Predicat de la langue 1

- informe_arg_langue1_I pointe vers les informations precisant les arguments du Predicat

- concept_langue2 pointe vers le Concept de la langue 2

- ajoute_trait_sem_langue1, ajoute_trait_sem_langue2 permet de "surcharger" le Predicat ou le Concept d'un point de vue semantique par des traits semantiques values (de la couche monolingue) en pointant vers des Trait_Sem_ValPond

- modifieur_langue1, modifieur_langue2 pointe vers des elements Modifieur qui precisent par modification le sens des elements en correspondance . Cela permet

de completer au niveau semantique l'ecart qui separe le sens des deux entrees en correspondance. -->

<!ELEMENT CorrespMult_Concept_Pred - O (SelectEtPreciseArg*)>

<!ATTLIST CorrespMult_Concept_Pred

id ID #REQUIRED

%pConstCorresp_Mult

%pConstFiltre_Ajoute

pred_langue2 IDREF #REQUIRED

concept_langue1 IDREF #REQUIRED>

<!-- Correspondance associant un Concept(objet descriptif de la langue1) a un Predicat (objet descriptif de la langue2), les divergences de descriptions sont ici importantes.

Le Predicat sera le plus souvent precise ou instancie par les InformeArg pointes par informe_arg_langue2_I En effet cette correspondance sera le plus souvent entre un Predicat ayant "instancie" ses arguments (tous ou non) et un Concept, associe ou non a des modifieurs.

- pred_langue2 pointe vers le Predicat de la langue 1
- informe_arg_langue2_I pointe vers les informations precisant les arguments du Predicat
- concept_langue1 pointe vers le Concept de la langue 2
- ajoute_trait_sem_langue1, ajoute_trait_sem_langue2 permet de "surcharger" le Predicat ou le Concept d'un point de vue semantique par des traits semantiques values (de la couche monolingue) en pointant vers des Trait_Sem_ValPond
- modifieur_langue1, modifieur_langue2 pointe vers des elements Modifieur qui precisent par modification le sens des elements en correspondance . Cela permet

de completer au niveau semantique l'ecart qui separe le sens des deux entrees en correspondance. -->

<!ELEMENT CorrespMult_Usem - O (CorrespMult_Arg_Arg*)>

<!ATTLIST CorrespMult_Usem

id ID #REQUIRED

%pConstCorresp_Mult

%pConstFiltre_Ajoute

usem_langue1 IDREF #REQUIRED

usem_langue2 IDREF #REQUIRED

autonomie (%pAutonomie_cor_m) C_AUTONOME>

<!-- Correspondance multilingue de Usem permet d'associer des Usem de chacune des langues independamment des contextes syntaxiques qui leur sont associes (Usyn associees dans la correspondance syntaxe-semantique). La correspondance d'Usem fait se correspondre des structures argumentaires quand les deux Usem sont liees a des Predicat, elles peuvent aussi etre l'une predicative et l'autre non.

CorrespMult_Usem fait implicitement se correspondre toutes les Usyn compatibles avec l'Usem langue1 avec toutes les Usyn compatibles avec l'Usem langue2. Si l'on souhaite filtrer sur la syntaxe, on utilisera des CorrespMult_Usyn_Usem qui s'appuient sur des CorrespMult_Usem .

On precise :

- CorrespMult_Arg_Arg : correspondances entre les arguments des Predicat associes aux Usem de chacune des langues. Lorsqu'il n'y a pas le meme nombre d'arguments (ou lorsque l'une des Usem est non-predicative), on peut preciser d'un point de vue semantique les arguments associes a l'une des Usem via son Predicat
- usem_langue1, usem_langue2 pointent vers les Usem en correspondance multilingue
- ajoute_trait_sem_langue1, ajoute_trait_sem_langue2 permet de "surcharger" les Usem d'un point de vue semantique par des traits semantiques values (de la couche monolingue) en pointant vers des Trait_Sem_ValPond
- modifieur_langue1, modifieur_langue2 pointe vers des elements Modifieur qui precisent par modification le sens des elements en correspondance. Cela permet de completer au niveau semantique l'ecart qui separe le sens des deux entrees en correspondance. -->

<!ELEMENT Modifieur - O (SelectEtPreciseArg*)>

<!ATTLIST Modifieur

id ID #REQUIRED

commentaire CDATA #IMPLIED

pred IDREF #REQUIRED

chemin_arg_modifie NUMBERS #REQUIRED>

<!-- Un Modifieur vient completer un element (Useur, Predicat ou Concept) en le modifiant.

La precision de sens se fait en s'appuyant sur un Predicat qui met en relation l'element modifie et le ou les autres arguments modifiant. Le Predicat permet de preciser la relation qui lie le modifie aux autres arguments qui ont par ailleurs un ou des roles semantiques. (A noter que les prepositions, par exemple, seront au niveau semantique associees a des Predicat a au moins deux arguments : les deux elements -minimalement, mais il peut y en avoir plusqu'elles mettent en relation)

- pred pointe vers le Predicat modifieur

- informe_arg_l pointe vers les Informe_arg precisant les arguments de ce predicat

- chemin_arg_modifie precise le chemin vers l'argument du Predicat qui est modifie et qui est donc associe a l'element que le Modifieur vient preciser, element principal de la correspondance qui peut etre Concept, Predicat, ou Useur suivant la correspondance multilingue dans laquelle le modifieur intervient. -->

<!ELEMENT CorrespMult_Usyn - O (CorrespMult_Pos_Pos*)>

<!ATTLIST CorrespMult_Usyn

id ID #REQUIRED

%pConstCorresp_Mult

usyn_langue1 IDREF #REQUIRED

usyn_langue2 IDREF #REQUIRED

autonomie (%pAutonomie_cor_m) C_AUTONOME

contr_descr_usyn1 IDREF #IMPLIED

contr_descr_usyn2 IDREF #IMPLIED

reecrit_lexicalise_synt_1 IDREFS #IMPLIED

reecrit_lexicalise_synt_2 IDREFS #IMPLIED

ajoute_position_usyn1 IDREFS #IMPLIED

ajoute_position_usyn2 IDREFS #IMPLIED>

<!-- Correspondance multilingue d'Usyn qui permet de se faire correspondre des Usyn independamment du ou des

sens associés. Cette correspondance d'Usyn est utile lorsqu'on ne souhaite pas associer une représentation sémantique à certaines Usyn que cependant l'on souhaite pouvoir mettre en correspondance multilingue (c'est par exemple le cas très probablement de la phraseologie, qui pourra être décrite au moyen d'Usyn composées auxquelles on n'associera pas nécessairement d'Usem). C'est intéressant également comme raccourci quand les langues distinguent différents sens pour les Usyn, mais que les Usyn se correspondent indépendamment de ces différents sens. Un module de traduction pourra utiliser ces correspondances sans avoir à choisir les Usem intervenant. Une approche syntaxiste de la traduction pourra s'appuyer sur ces correspondances.

On précise :

- CorrespMult_Pos_Pos : correspondances entre les positions des Descriptions associées aux Usyn de chacune des langues

- usyn_langue1, usyn_langue2 pointent vers les Usem en correspondance

multilingue

- contr_descr_usyn1, contr_descr_usyn2 pointent vers des ConstraintDescription afin de restreindre les réalisations des contextes syntaxiques associées aux descriptions (même mécanisme que lors de la correspondance syntaxe sémantique)

: contraint ou interdit la réalisation d'une position optionnelle, force une position à se réaliser par un certain de ses syntagmes occupant de position, rajoute des contraintes sous forme de traits à certain syntagmes... (voir le modèle et la DTD sémantique pour plus de détails)

- reecrit_lexicalise_synt_1, reecrit_lexicalise_synt_2 permet de contraindre structurellement par une réécriture avec lexicalisations ou non d'un syntagme non-terminal qui est présent dans la description sans être contraint structurellement. Les contraintes structurelles (réécriture sous forme de

listes de Position dont l'ordre est pertinent) devront être compatibles avec

les traits portés par le syntagme

- reecrit_lexicalise_synt_1 et reecrit_lexicalise_synt_2 pointent vers la liste des Reecrit_Lexicalise précisant les Description des Usyn de la langue1 ou de

la langue2

- ajoute_position_usyn1, ajoute_position_usyn2 permet de rajouter des positions "flottantes" aux éléments de la description qui s'expriment par une

liste de Position, liste dans laquelle s'insère la Position ajoutée

- ajoute_trait_sem_langue1, ajoute_trait_sem_langue2 permet de "surcharger" les Usyn par l'ajout de Position à fonction "modifieur", les réalisations de ces positions pouvant par ailleurs être précisées par des feuilles lexicalisées.

ajoute_position_usynx permet donc d'insérer des modifieurs entièrement, partiellement ou pas lexicalisés dans la structure syntaxique de référence qui est associée à l'Usyn.

ajoute_position_usynx et reecrit_lexicalise_synt_x jouent, au niveau

syntactique, le même rôle que les Modifieur et les Trait_Sem_ValPond au niveau sémantique : ils permettent de combler explicitement l'écart qu'il y a entre deux unités en correspondance. -->

<!ELEMENT AjoutePosition - O (CheminSyntagme?)>

<!ATTLIST AjoutePosition

id ID #REQUIRED

%pGlose

portee (%pPorteeCorresp) EXTERNE

nieme_composition NUMBER #IMPLIED

nieme_composante NUMBER #IMPLIED

position IDREF #REQUIRED

rang NUMBER #IMPLIED>

<!-- AjoutePosition pour ajouter une Position (non essentielle du point de vue monolingue, sinon on peut supposer qu'elle serait présente dans la couche monolingue) a un niveau quelconque de la Description

- portee, nieme_composition, nieme_composante ont le meme sens que dans la couche semantique : ils permettent d'identifier (en s'associant si necessaire a CheminSyntagme) ou doit s'insérer la position

- CheminSyntagme permet d'atteindre un Syntagme reecrit dans la liste de Position duquel va s'insérer la Position ajoutée.

- rang permet de préciser où cette Position s'insère. rang est optionnel car la Position n'est pas nécessairement contrainte du point de vue de sa place dans la surface.

Ex : portee = EXTERNE, pas de CheminSyntagme, rang = 2 signifie que la Position interviendra dans la liste de Position de la Construction, et qu'elle s'insère

en deuxième dans la liste. Afin d'éviter les problèmes que pose éventuellement l'insertion du SELF (insereself), le rang comptera SELF inséré comme un élément virtuel de cette liste.

Ainsi, dans l'exemple: si insereself = 1, et rang = 2, la Position insérée s'insère avant SELF, si rang = 3, la Position s'insère après SELF et avant la deuxième Position. Rang compte donc à partir de 1 (Il ne faut surtout pas confondre le rang d'insertion avec le rang des Positions dans la liste même si ce n'est pas sans rapport). -->

<!ELEMENT Reecrit_Lexicalise - O (CheminSyntagme)>

<!ATTLIST Reecrit_Lexicalise

id ID #REQUIRED

%pGlose

portee (%pPorteeCorresp) EXTERNE

nieme_composition NUMBER #IMPLIED

nieme_composante NUMBER #IMPLIED

position_I IDREFS #REQUIRED>

<!-- reecrit_lexicalise pour contraindre structurellement (par une liste de Position) un syntagme occupant de position non reecrit de la syntaxe et pour éventuellement le lexicaliser entièrement ou partiellement par des traits lex portés par les syntagmes entrant dans sa reécriture à un niveau quelconque. Dans le cas où une contrainte de structure est déjà présente mais que l'on veut ajouter une position "libre" (modifieur) à un certain niveau (le plus haut, construction ou non), on peut le faire par un Ajoute_position, la position précisera alors son ou ses occupants, le plus souvent lexicalisés

- portee, nieme_composition, nieme_composante ont le même sens que dans la

couche sémantique : ils permettent d'identifier (en s'associant si nécessaire à CheminSyntagme) le Syntagme à reécrire

- CheminSyntagme permet d'atteindre un Syntagme à reécrire. Il existe évidemment la contrainte que le Syntagme soit un Syntagme_NT dont la liste de Position est vide. Cette liste de Position (reécriture du Syntagme_NT) est donnée par

position_I qui pointe vers des Positions. -->

<!ELEMENT CorrespMult_Pos_Pos - O (CheminPos_langue1, CheminPos_langue2)>

<!-- CorrespMult_Pos_Pos fait correspondre une Position de la Description de

l'Usyn en langue 1 à une Position de la Description de l'Usyn en langue 2. -->

<!ELEMENT (CheminPos_langue1| CheminPos_langue2)- O (CheminPosition?)>

<!ATTLIST (CheminPos_langue1|CheminPos_langue2)

portee (SANS|%pPorteeCorresp) EXTERNE

nieme_composition NUMBER #IMPLIED

nieme_composante NUMBER #IMPLIED

ajoute_position IDREF #IMPLIED

reecrit_lexicalise_synt IDREF #IMPLIED>

<!-- le CheminPosition pour descendre dans la Construction, la Structure

interne, ou l'un des Composants (information donnée par portee = SANS, nieme_composition, nieme_composante) ajoute_position et reecrit_lexicalise_synt sont alors absents.

- ajoute_position pointe (s'il est présent) vers un Ajoute_position intervenant dans la CorrespMult_Usyn. Le CheminPosition se parcourt alors à partir de la reécriture de la Position ajoutée pour descendre au moins d'un niveau. (C'est le

seul cas où l'on a besoin de faire intervenir à la fois CheminPosition et Ajoute_position).

Si Ajoute_position est présent et CheminPosition absent, la Position pointée

est la Position ajoutée. C'est le seul cas où CheminPosition est absent.

Si Ajoute_position est présent, alors portee = SANS obligatoirement.

- reecrit_lexicalise_synt pointe (s'il est présent) vers un Reecrit_Lexicalise intervenant dans la CorrespMult_Usyn. Le

CheminPosition est alors a parcourir a partir de la liste de Position reecrivant le syntagme et contenue dans le Reecrit_Lexicalise. Dans ce cas, portee = SANS obligatoirement,

ajoute_position est vide, et CheminPosition est present. -->

4. CONTRAINTES Genelex Multilingue

<!-- CONTRAINTE GenelexMultilingue

genelex1 TYPE Genelex

genelex2 TYPE Genelex -->

<!-- CONTRAINTE %pConstFiltre_Ajoute

"ajoute_trait_sem_langue1 TYPE Trait_Sem_ValPond

ajoute_trait_sem_langue2 TYPE Trait_Sem_ValPond

modifieur_langue1_I TYPE Modifieur

modifieur_langue2_I TYPE Modifieur" -->

<!-- CONTRAINTE CorrespMult_Usyn_Usem

corresp_mult_usyn TYPE CorrespMult_Usyn

corresp_mult_usem TYPE CorrespMult_Usem -->

<!-- CONTRAINTE CorrespMult_Pred

%pConstFiltre_Ajoute

pred_langue1 TYPE Predicat

pred_langue2 TYPE Predicat -->

<!-- CONTRAINTE CorrespMult_Arg_Arg

informe_arg_langue1_I TYPE InformeArg

informe_arg_langue2_I TYPE InformeArg -->

<!-- CONTRAINTE CorrespMult_Concept

concept_langue1 TYPE Concept

concept_langue2 TYPE Concept -->

```
<!-- CONTRAINTE CorrespMult_Pred_Concept
pred_langue1 TYPE Predicat
concept_langue2 TYPE Concept -->
<!-- CONTRAINTE CorrespMult_Concept_Pred
pred_langue2 TYPE Predicat
concept_langue1 TYPE Concept -->
<!-- CONTRAINTE CorrespMult_User
usem_langue1 TYPE User
usem_langue2 TYPE User -->
<!-- CONTRAINTE Modifieur
pred TYPE Predicat -->
<!-- CONTRAINTE CorrespMult_Usyn
usyn_langue1 TYPE Usyn
usyn_langue2 TYPE Usyn
contr_descr_usyn1 TYPE ConstraintDescription
contr_descr_usyn2 TYPE ConstraintDescription
reecrit_lexicalise_synt_1 TYPE Reecrit_Lexicalise
reecrit_lexicalise_synt_2 TYPE Reecrit_Lexicalise
ajoute_position_usyn1 TYPE AjoutePosition
ajoute_position_usyn2 TYPE AjoutePosition -->
<!-- CONTRAINTE AjoutePosition
position TYPE (Position_C|Position_S) -->
<!-- CONTRAINTE Reecrit_Lexicalise
position_l TYPE (Position_C|Position_S) -->
<!-- CONTRAINTE (CheminPos_langue1|CheminPos_langue2)
ajoute_position TYPE (Position_C|Position_S)
reecrit_lexicalise_synt TYPE Reecrit_Lexicalise -->
```