

**Héla FEHRI<sup>1</sup>, Nouredine LOUKIL<sup>2</sup>, Kais HADDAR<sup>1</sup>, Laurent ROMARY<sup>3</sup> et Abdelmajid BEN HAMADOU<sup>2</sup>**

<sup>1</sup> Faculté des sciences de Sfax, BP 802, Sfax 3018, Tunisie

{hela.fehri,Kais.haddar}@fss.rnu.tn

<sup>2</sup> Institut Supérieur d'Informatique et Multimédia de Sfax, BP 1030, Sfax 3002, Tunisie

{nouredine.loukil, Abdelmajid.benhamadou}@isimsf.rnu.tn

<sup>3</sup> Laboratoire Lorrain d'Informatique et ses Applications, Nancy, France

Laurent.romary@loria.fr

### *Un système de projection du HPSG arabisé vers la plate-forme LMF*

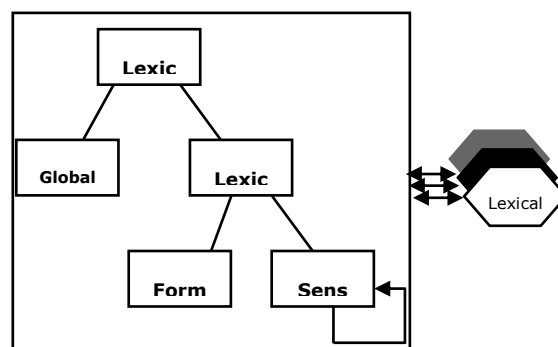
#### • Résumé étendu

Plusieurs chercheurs [1] [6] [7] dans le domaine de l'informatique linguistique ont proposé des grammaires HPSG dans le cadre de différentes applications telles que l'analyse syntaxique d'une langue naturelle et la construction d'un dictionnaire électronique. L'évaluation et la récupération des résultats fournis par ces applications ne sont pas des tâches faciles car celles-ci n'utilisent pas généralement des ressources lexicales normalisées telles que le lexique. En outre, les ressources utilisées peuvent contenir des catégories de données différentes. De ce fait, des langages pivot sont nécessaires afin de faciliter la création de telles ressources utilisant des mécanismes de fusion ou d'interopérabilité. Il s'agira donc éventuellement de choisir une norme de représentation et description pour le traitement automatique du langage naturel [4] vue l'existence de plusieurs grammaires HPSG non évaluées et assez développées.

C'est dans ce cadre que s'inscrit le présent article. En effet, l'objectif principal est de proposer une démarche de projection du HPSG vers un langage pivot normalisé et compatible avec le standard LMF (Lexical Markup Framework) [2] [3] [5] qui est validé par le comité ISO TC 37/SC4. Cela va nous permettre de vérifier le degré de couverture des grammaires HPSG déjà élaborées et de pouvoir fusionner les résultats obtenus. En effet, le même processus peut être appliqué sur des grammaires provenant de formalismes d'unification différents. Un autre objectif qui est nécessaire et qui va faciliter la projection est d'identifier les phénomènes linguistiques qui sont traités par les grammaires d'unification, en particulier HPSG, et ceux qui ne sont pas traités par le LMF (Lexical Markup Framework).

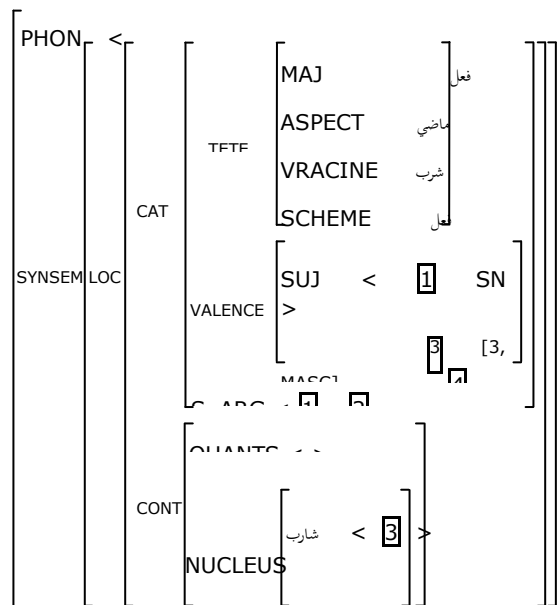
#### • Démarche proposée

En se basant sur le méta modèle de LMF et sur les extensions appliquées à ce modèle (voir figure 1), d'une part, et sur la notion d'intégration des connaissances de HPSG, d'autre part, nous élaborons une démarche de projection du HPSG vers LMF. Cette démarche est composée essentiellement de quatre étapes : la classification des traits de la structure attribut valeur (SAV), l'utilisation d'un système de règles de projection établi, l'élaboration d'un diagramme d'objets étendu selon le méta modèle de LMF et ses extensions et la génération d'un fichier XML en format LMF résultat de la projection.



**Figure 1** Méta modèle + extensions du LMF

En effet, la première étape consiste à classer les traits de la structure HPSG relative à une entrée lexicale en des traits morphologiques, syntaxiques et sémantiques. Les traits morphologiques dans une SAV adaptée à la langue arabe sont, par exemple, MAJ, ASPECT, VRACINE et NFORM. Les traits VALENCE, S\_ARG et VOIX en sont des exemples de traits syntaxiques alors que le trait NUCLEUS est considéré comme un trait sémantique.



**Figure 2** Entrée lexicale du verbe شرب

En effet, la SAV de la figure 2 montre que le verbe « شرب » est un verbe conjugué à l'accompli « ماضي » et en voix active. Ce verbe sous catégorise un sujet masculin conjugué avec la troisième personne du singulier et un complément qui doit être un syntagme nominal. Ainsi, la relation « شرب » comporte deux arguments représentés par les traits « شارب » « buveur » et « مشروب » « bu » ayant pour valeur respectivement les traits d'index du sujet et du complément. Les traits morphologiques dans la figure 5.1 sont ceux qui existent dans le trait TETE c'est-à-dire MAJ, ASPECT, VRACINE et SCHEME, les traits syntaxiques se trouvent dans les traits VALENCE et S\_ARG comme VOIX, SUJ et les traits sémantiques sont représentés dans le trait NUCLEUS (شارب et مشروب).

La quatrième étape nous fournit le fichier LMF suivant :

```

<struct type = "Lexical Entry">
  <feat type = "Part of speech"> verbe </feat>
  <feat type = "Language"> Arabe </feat>
  <struct type = "Form">
    <feat type = "Lemmatized Form"> شرب </feat>
    <struct type = "Inflected Form">
      <struct type = "Data Category 12620">
        <feat type = "Att"> SCHEME </feat>
        <feat type = "Val"> فعل </feat>
      </struct>
      <struct type = "Data Category 12620">
        <feat type = "Att"> PERSON </feat>
        <feat type = "Val"> 3 </feat>
      </struct>
      <struct type = "Data Category 12620">
        <feat type = "Att"> ASPECT </feat>
        <feat type = "Val"> ماضي </feat>
      </struct>
    </struct>
  </struct>

```

```

</struct>
<struct type = "Syntactic Behaviour">
  <struct type = "Self">
    <struct type = "Syntactic Feature">
      <feat type = "VOIX"> معلوم </feat>
    </struct>
  </struct>
</struct>
</struct>
<struct type = "Position">
  <feat type = "Function"> Subject </feat>
  <struct type = "Syntactic Actant">
    <feat type = "Syntactic Group" > N </feat>
  </struct>
</struct>
<struct type = "Argument">
</struct>
</struct>

```

**Figure 3** Fichier XML représentant le format LMF de شرب

Ainsi, le passage de HPSG vers LMF n'est pas une simple traduction : balise vers une autre balise. En effet, nous devons indiquer, pour chaque trait d'une entrée lexicale, la classe d'appartenance en se basant sur la nature de l'information qui l'exprime (morphologique, syntaxique ou sémantique). Aussi, nous pouvons affirmer qu'il n'y aura aucune perte d'informations mais plutôt gain d'informations.

- **Conclusion et perspectives**

Ce que nous venons de présenter dans cet article est une vue générale des concepts de base du formalisme HPSG et LMF, ainsi que leur application sur les entrées lexicales.

Tout d'abord, nous avons donné un aperçu sur le formalisme **HPSG**. Ensuite, nous avons présenté la plate-forme **LMF**. Puis, nous avons décrit la démarche proposée pour la projection de HPSG en LMF. Mais, il nous reste encore l'expérimentation de ce travail.

Le présent travail peut être généralisé sur d'autres formalismes; ainsi nous arrivons à les évaluer, à vérifier le degré de couverture de chacune d'eux et à fusionner les résultats obtenus. Nous pouvons aller plus loin et tester l'autre sens c'est-à-dire, essayer d'alimenter des applications conçues dans des grammaires d'unification à partir des bases de données lexicales en LMF. Ceci va favoriser la réutilisation et l'enrichissement des lexiques existants.

### **Bibliographie**

- [1] Detmar M. (2002), «Head-Driven Phrase Structure Grammar An Introduction as Background for Grammar Implémentation », OSU, LING795K, Spring.
- [2] Eagles (1996), « EAGLES Recommendations on subcategorisation », Eagles document EAG-CLWG-SYNLEX.
- [3] Gil FRANCOPOULO (2003), « Proposition de norme des lexiques pour le traitement automatique du langage », inria/loria-action syntaxe.
- [4] ISO/TC37/SC 4 N130 Rev.5, (2005) « Language Resource Management – Lexical Markup Framework (LMF) ».
- [5] Romary L. (2005) – Loria/CNRS, Nancy, Susanne Salmon-Alt-ATILF/CNRS, Nancy, « Modélisation de structures lexicales : Application aux lexiques ».
- [6] Pollard C., Sag, I.A. (1994), « Head-Driven Phrase Structure Grammar », publié par la presse dans l'université de Chicago, Edition Golgoldmittu, Chicago, LSLI.
- [7] Kasper R., Kiefer B., Netter K. ET K. Vijay-Shanker (1995), «Compilation of HPSG to TAG» proceeding of the 33rd ACL, p 92-99.