

Vers une Ressource Prédicative pour l'Extraction d'Information

Aurélien Bossard – aurelien.bossard@lipn.univ-paris13.fr – LIPN - Université Paris 13 - 93430 Villetaneuse

Problématique

- Ce travail tente de créer une ressource complète à partir de celles existantes, et l'appliquer à une tâche d'extraction.
- Des ressources prédicatives pour les verbes existent pour l'extraction d'information en anglais (PropBank, FrameNet ...).
- Aucune ressource de ce type dédiée à l'EI pour le français.
- Des ressources incomplètes existent (Volem, tables du LADL).

Volem (IRIT)

- Encode des informations sur les prédicats verbaux (alternances, rôles thématiques).
- Rôles thématiques trop généraux.
- Manque d'informations (prépositions, auxiliaires) mais bon point de départ pour l'extraction.

Description du verbe : acheter

GRILLE THEMATIQUE :
[[inic(agent),dest],[th],[src]]
LCS :
ALTERNANCES :
caus_2np_pp , anti_pr_np , anti_pr_np_pp , pas_être_part_np_2pp , pas_être_part_np_pp , caus_2np , caus_refl_pr_2np , caus_np_pp , caus_support_np
WN :
[13,2,3], [13,3,1], [13,3,8]
EXEMPLE :
Il a acheté ce livre à un brocanteur

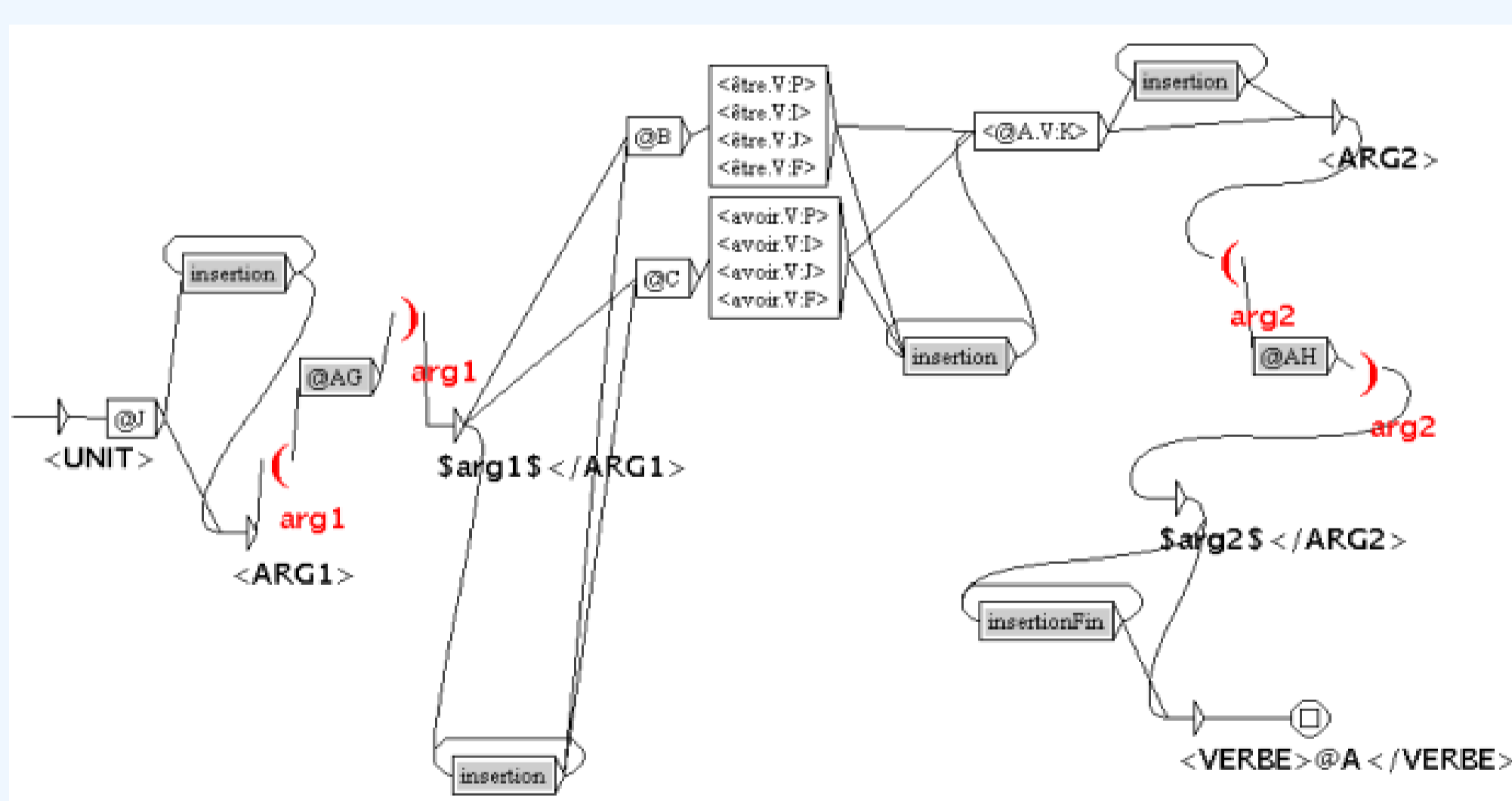
Développement d'une nouvelle ressource

- Ajout des auxiliaires, prépositions.
- Ajout de la forme nominale des verbes.
- Sélection d'un groupe de verbes correspondant à une tâche d'extraction.
- Spécialisation semi-automatique des rôles thématiques des verbes sélectionnés.

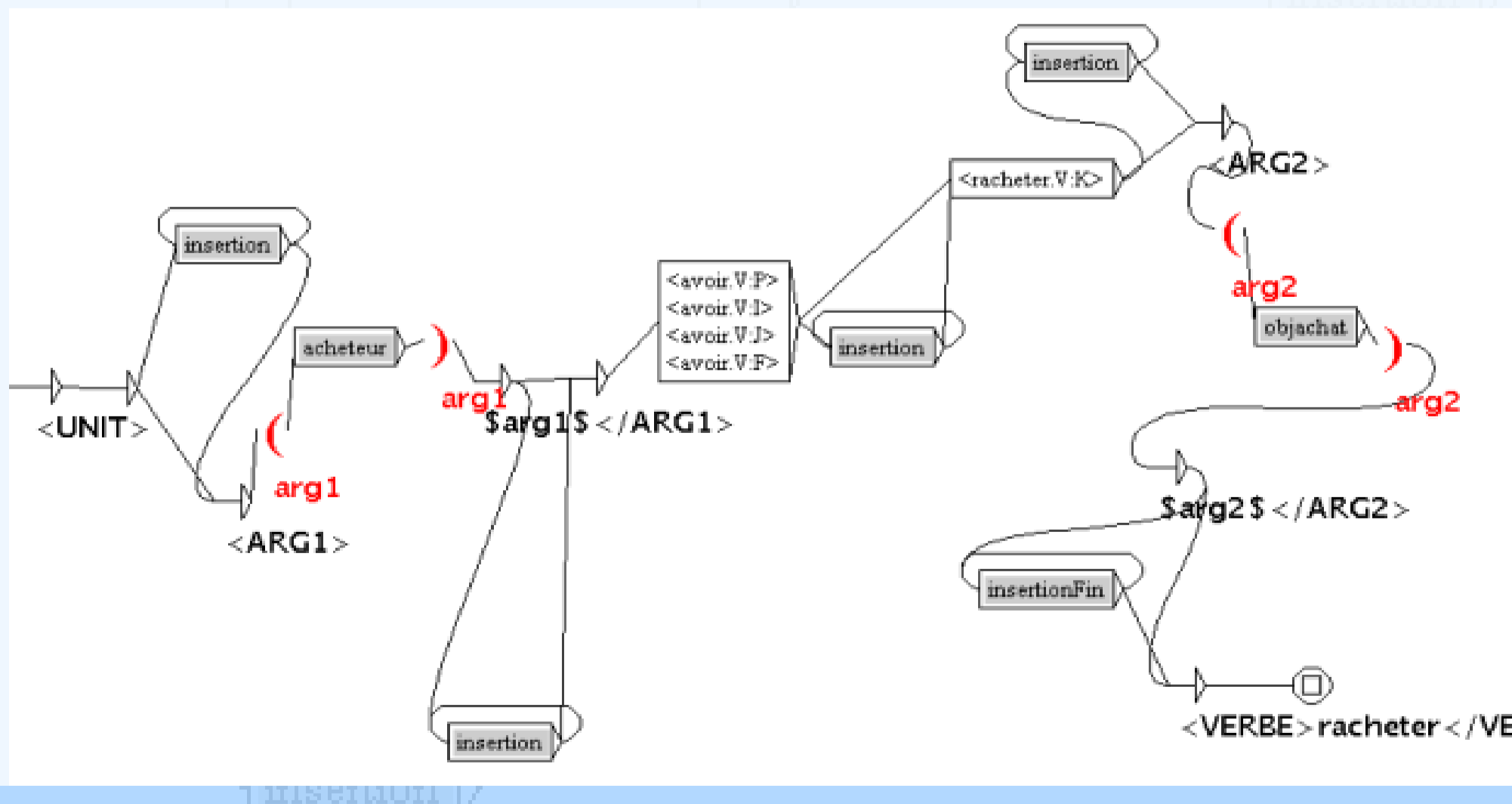
Verbe	Etre	Avoir	anti_2np	Alternances	state_2np_ppimp	Th1	Th2	Th3	Prep1	Prep2	Prep3	verbe_supponom
racheter	-	+	-	+	-	acheteur	objachat	vendeur	<E>	<E>	<à>+<au>+	rachat
revendre	-	+	-	-	-	vendeur	objachat	acheteur	<E>	<E>	<à>+<au>+	revente
se partager	+	-	-	-	-	initiateur	ftib		<E>	<E>	<E>	-
acheter	-	+	-	-	-	acheteur	objachat	vendeur	<E>	<E>	<à>+<au>+	achat
fusionner	-	+	-	+	-	initiateur	fusionne	fusionne	<E>	<avec>	<avec>+<e>+	fusion
fusionner	-	+	-	-	-	fusionne	fusionne	<E>	<E>	<avec>	<E>	fusion
céder	-	+	-	+	-	vendeur	objachat	acheteur	<E>	<E>	<à>	cession
vendre	-	+	-	-	-	vendeur	objachat	acheteur	<E>	<E>	<à>+<au>+	vente
acquérir	-	+	-	-	-	acheteur	objachat	vendeur	<E>	<E>	<à>+<au>+	acquisition
offrir	-	+	-	-	-	acheteur	objachat	acheteur	<E>	<E>	<E>	-
offrir	-	+	-	+	-	acheteur	montant	objachat	<E>	<E>	<pour>+<en>+	offre
reprendre	-	+	-	-	-	acheteur	objachat	acheteur	<E>	<E>	<à>+<au>+	reprise
détenir	-	+	-	-	-	acheteur	objachat		-	-	-	-

Graphes Unitex

- Création de graphes « générateurs » : des graphes qui reconnaissent des formes de surfaces des alternances de Volem.



- Création des graphes pour les verbes sélectionnés à partir de la ressource.



Bilan à travers une tâche d'extraction d'information

- Les graphes annotent le corpus au format XML. Les résultats sont extraits dans un tableau.

Acheteur	objachat	Vendeur	CCMontant
CPI	Fulmar		50 millions d'euros

- Performances d'extraction d'information correctes (65% sur un corpus économique relativement régulier).
- Etablissement de statistiques pour ajouter aux résultats les compléments circonstanciels jugés essentiels.
- Une partie manuelle subsiste (annotation en entités nommées, sélection de verbes). Le travail réalisé permet tout de même une systématisation des tâches d'extraction d'information.

Conclusions

- L'extraction d'information bénéficierait d'une ressource prédicative à large couverture.
- L'application de la ressource à l'extraction d'information nécessite une annotation correcte des entités nommées et un lien entre celles-ci et les rôles thématiques.
- Travaux à poursuivre sur l'acquisition semi-automatique de schémas prédicatifs.