

Comparaison de constructions verbales entre un corpus d'apprenants et un corpus d'experts

Sylvain Hatier

Rui Yan

LIDILEM, Université Grenoble Alpes

Rui.Yan@u-grenoble3.fr

Sylvain.Hatier@u-grenoble3.fr

8es Journées Internationales de Linguistique de Corpus
2-4 septembre 2015, Orléans

Plan

- Cadre de recherche
- Lexiques et Corpus
- Extraction des constructions verbales
- Comparaison entre corpus
- Conclusion et perspectives

Objet d'étude

- Sous-catégorisation des verbes du Lexique Scientifique Transdisciplinaire (LST)
- LST : lexique spécifique de l'écrit académique
 - Étudié à travers FOU (EAP)
 - Academic Vocabulary, Sub-technical vocabulary, LSG (, Coxhead, 2000, Paquot, 2010, Pecman 2005)
- Travaux sur les verbes du LST
 - Collocations V+N (Nesselhauf's, 2005; Cavalla, 2008)
→ analyse des erreurs
 - Formes et patrons verbaux (Hyland, 2008; Granger & Paquot, 2009) → sous-utilisation par des apprenants

Objet d'étude

En français, peu d'études sur les constructions syntaxiques des verbes dans les écrits d'apprenants

« *lexical verbs contribute significantly to some major EAP functions such as expressing personal stance, reviewing the literature, quoting, expressing cause and effects, summarizing and contrasting.* » (Granger&Paquot, 2009)

Intérêt de la maîtrise de constructions verbales remplissant des fonctions rhétoriques importantes

Analyse centrée sur 4 verbes : *expliquer, montrer, considérer et décrire*

Objectifs

- Examiner comment les apprenants chinois du français emploient les verbes les plus fréquents dans l'écrit scientifique
- Mettre en lumière les difficultés liées aux constructions verbales dans la rédaction scientifique.
 - quels types de constructions verbales les apprenants utilisent-ils le plus fréquemment?
 - quelles sont les constructions qui leur posent le plus problèmes ?

Hypothèses

- Moins de variétés dans les constructions verbales dans les écrits d'apprenants
 - Plusieurs cadres de sous-catégorisation possibles selon le verbe
 - Différentes alternances syntaxiques ou sémantiques (sujet métonymique) admises selon le verbe
- Les constructions verbales les plus fréquentes chez les experts sont les mieux maîtrisées par les apprenants, et inversement

LST

Définition (Tutin 2007) & propriétés (Drouin 2007) :

- Exprime/décrit les procédures scientifiques, le raisonnement, la filiation scientifique, le positionnement, le métatexte ...
- Majoritairement abstrait
- propre aux écrits scientifiques → Sur-représentation par rapport à d'autres genres
- Transdisciplinaire → traverse les disciplines → répartition forte

LST

- Extraction des unités du LST à partir des critères statistiques
 - Sur-représentation par rapport à un corpus de contraste hétérogène
 - Réparti dans les disciplines
- Ajout de traits sémantiques de classes et sous-classes basées
 - Propriétés sémantiques issues du DEM et LVF (Dubois & Dubois-Charlier)
 - Propriétés distributionnelles issues du corpus

Corpus

Corpus d'apprenants chinois

- Étudiants chinois de spécialité de français, B2 acquis
- 29 mémoires de Master 2 (20000 mots/mémoire)
 - Sujets concernés: traduction (9), littérature (12), linguistique (1) et socio-culturel (7)
- 600 000 mots

Corpus d'experts Scientext (<http://scientext.msh-alpes.fr/>)

- Articles de revues de 10 disciplines des SHS
- Annoté en parties textuelles
- 5 000 000 mots

Les deux corpus sont analysés syntaxiquement avec XIP (Aït-Mokhtar & al 2002) → homogénéité dans le calcul des relations de dépendance et dans la segmentation

Cadres de sous-catégorisation

- Étude effectuée sur des verbes très fréquents du LST :
 - Permet de récupérer un maximum d'occurrences
- Verbes essentiels du point de vue argumentatif

Verbe	occ/M - Experts	occ/M - Apprenants
<i>montrer</i>	649 (3114)	459 (270)
<i>considérer</i>	533 (2259)	460 (290)
<i>décrire</i>	236 (1134)	378 (238)
<i>expliquer</i>	341 (1637)	263 (166)

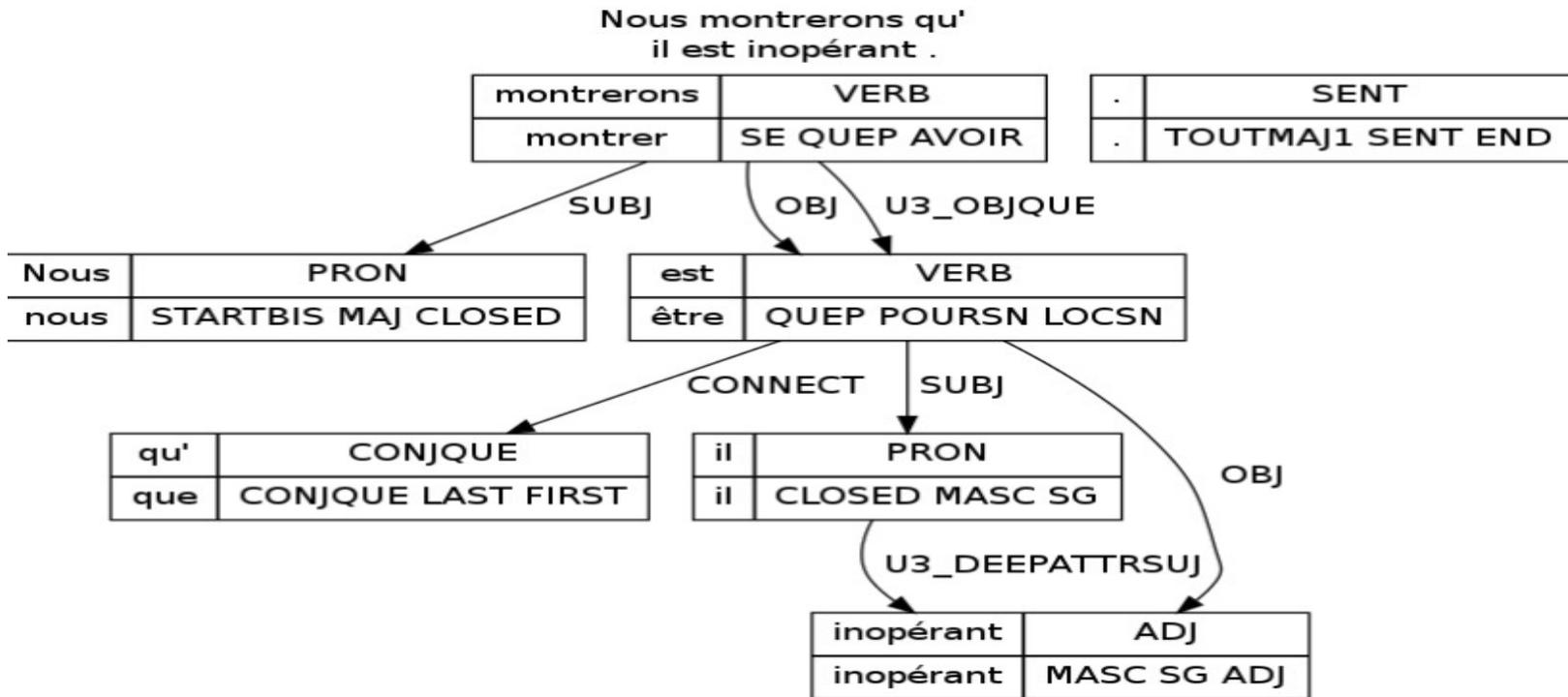
Nombre d'occurrences par millions de mots (fréquence absolue)

Extraction des cadres

- Pour chaque occurrence d'un verbe → extraction de l'ensemble des relations syntaxiques impliquant le verbe étudié (gouverneur ou dépendant)
 - Objectif : faire émerger du corpus des cadres sans a priori (Messiant & al 2010, Kupsc 2007)
- Identification des éléments les plus fréquents en tant que sujet et complément du verbe
 - Objectif : dégager des classes de mots en cooccurrence privilégiée pour ensuite repérer des patrons du type
 - SUBJ {communication/support} montrer OBJ {qualité/importance}

Dét (ouvrage|article|texte) montre Dét (rôle|importance)

Exemple de cadre



Phase de post-traitement syntaxique (U3_OBJQUE), de regroupement, suppression de certaines relations (AUX)

Formalisme : (SUBJ +hum) montrer_VERB (COMPLETIVE_QUE)

Extraction des cadres

Recherche tout champ

Afficher éléments

Verbe	Cadre	Occurrences	Exemples	Sujets	Objets
montrer	(SUBJ -hum)montrer_VERB(COMPLETIVE_QUE)	762	Texte_psycho_32_XipEmolex.xml : Les résultats montrent que le facteur qui influence plus particulièrement la performance des enfants est l'alignement de la position finale avec un axe implicite du dispositif (vertical ou horizontal) .	/ 91:résultat / 68:il / 55:analyse / 41:étude / 27:travail	/ 177:être / 53:pouvoir / 45:avoir / 12:exister / 10:permettre
montrer	(SUBJ -hum)montrer_VERB(OBJ -hum)	275	Texte_psycho_8_XipEmolex.xml : Les résultats montrent un effet signifi-catif de la présence de l'appareil , t (24) = - 2,27 ; p < 0,03 .	/ 29:analyse / 25:il / 16:résultat / 9:exemple / 8:étude	/ 13:importance / 13:effet / 10:différence / 10:existence / 6:rôle
montrer	(SUBJ +hum)montrer_VERB(COMPLETIVE_QUE)	150	Texte_economie_10_XipEmolex.xml : Nous montrons qu' il n' en est rien .	/ 103:nous / 39:on / 8:je	/ 39:être / 16:pouvoir / 8:permettre / 5:devoir / 4:dépendre
montrer	(SUBJ -hum)montrer_VERB(COMPLETIVE)	101	Texte_sociologie_23_XipEmolex.xml : En portant l'attention sur la nature et la constitution des choses , un mécanisme montre comment , sous certaines conditions , les effets d' une cause sont produits [10] .	/ 9:étude / 8:il / 8:travail / 6:analyse / 4:article	/ 19:être / 15:pouvoir / 6:permettre / 2:inscrire / 2:construire

Analyse des cadres

Pour les cadres les plus fréquents dans les deux corpus :

- Comparaison du rang d'usage
 - Est-ce la construction la plus fréquente, la moins fréquente pour le verbe étudié ?
- Comparaison des compléments intégrant le cadre, sujet ou objet
 - Distinction du trait *+hum* (pronom *nous, je, on*)
 - Repérage de classes sémantiques
- Identification des différences entre experts/apprenants

Cadres du verbe *montrer*

Cadre	Experts	Apprenants
(SUBJ -hum) montrer (COMPLETIVE_QUE)	1)	4)
(SUBJ +hum) montrer (COMPLETIVE_QUE)	2)	2)
(SUBJ -hum) montrer (OBJ -hum)	3)	1)
Comme le montrer (SUBJ +hum)	4)	-

- Corpus d'experts propose plus de diversité, des alternances passives, constructions infinitives avec *suggérer, nécessiter, permettre, etc.*
- Différence sur les sujets sélectionnés :
 - Proportion de pronom forte chez les apprenants
 - Plus de diversité, emploi métonymique de *article, étude, travail* chez les experts

Cadres du verbe *considérer*

Cadre	Experts	Apprenants
OBJ -hum considéré (Passif)	1)	-
(SUBJ +hum) considérer (COMPLETIVE_QUE)	2)	3)
(OBJ -hum) être considéré comme	3)	1)
(SUBJ +hum) considérer (OBJ -hum) comme	-	2)

- Plus grande diversité dans corpus SHS, avec :
 - *Considérons (OBJ -hum)*
 - *Si l'on considère (OBJ -hum)*
 - ...
- Cadre spécifique apprenant → *Il considère la théorie comme ...*
- Peu de différences au niveau des compléments sélectionnés, pronom *il* ou nom +hum

Cadres du verbe *expliquer*

Cadre	Experts	Apprenants
(SUBJ -hum) expliquer_VERB (OBJ -hum)	1)	1)
(OBJ -hum) se expliquer_VERB par (SUBJ -hum)	2)	-
(SUBJ -hum) expliquer_VERB COMPLETIVE_QUE	3)	-

Pour les experts

- Plus grande diversité des cadres avec les alternances en passif, pronominal passif
- Construction infinitive fréquente avec : *pouvoir, viser à, permettre,*

Pour les apprenants, trop peu de régularité pour regrouper les cadres → difficile à analyser

Cadres du verbe *décrire*

Cadre	Experts	Apprenants
(OBJ -hum) être décrit (Passif)	1)	3)
(SUBJ -hum) décrire_VERB (OBJ -hum)	2)	2)
(SUBJ +hum) décrire_VERB (OBJ -hum)	3)	1)

À nouveau, les alternances sont peu utilisées par les apprenants → une construction récurrente :

- *Il décrit* (OBJ +hum)

Analyse contrastive des cadres

- Les experts et les apprenants partagent les constructions de base (SVO et construction en complétive *que*).
- Mais plus d'alternances sémantiques (emploi métonymique) et syntaxiques (construction en passif) chez les experts.

*Plus fondamentalement , **cet article** vise à montrer comment a évolué l'appréhension...(E)*

*Cette différence **s'explique par** un effet générationnel classique. (E)*

*Elle **est** plutôt **décrite** en termes d' évolution historique générale (ce thème ayant des incidences sur les différents éléments). (E)*

- Certaines constructions sont peu présentes/absentes chez les apprenants.

– Considérer « prendre en considération »:

***Si l'on considère** l' ensemble de la population interrogée...(E)*

Analyse contrastive des cadres

Les compléments (sujet et objets) verbaux renvoient particulièrement :

- Aux noms du LST chez les experts (*article, résultat, étude*)
- À la langue générale / aux objets d'étude chez les apprenants

*D'un père poète à la fois professeur, Shan Sa a **montré son talent** poétique dès son enfance. (A)*

*En le traduisant par seulement « les nuages » ne peut **pas montrer l'image** décrite dans la version chinoise. (A)*

*L' **analyse** des vidéos **montre** bien d'ailleurs cette focalisation sur les appels liés au talkie . (E)*

Au niveau des pronoms sujets, les apprenants utilisent *il/ils* et les experts *on/nous* → différence de genre de texte

Conclusion

Hypothèses en partie vérifiées → + grande variété des cadres pour les experts

Mais les cadres les + fréquents dans le corpus d'experts ne sont pas les mieux maîtrisés ou les plus employés par les apprenants

- Certaines de ces constructions assurent des fonctions rhétoriques essentielles → importance dans l'encodage/décodage pour les non-natifs
- Intérêt d'enseigner les verbes du LST et leurs constructions en regroupant par fonction rhétorique

Conclusion

Les différences s'expliquent également par les différentes acceptions mobilisées (le verbe *considérer* dans son sens de *prendre en considération* très peu présent chez les apprenants)

Ces différences sémantiques peuvent être expliquées par :

- La nature des corpus
 - Taille (corpus experts 9 fois plus grand)
 - Sujet : mémoire de français vs article de recherche
- Les difficultés des apprenants à maîtriser les constructions appropriées

Perspectives

Élaboration d'une ressource didactique en fonction des résultats de la comparaison

Affiner l'extraction des cadres et améliorer les regroupements d'occurrences

Intégrer des informations sémantiques automatiquement dans les cadres → par exemple, le verbe *présenter* sélectionne des sujets de la classe {*support de communication*} : *figure, tableau, article, section, etc.*

Références

- Aït-Mokhtar, Salah, Jean-Pierre Chanod, and Claude Roux. "Robustness beyond shallowness: incremental deep parsing." *Natural Language Engineering* 8.3 (2002): 121-144
- Bolly, C. (2011). *Phraséologie et collocations. Approche sur corpus en français L1 et L2*. Bruxelles : Peter Lang.
- Cavalla, C. (2008). Les collocations dans les écrits universitaires : un lexique spécifique pour les apprenants étrangers. In Olivier Bertrand et Isabelle Schaffner. *Apprendre une langue de spécialité: enjeux culturels et linguistiques*, Editions Ecole Polytechnique, p.93-104.
- Drouin, P. (2007). Identification automatique du lexique scientifique transdisciplinaire. *Revue française de linguistique appliquée*, Vol. XII (2): 45-64.
- Granger, S., Paquot, M. (2009). Lexical verbs in academic discourse: a corpus-driven study of learner use. In Charles M.; Pecorari D.; Hunston S., *Academic Writing: At the Interface of Corpus and Discourse*, Continuum : London and New-York, p.193-214
- Hyland, K. (2008). Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18, (1), 41-62.
- Nesselhauf, N. (éd.). (2005). *Collocation in a Learner Corpus*. Amsterdam / Philadelphia : John Benjamins Publishing Company.
- Paquot, M. (2010). *Academic vocabulary in learner writing: From extraction to analysis*. Bloomsbury Publishing.
- Tutin, A. (2007). Autour du lexique et de la phraséologie des écrits scientifiques. *Revue française de linguistique appliquée*, Vol. XII (2): 5-14.