

L'apport d'un corpus annoté à l'étude du rapport entre vers et phrase en ancien français

Yannick Mosset
et T. M. Rainsford

Pratiques de corpus linguistiques
ENS Lyon, 29 mars 2012

Sommaire

- Introduction
- *Syntactic Reference Corpus of Medieval French*
- Étude

Nous remercions Achim Stein, Sophie Prévost et Alexei Lavrentiev de nous avoir fourni quelques diaporamas pour cette communication.



Introduction

Introduction

- La littérature médiévale en ancien français n'existe que sous forme versifiée (jusqu'au 13^e siècle)
- Les premiers documents linguistiques que nous possédons témoignent donc d'une langue informée par le vers
- Conséquence linguistique: coïncidence possible entre « regroupements syntaxiques » et « regroupements prosodiques » (Glikman, 2009)
- La métrique n'a toutefois pas de fonction syntaxique précise

Introduction

- Au plan littéraire, quatre notions s'entrecroisent: la syntaxe, la métrique, le genre et l'auteur
 - Les genres littéraires peuvent se définir par une (ou des) métrique(s) spécifique(s): le décasyllabe pour la chanson de geste, l'octosyllabe pour le roman
 - Les genres littéraires peuvent privilégier des formes syntaxiques
 - Chaque auteur peut (plus ou moins) s'individualiser au sein d'un genre
- Bérout et Chrétien de Troyes: étude de leur inscription générique et de leur singularisation personnelle



Présentation du *Syntactic Reference Corpus of Medieval French* (SRCMF)

Présentation du projet

■ Organisation et financement

- **Titre:** Syntactic Reference Corpus of Medieval French (SRCMF)
- **Financement:** Agence nationale de la recherche (ANR) et Deutsche Forschungsgemeinschaft (DFG), 1.3.2009–29.2.2012, ~ 400 K€
- **Staff:** coordinateurs**, chercheurs* et associés
 - **CNRS Lattice Paris (F):** Sophie Prévost**, Julie Glikman*
 - **ENS Lyon (F):** Céline Guillot, Serge Heiden, Alexei Lavrentiev*, Christiane Marchello-Nizia (emeritus), Tom Rainsford*
 - **ILR Universität Stuttgart (D):** Achim Stein**, Beatrice-Barbara Bischof*, Nicolas Mazziotta*

Présentation du projet

■ Objectifs

- production d'un « treebank » (corpus arboré) pour le français médiéval en rajoutant une couche d'annotation à deux corpus importants :

- Base de Français Médiéval

- <http://bfm.ens-lyon.fr>, resp. Ch. Marchello-Nizia → Céline Guillot
- ~ 80 textes, > 3 mln mots, IX^e – XV^e
- annotation POS vérifiée de ~ 20 textes, étiquettes Cattex

- Nouveau Corpus d'Amsterdam

- <http://uni-stuttgart.de/lingrom/stein/corpus>
- compilé par A. Dees (1987) → édité et publié par A. Stein (Stuttgart), P. Kunstmann (Ottawa) and M. Glessgen (Zurich, métadonnées) (2006)
- 300 échantillons des textes de l'anc. fr., > 3 mln mots, 12-13th s.
- annotation POS vérifiée partiellement, lemmatisation, étiquettes Dees/ILR

Corpus

■ Corpus en cours de finalisation

□ Textes

- 14 textes annotés du X^e au XIII^e siècles
- environ 250.000 mots
 - bientôt disponible à la communauté de recherche, excepté problèmes de droits d'éditeurs.

□ Couches d'annotation

- annotation morpho-syntaxique (Cattex)
- annotation syntaxique
 - modèle dépendentielle (par ex. Tesnière 1959, Mel'čuk & Polguère éd., 2009)
- annotation du discours direct
- annotation des débuts / fins de vers

Procédé d'annotation syntaxique

- Chaque texte annoté indépendamment par deux experts
 - logiciel d'annotation NotaBene (Mazziotta, 2010)
- Comparaison
 - identification des différences
 - élimination des erreurs évidentes
 - discussion des problèmes
- Vérification par les coordinateurs du projet

Exploitation du corpus

■ TigerSearch

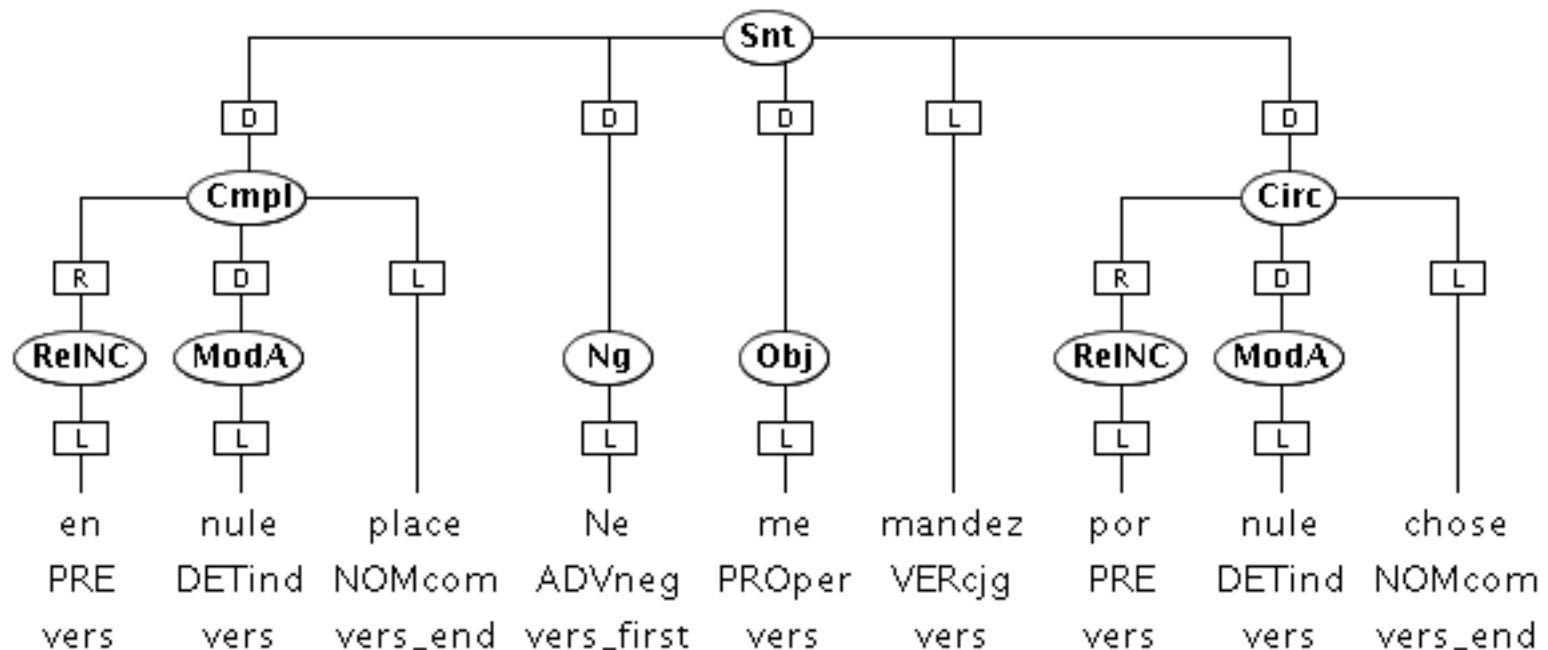
- <http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERSearch/>
- API open-source, non plus maintenu mais opérationnel
- installation locale
- langue de requête puissante pour l'investigation des structures arborescentes.

■ Export des résultats

- Export de TigerSearch en format Tiger-XML : visualisation difficile pour les utilisateurs
- XSL / scripts Groovy développés dans le cadre du projet SRCMF permettent l'export vers des concordances.

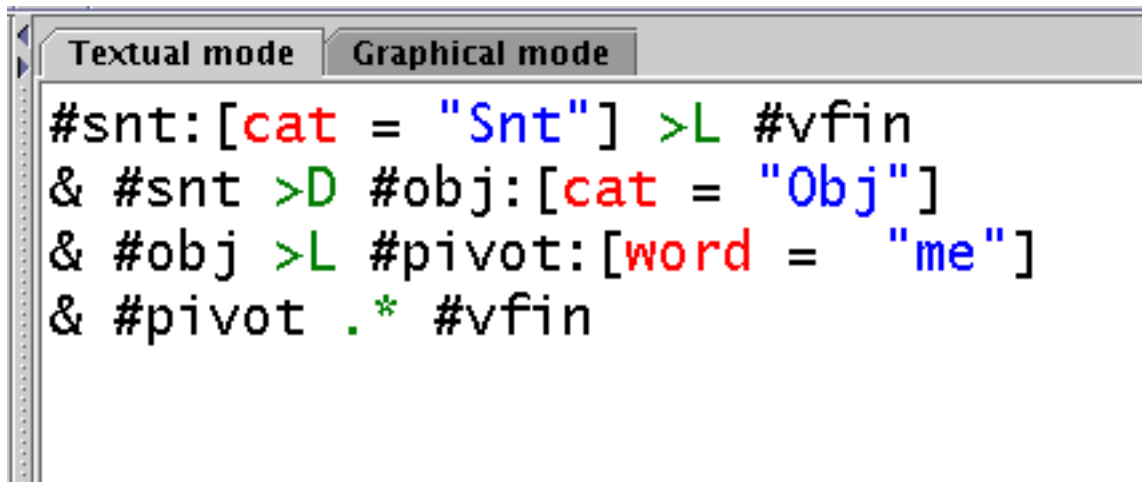
Exploitation du corpus

■ Visualisation de l'arbre syntaxique



Exploitation du corpus

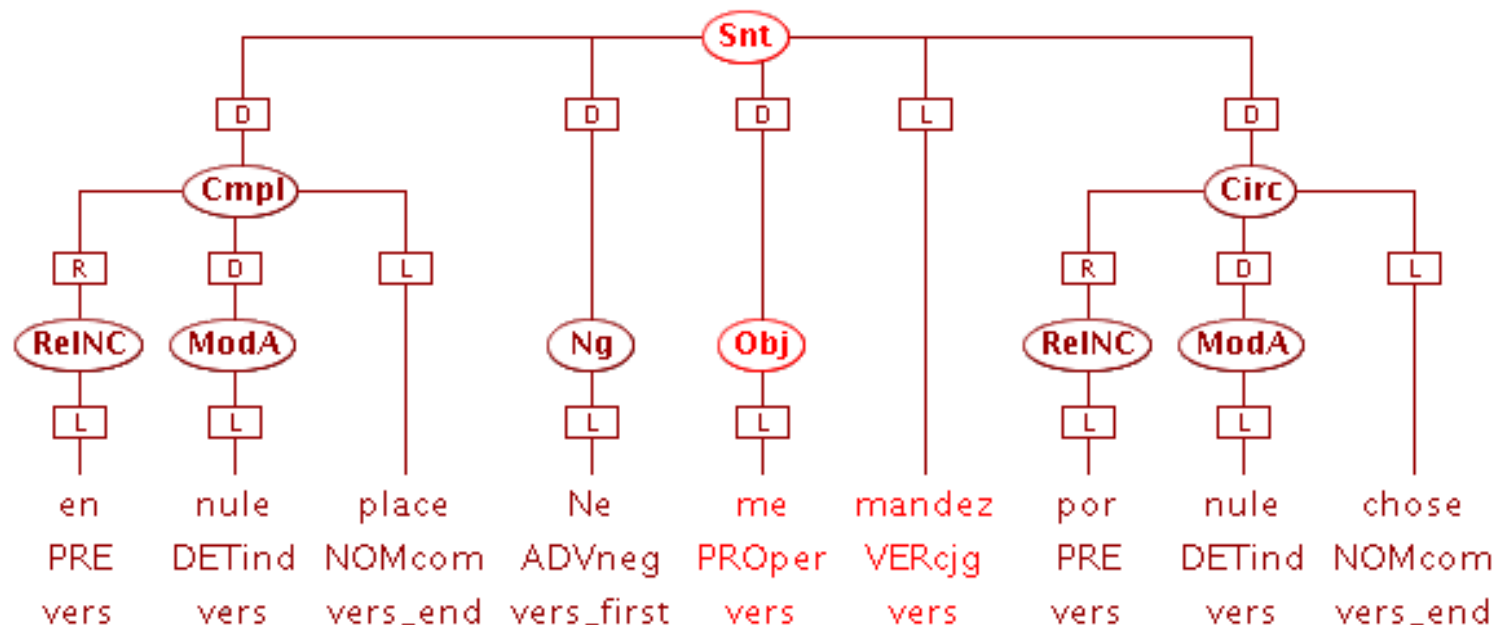
- Requête TigerSearch



```
Textual mode Graphical mode
#snt:[cat = "Snt"] >L #vfin
& #snt >D #obj:[cat = "Obj"]
& #obj >L #pivot:[word = "me"]
& #pivot .* #vfin
```

Exploitation du corpus

■ Graphique TigerSearch avec résultat



Exploitation du corpus

LeftCxInsideSnt	Pivot	RightCxInsideSnt
en nule place Ne	me	mandez por nule chose :
Il ne	me	lait sol escondire .
Se ja un poil en ai bruslé De la	me	laist tot ardoir u feu ;
haire qu' * avrai vestu , Si	me	porroit tenir por fole .
s' il or m' en ot parler , Bien	me	deconnut Qui * de ta feme me mescrut :
Beaus oncles , poi	me	fist monter ,
En cest arbre	me	pout plus ahonter .
Il ne	me	



Étude

Corpus

■ Textes utilisés

- *La Chanson de Roland*, chanson de geste (c.1100)
- *Tristan*, roman de Bérroul, (c. 1185)
- *Yvain*, roman de Chrétien de Troyes, (c. 1176-1180)
- *Le Miracle de Théophile* de Gautier de Coincy (genre narratif plus ou moins bref, c.1220-1230)
- extraits de la continuation du *Roman de la Rose*, roman allégorique et encyclopédique de Jean de Meun, (c.1270)

Méthode

- Utilisation de l'annotation des limites de phrases et de limites de vers.
 - Définition de la phrase : un seul verbe principal et ses dépendants (objet, attribut, circonstant, etc.)
 - Pas de coordination phrastique

Relations vers-phrases

	Roland	Beroul	Yvain	Coinci	Rose
Vers	3791	4433	6800	2006	3240
Vers-phrases	2043	1461	1172	277	424
	53.89%	32.96%	17.24%	13.81%	13.09%

Relations vers-phrases

	Roland	Beroul	Yvain	Coinci	Rose
Vers	3791	4433	6800	2006	3240
Vers-phrases	2043	1461	1172	277	424
	<i>53.89%</i>	<i>32.96%</i>	<i>17.24%</i>	<i>13.81%</i>	<i>13.09%</i>
Longueur moyenne des phrases (en vers)	0.97	1.34	1.75	1.96	2.25

Relations vers-phrases

	Roland	Beroul	Yvain	Coinci	Rose
Vers	3791	4433	6800	2006	3240
Vers-phrases	2043	1461	1172	277	424
	53.89%	32.96%	17.24%	13.81%	13.09%
Longueur moyenne des phrases (en vers)	0.97	1.34	1.75	1.96	2.25
Subordonnées / phrase (moyenne)	0.11	0.19	0.27	0.30	0.47

Rejet

■ Définitions

- Définition scolaire du rejet: « élément verbal bref, placé au début d'un vers ou d'un hémistiche, [qui] se trouve étroitement lié par la construction au vers [...] précédent, et prend par sa position une valeur particulière » (Mazaleyrat, 1995, p. 119).
- Définition linguistique adoptée: la phrase remplit (au moins) un vers entier et déborde sur le vers suivant, sans toutefois le remplir jusqu'au bout.

Rejet

■ Exemples:

- Et cil vient la qui mout covoite
Le pain / sel prant / et si i mort (Yvain, v. 2842–3)
- Li corbel qui sont plain de rage
Li font ennui / et il est sage (Bérroul, v. 3647–8)

Rejets

	Roland	Beroul	Yvain	Coinci	Rose
Phrases	3890	3319	3895	1021	1441
Rejets	28	98	181	23	36
	<i>0.72%</i>	<i>2.95%</i>	<i>4.65%</i>	<i>2.25%</i>	<i>2.50%</i>

Rejets

- Cette définition a généré des résultats qui ont dû être triés manuellement afin d'éliminer certains cas de figure:
 - Coordination
 - **Quant la chambre fu desfermee**
Si se leva / et oï messe (Yvain v. 4022–3)
 - Parallélisme
 - **Douce dame s'il ne t'anuie**
Je te requier / je te depri (Coinci, I Mir 10, v. 1328–9)
 - Introduction du discours direct
 - **La premiere foiz a consoil**
Li dist : « Dame mout me mervoil » (Yvain, v. 1599–1600)
 - Proposition incidente
 - **C'une tel plaie el vis avoit**
Messire Yvains / bien le savoit (Yvain, v. 2899–2900)

Rejets

	Roland	Beroul	Yvain	Coinci	Rose
Phrases	3890	3319	3895	1021	1441
Rejets	28	98	181	23	36
	<i>0.72%</i>	<i>2.95%</i>	<i>4.65%</i>	<i>2.25%</i>	<i>2.50%</i>
« Vrais » rejets	8	58	114	2	21
	<i>0.21%</i>	<i>1.75%</i>	<i>2.93%</i>	<i>0.20%</i>	<i>1.46%</i>

Rejets

■ Rejets « intra-phrastiques »

- Les statistiques permettent de définir un type particulier de rejet: le rejet « phrastique » (un élément de la phrase est rejeté dans un vers qu'il ne remplit pas entièrement)
- Toutefois, on peut définir un rejet « intra-phrastique »: un syntagme normalement solidaire est séparé par la frontière du vers et un de ses éléments est rejeté au vers suivant
 - **Jamés si riche mariage**
N'avroiz, se vos cestui n'avez (Yvain, v. 5710-1)

Rejets

- Résultat du dépouillement manuel partiel (1000 vers d'*Yvain* et du *Tristan*)
 - Dans *Yvain*, 15 rejets phrastiques et 47 rejets intraphrastiques
 - Dans *Tristan*, 14 rejets phrastiques et 12 rejets intraphrastiques

Rejets

■ Nombre de syllabes de l'élément rejeté

	Roland	Beroul	Yvain	Coinci	Rose
« Vrais » rejets	8	58	114	2	21
2 syll.	<i>0.00%</i>	<i>12.07%</i>	<i>14.04%</i>		<i>4.76%</i>
3 syll.	<i>0.00%</i>	<i>3.45%</i>	<i>28.95%</i>		<i>47.62%</i>
4 syll.	<i>100.00%</i>	<i>75.86%</i>	<i>50.88%</i>		<i>38.10%</i>
5 syll.	<i>0.00%</i>	<i>5.17%</i>	<i>6.14%</i>		<i>9.52%</i>
6 syll.	<i>0.00%</i>	<i>3.45%</i>	<i>0.00%</i>		<i>0.00%</i>

Conclusion

■ Classement des textes

- confirmation de la singularité du roman de *Bérout*:
 - proportion élevée des vers-phrases par rapport aux autres textes octosyllabiques;
 - statut plus « archaïque » plus proche des chansons de geste
- confirmation de l'usage fréquent du rejet chez Chrétien de Troyes
 - une innovation stylistique
 - pas utilisée avec la même fréquence dans les textes du XIIIe que nous avons examinés.

Conclusion

- Le corpus nous fournit un appui quantitatif sur une large base de données à des constats plus qualitatifs, voire intuitifs, qui existent dans la critique littéraire.
- De plus, le travail de corpus nous oblige à préciser nos définitions des phénomènes stylistiques.
- Une étude d'ampleur de la question de la métrique des genres narratifs médiévaux devient de plus en plus nécessaire.



MERCI!

Relations vers-phrases

	Roland	Beroul	Yvain	Coinci	Rose
Vers	3791	4433	6800	2006	3240
Vers-phrases	2043	1461	1172	277	424
	<i>53.89%</i>	<i>32.96%</i>	<i>17.24%</i>	<i>13.81%</i>	<i>13.09%</i>
Phrases	3890	3319	3895	1021	1441
Phrases en vers entiers	2499	2311	2851	781	1151
	<i>64.24%</i>	<i>69.63%</i>	<i>73.20%</i>	<i>76.49%</i>	<i>79.88%</i>

Rejets

- Différents éléments stylistiques peuvent ensuite être interrogés, de façon plus qualitative
 - Singularisation de Chrétien
 - Hémistiche formulaire rejeté « Messire Yvain » (6 occurrences + 1 de « Monseigneur Yvain »; absent de Béroutl)
 - Plus d'enui feire ne li quiert
 - Mes sire Yvains, ençois descent (v. 2258-9)
 - Singularisation de Béroutl
 - Tendance au commentaire post-rejet (9 occurrences, deux chez Chrétien)
 - Mot est li rois acoragiez
 - De destruire: c'es granz pechiez. (v. 1951-2)

Rejets

	Roland	Beroul	Yvain	Coinci	Rose
Rejets	28	98	181	23	36
« Vrais » rejets	8	58	114	2	21
	<i>28.57%</i>	<i>59.18%</i>	<i>62.98%</i>	<i>8.70%</i>	<i>58.33%</i>