



HAL
open science

Phraséologie du roman médiéval : une approche sur corpus outillée

Julie Sorba, Corinne Denoyelle

► **To cite this version:**

Julie Sorba, Corinne Denoyelle. Phraséologie du roman médiéval : une approche sur corpus outillée. Global Challenges Science Week. International interdisciplinary days of Grenoble Alpes, Jun 2019, Grenoble, France. . hal-02147699

HAL Id: hal-02147699

<https://hal.science/hal-02147699>

Submitted on 4 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Phraséologie du roman médiéval : une approche sur corpus outillée



Projets financés & porteuses

- Phraséologie & genres textuels : le cas du roman médiéval (Démarre SHS ! & CORLI ILF CNRS 2018)
Ce travail bénéficie d'une aide de l'État gérée par l'Agence Nationale de la Recherche au titre du programme « Investissements d'avenir » portant la référence ANR-15-IDEX-02.
- Phrases formulaires & formules rituelles en ancien et moyen français (IDEX UGA IRS 2018-2019)
- Porteuses des projets : Julie Sorba (LIDILEM, UGA) & Corinne Denoyelle (Litt&Arts UMR 5316, UGA)

Corpus (à ce jour)

	œuvres	Nb mots
Corpus romans arthuriens	<i>Lancelot T1-2</i>	228 934
	<i>Tristan T2-3 (Curtis); T1 (Ménard)</i>	300 728
	<i>Artus de Bretagne</i>	184 648
Corpus de contraste	<i>Roman de Silence</i>	46 699
	<i>Roman de Guillaume de Dôle</i>	32 255
Total		793 544

Méthodologie

- Opérations en cours :
 - Lemmatiser les textes avec LGerM (ATILF, U. de Lorraine)
 - Étiqueter les textes avec la version de TreeTagger développée pour le français médiéval (BFM, IRHIM, ENS-Lyon)
 - Entraîner le parseur (LIDILEM, UGA)
 - Intégrer les textes annotés dans le Lexicoscope (LIDILEM, UGA)
- Le Lexicoscope (Kraif 2016) :
 - un outil de fouille de corpus syntaxiquement arborés (analyseur XIP, Aït Mokhtar, Chanod J.-P. & Roux 2001)

Résultats 1

DENOYELLE C. & SORBA J. (2018). « Le Lexicoscope : première expérimentation sur les formules de congé ». Journées d'étude *Nouvelles perspectives autour de la phraséologie de la langue littéraire médiévale*. Grenoble (30 novembre 2018).

	Lancelot	Tristan	Artus	Silence	G. de Dôle
Prendre c.	28	11	81	9	6
- p. simple	- 5	- 4	- 58	- 0	- 0
- présent	- 12	- 2	- 0	- 5	- 1
Demander c.	3	2	4	0	
Donner c.	0	6	12	0	
Avoir c.	3	1	2	0	
Par son c.	8	1	1	1	1
Sans son c	9	6	2		
Autre				1	
Total	50	27	106	11	7



Références bibliographiques

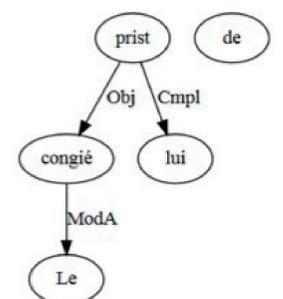
- Aït Mokhtar S., CHANOD J.-P. & ROUX C. (2001). Robustness beyond Shallowness: Incremental Deep Parsing. *Natural Language Engineering*, 8, 121-144.
- KRAIF O. (2016). Le lexicoscope : un outil d'extraction des séquences phraséologiques basé sur des corpus arborés. *Cahiers de lexicologie*, 108, 91-106.
- LEGALLOIS D. & TUTIN A. (2013). Présentation : vers une extension du domaine de la phraséologie. *Langages*, 189, 3-25.
- PRÉVOST S. & STEIN A. (2013). *Syntactic Reference Corpus of Medieval French (SRCMF)* [version number]. ENS de Lyon/ILR Stuttgart.
- SIEPMANN D. (2015). A corpus-based investigation into key words and key patterns in post-war fiction. *Functions of language*, 22(3), 362-399.
- SIEPMANN D. (2016). Lexicologie et phraséologie du roman contemporain : quelques pistes pour le français et l'anglais. *Cahiers de lexicologie*, 108, 21-41.

Objectif & intérêt scientifique

- Un objectif : étudier la phraséologie des romans en prose du XIII^e s.
- Intérêt scientifique : définir les genres textuels sur le plan linguistique par la phraséologie (*phraséologie* : ensemble des séquences lexicales perçues comme préconstruites, Legallois & Tutin 2013)
- Hypothèse : la langue littéraire se caractérise par la surreprésentation statistiquement significative de phraséologismes (lexèmes, collocations etc., voir Siepmann 2015 & 2016)

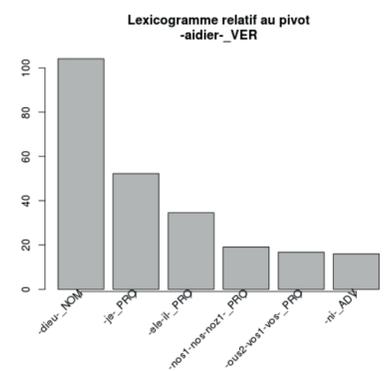
- Arbre Lexicosyntaxique Récurrent (ALR) <prendre_congié>

le congié prist de lui si l'ot.
Governal vint à son ostel...
(*Roman de Tristan*, Beroul)



- Analyse statistique des cooccurrents (calcul de spécificité du *log-likelihood ratio*)

- Ex du lexicogramme *aidier* (*Syntactic Reference Corpus of Medieval French*)



Résultats 2

DENOYELLE C. & SORBA J. (2019). « La phraséologie appliquée aux dialogues ». Colloque *Diachro IX*. Salamanque (28-30 mars 2019).

- La phraséologie pour contraster récit / dialogue :
 - Emploi spécifique marqueurs discursifs *si m'aïst Dieus / Dieus doint* (dialogue)
 - Unités phraséologiques privilégiées récit vs dialogue (*donner DET coup ; donner conseil*)
- La phraséologie pour contraster les genres littéraires :
 - *Conquête de Constantinople* se distingue avec des unités phraséologiques particulières (*aider à V inf*)
 - Roman en vers *Yvain* plus proche des romans en prose *Lancelot/Tristan* que la *Queste* (roman en prose)