



Phraséologie & genres textuels : le cas du roman médiéval

Julie Sorba (UGA, LIDILEM)

Journée de retours sur les projets soutenus par Démarre SHS! en 2018
10 décembre 2018

Le projet



- Un objectif : étudier la phraséologie des romans en prose du XIII^e s.
- Une technique : la fouille de corpus syntaxiquement arborés en vue d'extraire des séquences spécifiques.
- 2 laboratoires :
 - LIDILEM (Julie Sorba)
 - Litt&Arts UMR 5316 (Corinne Denoyelle)



Le projet



- Un objectif : étudier la phraséologie des romans en prose du XIII^e s.
- Intérêt scientifique : définir les genres textuels sur le plan linguistique (phraséologie).
- Phraséologie : ensemble des séquences lexicales perçues comme préconstruites.

Le projet



- Une technique : la fouille de corpus syntaxiquement arborés en vue d'extraire des séquences spécifiques.
- Intérêt : résultats moins bruités et mieux structurés.
- Outil lexicoscope (Kraif 2016) : concordances et cooccurrences

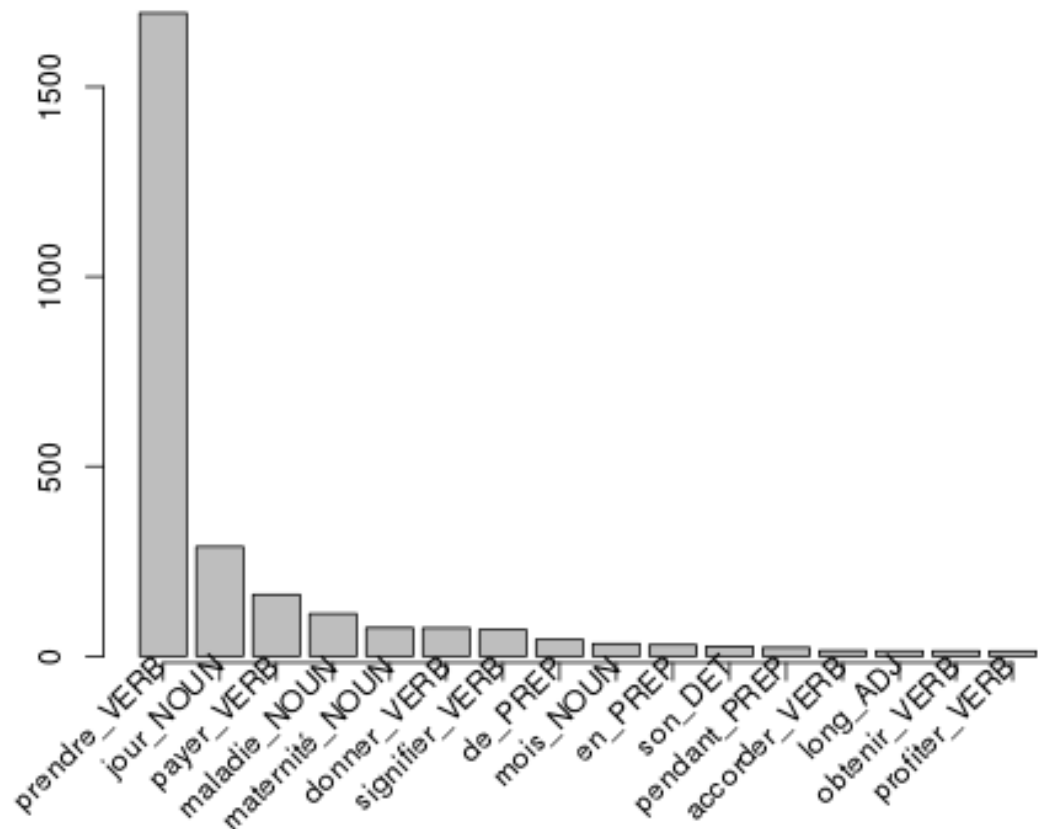
<http://phraseotext.univ-grenoble-alpes.fr/lexicoscope/>

Méthodologie



Lexicogramme relatif au pivot
congé_NOUN

- Requête *congé*
sur corpus
romanesque
français post
1950
(34 millions de
mots, 444 textes)



Cooccurents



Méthodologie



Lexicogrammes : Tableau de contingence et mesures d'association

Lexicogramme Graphiques

Copy CSV Excel PDF Print

Show 25 entries

Search: am.log.likelihood r.log.likeli

I1	I2	f.deprels	f	f1	f2	N	f.disp	am.log.likelihood	r.log.likeli
congé_NOUN	prendre_VERB	~OBJ ~SUBJ ~VMOD_POSIT1 ~VMOD_POSIT2	319	2737	161134	49064026	89	1694,3833	1
congé_NOUN	jour_NOUN	~NMOD_POSIT1 ~COORDITEMS NMOD_POSIT1	78	2737	87302	49064026	47	288,4581	2
congé_NOUN	payer_VERB	NMOD_POSIT1	32	2737	17305	49064026	23	162,4040	3
congé_NOUN	maladie_NOUN	NMOD NMOD_POSIT1	18	2737	5413	49064026	11	111,9385	4
congé_NOUN	maternité_NOUN	NMOD NMOD_POSIT1	8	2737	497	49064026	5	74,8452	5
congé_NOUN	donner_VERB	~OBJ ~SUBJ	37	2737	105647	49064026	25	74,0955	6
congé_NOUN	signifier_VERB	~OBJ ~SUBJ	13	2737	5977	49064026	10	69,9981	7
congé_NOUN	de_PREP	PREPOBJ	155	2737	1569220	49064026	65	43,9311	8
congé_NOUN	mois_NOUN	~NMOD_POSIT1 NMOD_POSIT1	12	2737	22611	49064026	10	32,6356	9
congé_NOUN	en_PREP	PREPOBJ	34	2737	201053	49064026	21	30,0405	10
congé_NOUN	son_DET	DETERM_POSS	47	2737	366436	49064026	32	25,4084	11
congé_NOUN	pendant_PREP	PREPOBJ	7	2737	9770	49064026	7	22,8495	12
congé_NOUN	accorder_VERB	~OBJ ~SUBJ	5	2737	8565	49064026	5	14,4456	13
congé_NOUN	long_ADJ	NMOD_POSIT1	6	2737	14212	49064026	5	13,8847	14
congé_NOUN	obtenir_VERB	~OBJ ~VMOD_POSIT1	5	2737	9124	49064026	5	13,8755	15
congé_NOUN	profiter_VERB	~VMOD_POSIT1	5	2737	9818	49064026	5	13,2195	16

Showing 1 to 16 of 16 entries

First Previous 1 Next Last

- I1 pivot
- I2 collocatif
- Indice statistique de spécificité (LLR)

Cooccurents





Méthodologie

FERMER - FR EN

Lexicoscope

Concordances

Afficher la liste des corpus XML

Nombre total d'occurrences 37
Dispersion 25

Requêtes

- Requête : <l=congé,c=NOUN,#1> && <l=donner,c=VERB,#co> :: (*,1,co)

Copy CSV Excel PDF Print

Show 10 entries

Search:

Identifiant	Contexte gauche	Pivot	Contexte droit
s10614	Une fois ses différents messages délivrés, M. Bolelli nous donna	congé	, en recommandant à Octave de suivre la voie que lui avait tracée son père, « dell'illustrissima Sorbonna ».
s11339	Elle lui apprend qu'elle a donné	congé	à l'hôtel où elle habite d'habitude pendant l'été, que ça faisait trop d'endroits.
s1273	Zénon lui donna	congé	pour la journée après lui avoir recommandé d'apporter d'abord à boire.
s14431	Mac Leod ne leur donnait pas	congé	.
s1452	Voilà tout ce que je pus tirer de ce méchant homme, qui s'enfuit sans me saluer quand le gouverneur lui donna	congé	.n
s16314	Il donna son	congé	à Elisabeth avant de verrouiller la porte vitrée, de baisser électriquement le rideau de fer puis de marcher, voûté contre le vent assez violent ce jour-là, jusqu'au métro Saint-Lazare.
s1695	Quelques semaines avant la Saint-Jean d'été, il se fit donner	congé	de remonter vers le Nord, pour observer par soi-même les effets du jour polaire.
s17945	Mon père avait certainement senti la rupture de ton, qui lui donnait	congé	, mais ne bougeait pas.
s2081	« J'ai dû leur donner	congé	.
s2124	Elle a donné son	congé	dans les premiers jours de mars, disant que ses amis lui avaient procuré du travail et l'hébergeraient le temps de se trouver un logis.

Showing 1 to 10 of 37 entries

First Previous 1 2 3 4 Next Last

Les concordances

Étape 1



- Constitution du corpus (inédit)
 - Partenariat avec ATILF (UMR 7118, U. Lorraine)
 - Partenariat avec IRHIM (UMR 5317, ENS-Lyon)
 - Numérisation et vérification des textes

	Déjà traités	En cours
Corpus romans prose 13 ^e s. arthurien	Tristan T1-2-3, Lancelot T1, Artus, Merlin	Lancelot T2-4-7-8
Corpus de contraste	Roman de Silence, Roman de Guillaume de Dôle	Couronnement de Louis, Erec et Enide, Roman de la Rose, Roman de Renart etc.
Nb mots	793 544	...

Merci à nos stagiaires Joëlle, Julia & Farida !

Étape 2



- Lemmatiser les textes : lemmatiseur LGerM (Gilles Souvay, ATILF, U. de Lorraine)



- Étiqueter les textes : étiqueteur TT (Alexey Lavrentev, IRHIM, ENS Lyon)



Étape 3



- Constitution de la treebank et entraînement du parseur

Sylvain Hatier & Olivier Kraif (LIDILEM, UGA)

Financier du contrat de l'ingénieur de recherche : ILF CNRS CORLI.

Gestion du contrat : ATILF, U. de Lorraine.

Durée : octobre-décembre 2018, 60%



Deux journées d'étude



- PhraseoMédiéval : Nouvelles perspectives autour de la phraséologie de la langue littéraire médiévale 29-30/11/2018 (MSH Alpes)
 - 1 conférencière invitée (U. Complutenses Madrid)
 - 7 communications (équipe, partenaires et appel ouvert)

<https://phraseomedieval.sciencesconf.org/>

- Autres financeurs : UGA INP & LIDILEM





Premiers résultats Lexicoscope

- Corpus d'entraînement (corpus srcmf.org)

Prévost, Sophie; Stein, Achim. 2013. Syntactic Reference Corpus of Medieval French (SRCMF) [version number]. ENS de Lyon/ILR Stuttgart.

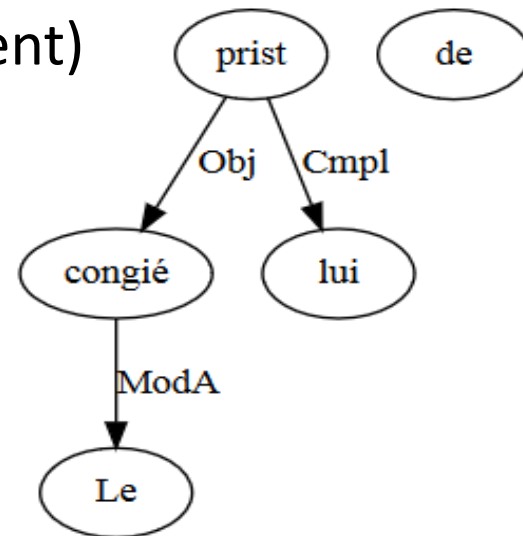
ALR (Arbre Lexicosyntaxique Récurrent)

le congié prist de lui

si l'ot Governal vint

à son ostel...

(Roman de Tristan, Beroul)

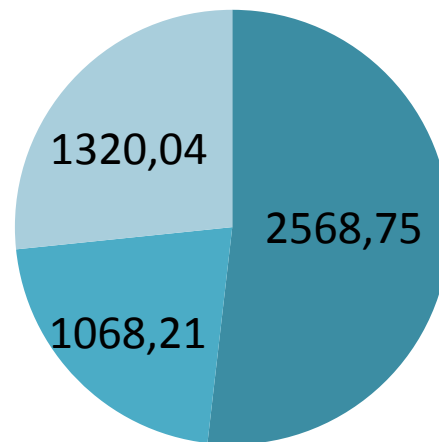


Ventilation budgétaire



Budget Démarre SHS!

■ Stagiaires ■ Missions ■ Equipement



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.

Et après ?



- Continuer le travail : Projet *Phrases formulaires et formules rituelles en ancien et moyen français* (Corinne Denoyelle, IRS 2018-2019)
- Rendre le corpus accessible à tous (plateforme Ortolang).
- Présenter notre projet à l'international :
 - Colloque *Diachro IX Le français en diachronie* (U. Salamanque, 28-30/03/2019)
 - Colloque *Viellissement des faits de langue, innovation et changement linguistique* (U. Sherbrooke, 19-21/06/2019)