

Constructions lexico-syntaxiques spécifiques dans le roman policier

Olivier KRAIF, Iva NOVAKOVA, Julie SORBA
(Université Grenoble Alpes - LIDILEM)

olivier.kraif@univ-grenoble-alpes.fr
iva.novakova@univ-grenoble-alpes.fr
Julie.sorba@univ-grenoble-alpes.fr

**Projet PHRASEOTEXT
AGIR POLE 2015-2016**

**Liste des
participants**

- **LIDILEM**
 - Francis Grossmann
 - Olivier Kraif
 - Iva Novakova (responsable)
 - Julie Sorba

- **LITT&ARTS**
 - Louis Autin
 - Isabelle Cogitore

Introduction : hypothèse

- Le langage littéraire se caractérise par la surreprésentation de lexèmes, collocations, patrons valenciels et colligations spécifiques (Siepmann 2015):
ex. : *Il en était là de ses réflexions quand...*
→ PRONOM + *en* + être + *là* + *de* + PRONOM POSS. + NOM
>> un processus cognitif + *quand/lorsque*
- Hypothèse : Certaines constructions lexico-syntaxiques (CLS) récurrentes permettent d'identifier des structures lexico-grammaticales prototypiques d'un certain (sous-)genre textuel. Ces structures peuvent, dans certains cas, s'apparenter à des *motifs* de mots co-occurents:
 - *Un ensemble d'éléments fixes et variables susceptibles d'accompagner la structuration textuelle, et simultanément, de caractériser des textes de genres divers* (Longrée & Mellet 2013 : 66).
 - *Une unité multidimensionnelle, constituée à la fois d'associations lexicales et grammaticales, d'appariements entre forme et sens ou fonction pragmatique/discursive* (Legallois 2012 : 45).

Introduction : cadre théorique

- Le motif: une construction lexico-grammaticale récurrente qui a la fonction de marqueur discursif structurant (Mellet & Longrée, 2012: 718)
- Modèles fonctionnels et contextualistes (Sinclair 1991, Biber 2009)
- 4 niveaux d'analyse des unités linguistiques: lexical, sémantique, syntaxique et discursif dans la perspective de l'organisation du texte littéraire (Sinclair 2004).
- Approche *corpus-driven*: observations guidées par les données
 - les outils numériques d'exploration du corpus permettent de faire apparaître des régularités et des phénomènes attendus mais parfois aussi inattendus à partir desquelles on peut définir ou caractériser les genres littéraires.

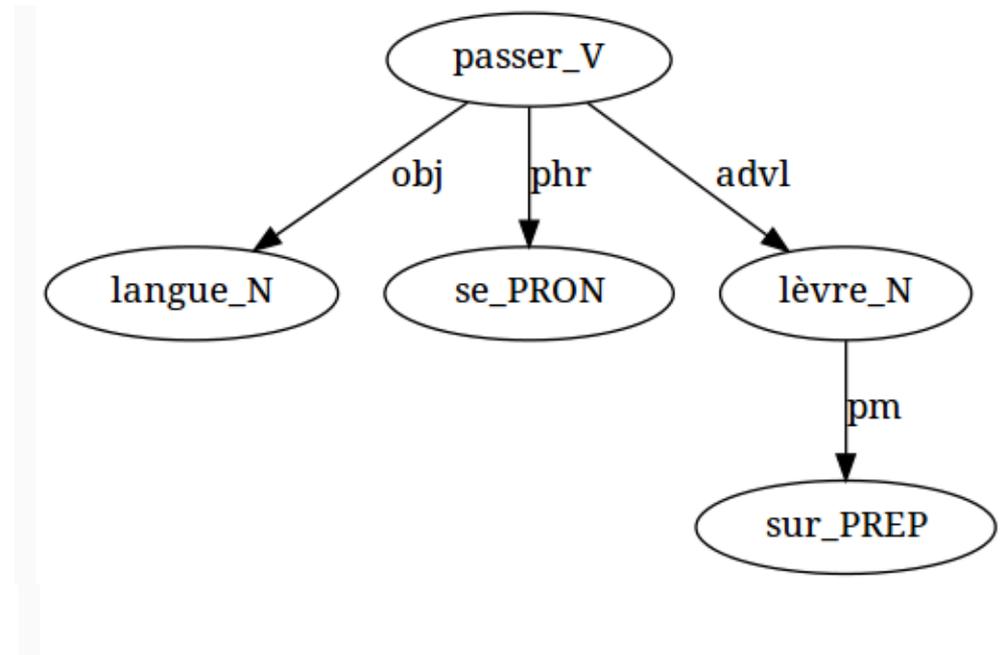
Introduction : corpus

Présentation des données	
Corpus LIT_MOD	16 M de mots
Sous-corpus POL	4 M de mots
Nombre de CLS	3185
TEXTES	35
AUTEURS	Aubert, Bannel, Chattam, Dard, Grangé, Simenon, Vargas.

- Corpus LIT_MOD partitionné en sous-genres : roman policier, science-fiction, roman historique, roman sentimental, autres.

Méthodologie - 1

- Outils : Lexicoscope (Kraif & Diwersy, 2012)
- Extraction automatique des CLS spécifiques au sous-corpus POL sous la forme d'Arbres Lexico-syntaxiques Récurrents (ALR) :



Méthodologie - 2

- Répartition de l'expression *se passer la langue sur les lèvres* :
POL : 12 occ., HIST : 2 occ., SF : 1 occ., AUTRE : 1 occ., SENT 0 occ.
- >> Il semble qu'elle soit spécifique à POL. Comment le mesurer ?
- Calcul de *spécificité* des CLS : fréquences relatives dans un sous-corpus par rapport aux fréquences dans l'ensemble du corpus. Tableau de contingence:
 - f1 : la fréquence dans le sous-corpus POL
 - f2 : la fréquence dans tout le corpus LIT_MOD
 - T1 : le nombre total de mots du sous-corpus POL
 - T2 : le nombre total de mots de tout le corpus LIT_MOD
 - > Calcul du *loglikelihood ratio* (rapport de vraisemblance)
 - > Seuil de spécificité ≥ 10.83 , dispersion ≥ 3
- Extraction d'ALR *émergents* (cf. motifs séquentiels émergents, Quiniou et al. 2012)

Méthodologie - 3

Sous-arbres	f1	f2	Disp	Loglike
<l=crime, c=N, #1>&&<l=scène, c=N, #2>&&<l=de, c=PREP, #3>:: (mod, 2, 1) (pm, 2, 3)	161	165	3	408,05
<l=un, c=DET, #1>&&<l=regard, c=N, #2>&&<l=lancer, c=V, #3>:: (det, 2, 1) (obj, 3, 2)	121	158	6	182,75
<l=un, c=DET, #1>&&<l=coup, c=N, #2>&&<l=jeter, c=V, #3>:: (det, 2, 1) (obj, 3, 2)	207	371	6	156,13
<l=champ, c=N, #1>&&<l=vision, c=N, #2>&&<l=de, c=PREP, #3>&&<l=(?:son leur mon ton notre votre), c=PRON, #4>:: (det, 1, 4) (pm, 1, 3) (mod, 1, 2)	23	7	4	118,06
<l=mot, c=N, #1>&&<l=sans, c=PREP, #2>&&<l=(?:un une des d de), c=DET, #3>:: (det, 1, 3) (pm, 1, 2)	142	248	6	114,15
<l=jeune, c=A, #1>&&<l=inspecteur, c=N, #2>:: (attr, 2, 1)	42	47	4	86,72
<l=quai, c=N, #1>&&<l=orfèvre, c=N, #2>:: (mod, 1, 2)	38	47	5	63,98
<l=tasse, c=N, #1>&&<l=café, c=N, #2>:: (mod, 1, 2)	65	112	5	53,98
<l=crâne, c=N, #1>&&<l=son, c=PRON, #2>:: (det, 1, 2)	135	304	6	52,45
<l=vitesse, c=N, #1>&&<l=à, c=PREP, #2>&&<l=tout, c=PRON, #3>:: (det, 1, 3) (pm, 1, 2)	78	151	6	48,1
<l=meurtre, c=N, #1>&&<l=série, c=N, #2>:: (mod, 2, 1)	22	24	3	47,98
<l=cadrer, c=V, #1>&&<l=pas, c=ADV, #2>&&<l=ne, c=ADV, #3>:: (neg, 1, 2) (ad, 2, 3)	25	30	3	44,72
<l=mise, c=N, #1>&&<l=scène, c=N, #2>:: (mod, 1, 2)	70	135	6	43,61
<l=chance, c=N, #1>&&<l=tomber, c=V, #2>&&<l=de, c=PREP, #3>:: (pm, 1, 3) (mod, 1, 2)	10	5	3	41,4
<l=point, c=N, #1>&&<l=suspension, c=N, #2>&&<l=de, c=PREP, #3>:: (pm, 1, 3)) (mod, 1, 2)	22	30	3	30,44
<l=instant, c=N, #1>&&<l=à, c=PREP, #2>&&<l=même, c=A, #3>:: (attr, 1, 3) (pm, 1, 2)	42	77	5	29,91

Récurrance thématique : *Scène de crime*

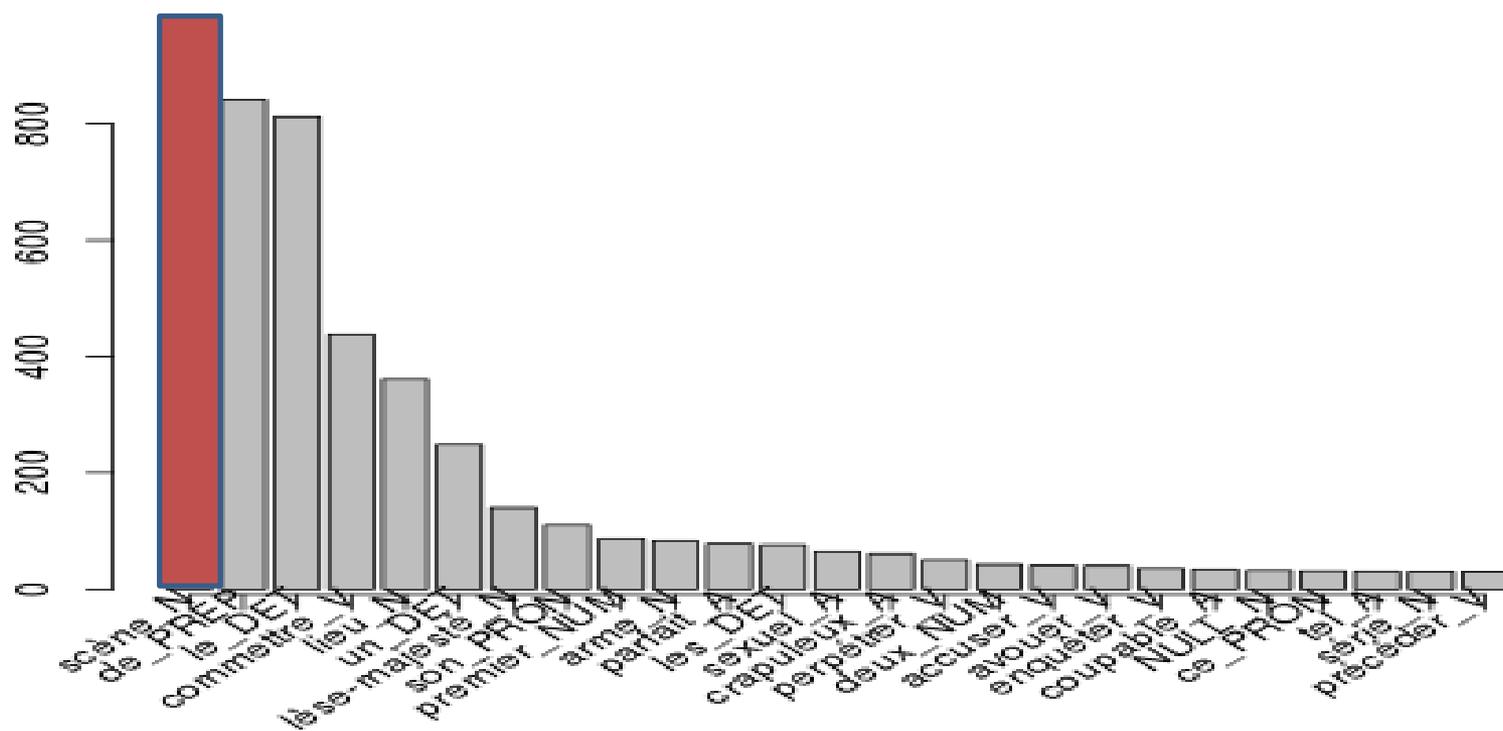
- **Collocation** la plus spécifique pour le POL fr (log 408) : affinité de *crime* pour le pivot *scène* et inversement (en revanche, *meurtre* se combine préférentiellement avec *série* dans *série de meurtres* et manifeste donc plutôt une aversion pour le collocatif *scène* (3 occ., cf. Hoey, 2005),

Sous-corpus	Nb d'occurrences
POL	116
AUT	2
SF	1

(1) **Sur la scène de crime**, Langleber me fait signe de venir. (Grangé, POL)

Récurrance thématique : *Scène de crime*

Lexicogramme relatif au pivot
crime_N



Récurrance thématique : *Scène de crime*

- **Collocation** (Sinclair 2004) : cooccurrence statistiquement significative d'items lexicaux.
- **Collocation grammaticale** (Bolly 2010) : combinaison lexicalement contrainte constituée d'un mot lexical qui sélectionne un mot grammatical (par ex., une préposition) :
 - > la plus spécifique pour *scène de crime* (log 179) :
Prép. **sur** + Dét. (*la, les*) + *scène de crime* (SC),
- Extension syntagmatique à gauche (préférence pour les V de mouvement) :
Vmouv (*arriver*) + Prép (*sur*) + Dét (*la*) + *scène de crime*
(2) C'était la première chose à faire **en arrivant sur une scène de crime**.
(Chattam, POL)
 - > **colligation** [Vmouv+CC] : préférence des lexies pour certains environnements grammaticaux et, de là, pour certaines fonctions grammaticales (Hoey 2005).

Récurrance thématique : *Scène de crime*

- Variations paradigmatiques du motif de la localisation :
 - Variations sur la PREP : *sur, autour de, (tout) près de, au seuil de, sur le seuil de, dans les environs de, à proximité de* + Dét. + SC
 - (3) Du temps où il était inspecteur, Brodin demandait régulièrement qu'on filme la foule présente **autour d'une scène de crime**. (Chattam, POL)
 - Variations sur le Vmouv : *aller, se rapprocher, partir, arriver, revenir*
 - (4) Il attrapa sa veste et **partit** à fond **vers la scène de crime**. (Grangé, POL)

=> motifs thématiques de la localisation

(variations syntagmatique et paradigmatique des constituants)

Collocation grammaticale (Prep+la SC) -> motif 'localisation statique'

Colligation (Vmouv + *sur la SC*) -> motif 'localisation dynamique'

Récurrance thématique : *Scène de crime*

- Vers une interprétation stylistique ?
 - >> Glissement du motif de la 'localisation dynamique' au motif de la 'localisation statique' > glissement de la narration à la description (Adam 2005).
 - Introduire l'élément initial de la diégèse (narration)
 - Créer un support de la remémoration (description)

Récurrance stylistique : *Lancer un regard*

(résultat inattendu)

- Patron lexico-syntaxique 1 : *Lancer un regard* + ADJ (*circulaire, incertain autour de lui, interrogatif vers qn*).
- Patron lexico-syntaxique 2 : *Lancer un regard* ADJ à **N** (expansion syntagmatique), présence d'un 3^e actant sous forme de N :

(5) Il lance à Oscar Avane un long regard d'épagueul pour le remercier de son pieux mensonge. (Dard, POL)

> CLS fort spécifique au sous-corpus POL comparé au corpus LIT_MOD (loglike 182)

- Autres récurrances stylistiques : *trembler de tous ses membres, se passer la langue sur les lèvres, allumer sa cigarette, caresser sa barbe...*

→ Motif des 'menus comportements'

Fonction(s) textuelle(s) des récurrences stylistiques

- Installer le lecteur dans un horizon de sens familier,
- Ancrer le récit dans une constellation générique de significations (cf. Dufays 2010, 230-232) /

→ les CLS servent de cadrage stéréotypique du dialogue (expressions désignant des comportements parfaitement interchangeables), par ex. :

1) Comportement : *X se recula sur sa chaise et alluma une cigarette / X se recula sur sa chaise et caressa sa barbe / ...:*

2) discours direct : - *Et vous savez pourquoi ?*

3) Comportement : *Y haussa les épaules / inclina la tête sur les côtés et dit avec effort / murmura ...*

Réurrence stylistique : *Ses lèvres trop rouges*

- Patron lexico-syntaxique : N + *trop* + ADJ couleur (expansion paradigmatique) :
 - POL : N = noms de parties du visages ou du corps (*peau, yeux, bouche, lèvre, paupière, cheveux, dents, joues, mains...*) (spécifique au POL: 11/17 occ.)

(6) Fascinant, dit-il, en léchant **ses lèvres trop rouges**. (Aubert, POL)
 - SF : N= éléments ou objets naturels (6/6 occ.) :

(7) Ce paysage trop calme et trop accueillant suintait l'angoisse par **ses pierres trop blanches, ses feuilles trop vertes, son ciel trop bleu** (Jeury, SF).

→ Motif stylistique de l'excès

Conclusion

- L'approche **quantitative** (statistique) et la méthodologie d'extraction des ALR a permis de faire émerger des :
- Récurrences thématiques (plutôt attendues) :
 - *Sur la scène de crime, Vmouv sur la scène du crime*
- Récurrences stylistiques (plutôt inattendues) :
 - *Lancer un regard*
 - *Ses lèvres trop rouges*
- Affinité de certaines CLS pour certains (sous-)genres.
- Clarifier le statut des unités phraséologiques : le motif comme attraction mutuelle entre des termes, récurrence de la cooccurrence ainsi formée mesurée par des outils lexicométriques, variations syntagmatiques et paradigmatisées au sein de cette séquence qui a également une fonction discursive, en l'occurrence de caractérisation d'un sous-genre textuel.
 - >> des unités structurantes pour la construction du sens (par ex. motifs de la localisation, de l'excès).

Conclusion et perspectives

- L'approche qualitative (analyse **fonctionnelle**) > analyser de manière fine les CLS statistiquement spécifiques sur le plan *sémantique* (sous-classes sémantiques de CLS, lieu, personnages, comportements), *syntactique* (colligations) et *discursif* (fonctions textuelles)
- En perspective, étudier le rôle des motifs, en tant qu' « unités discursives observables et dénombrables » dans l'organisation textuelle des romans (dimension réticulaire; structures phrastiques, relations inter-phastriques, dynamiques temporelles, séquences narratives et descriptives) (Longrée & Mellet, 2013)
- Grâce au repérage automatique des « colligations textuelles » (Hoey, 2005), analyser les préférences ou les aversions des CLS pour des positions / fonctions syntaxiques spécifiques dans la phrase et au-delà de la phrase (paragraphe), utile aussi pour les études stylistiques.
- Approche globale pour appréhender la grammaire des genres ?

→ Projet franco-allemand ANR DFG 2016-2020 PHRASEOROM

Bibliographie

- Adam, J.-M.** (2005), *Les Textes: types et prototypes. Récit, description, argumentation et dialogue*. Paris: Armand Colin.
- Biber, D.** (2009). A corpus-driven approach to formulaic language in English. Multiword patterns in speech and writing. *International Journal of Corpus Linguistics* 14:3, 245-311, J. Benjamins
- Bolly, C.** (2010). Flou phraséologique, quasi-grammaticalisation et pseudo-marqueurs de discours, un *no man's land* entre syntaxe et discours. *Linx*, n° 62-63, p, 11-38,
- Dufays, J.-L.** (2010), *Stéréotype et lecture*. Bruxelles: Peter Lang.
- Hoey, M.** (2005). *Lexical priming: a new theory of words and language*, London/New York: Routledge.
- Kraif, O. & Diwersy, S.** (2012). Le Lexicoscope: un outil pour l'étude de profils combinatoires et l'extraction de constructions lexico-syntaxiques », Antoniadis G., Blanchon H. Sérasset G. (dir) Actes de la conférence JEP-TALN-RECITAL 2012, vol.2: TALN, Grenoble, ATALA&AFCP, 399- 406
- Kraif, O., Novakova I., Siepmann D.** (2015) Unités polylexicales et construction du sens dans le roman policier français et anglais, Colloque international de Linguistique, GRéG IV, Nanterre, 6-7 novembre 2015
- Legallois, D.,** (2012). La colligation : autre nom de la collocation grammaticale ou autre logique de la relation mutuelle entre syntaxe et sémantique?, *Corpus n°11*, p. 31-54.
- Longrée, D. & Mellet, S.** (2013). Le motif : une unité phraséologique englobante? Étendre le champ de la phraséologie de la langue au discours, *Langages n°189* : 68-80.
- Mellet S. & Longrée D.** (2012) Légitimité d'une unité phraséologique englobante? : le motif », A. Dister, D. Longrée; G. Purnelle (dir) Actes des JADT 2012, Liège, p, 715-728,
- Novakova I., Sorba J., Kraif O.,** Expressions polylexicales spécifiques dans deux sous-genres littéraires: le roman policier et la science fiction, Colloque international GRATO (Grammaire & Texte), Lisbonne 2-4 juillet 2015
- Quiniou, S., Cellier, P., Charnois, T. & Legallois, D.** (2012). « Fouille de données pour la stylistique : cas des motifs séquentiels émergents ». In *Actes des 11^{es} Journées Internationales d'Analyse Statistique des Données Textuelles (JADT'12)*, Liège, p. 821-833
- Siepmann, D.** (2015), A corpus-based investigation into key words and key patterns in post-war fiction, *Functions of Language n°22*: 3, p, 362-399.
- Sinclair, J. M.** (1991) *Corpus, concordance, collocation*. Oxford: OUP,
- Sinclair, J. M.** (2004). *Trust the Text: Language, Corpus and Discourse*. London: Routledge.