

***Étude contrastive des expressions
d'émotions formant des motifs
textuels dans la littérature
contemporaine française et thaï***

Jirawan KIATPHOTHA – 1ère année
Sous la direction de Madame Iva NOVAKOVA

Séminaire DéLiCorTal, 14 février 2020

Plan de la présentation

- Origine du sujet de thèse
- Disciplines concernées
- Présentation des corpus
- Résultats escomptés

*Origine du
sujet de
thèse*

EMOLEX

ANR

DFG

LIDILEM



UNIVERSITÄT



STRASBURG

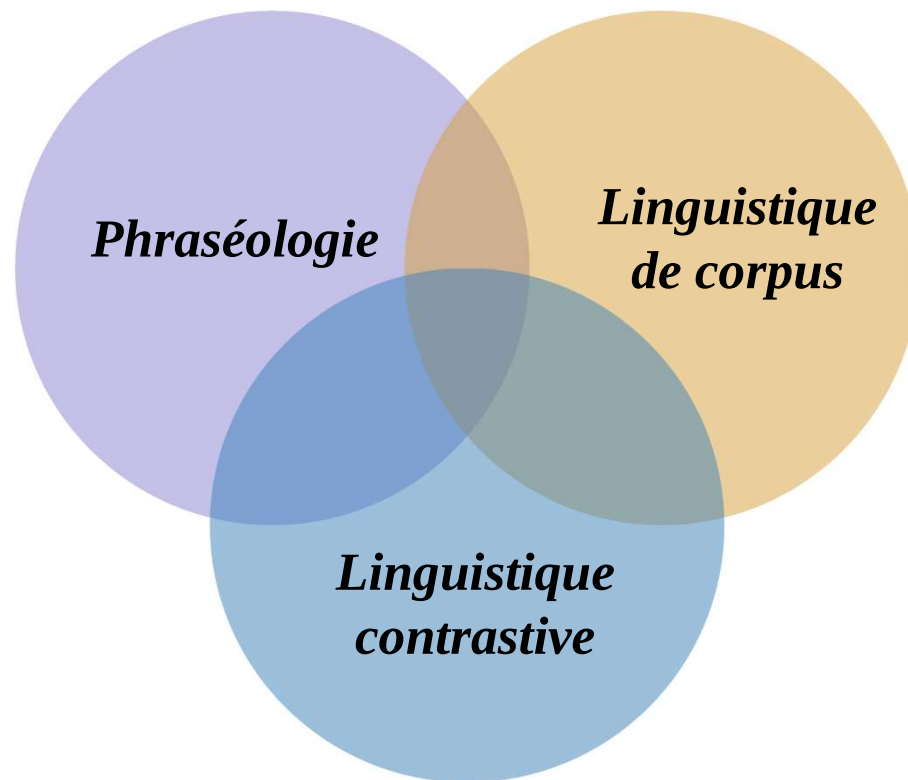
Le lexique des émotions dans 5 langues européennes.





*Contraster les expressions
d'émotions en français
et en thaï*

*Disciplines
concernées*



1

Phraséologie

Phraséologie

- Point de départ : *pour savoir si et comment la phraséologie étendue permet-elle d'étudier les genres et la langue littéraires.*

Phraséologie étendue
(Legallois et Tutin 2013)

Expressions polylexicales

Expressions regroupées en
motifs textuels

Constructions lexico-syntaxiques
récurrentes
« il gardait la nostalgie »

2

Linguistique de corpus

Une approche inductive *corpus-driven*
(Sinclair 2004, Hoey 2005, Biber 2009)

“

Le corpus est une collection de données langagières qui sont sélectionnées et regroupées selon les critères linguistiques explicites pour servir d'échantillon du langage.

Sinclair 2004

3

Linguistique contrastive

Le français et le thaï :
deux langues typologiquement différentes



Présentation des corpus

Corpus littéraire français

- ▶ PhraseoBase (Lexicoscope)
- ▶ 34 M de mots

The screenshot displays the Lexicoscope web interface. At the top, there is a navigation bar with a 'Déconnexion' button and a language selector set to 'FR EN'. The main title is 'Lexicoscope' with the subtitle '- Exploration des profils combinatoires -'. Below this, there are four tabs: 'Sélection du Corpus', 'Requête', 'Paramètres', 'Sessions sauvegardées', and 'Guide'. The 'Sélection du Corpus' tab is active, showing two radio buttons for 'Corpus monolingue' (selected) and 'Corpus parallèle'. A 'Langue' dropdown menu is set to 'fr'. A list of corpora is shown, with 'GEN : Corpus général de phraseorom' highlighted in blue. Below the list, a 'Description' section shows 'Corpus sélectionné : nombre de mots = 34 334 554 (444 textes)'. At the bottom, the footer reads '2013-2018 (cc) LIDILEM - Crédits'.

*Corpus
littéraire
thai*

8 M de mots

- *Thai National Corpus*



26 M de mots

- *Textes littéraires contemporains en ligne*





Domain	Imaginative	254	7330111	22.00 %	36.65 %
Domain	Informative: applied science	126	1100283	3.30 %	15.28 %
Domain	Informative: arts	72	1170177	3.51 %	16.25 %
Domain	Informative: belief & thought	33	692307	2.08 %	9.62 %
Domain	Informative: commerce & finance	378	637581	1.91 %	8.86 %
Domain	Informative: leisure	448	2542499	7.63 %	35.31 %
Domain	Informative: natural & pure science	142	1227040	3.68 %	17.04 %
Domain	Informative: social science	307	7691292	23.08 %	106.82 %
Domain	Informative: world affairs	47	214969	0.65 %	2.99 %
Domain	others	2430	10715737	32.16 %	0.00 %
Genre	Academic: Humanity arts	76	2564429	7.68 %	0.00 %
Genre	Academic: Medicine	0	0	0.00 %	0.00 %
Genre	Academic: Natural science	2	12094	0.04 %	0.00 %
Genre	Academic: Political science - law educati	38	1435075	4.30 %	0.00 %
Genre	Academic: Social science	136	4925516	14.75 %	0.00 %
Genre	Academic: Technology - computing engineer	15	477749	1.43 %	0.00 %
Genre	Advertisement	0	0	0.00 %	0.00 %
Genre	Biography-Expereinces	68	1700974	5.09 %	0.00 %
Genre	Blog	0	0	0.00 %	0.00 %
Genre	Commerce - finance economic	1	49919	0.15 %	0.00 %
Genre	Essay: University	480	368159	1.10 %	0.00 %
Genre	Essay: school	0	0	0.00 %	0.00 %
Genre	Fiction: Drama	1	4119	0.01 %	0.00 %
Genre	Fiction: Poetry	25	110123	0.33 %	0.00 %
Genre	Fiction: Prose	165	6224085	18.64 %	0.00 %
Genre	Fiction: Short Story	63	1097308	3.29 %	0.00 %
Genre	Institution documents	459	676961	2.03 %	0.00 %
Genre	Instructional diy	62	207769	0.62 %	0.00 %
Genre	InterviewQA	17	102098	0.31 %	0.00 %
Genre	Law regulation	562	1248221	3.74 %	0.00 %
Genre	Letter: Personal	4	41984	0.13 %	0.00 %
Genre	Letter: Professional	0	0	0.00 %	0.00 %
Genre	Miscellaneous	0	0	0.00 %	0.00 %
Genre	Newspaper: agriculture	84	115608	0.35 %	0.00 %
Genre	Newspaper: crimes	129	244989	0.73 %	0.00 %
Genre	Newspaper: economics	117	507409	1.52 %	0.00 %
Genre	Newspaper: editorial - views	444	2435521	7.29 %	0.00 %
Genre	Newspaper: education	83	187713	0.56 %	0.00 %



Textes littéraires thâi en ligne



ตายแดง
เรื่องราวยุคใหม่ที่เต็มไปด้วยความทุกข์ทรมานที่โหดร้าย...
[READ](#)



ให้หัวใจได้น้ำทาง
"ให้รักได้บ้าง" ... ใจของคนคนหนึ่ง...
[READ](#)



ขอบน้าจรดขอบฟ้า
"ขอบฟ้าที่ไกลออกไป... ใจที่รอคอย..."
[READ](#)



MESSY BUDDY คุณบิตเตอร์ ไข่เรอหรือเปล่า?
MESSY BUDDY คุณบิตเตอร์ ไข่เรอหรือเปล่า?...
[READ](#)



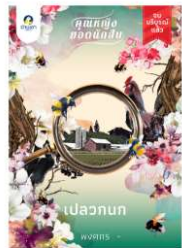
แผนร้ายเที่ยวรัก
แผนร้ายเที่ยวรัก... เรื่องราวความรัก...
[READ](#)



มนตราตะเกียงแก้ว
มนตราตะเกียงแก้ว... เรื่องราวของมนตรา...
[READ](#)



เพื่อนเดี่ยวเสียอีก
เพื่อนเดี่ยวเสียอีก... เรื่องราวของเพื่อน...
[READ](#)



เปลวทนก
เปลวทนก... เรื่องราวของเปลวทนก...
[READ](#)

File Home Insert Page Layout Formulas Data Review View Help Tell me what you want to do

Clipboard: Cut, Copy, Paste, Format Painter

Font: Calibri, 12, Bold, Italic, Underline, Text Color, Background Color

Alignment: Wrap Text, Merge & Center

Number: General, Percentage, Decimals

Conditional Formatting: Normal, Bad, Good, Neutral, Calculation, Check Cell

Cells: Insert, Delete, Format, AutoSum, Fill, Clear

Editing: Sort & Filter, Find & Select

A71: venu à leur rencontre

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	Example	Genres	Query	size	PivotN	PivotV	QueryN	QueryV	Motif	Pattern	Freq-FY	Disp-FY.fr	Loglike-FY.fr	Sent-FY.fr	CorpusSiz	Freq-t
69	porter à rencontre	FY HIST	<l=porter,c=	3	rencontre	porter	<l=porter,c=	<c=VERB,#1>	<c=VERB,#1> VERB PREP NC		27	11	20,86681005	Albert Raphaël - "Avant le	13323976	
70	va à sa rencontre	FY	<l=aller,c=VE	4	rencontre	aller	<l=aller,c=	<c=VERB,#1>	<c=VERB,#1> VERB PREP DE		22	14	11,68078917	Ange - "Les trois lunes de Tanjor 2 La	13323976	
71	venu à leur rencontre	FY	<l=leur,c=DE	4	rencontre	venir	<l=leur,c=	<l=leur,c=DE	<l=leur,c=DE VERB PREP DE		37	17	24,46090657	Boisset Eric - "La trilogie des	13323976	
72	venu à rencontre	FY	<l=rencontre	3	rencontre	venir	<c=NOUN, <l=rencontre	<c=NOUN,#1	VERB PREP NC		117	32	25,01727982	Albert Raphaël - "Avant le	13323976	
73	vint à sa rencontre	FY	<l=rencontre	4	rencontre	venir	<c=NOUN, <l=rencontre	<c=NOUN,#1	VERB PREP DE		55	23	18,54178771	Ange - "Les trois lunes de Tanjor 1 Le	13323976	
74	dit d' un ton	HIST	<l=de,c=PREI	4	ton	dire	<l=de,c=PF	<l=de,c=PREI	<l=de,c=PREI VERB PREP DE		9	5	-12,95640102		13323976	
75	ma mère ou	GEN	<l=avoir,c=VI	2	mère	avoir	<l=avoir,c=	<c=VERB,#1>	<c=VERB,#1> DET NOUN VE		15	7	22,76540126		13323976	

ALR-semantic-fr - Excel

File Home Insert Page Layout Formulas Data Review View Help Tell me what you want to do

Clipboard Font Alignment Number Styles Cells Editing

venu à leur rencontre

1	Example	Genres	Query-Dic	Catégorie sémantique	AND & OI	Catégorie	ID ALR	ID Group	Motif stylistique
69	porter à rencontre	FY HIST	=VERB,#1>&&<=renc ontre,c=NO	action:déplacement			6870	22	
70	va à sa rencontre	FY	e,c=NOUN, #1>&&<=à, c=PREP,#2>	action:déplacement			2675	22	
71	venu à leur rencontre	FY	e,c=NOUN, #1>&&<=v enir,c=VER	action:déplacement			5243	22	
72	venu à rencontre	FY	e,c=NOUN, #1>&&<=v enir,c=VER	action:déplacement			13654	22	
73	vint à sa rencontre	FY	e,c=NOUN, #1>&&<=v enir,c=VER	action:déplacement			6541	22	
74	dit d' un ton	HIST	OUN,#1>&&<=un,c=D ET,#2>&&<	communication:verbale	AND	qualia:affec t	3494	24	dire_voix
75	ma mère ou	CEN	NOUN,#1>&&<=mon, c=DET,#2>	autre			624	614	

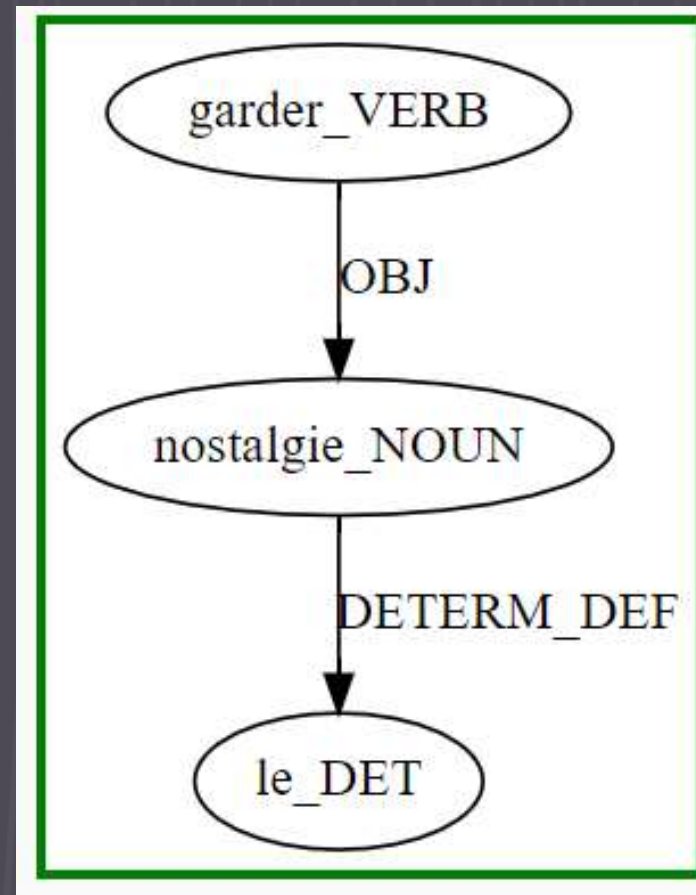
DimSem ALRs

17

*Annotation
des données*

Arbres lexico-syntaxiques
récurrentes (ALRs)

Garder la nostalgie



Annotation des données

Arbres lexico-syntaxiques
récurrentes (ALRs)

Corpus : littérature générale

1689 / 8415 ALRs

Annotés à l'aide d'une grille
d'analyse sémantique

Dimensions	Valeurs
action	indéfini
	mouvement
	déplacement
	autre
communication	indéfini
	verbale
	physique
état	indéfini
	animé
	inanimé
	abstrait
cognition	
qualia	indéfini
	sensation
	perception
	affect
temps	
lieu	
événement	
autre	ALR qui ne correspondent à aucune des catégories ci-dessus
-	bruit (ALR sans verbes, traces de copyright)

(Manuels de codage sémantique, PhraseoRom)

Bienvenue sur PhraseoBase



PhraseoBase est une base de données qui regroupe deux applications. Une première application donne accès, grâce au **lexicoScope**, au corpus de romans contemporains constitué dans le cadre du projet ANR-DGFi PhraseoForm (<https://pharostorm.univ-grenoble-alpes.fr/accueil>). La seconde application donne accès à un ensemble d'**analyses linguistiques et stylistiques** menées sur les données extraites de ce corpus.

Que trouverez-vous dans PhraseoBase ?



Un accès au **corpus** constitué dans le cadre du projet ANR-DGFi PhraseoForm. Il s'agit d'un corpus de romans contemporains français, anglais et allemands classés en six sous-genres : des romans policiers, sentimentaux, historiques, de science-fiction, de fantasy et des romans de littérature générale. Ce corpus est été enrichi linguistiquement : il s'agit de lemmatiser, étiqueter morphologiquement et syntaxiquement. L'interface du lexicoScope en permet une interrogation directe.



Un accès via une interface ergonomique aux **Arbres Lexico-syntaxiques** (RLS) qui portent sur tous genres romanesques contemporains et sur analyses linguistiques qui leur sont associées : données statistiques, lexicales, syntaxiques et codages sémantiques. Une seconde interface permet d'accéder aux mots stylistiques associés aux romans contemporains et à leur analyse discursive.



Un ensemble de **ressources** à télécharger : tableaux contenant les ARS et leur données linguistiques et stylistiques associées, guides détaillant les analyses sémantique et stylistique proposées, liste des étiquettes morpho-syntaxiques associées aux tokens du corpus, etc.

Accès au corpus

Ressources téléchargeables



Pour les chercheurs
Dans le domaine de la linguistique comme de la stylistique, les chercheurs trouveront en un accès à des corpus de grande taille composés de textes rigoureusement sélectionnés et enrichis linguistiquement, ainsi qu'à des données linguistiques et stylistiques riches, documentées et immédiatement exploitables.



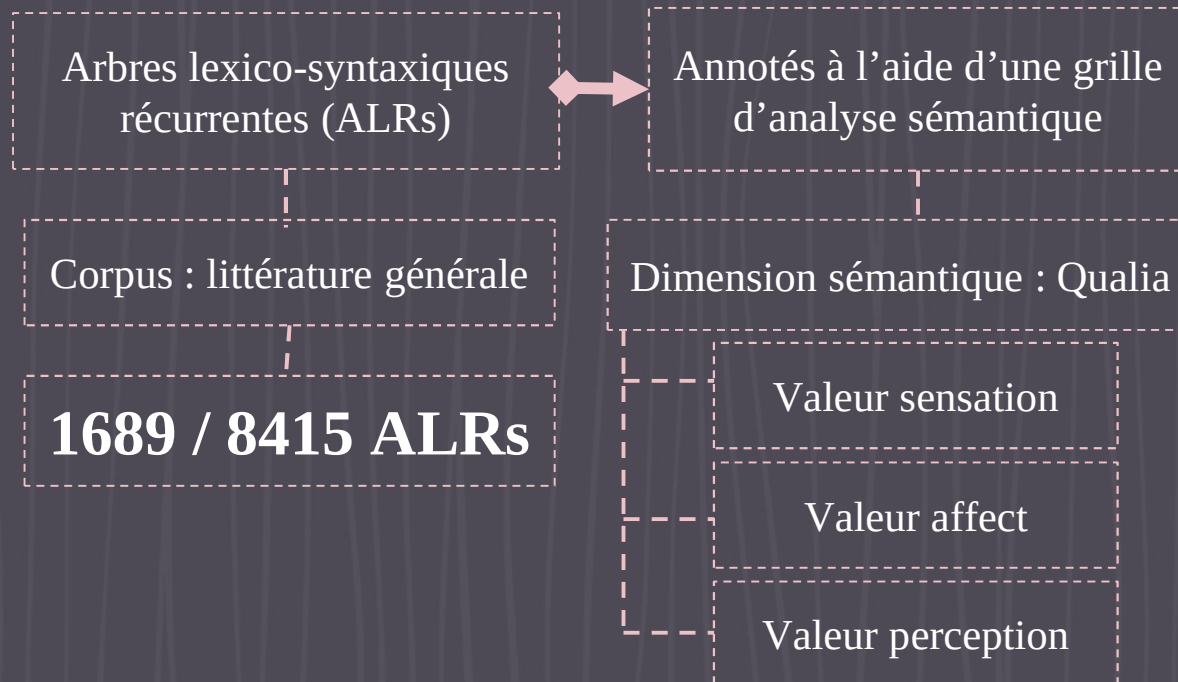
Pour les enseignants
Les données, analyses, propositions, permettent d'élaborer des parcours didactiques pour accéder aux textes romanesques. Le classement sémantique des ARS permet par exemple un accès chronologique aux données textuelles. De plus, dans le volet stylistique les exemples ont été sélectionnés un à un pour illustrer précisément les fonctions stylistiques des motifs textuels.



Pour les étudiants
Les étudiants trouveront dans la PhraseoBase à la fois des corpus, une interface d'interrogation et un ensemble de données brutes de ce corpus avec plusieurs niveaux d'analyse qui leur fournissent une base idéale pour se lancer dans l'analyse du corpus.

[Accéder au corpus PhraseoForm](#) [Accéder à PhraseoLang](#) [Ressources téléchargeables](#)

Annotation des données



*Dimension
sémantique
Qualia*

Les propriétés de la perception et généralement de l'expérience sensible

Sensation

Sensation physique par le biais des cinq sens

- ***J'ai le vertige***
- ***Le regard me donnait le vertige***
- ***Il avait les larmes dans les yeux (sensation et affect)***

Affect

Etats psychologiques généralement marqués positivement ou négativement, éprouvés par un sujet animé, liés à une source (cause ou objet), inscrits dans une dimension temporelle et ne pouvant être observés directement

- ***Sauter à la figure***
- ***Garder la nostalgie***
- ***Arracher les cheveux (affect ou action)***

Perception

Représentation mentale d'une sensation, organisation des données sensorielles. Activité par laquelle un sujet fait l'expérience d'objets ou de propriétés présents dans son environnement et qui repose sur des informations délivrées par les sens : perceptions auditive, visuelle, tactile, gustative et olfactive.

- ***Voir dans la glace***
- ***Je sens l'odeur***
- ***Voir les traces (perception et cognition)***

Annotation des données

Arbres lexico-syntaxiques
récurrentes (ALRs)

Corpus : littérature générale

1689 / 8415 ALRs

Annotés à l'aide d'une grille
d'analyse sémantique

Dimension sémantique : Qualia

Valeur sensation

Valeur affect

Valeur perception

- Le regard me donnait le vertige
- Sauter à la figure
- Je sens l'odeur

229 ALRs

Regroupés par rapport à leur
spécificité (Log-likelihood,
fréquence, dispersion)

Motifs textuels

(Legallois 2006;
Longrée & Mellet 2013,
Novakova & Siepmann 2020)

Extensions syntagmatiques et
paradigmatiques

*Approches
fonctionnelles et
contextualistes
(Sinclair 2004)*

Liste des expressions les
plus spécifiques en français

Equivalents
fonctionnels en thaï

Analyse
lexicale

Analyse
sémantique

Analyse
syntaxique

Analyse
discursive

Entendre le bruit

X entendre (distinctement) le bruit

(Bazin *L'huile sur le feu*, 1954)

X entendre (soudain) le bruit

(Laferrière *L'énigme du retour*, 2009)

X entendre le bruit (sec) (de la porte du four)

(Bazin *Au nom du fils*, 1960)

X entendre le (joli) bruit (d'un tissu)

(Ben Jelloun *Cette aveuglante absence de lumière*, 2001)

X entendre le (petit) bruit (patient de la râpe sur le bois)

(Merle *l'île*, 1962)

X entendre le bruit (de la pluie sur la ville)

(Foenkinos *Je vais mieux*, 2013)

Entendre le bruit

▷ ได้ยิน	เสียง	เอะอะ
<i>dai yin</i>	<i>sieng</i>	<i>ey a</i>
entendre	bruit	bruyant

(Jamkum Mae Jum Pen, 2008)

▷ ได้ยิน	เสียง	คน	ร้อง
<i>dai yin</i>	<i>sieng</i>	<i>khon</i>	<i>rong</i>
entendre	bruit	personne	pleurer

(Saowapong Pi Sat, 2001)



Résultats escomptés

- *Moyens morphosyntaxiques spécifiques disponibles dans chacune de ces langues pour exprimer les émotions*
- *Motifs textuels distingués de la littérature générale*
- *Fonctions discursives spécifiques de la littérature générale*

Merci pour votre attention