



HAL
open science

Rapport intermédiaire du Projet ANR-15-FRAL-0009 PHRASEOROM

Iva Novakova

► **To cite this version:**

Iva Novakova. Rapport intermédiaire du Projet ANR-15-FRAL-0009 PHRASEOROM. [Rapport de recherche] Université Grenoble Alpes (France); Université d'Osnabrück (Allemagne). 2018. hal-01874296

HAL Id: hal-01874296

<https://hal.univ-grenoble-alpes.fr/hal-01874296>

Submitted on 14 Sep 2018

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.

Projet ANR-15-FRAL-0009

PHRASEOROM

Programme FRAL 2015

A	IDENTIFICATION	2
B	LIVRABLES ET JALONS	2
C	RAPPORT D'AVANCEMENT.....	6
C.1	Objectifs initiaux du projet	6
C.2	Travaux effectués et résultats atteints sur la période concernée...	7
C.3	Difficultés rencontrées et solutions	8
C.4	Faits et résultats marquants	9
C.5	Travaux spécifiques aux entreprises (le cas échéant).....	9
C.6	Réunions du consortium (projets collaboratifs)	9
C.7	Commentaires libres.....	10
D	VALORISATION ET IMPACT DU PROJET DEPUIS LE DEBUT	11
D.1	Publications et communications	11
D.2	Autres éléments de valorisation	14
D.3	Pôles de compétitivité (projet labellisés)	15
D.4	Personnels recrutés en CDD (hors stagiaires)	16
D.5	État financier.....	17
E	ANNEXES EVENTUELLES	17

Ce document est à remplir par le coordinateur en collaboration avec les partenaires du projet. Il doit être transmis par le coordinateur aux échéances prévues dans les actes attributifs :

- 1. à l'ANR*
- 2. aux pôles de compétitivité ayant accordé leur label au projet.*

L'ensemble des partenaires doit avoir une copie de la version transmise à l'ANR.

Il doit être accompagné d'un résumé public du projet mis à jour, conformément au modèle associé à ce document.

Ce modèle doit être utilisé uniquement pour le(s) compte(s)-rendu(s) intermédiaire(s) défini(s) dans les actes attributifs de financement, hors rapport T0+6 pour lequel il existe un modèle spécifique. Il existe également un modèle spécifique au compte-rendu final.

A IDENTIFICATION

Acronyme du projet	PHRASEOROM
Titre du projet	Phraséologie du roman
Coordinateur du projet (société/organisme)	Laboratoire Lidilem E.A. 609, Université Grenoble Alpes UGA)
Date de début du projet	01.05.2016
Date de fin du projet	01.05.2019
Labels et correspondants des pôles de compétitivité (pôle, nom et courriel du corresp.)	
Site web du projet, le cas échéant	https://phraseorom.univ-grenoble-alpes.fr/

Rédacteur de ce rapport	
Civilité, prénom, nom	Mme Iva NOVAKOVA
Téléphone	06 51 52 62 46
Courriel	iva.novakova@univ-grenoble-alpes.fr
Date de rédaction	30.11.2017
Période faisant l'objet du rapport d'activité	01.05.2016 au 30.11.2017

B LIVRABLES ET JALONS

Quand le projet en comporte, reproduire ici le tableau des jalons et livrables fourni au début du projet. Mentionner l'ensemble des livrables, y compris les éventuels livrables abandonnés, et ceux non prévus dans la liste initiale.

Tableau 1 : **Planification des tâches liées au traitement des corpus**

	1-2	3-4	5-6	7-8	9-10	11-12	13-14	15-16	17-18	19-20	21-22	23-24
T1.1 Collecte												
T1.2 Reformatage												
T1.3 Annotation morphosyntaxique												
T1.4 Annotation sémantique												
T1.5 Indexation												
T2.1 Dossier de conception												
T2.2 Scripts d'indexation												
T2.3 Intégration partition corpus												
T2.4 Ajout contraintes métadonnées												
T2.5 Calcul de spécificité												
T2.6 Interface pour définir la partition												
T3.1 Motifs lexico-syntaxiques												
T3.2 Interface pour définir les motifs séquentiels												
T3.3 Extraction des motifs séquentiels												
Tests et débogages												

Responsables O. Kraif (UGA) et S. Diwersy (U. de Montpellier)

Le temps consacré à la réalisation de ce volet : 24 mois.

Les livrables dans la liste initiale du projet : Corpus multilingues annotés, scripts d'indexation et de recherche, interface d'interrogation et d'affichage des résultats, scripts d'extraction de motifs séquentiels, interface de définition de ces motifs et d'affichage des extractions ; évolution et

enrichissement de l'interface de l'EmoBase afin d'y intégrer un nouveau module *PhraséoBase* (en accès libre) qui regroupera les données issues du projet .

Tableau 2 : Planification des tâches liées à l'analyse linguistique

Mois (au total 36)	1-9	10-12	13-15	16-18	19-24	24-30	30-36
T1a Étude pilote sur corpus existants							
T1b Recensement des CLS sur les corpus définitifs							
T2 Choix théoriques							
T3 Grille sémantico-fonctionnelle et syntaxique							
T4 Analyse discursive							
T5 Équivalences interlinguistiques							
T6 Comparaisons traductologiques							
T7 Mise en commun des résultats							

Responsables : Iva Novakova (UGA) et Dirk Siepmann (U. d'Osnabrück)

Le temps consacré à la réalisation de ce volet : 36 mois.

Les livrables (dans la liste initiale du projet) : étude pilote sur les corpus, recensement des données des corpus comparables et parallèles, élaboration des fichiers Excel pour le traitement linguistique des données : codages syntaxico-sémantiques des résultats extraits des corpus sous forme de grilles ; rapports intermédiaires suite aux réunions rendant compte des avancées et des difficultés rencontrées ; rapport final, participations à des colloques internationaux et publications pour la valorisation du projet ; organisation d'un colloque international ; publication d'un ouvrage issu du projet.

Tableau 3 : Planification des tâches liées à l'analyse stylistique

Mois	1-16	16-22	22-24	24-30	30-36
T1 Constitution des corpus TAL					
T2 Repérage des récurrences au sein des corpus littéraire et paralittéraire					
T3 Repérage des données spécifiques aux discours littéraires par comparaison avec les discours non littéraires					
T4 identification des récurrences par sous-genres					
T5 interprétation de la valeur des formes récurrentes (poétique du roman)					

T6 identification de l'éventuelle spécificité des sous-genres paralittéraires et interprétation des données					
T7 identification et interprétation de la circulation des formes entre discours littéraires et discours non littéraires					
T8 Mise en commun des résultats					

Responsables : Julien Piat (UGA) ; Marion Gymnich (U. de Bonn)

Le temps consacré à la réalisation de ce volet : 36 mois.

Les livrables : [dans la liste initiale du projet](#)): participation à la constitution des corpus français et anglais ; vérifications et analyses des données extraites présentées sous forme de fichiers Excel dans la perspective stylistique, participation à des colloques internationaux, articles à valeur de rapport final, participation à l'ouvrage issu du projet.

N°	Intitulé	Nature*	Date de fourniture			Partenaires (souligner le responsable)
			Prévue initialement	Replanifiée	Livrée	
Corpus	T1.1 Collecte de nouveaux textes en sus du corpus Emolex, afin d'équilibrer les quantités au niveau des genres, pour chaque langue.	Volet TAL	Fin 2016	non	Oui	<u>O. Kraif</u> , S. Diwersy, J. Didier, M. Gymnich, D. XX, S. Dyka, L. Gonon, J. Piat
T1.2	2 Reformatage en suivant le schéma XML défini pour Emolex. Regroupement des textes par blocs de fichiers XML suffisamment grands (~ 2 go) pour optimiser l'utilisation du cache dans l'architecture d'EmoConc.	Volet TAL	12.2016	non	oui	<u>O. Kraif</u> , S. Diwersy, J. Didier, M. Gymnich, D. XX, S. Dyka, L. Gonon, J. PiatD.
T1.3.	3 Annotation morpho-syntaxique (étiquetage morphosyntaxique, lemmatisation et balisage des dépendances).	Volet TAL	02.2017	05.2017	oui	<u>O. Kraif</u> , J. Didier
T1.4	4 Annotation sémantique (à partir d'un thésaurus à large couverture). Ajout, pour chaque forme, d'un descripteur thématique général (issu du Thésaurus de Larousse).	Volet TAL	02.2017	02.2018	En cours	S. Diwersy
T1.5	5 Création des hachages enregistrant les phrases ainsi que les cooccurrences lexico-syntaxiques, et dotés d'index par lemme, par relation de dépendance, descripteur thématique, et par méta-données (genre, auteur, année).	Volet TAL	05.2017	Non	Oui	<u>O. Kraif</u> & J. Didier, S. Dyka
Tâche 2 :	1 Dossier de conception du nouvel outil. Design des interfaces et	Volet TAL	06.2016	11.2016	Oui	<u>O. Kraif</u> , <u>S. Diwersy</u> , <u>J. Didier</u>

N°	Intitulé	Nature*	Date de fourniture			Partenaires (souligner le responsable)
			Prévue initialement	Replanifiée	Livrée	
Réingénieur T2.1.	définition des nouvelles fonctionnalités d'interrogation du corpus.					
T2.2	2 Modification de la structure des index, intégration de tous les niveaux de balisage structurel (paragraphe, texte, etc.).	Volet TAL	02.2017	Oui	En cours	<u>O. Kraif, J</u> ; Didier
T2.3.	3 Intégration des critères de partitions du corpus de travail.	Volet TAL	03.2017	Oui	En cours	<u>O. Kraif, J</u> ; Didier
T2.4	4 Intégration de contraintes sur les méta-données dans le langage de requête.	Volet TAL	03 2017	Oui	En cours	<u>O. Kraif, J</u> ; Didier
T2.5	5.Modification des sorties en affichant des calculs de spécificité pour chaque sous-ensemble de la partition définie sur le corpus de travail.	Volet TAL	05.2017	Oui	En cours	<u>O. Kraif, J</u> ; Didier
T2.6.	6 Réalisation de l'interface de définition de la partition du corpus de travail (en fonction des critères de genre, auteur, année).	Volet TAL	05.2017	Oui	Non (en cours)	<u>O. Kraif, J</u> ; Didier, S. Diwersy
Linguistique	T.1. Étude pilote sur corpus existants	Volet linguistique	01.2017	Non	Oui	<u>I Novakova, D. Siepmann, V. Goossens, S. Dyka</u>
T1.2	Recensement des CLS sur les corpus définitifs	Volet linguistique	05.2017	Oui	En cours	<u>I Novakova, D. Siepmann, V. Goossens, S. Dyka, F. Grossmann, M. Gymnich, S. Koch, D. Legallois, J. Sorba,</u>
T2	Choix théoriques et méthodologiques	Volet linguistique	09.2017	Non	Oui	<u>I Novakova, D. Siepmann, V. Goossens, S. Dyka,</u>
T.3	Grille sémantico-fonctionnelle et syntaxique	Volet linguistique	12.2017	Non	Oui	<u>I Novakova, D. Siepmann, V. Goossens, S. Dyka, O. Kraif, S. Diwersy</u>
Stylistique T1	Constitution qualitative des corpus (choix des œuvres fr, angl, all et des corpus parallèles)	Volet Stylistique	04.2018	Non	En cours	<u>M. Gymnich, I. Piat, L. Gonon, D. XX, S. Koch, S. Dyka</u>
T2	T2 Repérage des récurrences au sein des corpus littéraire et paralittéraire	Volet stylistique	05.2018	Non	En cours	<u>M.Gymnich, L. Gonon, D. Burkhard, V. Goossens, I. Novakova, D.</u>

N°	Intitulé	Nature*	Date de fourniture			Partenaires (souligner le responsable)
			Prévue initialement	Replanifiée	Livrée	
						<u>Siepmann, S.</u> <u>Dyka, S. Koch, L.</u> <u>Fesenmeir</u>
T3	Repérage des données spécifiques aux discours littéraires par comparaison avec les discours non littéraires	Volet Stylistique	05.20 18	non	En cours	<u>M.Gymnich,</u> <u>L. Gonon, D.</u> <u>XX, V.</u> <u>Goossens, I.</u> <u>Novakova,</u> <u>D.</u> <u>Siepmann, S.</u> <u>Dyka, S.</u> <u>Koch, L.</u> <u>Fesenmeiri</u>
T4	identification des récurrences par sous-genres	Volet Stylistique	05.20 18	Non En cours		<u>M.Gymnich,</u> <u>L. Gonon, D.</u> <u>XX, V.</u> <u>Goossens, I.</u> <u>Novakova, D.</u> <u>Siepmann, S.</u> <u>Dyka, S.</u> <u>Koch, L.</u> <u>Fesenmeiri, F.</u> <u>Grossmann, J.</u> <u>Sorba</u>

* jalon, rapport, logiciel, prototype, données, ...

C RAPPORT D'AVANCEMENT

C.1 OBJECTIFS INITIAUX DU PROJET

Maximum 10 à 20 lignes.

- a)** Extraire de façon inductive (approche *corpus-driven*) des motifs lexico-syntaxiques à partir des expressions complexes automatiquement repérées (sous forme d'arbres lexico-syntaxiques ALR) à partir de vastes corpus annotés syntaxiquement dans les trois langues du projet (français, anglais, allemand).
- b)** Développer des outils permettant de mesurer la spécificité des constructions relativement aux différents genres, sous-genres et de rechercher des motifs séquentiels transphrastiques (Lexicoscope : <http://phraseotext.u-grenoble3.fr/lexicoscope/>, outil d'interrogation de la future base de données open source **PhaseoBase**).
- c)** Elaborer une typologie structurelle et fonctionnelle des constructions lexico-syntaxiques spécifiques (CLS) au discours romanesque francophone, anglophone et germanophone formant des motifs textuels afin de mieux identifier et caractériser le genre littéraire par rapport à d'autres genres et sous-genres littéraires et paralittéraires (romans sentimentaux, policiers, science-fiction, fantasy etc).
- d)** Proposer des comparaisons interlinguistiques ponctuelles d'expressions lexico-syntaxiques (ELS), en travaillant sur des corpus comparables et parallèles d'œuvres françaises, anglaises et allemandes afin de mieux cerner les ressemblances et les divergences dans les sous-genres littéraires anglophones, francophones et germanophones et les traditions scripturales respectives.

d) Conjoindre les approches textométriques, linguistiques et stylistiques pour aboutir à une description des patrons stylistiques littéraires en termes de « motifs textuels ».

g) Jeter les fondements d'un « lexique-grammaire » des unités lexicales étendues qui contribuent à la construction du discours romanesque grâce au moyens modernes de la linguistique de corpus et de la textométrie.

f) Créer une base de données open source PhraséoBase ouverte aux chercheurs et aux étudiants à usage interdisciplinaire (linguistique de corpus, TAL, stylistique, traductologie, phraséologie, linguistique contrastive) en Humanités numériques.

C.2 TRAVAUX EFFECTUES ET RESULTATS ATTEINTS SUR LA PERIODE CONCERNEE

Maximum 1 page. Travaux et résultats obtenus pendant la période concernée, conformité de l'avancement des travaux avec le plan initialement prévu. Prévision de travaux pour la (les) prochaine(s) période(s).

Volet Corpus et TAL

La première phase de collecte et de traitement du corpus est achevée, et concerne essentiellement le français et l'anglais. Pour le français, 1164 œuvres de 425 auteurs ont été collectées, pour un total d'environ 112 millions de mots répartis en 6 catégories (littérature générale, policier, fantasy, science-fiction, sentimental et historique). Pour l'anglais, on dénombre 726 œuvres de 299 auteurs pour un total de 94 millions de mots (même répartition). Un échantillon de corpus littéraire équilibré de 30 millions de mots par langue a été constitué pour confronter la langue littéraire (tous sous-genres confondus) à d'autres types de textes. Ainsi, des corpus de contraste (variés et non littéraires) ont été constitués pour faire ressortir les spécificités de ces échantillons littéraires. Ces corpus de contraste sont constitués d'un mélange de presse, d'articles scientifiques, de sous-titres d'émission TV, de commentaires d'actualités, de débats parlementaires, et représentent environ 60 millions de mots pour le français et 78 millions de mots pour l'anglais.

Les tâches T1.1 à T1.5 sont terminées pour tous ces corpus. Concernant les travaux sur les outils du Lexicoscope, après rédaction d'un dossier de conception (T2.1), la réalisation de la nouvelle interface basée sur une architecture AJAX a été initiée en septembre 2017. L'interface est en cours de réalisation (T2.6). Les tâches de modification des scripts d'indexation, T2.2 à T2.4 sont commencées, mais ne seront finalisées qu'au printemps 2018. Cela n'a pas empêché de commencer les tâches d'extraction de motifs lexico-syntaxiques (ALR). L'algorithme a été revu, corrigé et optimisé, et des séries d'ALR ont été extraites corpus par corpus (tâche T3.1).

Volet Linguistique

Les tâches de T1 à T3 incluse ont été réalisées (18 mois programmés). L'élaboration de la grille sémantico-fonctionnelle et syntaxique est en cours de finalisation (décembre 2017). Elle permettra la réalisation des tâches de 4 à 7 et, plus particulièrement, les codages syntaxico-sémantiques et discursifs des résultats extraits des corpus sous forme de grilles prévues dans les livrables. Les données (des expressions lexico-syntaxiques (ELS) récurrentes spécifiques aux corpus littéraires par rapport aux corpus de contraste (textes non littéraires) ainsi qu'aux différents sous-genres (policier, sent, science-fiction, littérature générale etc.) ont été extraites grâce aux nouvelles fonctionnalités du Lexicoscope (outil créée par O. Kraif et S. Diwersy) et intégrées dans des fichiers Excel dont l'architecture a été mis en place en commun avec les équipes allemandes. Toutes les données extraites sont stockées dans un espace collaboratif dédié.

Les codages des ALR (arbres lexico-syntaxiques récurrents) spécifiques du corpus littéraire général anglais par rapport au corpus de contraste ont été terminés (plus de 6500 ALR ont été

classés selon la grille sémantique élaborée en commun par les équipes françaises et allemandes). Les codages des ALR du corpus LITT français comparé au corpus de contraste CONT (6500 ALR) est en cours. Il va se terminer fin décembre 2017.

Une deuxième phase essentielle pour le traitement et l'analyse des données a commencé pour l'anglais : les codages syntaxico-sémantiques et discursifs des ALR déjà classés (cf. phase précédente)

Des groupes de travail mixtes ont été constitués (composés de membres des équipes françaises et allemandes) selon 3 thématiques :

- Caractérisation des personnages/état/setting (temps-espace)
- Cognition/communication/qualia (affects)
- Actions

Ces groupes de travail auront pour mission d'identifier à partir des données (approche *corpus driven*) les motifs textuels récurrents dans les différents sous-genres selon une méthodologie commune en voie de finalisation. Toutes les données seront présentées sous forme de fichiers Excel, classées et intégrées à la PhraséoBase.

Volet stylistique

Les tâches de 1 à 4 ont été réalisées (la T3 en avance par rapport au planning). La T4 est en cours de réalisation. Le choix et la récolte des textes anglais et français sont complètement achevés. Les métadonnées des œuvres ont été complétées. Ils sont tous déjà intégrés à la base de données (Lexicoscope/PhraséoBase). La constitution des corpus allemands comparables et des corpus parallèles est en cours.

Une réflexion méthodologique sur les tâches T5 et T6 du volet a été entamée en commun avec les collègues linguistes. Elle sera présentée à la 3^{ème} réunion générale à Bonn (janvier 2018).

C.3 DIFFICULTES RENCONTREES ET SOLUTIONS

Maximum 10 à 20 lignes. Difficultés éventuelles rencontrées et solutions de remplacement envisagées ex : impasse technique, abandon d'un prestataire, maîtrise des délais, maîtrise des budgets. Faut-il revoir le contenu du projet ? Faut-il revoir le calendrier du projet ?

1. Le traitement de très grandes masses de données linguistiques. Nécessité de créer des filtres et des seuils (par ex. 30 000 expressions ont été extraites à partir du corpus LITT anglais), afin de pouvoir les classer et analyser ;
2. Les analyseurs syntaxiques (Xip, Connexor) ne sont pas fiables à 100% ce qui introduit du bruit dans les données extraites ;
3. Quelques bugs informatiques dans le traitement des corpus ;
4. La collecte des corpus allemands a pris plus de temps que prévu, étant donné le nombre restreint de textes datant de projets antérieurs
5. La collecte de corpus parallèles (anglais- français, anglais-allemand, français – allemand) s'est également avérée plus complexe que prévu (manque de textes disponibles au bon format et de bonne qualité) ;

Ces deux tâches seront cependant réalisées courant 2018 ;

6. Le congé de formation de 1 an de J. Piat (Litt&Arts). Renforcement du volet stylistique par l'intégration récente (octobre 2017) de Clémence Jacquot (Univ. de Montpellier, stylisticienne) à l'équipe française (Litt&Arts)) et par l'arrivée d'un post-doctorant et de stagiaires de Master à la rentrée 2018 (UGA, Litt&Arts).

C.4 FAITS ET RESULTATS MARQUANTS

En quelques lignes pour chaque fait ou résultat marquant. Cet élément pourrait donner lieu à communication, après accord du coordinateur du projet.

1. Constitution de très larges corpus de textes littéraires français et anglais (allemands en cours), annotés syntaxiquement et intégrés dans la Base de données PhraseoBase du Lexicoscope qui seront mis à la disposition de la communauté scientifique à la fin du projet. Ces textes sont partitionnés et classés en sous-genres (littéraires et paralittéraires).
2. Création de nouvelles fonctionnalités spécifiques de l'outil Lexicoscope pour l'interrogation des corpus littéraires multilingues et l'extraction automatique d'unités polylexicales (motifs textuels), spécifiques aux différents sous-genres littéraires. A notre connaissance, un tel outil complexe et puissant n'existe pas pour l'interrogation des bases de données existantes. Ce projet ambitieux est en cours et sera terminé avec l'inauguration de la PhraséoBase à la fin du projet. Il sera le résultat d'échanges systématiques entre informaticiens, linguistes et stylisticiens.
3. Elaboration d'une méthodologie multidimensionnelle pour l'extraction, le traitement et l'analyse de grosses masses de données linguistiques commune pour les trois langues du projet (en cours de finalisation).
4. Elaboration d'une typologie trilingues (français, anglais, allemand) des motifs textuels statistiquement spécifiques du genre littéraire en général et des sous-genres retenus (POL, SF, SENT, HIST, FANT, GEN) permettant de mieux les caractériser et identifier. Plusieurs études pilotes ont été déjà réalisées ou sont en cours dans ce sens (Colloque phraséologie française, Artois 2017, Europhras, Londres 2017, CMLF, Mons 2018. Trois nouvelles soumissions sont en préparation (Colloque de Pau sur les genres textuels, mars 2018, Colloque Cognition Paris juillet 2018, JADT 2018 Rome).

C.5 TRAVAUX SPECIFIQUES AUX ENTREPRISES (LE CAS ECHEANT)

Entreprise xxx

Maximum 10 à 20 lignes par entreprise. Pour chaque entreprise du consortium, décrire les activités dans le projet, en se concentrant sur les apports, collaborations et perspectives liés au projet. Préciser notamment les perspectives d'application industrielle ou technologique, de potentiel économique et commercial, d'intégration dans l'activité industrielle, etc.

Entreprise	Xxx
Rédacteur (nom + adresse mél)	

C.6 REUNIONS DU CONSORTIUM (PROJETS COLLABORATIFS)

Date	Lieu	Partenaires présents	Thème de la réunion
20-22.06.2016	Grenoble	Toutes les équipes françaises et allemandes	Réunion générale de démarrage du projet PhraseoRom. Présentation des équipes, des objectifs et des tâches.
28.10.2016s	Osnabrück	Marion Gymnich, Dirk Siepman, Ludwig Fesenmeier, Denise Burkhard, Susanne Dyka	Réunion partielle. Discussion sur l'élaboration de la grille sémantique (catégories et sous-catégories à retenir). Propositions sur

Date	Lieu	Partenaires présents	Thème de la réunion
			la structuration des données extraites.
25.01.2017	Grenoble	Olivier Kraif, I. Novakova, Johan Didier, Julie Sorba (par skype), Agnès Tutin, Sascha Diwersy, Dominique Legallois, Laetitia Gonon, Manon Lalanne (stagiaire)	Réunion partielle. État des lieux sur les corpus : compositions, métadonnées, textes à retenir, classement des textes par sous-genres, intégration des textes à la BDD. Colloques à venir, nouvelles du consortium « Motifs », recrutement du post-doc en linguistique (profil du poste).
10-12.05.2017	Osnabrück	Toutes les équipes françaises et allemandes	Réunion générale. Etat des lieux sur les corpus et les outils d'interrogation avancées informatiques) : état des lieux sur les analyses linguistiques et stylistiques, discussion sur les méthodologies ; site Internet du projet, colloques à venir, programmation des tâches pour les 6 mois à venir.
Octobre 2017	Grenoble	L. Fesenmeir, Stefan Koch (Erlangen), le 06.10.2017 Susanne Dyka (Osnabrück) (11-13.10.2017) I.Novakova, V. Goossens, F. Grossmann, O. Kraif	Réunions partielles entre l'équipe française du Lidilem et des membres des équipes allemandes (Erlangen et Osnabrück). Codages des ALR en français et en anglais ; travail sur la mise en place des méthodologies, état des lieux des corpus, des tâches.
03.11.2017	Osnabrück	Dirk Siepman, Susanne Dyka, Stefan Koch, Ludwig Fesenmeier, Denise Burkhard, Marion Gymnich	Réunion partielles des équipes allemandes. Corpus allemands et parallèles, codages des ALR anglais, répartition des tâches, colloques, publications. Constitution des groupes de travail par thématiques des motifs textuels.

C.7 COMMENTAIRES LIBRES

Commentaires du coordinateur

Commentaire général à l'appréciation du coordinateur, sur l'état d'avancement du projet, les interactions entre les différents partenaires...

Le projet a démarré le 01 mai 2016. Les embauches se sont faites progressivement. Selon le prévisionnel des tâches et du budget restants, une prolongation de 8 mois (du 01.05.2019 au 31.12.2019) sera demandée.

Commentaires des autres partenaires

Éventuellement, commentaires libres des autres partenaires

...

Question(s) posée(s) à l'ANR

Éventuellement, question(s) posée(s) à l'ANR...

...

D VALORISATION ET IMPACT DU PROJET DEPUIS LE DEBUT

Cette partie rassemble des éléments cumulés depuis le début du projet qui seront suivis tout au long de son avancée, et repris dans son bilan final.

D.1 PUBLICATIONS ET COMMUNICATIONS

Citer les publications résultant du projet en utilisant les normes habituelles du domaine. Si la publication est accessible en ligne, préciser l'adresse. L'ANR encourage, dans le respect des droits des co-auteurs et des éditeurs, à publier les articles résultant des projets qu'elle finance dans l'archive ouverte pluridisciplinaire HAL : <http://hal.archives-ouvertes.fr/>

Attention : éviter une inflation artificielle des publications, mentionner uniquement celles qui résultent directement du projet (postérieures à son démarrage, et qui citent le soutien de l'ANR et la référence du projet).

Liste des publications multipartenaires (résultant d'un travail mené en commun)		
International	Revues à comité de lecture	
	Ouvrages ou chapitres d'ouvrage	1. Dyka, S., Novakova I., & Siepmann D. (2017). A Web of Analogies: Depictive and Reaction Object Constructions in Modern English and French Fiction. In <i>Computational and Corpus-Based Phraseology, Second International Conference, Europhras 2017, London, UK, November 13-14, 2017, Proceedings</i> (pp. 87-101). Springer International Publishing.
	Communications (conférence)	1. Dyka, S., Novakova I., & Siepmann D. (2017). Computational and Corpus-based Phraseology – recent advances and interdisciplinary approaches. <i>Europhras</i> , Londres, 13-14 novembre 2017.
France	Revues à comité de lecture	1. 2.
	Ouvrages ou chapitres d'ouvrage	1. Kraif Olivier, Novakova Iva, Siepmann Dirk (à paraître) « Multi-word expressions and the Construction of Meaning in french and English Crime Fiction » (Actes du Colloque <i>Marqueurs et structures: quelles articulations dans la (re)construction du sens</i> , U. de Nanterre, Paris. Lambert-Lucas 2.
	Communications (conférence)	2. 3.
Actions de diffusion	Articles de vulgarisation	1. 2.
	Conférences de vulgarisation	1.

		2.
	Autres	1. 2.

NB Les publications sont classées par ordre alphabétique des contributeurs

Liste des publications monopartenaire (impliquant un seul partenaire)		
International	Revue à comité de lecture	<ol style="list-style-type: none"> 1. Delente, É., & Legallois D. (2016). La répétition littérale dans les Rougon-Macquart : présentation et interprétation d'un phénomène peu connu. <i>Excavatio</i>. XXVIII : https://sites.ualberta.ca/~aizen/excavatio/articles/v28/DelenteLegallois.pdf 2. Grossmann, F. (à paraître). Du rituel social aux motifs : la scène de présentation interpersonnelle dans le roman. <i>Revue Gerflint - Synergies Pologne</i>. 3. Kraif, O. (à paraître). Traduire le polar : une étude textométrique comparée de la phraséologie du roman policier en français source et cible. <i>Revue Gerflint - Synergies Pologne</i>. 4. Muryn T. & Novakova I. Eds (2018, à paraître). De la phraséologie aux genres textuels : état des recherches et perspectives méthodologiques. <i>Synergies Pologne no 14 / 2018</i>. 5. Novakova, I., & Sorba J. (à paraître). À la recherche de motifs autour de meurtre, crime, assassinat dans le roman policier contemporain. <i>Revue Gerflint - Synergies Pologne</i>. 6. Siepmann, D. (2015). A corpus-based investigation into key words and key patterns in post-war fiction. <i>Functions of language</i>, 22(3), p. 362-399.
	Ouvrages ou chapitres d'ouvrage	<ol style="list-style-type: none"> 1. Gonon, L., Goossens V., & Novakova I. (à paraître). Les phraséologismes spécifiques à deux sous-genres de la paralittérature : le roman sentimental et le roman policier. In <i>Actes du colloque Phraséologie Française</i>. Hermann. 2. Gonon, L., Kraif O., Novakova I., Piat J., & Sorba J. (2016). Sur la scène de crime : Enquête sur les enjeux linguistiques et stylistiques de motifs récurrents dans le thriller contemporain. In Neveu, F., Bergounioux G., Côté M-H., Fournier J-M., Hriba L., & Prévost S., (éds.), <i>Congrès Mondial de Linguistique Française, 4-8 juillet 2016, Tours</i>. https://www.shs-conferences.org/articles/shsconf/abs/2016/05/shsconf_cmlf2016_06006/shsconf_cmlf2016_06006.html 3. Grossmann, F. (à paraître). La fragmentation de l'identité dans les rituels de présentation : le cas Modiano. In <i>Actes du colloque Texte Fragmentation Créativité, Lublin</i>. 4. Legallois, D., & Diwersy S. (à paraître). Comment l'exploration de motifs permet-elle de repenser la phraséologie ? In <i>Actes du Colloque International Approches théoriques et empiriques en phraséologie</i> (Nancy, décembre 2014).
	Communications (conférence)	<ol style="list-style-type: none"> 1. Chambre, J., & Kraif O. (2017). Séquences récurrentes typiques du roman policier et de science-fiction : une étude préliminaire sur corpus. <i>Journées internationales de la linguistique de corpus</i>, Grenoble, 4-6 juillet 2017. 2. Gonon, L., Goossens V., & Novakova I. (2017). Les phraséologismes spécifiques à deux sous-genres de la paralittérature : le roman sentimental et le roman policier. <i>Colloque Phraséologie française</i>, Arras, 21-22 septembre 2017. 3. Gonon, L., Kraif O., Novakova I., Piat J., & Sorba J. (2016). Sur la scène de crime : Enquête sur les enjeux linguistiques et stylistiques de motifs récurrents dans le thriller contemporain. <i>Congrès Mondial de Linguistique Française, Tours, 4-8 juillet 2016</i>. 4. Grossmann, F. (2016). « Se présenter » et « présenter quelqu'un à autrui » : formes et fonctions des motifs phraséologiques de l'auto-présentation et de la présentation d'autrui dans le roman : l'exemple de Simenon et de Modiano.

		<p><i>Journée phraséologique franco-polonaise, Cracovie, 18 octobre 2016.</i></p> <ol style="list-style-type: none"> Grossmann, F. (2016). La fragmentation de l'identité dans les rituels de présentation : le cas Modiano. <i>Colloque international Texte Fragmentation Créativité</i>, Lublin, 21-22 octobre 2016. Kraif, O. (2016). Traduire le polar : une étude textométrique comparée de la phraséologie du roman policier en français source et cible. <i>Journée phraséologique franco-polonaise, Lublin, 18 octobre 2016.</i> Kraif, O. (2017). Sous-genre littéraire et traduction : une approche textométrique. <i>Congrès mondial de traductologie, Atelier textométrie et traduction</i>, Paris, 17 avril 2017. Legallois, D. (2017). « Je vais partir pour Rome et je n'ai pas, malheureusement, le temps de répondre à vos questions » – la traduction des répétitions dans <i>Les Rougon-Macquart</i>. <i>Colloque international « Traduire Zola, du XIXe siècle à nos jours »</i>, Rome, 27-29 avril 2017. Legallois, D. (2017). Pour en revenir au texte : aspects de la lecture à « bonne distance » en humanités numériques littéraires ». <i>Colloque Des humanités numériques littéraires ?</i>, Cerisy, 15-22 juin 2017. Novakova, I., & Sorba J. (2016). À la recherche de motifs autour de meurtre, crime, assassinat dans le roman policier contemporain. <i>Journée phraséologique franco-polonaise, Lublin, 18 octobre 2016.</i> Siepmann, D. (2017). Tolkien as a Stylist: a Corpus-Based Investigation into 'The Lord of the Rings'. <i>9th International Corpus Linguistic Conference</i>, Birmingham, 24-28 juillet 2017. Siepmann, D., & Dyka S. (2017). A web of analogies: key descriptive constructions in modern fiction. <i>9th International Corpus Linguistic Conference</i>, Birmingham, 24-28 juillet 2017.
France	Revue à comité de lecture	<ol style="list-style-type: none"> Grossmann, F. (à paraître). Les phraséologismes montrés par comme on dit dans les romans contemporains français. <i>Le Français Moderne</i>. Kraif, O. (2016). Le lexicoscope : un outil d'extraction des séquences phraséologiques basé sur des corpus arborés. <i>Cahiers de lexicologie</i>, 1(108), p. 91-106. Kraif, O., Novakova I., & Sorba J. (2016). Constructions lexico-syntaxiques spécifiques dans le roman policier et la science-fiction. <i>Lidil</i>, 53, p. 143-159 : http://journals.openedition.org/lidil/3976 Legallois, D., Charnois T., & Poibeau T. (2016). Repérer les clichés dans les romans sentimentaux grâce à la méthode des "motifs". <i>Lidil</i>, 53, p. 95-117 : http://journals.openedition.org/lidil/3950 Siepmann, D. (2016). Lexicologie et phraséologie du roman contemporain : quelques pistes pour le français et l'anglais. <i>Cahiers de lexicologie</i>, 1(108), p. 21-41.
	Ouvrages ou chapitres d'ouvrage	<ol style="list-style-type: none"> Legallois, D. (à paraître). Caractériser le style d'un auteur par des patrons lexico-grammaticaux : une nouvelle approche en stylistique. In <i>Méthodes stylistiques : Unités et paliers de pertinence ?</i>. Lyon : Presses Universitaires de Lyon.
	Communications (conférence)	<ol style="list-style-type: none"> Grossmann, F. (2016). Les phraséologismes montrés dans les romans contemporains. <i>Journée de valorisation Agir Phraséotext</i>, Grenoble, 8 décembre 2016. Kraif, O. (2016). Autour de l'extraction d'arbres lexico-syntaxiques récurrents spécifiques : corpus, outils et méthodes. <i>Journée de valorisation Agir Phraséotext</i>, Grenoble, 8 décembre 2016. Piat, J., & Gonon L. (2016). Analyse stylistique de phraséologismes spécifiques

		dans le roman contemporain : réflexions méthodologiques. <i>Journée de valorisation Agir Phraséotext</i> , jeudi 8 décembre 2016.
Actions de diffusion	Articles de vulgarisation	1. 2.
	Conférences de vulgarisation	1. I. Novakova (2016) « Constructions lexico-syntaxiques spécifiques dans le roman policier et la science-fiction », Présentation du projet AGIR Pôles Phraséotext et du projet ANR DFG Phraséorom (le 11 mai 2016, Séminaire doctorants et post-doctorant, Section de français, Université de Lausanne). 2. I. Novakova & O. Kraif (2016) : « Analyse linguistique du lexique sur grands corpus informatisés » (cycle de conférences Recherches en SHS & Collège doctoral <i>Les méthodes évoluent, et vous?</i> Le 14 juin 2016, MSH Alpes, Grenoble) https://www.youtube.com/watch?v=Hg75j2TLOYI http://cyan1.grenet.fr/podcastmedia/evolution-recherches-SHS/confNovakova.m4v 3. I. Novakova (2017) 'La phraséologie spécifique du roman contemporain français. Conférence donnée le 7 juin 2017 au Centre d'études francophones de l'Université de Sofia, Bulgarie. Présentation des premiers résultats du projet ANR DFG Phraséorom, suivie d'une démonstration et d'une formation à l'utilisation des bases de données EmoBase et Lexicoscope.
	Autres	1. Réunion inaugurale du Consortium « Motifs », organisée par Dominique Legallois (U. Paris 3, Sorbonne nouvelle), le 27 mai 2016 réunissant une trentaine de chercheurs (France, Belgique, Pologne, Allemagne) en vue de l'élargissement du réseau de chercheurs travaillant sur les motifs phraséologiques dans la littérature 2. 2 ^{ème} Rencontre du groupe « Motifs », Grenoble le 07.07.2017, MSH Alpes. Réflexion autour du dépôt d'un projet européen

D.2 AUTRES ELEMENTS DE VALORISATION

Les éléments de valorisation sont les retombées autres que les publications. On détaillera notamment :

- *brevets nationaux et internationaux, licences, et autres éléments de propriété intellectuelle consécutifs au projet.*
- *logiciels et tout autre prototype*
- *actions de normalisation*
- *lancement de produit ou service, nouveau projet, contrat,...*
- *le développement d'un nouveau partenariat,*
- *la création d'une plate-forme à la disposition d'une communauté*
- *création d'entreprise, essaimage, levées de fonds*
- *autres (ouverture internationale,..).*

Ce tableau détaille les brevets nationaux et internationaux, licences, et autres éléments de valorisation consécutifs au projet, du savoir-faire, des retombées diverses en précisant les partenariats éventuels. Voir en particulier celles annoncées dans l'annexe technique.

Liste des éléments. Préciser les titres, années et commentaires	
Brevets internationaux obtenus	1. 2.
Brevet internationaux en cours d'obtention	1. 2.
Brevets nationaux obtenus	1. 2.

Brevet nationaux en cours d'obtention	1. 2.
Licences d'exploitation (obtention / cession)	1. 2.
Créations d'entreprises ou essaimage	1. 2.
Nouveaux projets collaboratifs	<p>Projet PHRASEOLIT déposé dans le cadre de l'AAP SFR Création de la maison de la création de l'UGA http://maisondelacreation.univ-grenoble-alpes.fr/fr/recherche/projets/ Résultat attendu pour le 20.12.2017</p> <p>Ce nouveau projet vise à explorer les <i>motifs textuels</i>, spécifiques à la littérature générale (GEN) en langue française et anglaise, en vue de leur exploitation dans le cadre d'ateliers d'écriture littéraire pour adultes. Pour ce faire, il réunira les efforts de linguistes et de stylisticiens de deux laboratoires de l'UGA (Lidilem et Litt&Arts) et de professionnels qui animent des ateliers et des stages d'écriture littéraire au sein de l'association Les Mots Voyageurs http://lesmotsvoyageurs.com</p> <p>Les résultats du projet ANR DFG Phraséorom serviront de base au développement de cette nouvelle collaboration.</p>
Colloques scientifiques	1. Contacts établis par Dirk Siepmann et Iva Novakova avec l'équipe de Michaela Mahlberg de l'Université de Birmingham (GB) en vue de l'organisation d'une Journée d'étude internationales sur les motifs littéraires en 2018 ou 2019
Autres (préciser)	1. 2.

D.3 POLES DE COMPETITIVITE (PROJET LABELLISES)

Pour les projets labellisés par un ou plusieurs pôles de compétitivité,

Collaboration du projet avec le(s) pôle(s) ayant labellisé

Quelles collaborations y a-t-il eu entre votre projet et le(s) pôle(s) de compétitivité l'ayant labellisé ?

...

Activités financées par le complément de pôle (laboratoires publics uniquement)

Détailler les activités réalisées par les laboratoires publics avec le complément de financement accordé au titre de la labellisation. Préciser notamment les partenaires impliqués et la collaboration menée avec le ou les pôles.

Montant du complément accordé par l'ANR (pour chaque labo public)	<ul style="list-style-type: none"> - Partenaire XXX : xxx € - Partenaire YYY : yyy €
--	--

Type d'action menée	Détails (exemples non limitatifs)	Dépenses complément de pôle*
Actions contribuant à la réflexion stratégique et à la	Ex : Participation aux journées thématiques organisées par le pôle	Xxx : xxy € Yyy : yyy €

programmation scientifique du pôle		
Actions de communication scientifique et publique bénéficiant à la notoriété du pôle	Ex : colloque de projets	Xxx : xxy € Yyy : yyy €
Développement de la recherche partenariale (recherche de partenaires, frais de gestion du partenariat, ingénierie de projets,...)	Ex : accord de consortium, frais de formation à la propriété intellectuelle, à la gestion de projets, dépenses relatives au montage du projet	Xxx : xxy € Yyy : yyy €
Valorisation de la recherche et transfert vers le monde industriel	Ex : étude de brevetabilité	Xxx : xxy € Yyy : yyy €

* Estimation des dépenses imputées sur le complément de financement accordé au titre de la labellisation par un pôle de compétitivité, partenaires publics seulement.

D.4 PERSONNELS RECRUTES EN CDD (HORS STAGIAIRES)

Ce tableau dresse le bilan du projet en termes de recrutement de personnels non permanents sur CDD ou assimilé. Renseigner une ligne par personne embauchée sur le projet quand l'embauche a été financée partiellement ou en totalité par l'aide de l'ANR et quand la contribution au projet a été d'une durée au moins égale à 3 mois, tous contrats confondus, l'aide de l'ANR pouvant ne représenter qu'une partie de la rémunération de la personne sur la durée de sa participation au projet.

Les stagiaires bénéficiant d'une convention de stage avec un établissement d'enseignement ne doivent pas être mentionnés.

Des données complémentaires sur le devenir professionnel des personnes concernées seront demandées à la fin du projet. Elles pourront faire l'objet d'un suivi jusqu'à 5 ans après la fin du projet.

Identification				Avant le recrutement sur le projet			Recrutement sur le projet			
Nom et prénom	Sexe H/F	Adresse email (1)	Date des dernières nouvelles	Dernier diplôme obtenu au moment du recrutement	Lieu d'études (France, UE, hors UE)	Expérience prof. antérieure (ans)	Partenaire ayant embauché la personne	Poste dans le projet (2)	Date de recrutement	Durée missions (mois) (3)
Johan DIDIER	H	johan.dider@yahoo.fr		Master Industries de la langues à l'UGA	France	-	Lidilem	Ingénieur d'étude	01.10.2016	24 mois
Vannina GOOSSE NS	F	vannina.goossens@univ-grenoble-alpes.fr		Doctorat en Sciences du Langage	France	5 ans CDD	Lidilem	Post-doc	01.03.2017	12 mois (renouvelable)

Aide pour le remplissage

(1) **Adresse email** : indiquer une adresse email la plus pérenne possible

(2) **Poste dans le projet** : post-doc, doctorant, ingénieur ou niveau ingénieur, technicien, vacataire, autre (préciser)

(3) **Durée missions** : indiquer en mois la durée totale des missions (y compris celles non financées par l'ANR) effectuées ou prévues sur le projet

Les informations personnelles recueillies feront l'objet d'un traitement de données informatisées pour les seuls besoins de l'étude anonymisée sur le devenir professionnel des personnes recrutées sur les projets ANR. Elles ne feront l'objet d'aucune cession et seront conservées par l'ANR pendant une durée maximale de 5 ans après la fin du projet concerné. Conformément à la loi n° 78-17 du 6 janvier 1978 modifiée, relative à l'Informatique, aux Fichiers et aux Libertés, les personnes concernées disposent d'un droit d'accès, de rectification et de suppression des données personnelles les concernant. Les personnes concernées seront informées directement de ce droit lorsque leurs coordonnées sont renseignées. Elles peuvent exercer ce droit en s'adressant l'ANR (<http://www.agence-nationale-recherche.fr/Contact>).

D.5 ÉTAT FINANCIER

Donner un état indicatif de la consommation des crédits par les partenaires. Indiquer la conformité par rapport aux prévisions et expliquer les écarts significatifs éventuels.

Nom du partenaire	Crédits consommés (en %)	Commentaire éventuel
LIDILEM	33%	Le projet a démarré le 01.05.2016 ce qui explique l'écart budgétaire par rapport aux prévisions. Le recrutement du personnel côté allemand s'est fait en avril 2016 et le lancement officiel du projet (devant se faire en parallèle) a été repoussé de quelques mois). Cet écart sera compensé en 2018 et 2019. Une prolongation de 8 mois sera demandée à l'ANR.
Litt&Arts	10%	Même raison. A cela s'ajoute le congé de formation de 1 an de Julien Piat. De plus, le post-doc (1 an), prévu sur ce budget, sera recruté en 2018 selon le planning des embauches et la spécificité des tâches. Pour renforcer l'équipe ; Clémence Jacquot, stylisticienne de l'U. de Montpellier a été intégrée à l'équipe de Litt&Arts. La prise en charge de ses missions liées à Phraséorom se fera donc sur ce budget. Le recrutement d'un stagiaire est également prévu. Le budget sera consommé dans la deuxième moitié du projet.

E ANNEXES EVENTUELLES