

Appel à projets franco-allemand
en sciences humaines et sociales

Ausschreibung eines deutsch-französischen
Programms in den Geistes- und Sozialwis-
senschaften

Programme non-thématique

Ohne thematische Vorgaben

2015

2015

Titre du projet

Projekttitel

La phraséologie du roman

Phraseologie des Romans

Acronyme

Akronym

PHRASEOROM

PHRASEOROM

Nom du coordinateur et institutions de ratta-
chement

Name des Projektleiters und seiner Instituti-
on

Iva NOVAKOVA
Université Stendhal, Grenoble 3, LIDILEM,
EA (609)

Dirk SIEPMANN
Universität Osnabrück – Fachbereich 7
Institut für Anglistik und Amerikanistik

1. Allgemeine Angaben

1.1. Antragsteller, Mitglieder der Arbeitsgruppe und Kooperationspartner

Deutsche Arbeitsgruppe:

Vorname	Nachname	Status	Institution	Adresse	Rolle im Projekt	
Dirk	Siepmann	Professor	Universität Osnabrück, Institut für Anglistik und Amerikanistik	Universität Osnabrück, Neuer Graben 40, D-49069 Osnabrück, dirk.siepmann@uos.de	Antragsteller	
Ludwig	Fesenmeier	Professor	Universität Erlangen, Institut für Romanistik	Bismarckstr 1, D-91054 Erlangen ludwig.fesenmeier@fau.de	Mitglied der Arbeitsgruppe	
Marion	Gymnich	Professorin	Universität Bonn, Institut für Anglistik, Amerikanistik und Keltologie	Universität Bonn, Regina-Pacis-Weg 5 D-53113 Bonn, mgymnich@uni-bonn.de	Mitglied der Arbeitsgruppe	
Sascha	Diwersy	wissenschaftlicher Mitarbeiter	Universität Köln, Romanisches Seminar	Universität zu Köln, Albertus-Magnus-Platz, D-50923 Köln sascha.diwersy@uni-koeln.de	Mitglied der Arbeitsgruppe	
N.N.		wiss. Mitarbeiter	Universität Osnabrück		Mitglied der Arbeitsgruppe	
N.N.		wiss. Mitarbeiter	Universität Bonn		Mitglied der Arbeitsgruppe	
N.N.		wiss.	Universität		Mitglied der	

		Mitarbeiter	Erlangen		Arbeitsgruppe	
--	--	-------------	----------	--	---------------	--

Kooperationspartner (der deutschen Arbeitsgruppe):

Vorname	Name	Status	Institution	Adresse
John Desmond	Gallagher	Lehrbeauftragter	Universität Münster	Erlenallee 3, 48155 Münster, jdgallagher@gmx.de
Peter	Blumenthal	Professor i.R.	Universität Köln	Universität zu Köln, Albertus-Magnus-Platz, D-50923 Köln, peter.blumenthal@uni-koeln.de

Französische Arbeitsgruppe:

Prénom	NOM	Statut	Rattachement	Coordonnées	Rôle dans le projet	Temps d'engagement
Iva	Novakova	Professeur	Lidilem E.A. 609	Université Stendhal, Grenoble 3, BP 25 F-38040 Grenoble cedex Iva.Novakova@u-grenoble3.fr	Coordinatrice	30
Olivier	Kraif	MCF, HDR	Lidilem	Olivier.Kraif@u-grenoble3.fr	Participant direct au projet	30
Francis	Grossmann	Professeur	Lidilem	Francis.Grossmann@u-grenoble3.fr	Participant direct au projet	18
Agnès	Tutin	Professeur	Lidilem	Agnes.Tutin@u-grenoble3.fr	Participant direct au projet	12
Julie	Sorba	Docteur en SCL, PRAG	Lidilem	Julie.Sorba@u-grenoble3.fr	Participant direct au projet	12
Julien	Piat	MCF	LITT&ARTS E.A. 7355 Equipe Traverses 19-21	Julien.Piat@u-grenoble3.fr	Participant direct au projet	18
Laetitia	Gonon	MCF	LITT&ARTS E.A. 7355 Equipe Traverses 19-21	Laetitia.Gonon@u-grenoble3.fr	Participant direct au projet	18
Dominique	Legallois	MCF, HDR	Crisco EA 4255	Université de Caen, Campus 3 Boulevard Yitzhak Rabin , 14123 IFS dominique.legallois@unicaen.fr	Participant direct au projet	18

Kooperationspartner (der französischen Arbeitsgruppe):

Vorname	Nachname	Status	Institution	Adresse
Gilles	Philippe	Professeur ordinaire	Université de Lausanne, Centre de recherche en langues et littératures comparées (CLE)	Unil-Dorigny, Bâtiment Anthropole, CH-1015 Lausanne gilles.philippe@unil.ch
François	Maniez	Professeur	Université Lyon 2, Directeur du Centre de recherche en terminologie et en traduction (E. A. 4162),	6 rue Pasteur, 69365 Lyon cedex 7 francois.maniez@univ-lyon2.fr
Teresa	Muryn	Professeur	Pädagogisches Institut Krakau, Polen ILLM ; Lehrstuhl für romanistische Linguistik	Instytut Neofilologii Uniwersitet Pedagogiczny Ul. Podchorazych 2 30-084 Krakow teresa.muryn@gmail.com
Małgorzata,	Niziolek	Maître de conférences	Pädagogisches Institut Krakau, Polen ILLM ; Lehrstuhl für romanistische Linguistik	Instytut Neofilologii Uniwersitet Pedagogiczny Ul. Podchorazych 2 30-084 Krakow mniziolek1@gmail.com

1.2. Thema des Vorhabens und Akronym

Phraseologie des Romans
PHRASEOROM

1.3. Fach- und Arbeitsrichtung

Linguistik; Korpuslinguistik; Phraseologie; maschinelle Sprachverarbeitung; Literaturwissenschaft und Stilistik; Textlinguistik; Kontrastive Linguistik; Übersetzungswissenschaft

1.4. Antragszeitraum

36 Monate

1.5. Zusammenfassung

Hauptanliegen des Projekts ist es, im Rahmen einer induktiven Vorgehensweise bzw. eines *corpus-driven approach* eine strukturelle und funktionelle Typologie lexikalisch-syntaktischer Konstruktionen (LSK) zu erarbeiten, die spezifisch für den deutschen, britischen und französischen Roman des 20. Jahrhunderts als meistrezipierter literarischer Gattung sind. Auf der Grundlage dieser Typologie werden zwei Arten vergleichender Analysen durchgeführt:

- a) zwischen der sogenannten Hochliteratur einerseits und der sogenannten Trivialliteratur andererseits (frz. *para-littérature*, engl. *popular literature*; Science-Fiction; Kriminalroman; Liebesroman);
- b) zwischen den stilistischen Praktiken, die sich in den literarischen Traditionen unterschiedlicher Länder (Vereinigtes Königreich, Frankreich und Deutschland) beobachten lassen.

Zunächst sollen statistische Berechnungen durchgeführt werden, die der Identifizierung von LSK dienen, die im Vergleich zu einem Referenzkorpus journalistischer und wissenschaftlicher Texte für literarische Texte spezifisch sind. Im Anschluss soll anhand großer Textkorpora untersucht werden, in welchem Umfang die ermittelten LSK eine Rolle bei der Konstruktion literarischer Texte spielen. Die linguistische Analyse auf semantischer, syntaktischer und textstruktureller Ebene wird dabei verbunden mit einer literaturvergleichenden stilistischen Untersuchung innerhalb einzelner literarischer Gattungen. Damit sollen die Grundlagen einer „Lexiko-Grammatik“ phraseologischer Ausdrücke in der Literatursprache gelegt werden, die ihren Niederschlag dann im Bereich der kontrastiven Linguistik und Stilistik ebenso wie in der Übersetzungswissenschaft finden soll.

Es handelt sich also um ein interdisziplinär angelegtes Projekt an der Schnittstelle zwischen Sprach- und Literaturwissenschaft, insbesondere zwischen Phraseologieforschung, Stilistik, Gattungstheorie, Korpuslinguistik und maschineller Sprachverarbeitung. Aufgrund seines Untersuchungsgegenstands (Phraseologismen im Roman) und der angewandten Methode (Korpuslinguistik im Sinne einer *linguistique outillée*, vgl. hierzu Habert 2005) ordnet sich das Projekt in den Bereich der *Digital Humanities* in den Geistes- und Sozialwissenschaften ein.

2. Stand der Forschung, eigene Vorarbeiten

2.1. Stand der Forschung

2.1.1 Wissenschaftlicher Kontext: Konvergenz zwischen literarischer Stilistik und Linguistik

Die Phraseologie erfährt derzeit zunehmende Aufmerksamkeit innerhalb zweier traditionell getrennter Forschungsstränge: der Korpuslinguistik und der stilistischen Analyse literarischer Texte.

Zum einen hat die Korpuslinguistik (im Sinne einer *linguistique outillée*, vgl. Habert 2005) Verfahren zur automatisierten Extraktion lexikalischer Einheiten entwickelt, die einen verbesserten Zugang zum Phänomen der Polylexikalität erlauben. Hinzu kommt eine breitere Reflexion in der Phraseologismus-Forschung zur Funktionsweise von Phraseologismen, zu deren Beschreibung auf semantischer (Kompositionalität, Durchsichtigkeit, Motiviertheit), syntaktischer (Integrierbarkeit in die phrastische und transphrastische Kombinatorik) und pragmatischer Ebene (Verwendungsrestriktionen auf Äußerungsebene) sowie zu Aspekten

des Phänomens ‚Phraseologismus‘ in einzelsprachbezogener und sprachvergleichender Hinsicht (vgl. Europhras 2014). Angestrebt wird ein umfassender Ansatz, der bei der Auseinandersetzung mit Phraseologismen die lexikalisch-semantische, die syntaktische und die textstrukturelle Ebene integriert (vgl. *Rencontres phraséologiques*, Grenoble 2013). Zum anderen haben neuere Forschungsansätze in der Stilanalyse die Existenz stilistischer Muster (*patrons stylistiques*) nachweisen können; damit steht ein nützliches Werkzeug zur Verfügung, um die Historizität nicht nur der literarischen Formen selbst, sondern auch die ihrer Rezeption zu erfassen (Maingueneau & Philippe 1999, Philippe & Piat 2009, Vaudrey-Luigi 2013).

Darüber hinaus widmen sich allein im Jahr 2015 mehrere internationale Symposien der Beziehung von Gattung bzw. Textsorte und Stil bzw. Grammatik: *Grammaire des genres et des styles*, Journée ConsCiLa, Januar 2015, Paris; GRATO-Symposium, Juli 2015, Lissabon; *Marqueurs et structures dans la (re)construction du sens*, November 2015, Paris-Nanterre; *Text and discourse in confrontation in the European context : towards an epistemological and heuristical renewal*, September 2015, Metz.

Im gegenwärtigen wissenschaftlichen Kontext bietet sich die Möglichkeit, die Synergieeffekte zu nutzen, die sich aus den Stärken dieser einander wechselseitig befruchtenden Ansätze ergeben. So müssen die Forschungsfragen und die stichprobenartigen Ergebnisse aus der Stilistik einer statistisch aussagekräftigen Überprüfung im Rahmen großer Korpora unterzogen werden. Umgekehrt ist literaturwissenschaftliche Expertise unerlässlich, um die Relevanz bestimmter Phänomene zu bestätigen, die im Rahmen der automatischen Textanalyse identifiziert worden sind. Die Ziele des vorliegenden Projekts knüpfen an diese Aspekte an und schließen die Lücke einer breit angelegten Analyse im Rahmen eines internationalen Projekts.

2.1.2 Linguistische Herangehensweisen an literarische Texte

Zahlreiche Arbeiten im Bereich der Stilistik (Barthes 1966, Greimas 1972/1982, Leech & Short 1981), der Stilometrie (Magri-Mourgues 2006), der Korpusstilistik (Stubbs 2005, Fischer-Starcke 2010, Malberg 2012) und der Textometrie (Guiraud 1954, Brunet 1981) befassen sich systematisch mit den Stilelementen und lexikalischen und grammatischen Besonderheiten bei verschiedenen Autoren (z.B. Flaubert, Proust, Dickens, Jane Austen) bzw. identifizieren rekurrente stereotype Schemata, z.B. für den Kriminalroman (Todorov 1971, Marion 2009, Lits 2011), um zu einer Typologie literarischer Texte zu gelangen. Mit Ausnahme von Legallois (2012) und Siepmann (2015b, 2016) sind die lexikalisch-syntaktischen Konstruktionen, die in literarischen Texten verwendet werden, bislang hingegen keiner systematischen, strikt linguistischen Untersuchung (auf semantischer, syntaktischer und diskursiver Ebene) unterzogen worden; dies ist insbesondere nicht auf der Grundlage großer Korpora und unter Verwendung des Instrumentariums der maschinellen Sprachverarbeitung erfolgt.

Überdies herrscht in der Linguistik weitgehend Einigkeit darüber, dass literarischen Texten insofern kein Sonderstatus zukommt, als die formalen Eigenschaften, die oftmals als spezifisch „literarisch“ kategorisiert werden, wie Metaphern oder Ironie, in gleicher Weise auch in anderen Texttypen auftreten (Burton & Carter 2006: 273). Diese Sicht liegt wiederum Versuchen zu Grunde, die Spezifität literarischer Texte über einen funktionellen Ansatz zu bestimmen, anhand von Kriterien wie dem der „Duplizität“ (Scholes 1982) der verschiedenen am Kommunikationsprozess beteiligten Faktoren (z.B. die Unterscheidung zwischen Autor und Erzähler). In dieser Sicht wäre der subjektive Eindruck von „Literarizität“ (Jakobson 1960), der sich bereits aus einem kurzen Romanausschnitt ergibt, nur eine Art zufälliges Epiphänomen, das sich dem schriftstellerischen Produkt dazugesellt. Diese Behauptung eines

funktionalen Unterschieds ohne Korrelat auf der formalen Ebene dürfte in der Linguistik einmalig sein und damit zusammenhängen, dass sich die Aufmerksamkeit eher auf bestimmte Autoren oder bestimmte Texte konzentriert hat als auf große literarische Korpora.

Unser Forschungsansatz beruht also auf der Hypothese, dass die Sprache des Romans, sowohl als Gesamtheit als auch in ihren gattungsspezifischen Ausprägungen, sich durch eine statistisch signifikante Überrepräsentation bestimmter sprachlicher Phänomene auszeichnet (Lexeme, Kollokationen, Kolligationen, Valenzschemata), welche eine wesentliche Rolle in der literarischen Konstruktion des Textes spielen. Dies schließt keineswegs aus, dass mit bestimmten Mustern auch gebrochen wird, wie bereits im russischen Formalismus postuliert wurde; in der Regel jedoch stellen Brüche entweder elaboriertere Varianten bestimmter referentieller Stereotypen dar (*a silence fell* → *a blanket of strained silence instantly descended*) oder stellen Varianten („stéréotypes du deuxième degré“, Dufays 2010: 239, z.B. *an economically dressed nymphet* statt *a scantily dressed young girl*) dar, die sich auf verschiedene Weise von diesen Stereotypen absetzen (Siepmann 2015a: 106-107).

Die hier formulierte Hypothese hat auf sehr allgemeiner Ebene bereits eine erste Bestätigung erfahren durch die Pionierarbeit von Stubbs & Barth (2003) zum Englischen, die zu folgendem Ergebnis gelangen: „text types are distinguished by lexical and grammatical patterns“ (2003: 79). Allerdings stützt sich diese Studie nur auf ein quantitativ begrenztes Korpus und auf die Ermittlung der 200 häufigsten n-grams in jeder Textsorte, wodurch gerade die stärker lexikalisch geprägten Einheiten, welche wiederum die Wahrnehmung des Lesers stark beeinflussen, mehr oder weniger ausgeschlossen bleiben.

Über die Besonderheiten des englischsprachigen literarischen Diskurses informiert ferner die deskriptive Grammatik von Biber u.a. (1999), in der spezifisch literatursprachliche grammatische Eigenschaften identifiziert werden, wie z.B. der doppelte Genitiv oder Existenzsätze mit *there* in Verbindung mit anderen Verben als *to be*.

So wichtig eine solche Erfassung grammatischer Eigenschaften ist, so erlaubt sie doch nicht die Identifizierung spezifischer Kombinationen bestimmter grammatischer Kategorien, d.h. es mangelt an einer Erfassung der syntagmatischen Sequenzen, die in einer bestimmten Gattung zu Lasten anderer Konstruktionen bevorzugt werden. Um diese Lücke im Bereich des Französischen zu schließen, verwendet Legallois (2012) die Methode der sogenannten *motifs séquentiels* (vgl. 2.1.4), mit der die syntagmatischen Sequenzen eines Texts erfasst und die lexikalisch-grammatischen Besonderheiten von Gattungen oder Autoren beschrieben werden können. Legallois kann so bei acht verschiedenen Autoren des 19. Jahrhunderts das *motif le N1 du N2 qui V dans* nachweisen, das als Rahmen für lexikalische Füllungen u.a. im Bereich der folgenden Isotopien dient: Wind {*vent, brise, bise*}, Nacht {*nuit, soir*}, Tod {*mourir, gémir, pleurer, tomber, se briser, abîme*}. Im Lichte solcher Beispiele können *motifs séquentiels* als diskursive Einheiten verstanden werden, die an einen bestimmten Typ von phraseologischen Einheiten gebunden sind und als Hinweise auf intra- und intertextuelle Beziehungen fungieren, die ihrerseits wiederum einer genaueren Untersuchung im Rahmen der literarischen Stilistik sowie Gattungstheorie und -geschichte bedürfen. Die Vorgehensweise von Legallois stellt dabei eine Weiterentwicklung und Anwendung auf grammatische Konstruktionen der von Greaves (2005) entwickelten Methode der *congrams* dar.

In der Studie von Siepmann (2016, erscheint), die an diese Arbeiten unmittelbar anschließt, wird ein Korpus englischsprachiger Romane im Umfang von 160 Mio. Wörtern induktiv mit Blick auf drei Aspekte analysiert: a) *key word analysis*, b) Analyse von wiederholt auftretenden *n-grams* (Lebart & Salem 1988) bzw. *lexical bundles* (vgl. Biber et al. 1999), c) detaillierte Analyse von drei *key words* unterschiedlichen Typs (*thought, sun, jerk*).

Siepmann identifiziert 12 Hauptklassen englischer *key words* (z.B. Körperteilbezeichnungen, Natur und Umwelt, Bewegungsverben). Ausgehend von *key words* kann man zeigen, wie diese dann in die Bildung längerer Ketten eingehen: Fast alle der häufigsten *lexical bundles*, die mit *like* gebildet sind, basieren auf *key words* (z.B. *sun was like a, room was like a, voice was like a*). Siepmann (2015b) zeigt u.a., dass die Frequenz bestimmter äquivalenter *key words* (z.B. die Substantive *breasts* und *seins*, das Verb *nod* und seine französischen Entsprechungen *faire un signe de la tête, acquiescer de la tête, hocher le front dodeliner de la tête* usw.) deutlich je nach Einzelsprache variieren.

Eine genauere Analyse von drei *key words* (*thought, sun, jerk*) macht schließlich deutlich, dass eine beachtliche Zahl von lexikalisch-grammatischen Sequenzen spezifisch für den Roman sind; von diesen wiederum können viele als *extended lexical items* im Sinne Sinclairs beschrieben werden (z.B. der Übergangsmarker „Genitiv + *thoughts were on* + NP“). Kontrastive Analysen in Siepmann (2015b) zeigen, dass englische Strukturen des Typs *the sun trickled through the trees* vielfach über Veränderungen im verbalen Slot (vgl. *slant/pour/lick/seep/... through*) variiert werden, während im Französischen der Ausdruck desselben Bedeutungsinhalts wesentlich komplexer ausfällt (z.B. *des gouttes de soleil pénétraient à travers le feuillage*).

2.1.3 Ansätze der Idiomatizitäts- und der Phraseologieforschung

Während es also kaum Arbeiten zur lexikalisch-grammatischen Beschaffenheit literarischer Texte gibt, beschäftigen sich zahlreiche Studien mit idiomatischen Kombinationen in anderen Genres wie journalistischen und wissenschaftlichen Texten (Sinclair 1991/2004, Hunston & Francis 2000, Hoey 2005, de Beaugrande 2005). Idiomatizität stellt also ein allgemeines sprachliches Funktionsprinzip dar, das sich in Texten durch ein breites Spektrum von Phraseologismen manifestiert, die sich ihrerseits derzeit noch einer umfassenden und allgemein akzeptierten Definition entziehen: So ist die Rede von „extended units of meaning“ (Sinclair 2004), „constructions“ (Goldberg 1995), „collostructions“ (Stefanowitsch & Gries 2003), „collocations“ (Hausmann 1979, Mel’cuk et al. 1995, Siepmann 2005, Tutin 2010), „lexical bundles“ (Biber u.a. 1999), „motifs séquentiels“ (Quiniou u.a. 2012) oder „usuellen Wortkombinationen“ (Steyer 2013).

Gleichwohl ist derzeit eine zunehmende Konvergenz dieser Ansätze zu verzeichnen, deren Gemeinsamkeit in der Abkehr von der Unterscheidung zwischen Grammatik (Regeln) und Wortschatz (Wörter und Wendungen) besteht. Die dem Firthschen Ansatz verpflichtete Schule, deren fortgeschrittenste Richtung zweifellos die *Lexical Priming Theory* von Hoey (2005) ist, postuliert eine Lexikon-Grammatik, die Kombinationen eher grammatischer Natur („colligations“: z.B. NP + *to be* + *about* + V-ing: „probability is about calculating the likelihood of various outcomes from particular events“) und Kombinationen eher lexikalischer Natur umfasst („collocations“: z.B. *clear motorway, I couldn’t agree more*). Die Konstruktionsgrammatik hingegen fasst Sprache als ein Inventar von konventionalisierten Form-Funktions-Paaren auf, die auf einem Kontinuum vom Lexikon über idiomatische Sequenzen bis hin zu grammatischen Strukturen angeordnet sind (z.B. Kay & Fillmore 1999, Goldberg 1995, Croft 2001). Zwischen dem Kontextualismus und bestimmten Richtungen der Konstruktionsgrammatik, wie etwa den „collostructions“ von Stefanowitsch & Gries (2003), gibt es dabei offensichtliche Verwandtschaften. Während Vertreter der erstgenannten Richtung von Strukturen allgemeinen Charakters ausgehen, wie etwa der ditransitiven Konstruktion (V + NP + NP) und sich für deren lexikalische Füllung interessieren, schlagen die Letztgenannten den umgekehrten Weg vom Einzellexem zum abstrakten Aktantenschema (z.B. *give*) ein. Pragmatisch-soziolinguistische Ansätze wie Feilke (1994, 1996, 2003) schließlich verweisen durch Rückgriff auf soziologische Theorien auf die soziale Konstitution idiomatischen Sprachgebrauchs.

Sie rücken das Phänomen idiomatischer Prägung von der Peripherie in das Zentrum der Funktionsweise von Sprache und kennzeichnen Sprachkompetenz folgerichtig als „Common-Sense-Kompetenz“, die auf möglichen und im sozialen Handeln bestätigten Selektionen aufbaut.

Die starke Konvergenz dieser verschiedenen Ansätze führt dazu, dass sich die Aufmerksamkeit der Forschung von den festen Sequenzen der traditionellen Phraseologie-Forschung (Routineformeln, Sprichwörter, zweigliedrige Kollokationen; Hausmann 1979, Mel'čuk 1998, Tutin 2010) hin zu allen Typen von sprachlichen „Prägungen“ (Feilke 1996) und *extended lexical items* (Sinclair 2004) verlagert. In dieser Perspektive besonders vielversprechende Begriffe sind „collocational framework“ (Renouf & Sinclair 1991), „motif“ (vgl. oben; Legallois 2006, Longrée & Mellet 2013) oder „Probabem“ (Herbst & Klotz 2003), d.h. eine polylexikalische Sequenz, die Sprecher mit einer gewissen Wahrscheinlichkeit verwenden (z.B. *il n'avait pas de mots assez durs pour INF* (Hausmann 2007: 136); vgl. *?il avait des mots durs pour INF*).

Die umfassendste Modellierung kombinatorischer Phänomene findet sich in der *lexical priming*-Theorie von Hoey (2005), der sich des Begriffs des „Kollokationsnests“ bedient (nicht-kompositional, z.B. *say a word -> say a word against -> won't say a word against*, Hoey 2005: 11). Weiter ergänzt Hoey die Beschreibung „lokaler“, d.h. satzinterner lexikalisch-syntaktischer Beziehungen um die Analyse von lexikalisch-syntaktischen Beziehungen innerhalb größerer Einheiten auf Textebene.

Nichtsdestoweniger ist der Bereich der Kollokations- und Kolligationsnester noch nicht detaillierter auf der Grundlage größerer Korpora untersucht worden, vor allem nicht im Sprachkontrast. Im Rahmen von EMOLEX sind aber gleichwohl automatisierte Verfahren entwickelt worden, die es erlauben, die Ebene von Zweiwort-Kombinationen zu überschreiten (vgl. unten).

2.1.4 Maschinelle Sprachverarbeitung und Extraktion lexikalisch-syntaktischer Konstruktionen

Neuere Verfahren aus dem Bereich der maschinellen Sprachverarbeitung erlauben es nunmehr, die gängige Ebene der Verarbeitung einfacher N-Gramme von Wortformen zu überschreiten. Ein vielversprechender Ansatz in diesem Bereich ist derjenige von Quiniou *u.a.* (2012) und Legallois (2012), die Oberflächensequenzen mittels einer aus dem *Data Mining* bekannten Technik erschließen: die Extraktion sogenannter *motifs séquentiels*. Solche *motifs séquentiels* werden hier zunächst als kontinuierliche oder diskontinuierliche Einheiten aufgefasst, die sich aus in einer linearen Abfolge stehenden Elementen zusammensetzen, bei denen es sich um sog. *items* oder um *item-sets* handeln kann, wobei ein *item* eine einzelne Information, z.B. eine Wortform, repräsentiert, während bei *item-sets* eine Kombination mehrerer Typen von Informationen vorliegt, z.B. Wortformen, Lemmata oder morphosyntaktische Kategorien. Solche *motifs* können als Erweiterung rekurrenter Wortformenabfolgen (sog. *segments répétés*) oder N-Gramme aufgefasst werden: Während es sich bei letzteren um kontinuierliche Wortfolgen handelt (z.B. *il était une fois*), erlauben die *motifs séquentiels*, die überdies auf verschiedenen Generalisierungsstufen ermittelt werden können, auch die Berücksichtigung frei bestimmbarer Lücken, sog. *gaps*, zwischen den Elementen. Die *motifs* stehen damit in einer Reihe mit den von Fletcher (2007) vorgeschlagenen *phrase-frames* (mit dem Unterschied, dass *phrase-frames* lediglich eine Lücke aufweisen, z.B. ** was like a*) und den von Greaves (2005) vorgeschlagenen *congrams* (alle Kookkurrenzen mehrerer Wörter unabhängig von ihrer Morphologie oder ihrer Position im Satz). Nach Legallois (2012) ist z.B. das folgende *motif* spezifisch für den Stil Zolas (auf der Grundlage der in der Textometrie verwendeten Spezifizitätsberechnung und im Vergleich mit einem Korpus von Romanen des 19.

Jahrhunderts): *il V le N de DETPOSS N ADJ* (z.B. *elle écarta les toiles de ses mains tremblantes* oder *elle tournait les pages, de ses doigts minces*).

Ein weiterer innovativer Ansatz, der über die einfache lineare Wortfolge hinausgeht, liegt mit Kraif & Diwersy (2012) vor. Dieser Ansatz nutzt die aktuelle Entwicklung leistungsfähiger Parser (XIP, Malt Parser, Connexor, IDP Parser, Mate Tools, Talismane, FIPS, DeSR...), um aus großen Korpora syntaktische Abhängigkeitsbeziehungen automatisch und mit großer Genauigkeit zu extrahieren. Damit eröffnet sich für die Phraseologieforschung ein gänzlich neues Forschungsfeld, denn die auf syntaktischen Annotationen beruhende automatische Kookkurrenzanalyse birgt unter Präzisionsgesichtspunkten erhebliche Vorteile gegenüber fensterbasierten Ansätzen (vgl. Evert 2008). Im Anschluss an Seretan u.a. (2003) haben sich Charest u.a. (2010) zur Kollokationsextraktion für das Wörterbuch *Antidote RX* auf diesen Typ von Kookkurrenz gestützt, wobei sie das Feld um multilexematische Ausdrücke mit mehr als zwei Wörtern erweitert haben.

Unter Anwendung eines iterativen Extraktionsverfahrens haben Kraif und Diwersy ausgehend von dem Begriff mehrteiliger Nodes eine Methode entwickelt, die es erlaubt, ohne vorgängige Berechnungen rekurrente Ausdrücke der Länge n mit einem gegebenen Node zu extrahieren. Diese Ausdrücke überschreiten den einfachen Rahmen rekurrenter Wortformenabfolgen oder *lexical bundles* und beruhen auf rekurrenten syntaktischen Teilbäumen, die sich in unterschiedlicher Weise an der Textoberfläche manifestieren können (vgl. dazu Abb. 1; Extraktion mit dem Werkzeug EmoConc [<http://emolex.u-grenoble3.fr/emoBase/>]).

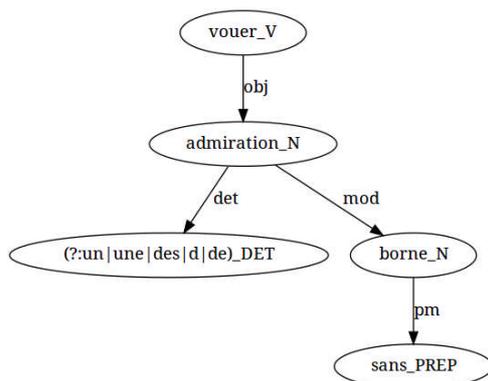


Abbildung 1: Automatische Extraktion des Teilbaums zum Ausdruck *vouer une admiration sans bornes*

2.1.5 Stilistische und gattungstheoretische Analysen

Die Beschäftigung mit der Literatursprache ist eine neuere Entwicklung im Bereich der Stilistik: G. Philippe und J. Piat haben gezeigt, dass seit der Verselbständigung des literarischen Feldes in Frankreich um 1850 (Bourdieu 1992, Vaillant 2003) im literarischen Kontext eine andere Sprechweise entstanden ist. Betrachtet man nämlich den Diskurs der Schriftsteller über ihre eigene Praxis und vergleicht man Generation für Generation diese Praktiken, so identifiziert man historisch relevante Rekurrenzen (etwa Formulierungen des Typs *la propriété des vitres* anstelle von *des vitres propres* gegen Ende des 19. Jahrhunderts oder die große Zahl metadiskursiver Zwischenbemerkungen um 1960); verfolgt man zusätzlich den auf die Literatur bezogenen Diskurs von Kritikern und Grammatikern, so wird deutlich, dass sich nach und nach eine spezifische Literatursprache konstituiert hat – verwiesen werden kann hier ebenso auf die Verbannung literarischer Beispiele aus den Grammatiken wie auf die Diskussion der 1920er Jahre über den „Stil“ Flauberts oder des Waadtländers Ramuz.

Die Identifizierung stilistischer Muster ist eines der Hauptverdienste dieser Forschungsrichtung (vgl. 2.1.1). Im Sinne einer Reihe kookkurrenter Merkmale, die in konventionalisierter (und damit historischer) Weise einen bestimmten Effekt auf der Ebene des Texts hervorrufen, bieten diese Merkmalskonfigurationen ein äußerst fruchtbares interdisziplinäres Forschungsfeld, das neue Perspektiven auf die Historizität literarischer Formen und die linguistisch basierte Kategorisierung von Gattungen eröffnet. So hat Piat (2011) gezeigt, dass der Stil des *Nouveau Roman* der 1950er Jahre auf der Aktivierung eines phänomenologischen Musters beruht: hier wurden extrem kurze oder lange orthographische Sätze mit syntaktisch-diskursiver Diskontinuität in Verbindung gebracht, um so verschiedene mentale Operationen zu repräsentieren.

Hinsichtlich der Historizität literarischer Formen haben Philippe & Piat (2009) ausgehend von einer Reihe generationenbezogener Studien nachgezeichnet, welche stilistischen Muster sich jeweils ausprägen und die französische Prosa zu einem bestimmten Zeitpunkt ihrer Geschichte dominieren. Dadurch gelingt es ihnen, Stil im Sinne einer historischen Praxis zu reflektieren und das literarische Feld zu kartographieren, indem die Orte bestimmt werden, an denen die Praxis bestimmter Schriftsteller übereinstimmt, sich aber von derjenigen anderer wiederum unterscheidet.

Ebenso konnten stereotypisierte lexikalische Kollokationen herausgearbeitet werden, die besonders häufig in bestimmten Gattungen wie dem Feuilleton-Roman auftreten; wird dort z.B. ein Rächer in Szene gesetzt, findet man vorrangig die Folge „NP, die den Arm oder die Hand bezeichnet + *d'acier*“ (z.B. „*Sous la peau délicate et douce de cette main qui vint le saisir brusquement à la gorge, le Chourineur sentit se tendre des nerfs et des muscles d'acier*“ [E. Sue]). Der Feuilleton-Roman entnimmt zu seiner Zeit viele seiner Themen und seiner Darstellungsweisen der Rubrik „Vermischtes“ in Zeitungen; diese beiden Diskursgattungen führen dann in den 1860er Jahren zum Kriminalroman, als dessen erster französischsprachiger Vertreter Émile Gaboriaus *L’Affaire Lerouge* von 1866 gilt. Der Feuilleton- und der Kriminalroman wurden als Subtypen von Trivilliteratur aufgefasst, aber diese Kategorisierung ist gleichwohl zu hinterfragen, und zwar sowohl hinsichtlich ihrer Relevanz als auch hinsichtlich der außerordentlichen (stilistischen wie literarischen) Heterogenität dessen, was man unter der Gattung ‚Feuilleton-Roman‘ bzw. ‚Kriminalroman‘ fasst.

Mit Bezug auf gattungstheoretische Fragen ist darüber hinaus festzustellen, dass bisherige Arbeiten in der Gattungstheorie meist recht deutlich zwischen Hoch- und Trivilliteratur unterscheiden sowie auf auf inhaltsseitige Kriterien der Demarkierung von Gattungen zurückgreifen, wie z.B. die Beschreibung bestimmter emotionaler Zustände in Liebesromanen (Gymnich, Neumann und Nünning, 2007, Zymner 2003, Frow 2006, Duff 2000, Monte & Philippe 2014). Mit dem Ziel einer Erweiterung der Gattungstheorie um neue Perspektiven soll diese rein inhaltliche Bestimmung von Generizität im Projektverlauf durch ausdrucksseitige ergänzt werden, wobei idealiter Korrespondenzen zwischen Inhalts- und Ausdrucksseite herausgearbeitet werden sollen (vgl. a. Beauvisage 2001, Rastier, 2011). Auf der Folie der Erkenntnisse zu literarischen Stereotypen können Neubewertungen bzw. -interpretationen bestimmter literarischer Strömungen sowie in der Folge auch von Autoren und Einzelwerken vorgenommen werden.

2.1.6 Kontrastive Linguistik und literarische Übersetzung

Trotz der zunehmenden Kooperation zwischen Kontrastivlinguisten und Übersetzungswissenschaftlern, wie sie aus den thematischen Schwerpunktsetzungen einer beachtlichen Zahl einschlägiger Zeitschriften (z.B. *Languages in Contrast*) und der regelmäßigen Durchführung von Tagungen deutlich wird (etwa *Using Corpora in Contrastive and Translation Studies* [UC-

CTS]), bleibt die Sekundärliteratur zur Interdependenz zwischen kontrastiver Linguistik und literarischer Übersetzung nach wie vor sehr überschaubar. Insbesondere gibt es kaum neuere Arbeiten zu systematischen Ähnlichkeiten und Unterschieden zwischen dem literarischen Sprachgebrauch bei franko-, anglo- und germanophonen Autoren. Tatsächlich kann die fehlende Vertrautheit mit diesen Ähnlichkeiten und Unterschieden insbesondere im Bereich der Phraseologismen u.U. zu unzutreffenden Deutungen der Übersetzungsdaten seitens der Übersetzungswissenschaftler führen (Granger et al. 2003: 25) und stellt somit auch eine häufige Fehlerquelle in der Übersetzungspraxis dar. Darüber hinaus stellt die kontrastive Linguistik insofern eine Hilfsdisziplin für die Übersetzungskritik dar, als die Ergebnisse kontrastiver sprachwissenschaftlicher Analysen es deutlich erleichtern, die Angemessenheit der von literarischen Übersetzern vorgeschlagenen Übersetzungsäquivalente zu beurteilen.

Wir gehen von der Hypothese aus, dass man zwischen zwei Sprachen systematische Entsprechungen literarischer Phraseologismen identifizieren kann (*X interrupted his thoughts* – *X le tira de ses réflexions*; vgl. Siepmann 2015b), von denen manche wiederum für bestimmte Gattungen typisch sind. Damit sind zwei Forschungsrichtungen eröffnet: Zunächst ist die Frage von Interesse, in welchem Umfang Schriftsteller als Angehörige unterschiedlicher Sprachgemeinschaften bei gleicher Ausdrucksabsicht analoge oder divergente syntagmatische Verbindungen kreieren; neben Entsprechungen zwischen rekurrenten Sequenzen wie *X interrupted his thoughts/X le tira de ses réflexions* lassen sich bei Romanen trotz unterschiedlicher Sprachen nämlich auch *a priori* überraschende Äquivalenzen bei Syntagmen geringer Frequenz feststellen, wie z.B. *un jour pluvieux se déversait dans la chambre* (Henri Troyat) et *a sunny rainy light filled the room* (Iris Murdoch) oder *qualmiges Fackellicht* (Thomas Mann) et *la lumière fuligineuse des torches* (Maurice Genevoix) (vgl. Gallagher 2007). Zum anderen existieren einzelsprachspezifische Konfigurationen wie die *extended lexical items* mit dem englischen Substantiv *sun* als Basis und Partikelverben, die das Einfallen des Lichts an einem bestimmten Ort beschreiben (z.B. *pour, slant, trickle [through the trees]*), wodurch bestimmte aspektuelle Differenzierungen möglich werden; die für das Französische typischen äquivalenten Konfigurationen hingegen sind eher abschwächend oder semantisch neutral (*le soleil filtrait/glissait entre les [Baumname]*; vgl. Siepmann 2015b). Die Auswirkungen solcher Ergebnisse auf die Übersetzungstheorie und die Übersetzungspraxis sind umso größer, als der Übersetzer zwischen einer normalisierenden, aber semantisch schwächeren und einer vollumfänglichen Übersetzung schwanken wird; letztere verlangt aber häufig komplexe Transpositionen, denen die aktuellen Theorien zu unterschiedlichen Lexikalisierungsgraden in verschiedenen Sprachen, seien es die vergleichende Stilistik oder Ansätze wie Talmy (2000), nicht wirklich gerecht werden.

Solche Analysen erlauben es, zahlreiche, in der vergleichenden Stilistik seit langem diskutierte Phänomene (Malblanc 1961, Vinay & Darbelnet 1958, Truffaut 1983, Grünbeck 1976, François 1989, Chuquet & Paillard 1987, Ballard 1992) präziser zu erfassen und dem Mangel an Interesse gegenzusteuern, das in diesem Bereich der Gattungsfrage und der Historizität sprachlicher und literarischer Formen immer noch entgegengebracht wird - wird hier doch immer noch davon ausgegangen, dass einzelsprachliche Unterschiede innerhalb bestimmter Textgattungen für die jeweiligen Sprachen insgesamt und unabhängig von deren historischem Entwicklungsstand gelten. Diese unhinterfragten Vorannahmen verhindern, dass bestimmte, für die Literatursprache spezifische Phänomene wie die o.g. stilistischen Muster oder spezifisch literarische Kollokationen (z.B. *nod in agreement, jerk awake*, vgl. Siepmann 2016, erscheint) in den Blick der Forschung rücken.

2.2 Eigene Vorarbeiten zum Thema

Grenoble

Die an diesem Projekt beteiligten Forscher des LIDILEM arbeiten seit geraumer Zeit im Bereich der Phraseologie (vgl. u.a. Tutin 2013). So wurden im ANR-DFG-Projekt EMOLEX ein funktionales Modell der Analyse von Kollokationen des Gefühlswortschatzes (Novakova & Melnikova 2013, Novakova 2015) sowie neue theoretische und methodische Zugänge zur Untersuchung des diskursiven Verhaltens derselben erarbeitet (Novakova & Sorba 2013a/b, 2014a et b).

Weitere Arbeiten beziehen sich auf die Phraseologie von Wissenschaftstexten (ANR-Projekt Scientext, <http://scientext.msh-alpes.fr>), darunter u.a. ein jüngst erschienener Sammelband, der einen konzentrierten Überblick über die Forschung zum Gebrauch von Phraseologismen zu Zwecken der Positionierung des Autors gegenüber Vorgängern und Peers bietet (Tutin & Grossmann 2014). Die dabei gewonnene Expertise in der Ermittlung von wissenschaftssprachlichen Diskursroutinen kann nutzbringend in der kontrastierenden Analyse von Romangattungen im aktuellen Projekt eingesetzt werden.

Im Bereich der maschinellen Sprachverarbeitung war Grenoble an zahlreichen Projekten beteiligt, bei denen die Entwicklung und Nutzung großer mehrsprachiger Korpora von entscheidender Bedeutung war (Carmel, Scientext, Emolex). Die Grenobler Forscher können somit auf eine reichhaltige Erfahrung bei der Erstellung von komplexen Verarbeitungsketten (Annotation, Alignment, Indexierung) sowie Abfragewerkzeugen und Schnittstellen zurückgreifen (Kraif 2008, Kraif & Diwersy 2012, 2014). Die Arbeit mit bilingualen Korpora und Parallelkorpora hat dabei zu methodischen Neuansätzen auf dem Gebiet der Korpuslinguistik geführt (Melnikova/Novakova/Kraif 2009). Von großem Nutzen für diesen Bereich des Projekt sind andererseits die Arbeiten von D. Legallois (Crisco, Universität Caen) zum Einsatz von *Data-Mining*-Techniken mit dem Ziel der Ermittlung kollokationeller und kolligationseller Frameworks („motifs séquentiels“; vgl. Legallois 2006, 2012). Legallois hat darüber hinaus in Zusammenarbeit mit A. Tutin u.a. einen Zeitschriftenband zum aktuellen Stand der Phraseologieforschung vorgelegt (Legallois & Tutin 2013).

Einen weiteren Anknüpfungspunkt für das hier in Rede stehende Projekt stellen die Grenobler Arbeiten der Gruppe EA Litt&Arts (Traverses 19-21) auf dem Gebiet der Stilistik und historischen Poetik da. Darunter fallen die Werke von J. Piat (2006, 2011, 2013) zur Literatursprache, in denen die theoretische und methodische Grundlage für den Begriff des „patron stylistique“ (etwa: Stilmuster) gelegt werden, sowie die Arbeiten von L. Gonon (2010, 2012) zu festen Wendungen und medizinischem, kriminalistischem und literarischem Interdiskursen in Zeitungsmeldungen des 19. Jhds. und zu Kollokationen, Phraseologismen und Gemeinplätzen in Serienromanen.

Osnabrück

Die Osnabrücker Arbeiten (D. Siepmann und B. Kern) sind im Bereich der Phraseologie bzw. Kollokationsforschung und der Kontrastivlinguistik angesiedelt. Siepmann (2005a) legt eine Taxinomie von Mehrworteinheiten mit diskursiver Funktion („Mehrwortdiskursmarker“) vor und analysiert deren textuelles Verhalten in drei Sprachen. Auf der Grundlage eines korpusgesteuerten kontextualistischen Ansatzes nimmt Siepmann (2002, 2004, 2005b) eine Neudefinition des Kollokationsbegriffs vor. Siepmann (2011) erweitert das *idiom principle* Sinclairs (1991) durch die Postulation eines Kreativitätsprinzips. Im Rahmen von EMOLEX hat Siepmann (2014b/2014c) zwei Studien zu semantischen Differenzen zwischen Kollokationen des deutschen, englischen und französischen Gefühlswortschatzes vorgelegt. B. Kern (2014) hat in Zusammenarbeit mit A. Grutschus zwei kontrastive Studien zur Polarität im

Gefühlswortschatz und zum diskursiven Verhalten von Gefühlsnomina verfasst. Siepmann (2015a/b; 2016, erscheint) sind grundlegende Arbeiten zur Thematik dieses Antrags. Siepmann (2012) fasst zahlreiche korpusbasierte Einzeluntersuchungen zur Phraseologie des englischen Wissenschaftswortschatzes und zu anderen Eigenschaften der englischen Wissenschaftssprache in einem Schreiblehrbuch für deutschsprachige Wissenschaftler zusammen. Im Bereich der maschinellen Sprachdatenverarbeitung verfügt Osnabrück über umfangreiche Erfahrungen im Bereich der Erstellung von Textarchiven und Korpora. In Zusammenarbeit mit Köln (S. Diwersy) und der romanistischen Fremdsprachendidaktik in Osnabrück ist in Osnabrück das erste Referenzkorpus der französischen Gegenwartssprache entstanden (Siepmann/Bürgel 2014).

Bonn

M. Gymnich hat umfangreiche Erfahrungen mit Fragestellungen, die an der Schnittstelle von Literatur-, Kultur- und Sprachwissenschaft angesiedelt sind. Das Interesse an diesem Bereich schlägt sich besonders deutlich in ihrer Habilitationsschrift zum Thema *Metasprachliche Reflexionen und sprachliche Gestaltungsmittel im englischsprachigen postkolonialen und interkulturellen Roman* (2007) nieder, in der auf linguistische wie auch auf literaturwissenschaftliche Analysekategorien zurückgegriffen wird und in der u.a. Fragen der Semantisierung einer Verwendung sprachlicher Varietäten durch verschiedene textuelle Sprecher (Figuren, unterschiedliche Typen von Erzählinstanzen) systematisch untersucht werden. Der englischsprachige Roman (aus Großbritannien, den USA und postkolonialen Literaturen) im 20. Jahrhundert bildet bereits seit der Dissertation (2000 veröffentlicht) einen Schwerpunkt von M. Gymnichts Forschung. Sie hat sich sowohl mit literaturgeschichtlichen Aspekten beschäftigt als auch mit Gattungen wie dem weiblichen Bildungsroman und dem Kriminalroman. Verschiedene Publikationen in den Bereichen Gattungstheorie (u.a. 2007, 2010, 2011) und Narratologie (u.a. 2002, 2013) unterstreichen zwei wichtige Schwerpunkte ihrer Forschung im Bereich Literaturtheorie.

Aufgrund seiner literatur- und sprachvergleichenden Ausrichtung und dem Interesse an sprachlichen Strukturen bietet das Projekt Anknüpfungspunkte zum Bonner Zentrum für Transkulturelle Narratologie, in dem M. Gymnich Gründungs- und Vorstandsmitglied ist. In diesem interdisziplinären Zentrum arbeiten Vertreter verschiedener Philologien zusammen.

Erlangen

Die Arbeiten von L. Fesenmeier fallen in den Bereich der distinktiven Synonymik und Kontrastivlinguistik (Französisch, Italienisch, Spanisch). Von der Feststellung ausgehend, dass der Großteil der Synonymie-Forschung nur die Paradigmatik in den Blick nimmt (typisch z.B. *savoir / connaître*), richtet L. Fesenmeier in seiner Habilitationsschrift sein Hauptaugenmerk auf Synonyme, deren kollokationelles Verhalten mehr oder weniger stark überlappt (z.B. *se souvenir / se rappeler*). Unter Verwendung eines korpusgesteuerten Ansatzes und auf der Grundlage großer journalistischer und literarischer Korpora zeigen Fesenmeier (2008a/b, 2010) und Grutschus/Fesenmeier (2013), dass Differenzen unter Synonymen auf unterschiedliche sprachliche Konzeptualisierungen zurückzuführen sind. Fesenmeiers (2008, 2010, 2013) kontrastive Arbeiten weisen darüber hinaus nach, dass die romanischen Sprachen sich hinsichtlich der Kombinatorik synonymischer Paare (*saber / sapere, conocer / conoscere*) recht unterschiedlich verhalten. Daraus ergeben sich zahlreiche neue translatorische Perspektiven und Forschungsfragen.

Das hier in Rede stehende Projekt hat Bezüge zur Erlangener Forschung am Interdisziplinären Zentrum für Dialekte und Sprachvariation und am Interdisziplinären Zentrum für Lexikographie, Valenz und Kollokation, dem sowohl L. Fesenmeier als auch D. Siepmann (als externes Mitglied) angehören.

Köln

Die Kölner Forschung befasst sich in erster Linie mit der Entwicklung von Methoden der statistischen Auswertung von Textdaten. Dabei liegt ein Schwerpunkt auf dem Aufbau von Großkorpora für verschiedene Sprachen sowie der maschinellen Verarbeitung derselben und der Entwicklung von geeigneten Abfrage- und Rechenwerkzeugen. Ein Teil dieser Arbeiten wurde im EMOLEX-Projekt geleistet (Diwersy et al. 2015). Daneben ist ein sehr umfangreiches Korpus der frankophonen Presse erstellt worden (Varitext: <http://syrah.uni-koeln.de/varitext/>). Weitere Arbeiten im Rahmen des DFG-ANR-Projektes PRESTO beziehen sich auf die Erarbeitung einer Methodik zur Ermittlung textueller Kolligationen. Schließlich hat die Kölner Seite entscheidende Beiträge zur Entwicklung der mehrsprachigen Plattform EmoBase im Rahmen des EMOLEX-Projektes geleistet (vgl. zu den Kölner Arbeiten: Diwersy (2012a, 2012b), Diwersy & Fesenmeier (2006), Diwersy & François (2011), Diwersy & Kraif (2013), Diwersy, Evert & Neumann (2014)).

Die Kompetenzen der deutschen und französischen Arbeitsgruppen erweisen sich in allen Bereichen als in höchstem Maße komplementär. Für den Erfolg des Vorhabens sprechen die zahlreichen bereits durchgeführten Projekte (Symposien und Tagungen, Herausgabe von Zeitschriftenbänden, EMOLEX) und die im Vorfeld angebahnte Zusammenarbeit mit Kollegen aus der literarischen Stilistik, der Bonner Anglistik und der Erlangener Romanistik. Hinzu kommt die für die verschiedenen Bereiche des Projekts eingeworbene Expertise von sechs Kooperationspartnern an anderen Standorten – P. Blumenthal (Linguistik), G. Philippe (Stilistik), F. Maniez und John D. Gallagher (Übersetzungswissenschaft), T. Muryn und M. Niziolek (kontrastive Phraseologie des Kriminalromans).

3. Ziele und Arbeitsprogramm

3.1 Ziele

Das Projekt verfolgt die im Folgenden genannten Ziele:

- a) im Bereich der maschinellen Sprachverarbeitung soll eine korpusinduzierte Extraktion lexiko-syntaktischer Mehrwortkonstruktionen (Drei-Wort-Konstruktionen und höher) erfolgen; diese geht von komplexen Ausdrücken aus, die automatisch in der Architektur der EmoBase-Datenbank ermittelt werden (vgl. Abb. 1);
- b) im Bereich der praktischen Anwendungen sollen dem Emoconc-Modul der EmoBase-Datenbank neue Funktionalitäten hinzugefügt werden: 1) Werkzeuge zur Messung der statistischen Spezifität der untersuchten Konstruktionen in Bezug auf verschiedene Gattungen und Diskurstypen sowie 2) Werkzeuge zur Ermittlung von satzübergreifenden Mustern;
- c) im Bereich der sprachwissenschaftlichen Forschung soll eine strukturelle und funktionale Typologie lexiko-syntaktischer Konstruktionen, die spezifisch für deutschsprachige, französischsprachige und englischsprachige Romane sind, erstellt werden; auf der Grundlage dieser Typologie werden die spezifischen diskursiven Vorkommenskontexte dieser Konstruktionen beschrieben (narrative Rollen, Wechsel der Erzählperspektive, deskriptive Sequenzen usw.; vgl. z.B. Adam 2005);

- d) im Bereich der literarischen Stilistik gilt es, 1) bisherige Zugänge zu literarischen Texten um eine neue Form korpusinduzierter, statistisch begründeter Methodik zu erweitern, 2) die Ermittlung und die Beschreibung literaturspezifischer stilistischer Muster (*patrons stylistiques*) weiter voranzutreiben, 3) die Reflektion über die historische Poetik der Formen durch die Identifizierung lexiko-syntaktischer Rekurrenzen in verschiedenen (Sub-)Gattungen zu vertiefen und eventuelle Konvergenzen aufzudecken;
- e) die unter c) und d) beschriebenen linguistischen und literarstilistischen Ansätze sollen miteinander vereint werden, um dadurch zu einer linguistischen Beschreibung stilistischer Muster (*patrons stylistiques*) zu gelangen, die diese als lexiko-syntaktische Konstruktionen bzw. *motifs* im Sinne von Legallois (2012) auffasst (*en être là de + DetPoss + GN [réflexions, pensées, discours] quand + proposition*);
- f) ausschnittsweise sollen auf der Grundlage von teilweise aus dem EMOLEX-Projekt verfügbaren Parallel- und Übersetzungskorpora zwischensprachliche Vergleiche von lexiko-syntaktischen Konstruktionen und Übersetzungsvergleiche vorgenommen werden;
- g) es sollen die Grundlagen einer Lexiko-Grammatik von *extended lexical items* gelegt werden, die der Handlungskonstruktion und der Konstruktion von Romandiskursen allgemein dienen (Skripte und Erzählgerüst, vgl. Baroni 2002).

3.1.1 Innovative Aspekte; wissenschaftliche und gesellschaftliche Relevanz des Projekts

Das Projekt betritt sowohl durch die Wahl des Forschungsgegenstandes als auch in theoretisch-konzeptueller und methodischer Hinsicht Neuland. Durch die Verbindung der Perspektiven der literarischen Stilforschung und der Linguistik werden Synergieeffekte entstehen, die zu einer vertieften Kenntnis der sprachlichen Struktur von (Sub-)Gattungen führen werden. Die wissenschaftliche und gesellschaftliche Relevanz eines solchen Projekts ist vielfältig:

- auf sprachwissenschaftlicher Ebene entsteht eine Typologie lexiko-syntaktischer Konstruktionen, die spezifisch für bisher kaum in dieser Hinsicht untersuchte Textsorten (Hoch- und Trivilliteratur) sind;
- auf der Ebene der maschinellen Sprachverarbeitung werden neue Techniken des Data-Mining in Linguistik und literarischer (Korpus-)Stilistik entwickelt, insbesondere solche, die die Extraktion von Mehrworteinheiten auf der Grundlage syntaktisch analysierter Korpora ermöglichen;
- im Bereich der Korpuslinguistik wird eine neue Methodik der Auswertung ein- und mehrsprachiger Daten angestrebt;
- auf der Ebene der literarischen Stilistik wird, ersten Ansätzen in Siepmann (2016, erscheint) folgend, eine korpuslinguistische Untersuchung großer Mengen literarischer Texte vorgenommen, die ohne entsprechende Werkzeuge manuell nicht zu leisten wäre; auf diese Weise wird eine verbesserte Beschreibung der Literatursprache in ihren spezifischen Ausprägungen und (ggf.) in ihrem Dialog mit nicht-literarischen Diskursen erreicht.
- im Bereich der Gattungstheorie werden Ansätze zur Unterscheidung von (Sub-)Gattungen sowie von Hochliteratur und Trivilliteratur erarbeitet; außerdem ergeben sich innovative Perspektiven auf die Frage der Gattungshybridität; nach Rastier (2011 : 72) besteht eine wachsende gesellschaftliche Nachfrage nach einer operativen Gattungstheorie.
- im Bereich der Kontrastivlinguistik und der Übersetzungswissenschaft wird in analoger Weise ein Beitrag zu einer soliden korpuslinguistischen Fundierung geleistet;
- auf gesellschaftlicher Ebene werden die im Projekt erarbeiteten Daten der interessierten Öffentlichkeit unter Beachtung des Copyrights und zur wissenschaftlichen Nutzung zur Ver-

fügung gestellt. Das Projekt wird außerdem einen positiven Transfer in die Lehre in Linguistik und Stilistik, die Übersetzerausbildung und die Übersetzungspraxis ermöglichen. Zusammenfassend sei festgehalten, dass das Projekt neue thematische und methodische Wege in bisher wenig erforschtem Gelände erkunden wird: « Beaucoup – énormément – reste à faire dans ce domaine, puisque le chantier est encore à peine ouvert » (Legallois 2012 : 50).

3.2 Methoden, Arbeitsprogramm, Zeitplan

3.2.1 Methoden der Korpuserstellung und –aufbereitung

Zusammensetzung der Korpora

Die Zusammensetzung unserer einsprachigen Korpora orientiert sich als Ausgangspunkt an der von den Autoren des Brown Corpus anhand externer Kriterien vorgenommenen Unterscheidung zwischen „belles lettres“ (Hochliteratur) und „general fiction“, die im Projekt im Lichte der aktuellen Diskussion über Gattungseinteilung sowie der linguistischen Analysen anhand interner Kriterien (vgl. Lee 2001) kritisch hinterfragt werden kann. Hinzu kommt eine feinere Unterscheidung zwischen verschiedenen (Sub-)Gattungen, wobei das Spektrum vom Kriminalroman über Science-Fiction bis zum Liebesroman reicht. Wir gehen von vergleichbaren Korpora gleichen Umfangs aus (ca. 10 Mill. Wörter Hochliteratur und „general fiction“ und ca. 20 Mill. Trivilliteratur für jede Sprache¹), die dieselbe Zeitspanne abdecken (20. Jhd.). Der Größenunterschied ergibt sich aus der Verfügbarkeit der jeweiligen Gattungen. Über die Metadaten lässt sich das Korpus in verschiedener Granularität bzw. Abstraktion (vom Einzeltext bis hin zu Subgattungen) partitionieren.

Der übersetzungswissenschaftliche Teil des Projekts stützt sich auf Parallelkorpora, die im Rahmen des Emolex-Projekts entstanden sind und die gemäß der gleichen Genreaufteilung wie die vergleichbaren Korpora erweitert werden sollen. Diese Parallelkorpora werden ca. 10 Mill. Wörter enthalten, die unter den untersuchten Sprachenpaaren (F-E, F-D, E-D) gleichmäßig aufgeteilt werden.

Als Referenzkorpora für die Ermittlung statistisch signifikanter lexiko-syntaktischer Konstruktionen dienen die Zeitungskorpora aus dem Emolex-Projekt (ca. 100 Mill. Wörter pro Sprache) sowie das Wissenschaftskorpus Scientext (ca. 5 Mill. Wörter, erweitert um deutsche Wissenschaftstexte) (s. Anhang I zu diesen Korpora).

Aufbereitung und Auswertung der Korpora

Nach der Erstellung eines strukturierten Korpus (XML-Markup und morphosyntaktische Annotation, s. Abschnitt „Programm“) erfolgt die Extraktion der lexiko-syntaktischen Konstruktionen auf der Grundlage einer innovativen Methodik und mithilfe hochentwickelter Informatik-Werkzeuge.

Ein besonderes Augenmerk liegt auf der Weiterentwicklung der Methodik der Extraktion von Mustern, die als Grundlage für die Identifikation von lexiko-syntaktischen Konstruktionen dienen. Dazu werden die Knoten rekurrenter Teilbäume nicht mehr mit lexikalischen Einheiten, sondern mit Klassen von Einheiten verbunden. Diese können morphosyntaktischer, aber auch semantischer Natur sein; letztere werden auf distributioneller Basis oder über externe Ressourcen wie WORDNET ermittelt werden.

Durch Partitionierungen des Korpus in verschiedene Gattungen sollen jene Konstruktionen identifiziert werden, die am stärksten zwischen Gattungen diskriminieren und damit der

¹ Die deutsch-, französisch- und englischsprachigen Korpora, über die wir in EMOBASE verfügen, werden ergänzt und in ein ausgewogenes Verhältnis gebracht werden.

Textklassifikation dienen können. Dazu werden multivariate Methoden und Berechnungen von spezifischen Frequenzabweichungen eingesetzt (vgl. Lebart & Salem 1994).

3.2.2 Methodik der sprachwissenschaftlichen Analyse

Die Analyse erfolgt in mehreren Schritten. Zunächst sollen die romanspezifischen lexiko-syntaktischen Konstruktionen ermittelt werden, und zwar auf der Grundlage statistischer Berechnungen auf zwei Ebenen: einerseits der statistisch signifikanten Rekurrenz von Konstruktionen innerhalb des gesamten Korpus der Hoch- und Trivalliteratur, andererseits der spezifischen Frequenz in den literarischen und nicht-literarischen Korpora. In einem zweiten Schritt werden die ermittelten Konstruktionen nach begrifflichen Kriterien (z.B. Zeit, Ort, Bewegung) in semantisch-funktionale Klassen aufgeteilt, die sich sowohl auf das Gesamtkorpus als auch auf die verschiedenen Gattungen beziehen. Schließlich wird die Rolle untersucht, die diese Klassen in der diskursiven Organisation des Romans einnehmen. Diese Analyse orientiert sich an funktional-kontextualistischen Modellen, darunter insbesondere dem Modell der *extended lexical items* von Sinclair (1998/2004), das Lexis, Syntax, Semantik und Pragmatik miteinander verknüpft; dieses Modell soll insbesondere mit Blick auf die Organisation literarischer Texte weiterentwickelt werden. Darüber hinaus sollen z.B. mit Blick auf die Unterscheidung der interpersonalen, ideationalen und textuellen Ebene weitere funktionale Theorien in den Blick genommen werden (Halliday 2004, Dik 1997, Hengeveld/Mackenzie 2010). Hinzu treten die von Hoey (2005) in seiner Theorie des *Lexical Priming* entwickelte Perspektive auf textuelle Kollokationen und Kolligationen und das in Emolex entwickelte funktionale Modell binärer Kollokationen.

Ein weiterer großer Bereich der linguistischen Analyse betrifft den Sprachvergleich und die übersetzungswissenschaftliche Untersuchung. Hier kommen zwei bewährte Methoden zum Einsatz. Ausgehend von den zuvor aufgestellten semantisch-funktionalen Klassen, die als Vergleichsbasis dienen, werden wir erstens die Äquivalenzrelationen zwischen einer Auswahl von Konstruktionen jeder Klasse untersuchen. Dadurch werden sich eine größere Zahl von Hypothesen der vergleichenden Stilistik auf eine solidere empirische Basis stellen lassen. Zweitens werden wir mithilfe der Übersetzungskorpora prüfen, inwieweit Übersetzer zu adäquaten Wiedergaben der in der kontrastiven Analyse verglichenen Konstruktionen gelangen. Durch die Verbindung der Daten, die die jeweiligen einsprachigen sowie die Übersetzungskorpora liefern, soll ermittelt werden, welchem Interdiskurs übersetzerische Lösungen in der Zielsprache entstammen. Der Sprachvergleich E-F sowie D-E wird von Osnabrück und Grenoble durchgeführt werden, der Sprachvergleich D-F von Erlangen.

3.2.3 Methodik der stilistischen Analyse

Die durch die korpuslinguistische Extraktion gewonnenen Rohdaten müssen in einer literaturwissenschaftlichen Analyse durch Kontextualisierung und Vergleich auf ihre Relevanz überprüft werden. In einem ersten Schritt gilt es, rekurrente Erscheinungen in den Korpora der Hochliteratur und Trivalliteratur zu ermitteln, um diese in einem zweiten Schritt einer Relevanzprüfung durch einen Abgleich mit nicht-literarischen Korpora zu unterziehen. So lassen sich zwei Typen von Beobachtungsfakten unterscheiden: jene, die aufgrund ihrer Überrepräsentation in literarischen Texten zur Konstruktion einer spezifisch literarischen Sprache beitragen, und jene, die keine im eigentlichen Sinne literarischen Merkmale aufweisen. Auf dieser Grundlage werden folgende methodische Schritte vorgenommen:

- erstens werden wir die kontextuelle Wertigkeit dieser Konstruktionen untersuchen, d.h. ihre Funktion und Relevanz für die Poetik des Romans. Zwei Bereiche sollen im Vorder-

grund der Betrachtung stehen: die „narrative Spannung“ (Baroni 2007; Adam 2005) sowie die Erzeugung perspektivischer Effekte (Rabatel 1998, 2008).

- zweitens werden wir durch einen internen Vergleich prüfen, ob charakteristische Merkmale bestimmter Subgattungen existieren und ob die Wertigkeit dieser Merkmale ggf. von ihrem Vorkommen in diesen Subgattungen abhängt. Dies soll perspektivisch zu einem Nachdenken über die Relevanz der vorgenommenen Klassifikationen führen.

- gleichzeitig wird die Frage zu untersuchen sein, welche Arten von Korrelationen zwischen Phraseologismen in Romanen für ein Massenpublikum sich die Phraseologismen der Literatursprache: liegt hier eher eine Unter- oder eine Überrepresentation vor? Die Antworten auf diese Fragen verweisen auf eine pragmatische Fragestellung: versucht z.B. der Poproman sich der literarischen Sprache anzunähern oder sich von ihr zu entfernen? Ist er auf linguistischer oder nur auf thematischer Ebene der Trivilliteratur zuzuordnen?

- schließlich werden wir auf der Grundlage des Vergleichs zwischen literarischen und nicht-literarischen Texten die Verbreitung bestimmter Formen gegenüberstellen; aus dieser Form der stilistischen Analyse soll hervorgehen, ob ein Diskurstyp einen Einfluss auf den anderen ausübt (und in welcher Richtung: werden literarische Diskurse nicht literarischen entlehnt oder umgekehrt?). Zur Beantwortung dieser Frage muss eine Periodisierung des Korpus über die Metadaten vorgenommen werden, um feststellen zu können, welcher Interdiskurs für welche Periode typisch war.

3.2.4 Arbeitsprogramm

Arbeitsprogramm maschinelle Sprachverarbeitung

Hier sind folgende Schritte durchzuführen:

Arbeitspaket 1: Korpusaufbereitung

AP1.1 Beschaffung neuen Textmaterials, welches das bereits vorhandene Emolex-Korpus ergänzen soll, um eine ausgewogene Gewichtung der Genres für jede Sprache vornehmen zu können.

AP1.2 Reformatierung nach dem in EMOLEX verwendeten XML-Schema. Einteilung der Textdaten in Datei-Blöcke von angemessener Größe (ca. 2 GB) zur Optimierung der Cache-Nutzung in der EmoConc-Architektur (EmoBase-Modul)

AP1.3 Automatische morpho-syntaktische Annotation (Annotierung von Wortarten und morphologischen Merkmalen, Lemmatisierung und Annotierung der Abhängigkeitsrelationen)

AP1.4 semantische Annotation sowie Hinzufügung allgemeiner thematischer Deskriptoren mithilfe von großen Thesauri

AP1.5 Erstellung einer Hash-Datenbank, die sämtliche Sätze des Korpus und sämtliche lexiko-syntaktischen Kookkurrenzen enthält und über entsprechende Indizierungen für Lemmata, Abhängigkeitsrelationen, thematische Deskriptoren und Metadaten (Genre, Autor, Jahr usw.) verfügt

Arbeitspaket 2: Re-Engineering von EmoConc

AP2.1 Erstellung einer Entwicklungsdokumentation für das neue Werkzeug. Schnittstellendesign und Definition der neuen Abfragefunktionalitäten

AP2.2 Anpassung der Indexstruktur unter Einbeziehung sämtlicher Textelementebenen (Absatz, Text usw.)

AP2.3 Integration der zur Partitionierung der Arbeitskorpora vorgesehenen Deskriptoren

AP2.4 Erweiterung der Abfragesprache um die Verarbeitung metadatenbezogener Constraints

AP2.5 Erweiterung der Datenausgabe um die Anzeige von Spezifitätswerten für jede Un-

tergruppe der in einem Arbeitskorpus jeweils definierten Partition

AP2.6 Entwicklung einer Schnittstelle für die Festlegung von Partitionen innerhalb der Arbeitskorpora (nach Genre, Autor, Jahr usw.)

Arbeitspaket 3: Extraktion lexikalisch-syntaktischer Muster (vgl. 2.1.4)

AP3.1 Entwicklung eines Algorithmus für die Extraktion lexikalisch-syntaktischer Konstruktionen auf der Grundlage der für einen bestimmten Node automatisiert extrahierten Mehrwortausdrücke;

AP3.2 Entwicklung einer Schnittstelle für die Definition von satzübergreifenden Kollokationen und Kolligationen

AP3.3 Implementierung eines Moduls für die Suche nach satzübergreifenden Kollokationen und Kolligationen

Arbeitsprogramm Sprachwissenschaft

AP1: Ermittlung der lexiko-syntaktischen Konstruktionen auf der Grundlage der extrahierten Sequenzen

AP2: erste Kategorisierung von Konstruktionen unter Berücksichtigung kontextualistischer und funktionaler Sprachtheorien

AP3: Erstellung eines semantisch-funktionalen und syntaktischen Rasters zur genauen Klassifikation der ermittelten Konstruktionen

AP4: Auswahl einer eingeschränkten Zahl von Kategorien (z.B. Zeit, Ort, Körper, Bewegung) mit dem Ziel einer detaillierten Untersuchung ihrer Funktionsweise im Romandiskurs verschiedener Genres

AP5: Untersuchung der Äquivalenzrelationen zwischen den unter 4 ausgewählten Klassen von lexiko-syntaktischen Konstruktionen in den drei Sprachen

AP6: Vergleich einer eingeschränkten Zahl von Übersetzungen, um übersetzerische Entscheidungen im Bereich der lexiko-syntaktischen Konstruktionen zu erfassen und zu beurteilen

AP7: Verknüpfung der Ergebnisse des linguistischen Teils und des literar-stilistischen Teils

Arbeitsprogramm literarische Stilistik und Gattungstheorie

AP1: Mitarbeit und Beratung bei der Korpuserstellung ([Sub-]Gattungen)

AP2: Identifikation rekurrenter Konstruktionen in den Korpora der Hoch- und Trivialliteratur

AP3: Vergleich des der literarischen Korpora mit nicht-literarischen Korpora zur Ermittlung des literaturspezifischen Konstruktionen

AP4: Vergleich des Korpus der Hochliteratur mit dem Korpus der Trivialliteratur mit dem Ziel der Identifikation von Rekurrenzen nach Sub-Gattungen sowie Anwendung der Ergebnisse auf hybride Romane;

AP5: Eruierung der kontextuellen Wertigkeit der identifizierten Konstruktionen aus der Perspektive der Poetik des Romans und der Narratologie (narrative Spannung? Konstruktion der Erzählperspektive ?) sowie ggf. Unterscheidung spezifischer Funktionsweisen in bestimmten Sub-Korpora bzw. -Gattungen;

AP6: Ermittlung der spezifischen Merkmale der Korpora der Trivialliteratur im Vergleich zu den literarischen Korpora ; Bestimmung des sprachlichen Status von Trivialliteratur

AP7: Interpretation des Verhältnisses zwischen literarischen und nicht-literarischen Diskursen im Bezugsrahmen der historischen Poetik und Gattungsgeschichte

3.2.5 Zeitplan und Aufteilung der Arbeitspakete nach Arbeitsgruppen

Monate (insgesamt 36)	1-2	3-4	5-6	7-8	9-10	11-12	13-14	15-16	17-18	19-20	21-22	23-24
AP1.1 Datensammlung	■	■										
AP1.2 Reformatierung		■	■									
AP1.3 morphosyntaktische Annotation			■	■								
AP1.4 semantische Annotation			■	■								
AP1.5 Indexierung					■							
AP2.1 Entwicklungsdokumentation	■											
AP2.2 Indizierungsskripte				■								
AP2.3 Integration Korpuspartitionierung					■							
AP2.4 Implementation metadatenbezogener Abfrage-Constraints					■							
AP2.5 Spezifitätsberechnungen						■						
AP2.6 Schnittstelle für Partitionierung						■						
AP3.1 Extraktionsalgorithmus							■	■				
AP3.2 Schnittstelle Definition Sequenzen									■			
AP3.3 Extraktion										■	■	
Tests und Debugging				■	■	■	■	■	■	■	■	■

Tabelle 1 : Zeitplan Korpuserstellung

Verantwortliche : Olivier Kraif und Sascha Diwersy

Für die betreffenden Arbeitspakete vorgesehene Mitarbeiter:

Fest angestellte Mitarbeiter:

LIDILEM : O. Kraif,

COLOGNE : S. Diwersy

Nicht fest angestellte Mitarbeiter:

LIDILEM : 1 Vertrag über 12 Monate + 1 Vertrag über 13 Monate (Ingénieur de recherche)

Zeitaufwand : 24 Monate (+ 1 Monat für die Inbetriebnahme der PhraséoBase)

Lieferbare Ergebnisse : Annotierte multilinguale Korpora, Indizierungs- und Suchskripte, Schnittstellen für Abfrage und Ergebnisanzeige, Skripte für die Extraktion von *motifs séquentiels*, Schnittstelle für die Definition dieser *motifs* und die Anzeige der extrahierten Sequenzen; Weiterentwicklung und Vergrößerung der EMOBASE-Schnittstelle mit Integration eines PhraséoBase-Moduls (freier Zugang), das die Projektdaten und Ergebnisse zur Verfügung stellt

Monate (insgesamt 36)	1-9	10-12	13-15	16-18	19-24	24-30	30-36
AP1a Pilotstudie auf der Basis der bestehenden Korpora	■						
AP1b endgültige Identifikation der LSK mithilfe der finalisierten Korpora		■					

AP2 Theoretische Entscheidungen							
AP3 Erstellung des syntaktisch-semantischen Rasters							
AP4 Diskursanalyse							
AP5 Herstellung der Äquivalenzen							
AP6 Übersetzungsvergleich							
AP7 Verbindung Resultate Linguistik – literarische Stilistik							

Tabelle 2: Zeitplan sprachwissenschaftliche Analyse

Verantwortliche : Iva Novakova und Dirk Siepmann

Für die betreffenden Arbeitspakete vorgesehene Mitarbeiter:

Fest angestellte Mitarbeiter:

LIDILEM : I. Novakova, A. Tutin, F. Grossmann, J. Sorba; D. Legallois (CRISCO)

OSNABRÜCK : D. Siepmann

ERLANGEN : L. Fesenmeier

BONN: M. Gymnich

COLOGNE : S. Diwersy

Nicht fest angestellte Mitarbeiter:

LIDILEM : 1 Post-Doc-Stelle 12 Monate ; 1 befristeter Mitarbeiter 18 Monate

OSNABRÜCK : 1 wissenschaftlicher Mitarbeiter/wissenschaftliche Mitarbeiterin 36 Monate

ERLANGEN : 1 wissenschaftlicher Mitarbeiter/wissenschaftliche Mitarbeiterin 36 Monate

Zeitaufwand : 36 Monate

Lieferbare Ergebnisse : Pilotstudie, Erfassung der Daten der vergleichbaren und Parallelkorpora ; Erstellung von Excel-Dateien zur linguistischen Verarbeitung der Daten in Form von Rastern ; syntaktisch-semantische Kodierung der Daten ; Zwischenberichte nach jedem Arbeitstreffen über Fortschritte und Schwierigkeiten ; Endbericht ; Teilnahme an internationalen Tagungen ; Organisation einer internationalen Tagung ; Buchveröffentlichung zu den Ergebnissen des Projekts

Monate (insgesamt 36)	1-16	16-22	22-24	24-30	30-36
AP1 Korpusaufbau					
AP2 Identifikation von Rekurrenzen					
AP3 Identifikation der literaturspezifischen Daten					
AP4 Identifikation der Rekurrenzen nach Sub-Genres					
AP5 Interpretation der Wertigkeit der Rekurrenzen (Poetik des Romans)					

AP6 Identifikation der spezifischen Merkmale der trivialliterarischen Subgattungen sowie Interpretation					
AP7 Identifikation und Interpretation des Zusammenspiels von literarischen und nicht-literarischen Diskursen					
AP8 Verbindung Resultate Linguistik – literarische Stilistik					

Tabelle 3 : Zeitplan stilistische Analyse

Verantwortlicher : Julien Piat; Marion Gymnich

Für die betreffenden Arbeitspakete vorgesehene Mitarbeiter:

Fest angestellte Mitarbeiter:

LITT & ARTS, (TRAVERSESES 19-21) : Laetitia Gonon und Julien Piat

Nicht fest angestellte Mitarbeiter: 1 Vertrag über 12 Monate

BONN : Marion Gymnich

BONN: 1 wissenschaftlicher Mitarbeiter/wissenschaftliche Mitarbeiterin 36 Monate

Zeitaufwand: 36 Monate

Lieferbare Ergebnisse : Beteiligung an der Korpuszusammenstellung ; Überprüfung und Analyse der Rasterdaten aus literaturwissenschaftlicher Perspektive ; Teilnahme an internationalen Konferenzen ; Zeitschriftenartikel ; Beteiligung an der Buchveröffentlichung

3.2.6 Verbreitung der Ergebnisse

Geplant sind mehrere Veröffentlichungen in Fachzeitschriften sowie Vorträge bei nationalen und internationalen Tagungen. Darüber hinaus soll ein internationales Symposium zur Literatursprache in Deutschland stattfinden. Aus diesen Aktivitäten soll zusätzlich die Herausgabe eines einschlägigen internationalen Sammelbandes erwachsen. Dieser wird sämtliche theoretischen, konzeptuellen und methodischen Ergebnisse des vorliegenden Projekts zusammenbringen; außerdem sollen Ergebnisse anderer internationaler Arbeitsgruppen einfließen, mit denen wir kooperieren. Der Sammelband soll zu einem Referenzwerk für Forscher verschiedener Disziplinen werden (Lexikologie, maschinelle Sprachverarbeitung, Stilistik, Übersetzungswissenschaft).

3.3. Literaturverzeichnis zum Stand der Forschung, zu den Zielen, dem Arbeitsprogramm und zu den projektbezogenen Publikationen

Adam, J.-M. ([2005] 2008). *La Linguistique textuelle. Introduction à l'analyse textuelle des discours*, Paris: Armand Colin.

Ballard, M. (1992). *La traduction de l'anglais au français*. Paris : Nathan.

Baroni, R. (2002). "Le rôle des scripts dans le récit", *Poétique* n°129: 105-126.

– (2007). *La Tension narrative. Suspense, curiosité et surprise*, Paris : Editions du Seuil.

Barthes, R. (1966). "Introduction à l'analyse structurale des récits", *Communications* n° 8, Paris, Seuil.

- Basseler, M., Nünning, A. & Schwanecke, C. (Eds.) (2013). *The Cultural Dynamics of Generic Change in Contemporary Fiction: Theoretical Frameworks and Model Interpretations*. Trier: WVT.
- Beauvisage, T. (2001) « Exploiter les données morphosyntaxiques pour l'étude statistique des genres –Application aux roman policier, TAL No 43, TAL No 43 <<http://www.revue-texto.net/Inedits/Beauvisage/index.html>>.
- Bourdieu P. (1992). *Les règles de l'art : genèse et structure du champ littéraire*, Seuil.
- Biber, D. (1993). "Using register-diversified corpora for general language studies." *Computational Linguistics* 19: 219-241.
- Biber, D., Johansson, S., Leech, G., Conrad S, Finegan, E.(1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- Blumenthal, P., Novakova, I. & Siepmann D. (Eds.) (2014). *Les émotions dans le discours. Emotions in Discourse*, Peter Lang: Berlin, 434 p.
- Blumenthal, P. (2006). *Wortprofil im Französischen (Beihefte zur ZrP 332)*, Tübingen: Niemeyer.
- Blumenthal, P., Diwersy, S. & Mielebacher, J. (2005). "Kombinatorische Wortprofile und Profilkontraste. Berechnungsverfahren und Anwendungen", *Zeitschrift für romanische Philologie* n°121: 49-83.
- Brunet E. (1981). *Le Vocabulaire français de 1789 à nos jours*, Slatkine-Champion, Genève-Paris.
- Burton, B. & Carter, R. (2006). "Literature and the Language of Literature". in: K. Brown, A. H. Anderson, L. Bauer, M. Berns, G. Hirst, & J. Miller (Eds.), *Encyclopedia of Language and Linguistics*, Vol. 7. Boston: Elsevier, (2nd ed., pp. 267-274).
- Charest, S., Brunelle E. & Fontaine J. (2010). "Au-delà de la paire de mots : extraction de cooccurrences syntaxiques multilexémiques", *Actes de TALN 2010*, Montréal: 19-23 juillet 2010.
- Chuquet, H. & Paillard M. (1987). *Approche linguistique des problèmes de traduction*. Paris : Ophrys.
- Croft, W. (2001). *Radical construction grammar*. Oxford: Oxford University Press.
- de Beaugrande, R. (2005). 'Corporate bridges' twist text and language: twenty arguments against corpus research and why they're a right load of Old Codswallop. Internet: <http://www.beaugrande.com/Corporate%20Bridges.htm>.
- Dik, S. (1997a). *The theory of functional grammar*. Part 1, *The structure of the clause*. 2nd ed., Berlin/New York: Mouton de Gruyter.
- (1997b). *The theory of functional grammar*. Part 2, *Complex and derived constructions*, 2nd ed., Berlin/New York: Mouton de Gruyter.
- Diwersy, S. (2012a). "Kookkurrenz, Kontrast, Profil. Korpusinduzierte Studien zur lexikalisch-syntaktischen Kombinatorik französischer Substantive (mit ergänzenden Betrachtungen zum Deutschen)", *Beihefte zur Zeitschrift für romanische Philologie*: 373.
- (2012b). "La francophonie multivariée. Ou: comment mesurer les français en Afrique?", *Le français en Afrique* n°27: 75-91. [URL: <http://www.unice.fr/ILF-CNRS/ofcaf/27/DIWERSY.pdf>]
- Diwersy, S., Evert, S. & Neumann, S. (2014). A semi-supervised multivariate approach to the study of language variation, in: B. Szmrecsanyi, & B. Waelchli, (Eds.), *Linguistic*

- variation in text and speech, within and across languages*, Berlin e.a.: de Gruyter, 174-204.
- Diwersy, S. & Fesenmeier, L. (2006). A (Twofold) Contrastive Analysis of Collocations: Fr. patience – It. pazienza, in: E. Corino, C. Marellò, & C. Onesti (Eds.), *Proceedings of the 12th EURALEX International Congress*, Alessandria: Edizione dell’Orso: 959-966.
- Diwersy, S. & François, J. (2011). “La combinatoire des noms d’affect et des verbes supports de causation en français: étude de leur attirance au niveau des unités et de leurs classes syntactico-sémantiques”, *Tranel* n°55: 139-161.
- Diwersy S., Goossens V., Grutschus A., Kern B., Kraif O., Melnikova, E. Novakova I. (2014) Traitement des lexies d’émotion dans les corpus et les applications d’*EmoBase* , revue Corpus No 13, 269-293.
- Diwersy, S. & Kraif, O. (2013). Observations statistiques de cooccurrents (lexico-) syntaxiques pour la catégorisation sémantique d’un champ lexical, in: Baider, F. & Cislaru, G. (Eds.), *Cartographie des émotions*, Paris: Presses universitaires de la Sorbonne, 55-70.
- Duff, D. (2000). *Modern Genre Theory*. London: Longman.
- Evert, S. (2008). Corpora and collocations. in: A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics. An International Handbook*, Berlin: Mouton de Gruyter: 1212-1249.
- Feilke, H. (1994). *Common sense-Kompetenz. Überlegungen zu einer Theorie des 'sympathischen' und 'natürlichen' Meinens und Verstehens*. Frankfurt/Main: Suhrkamp.
- (1996). *Sprache als soziale Gestalt. Ausdruck, Prägung und die Ordnung der sprachlichen Typik*. Frankfurt/Main: Suhrkamp.
- (2003). „Textroutine, Textsemantik und sprachliches Wissen.“ In: Linke, A., Ortner, H. & Portmann P. R. (Eds.): *Sprache und mehr. Ansichten einer Linguistik der sprachlichen Praxis*. (Reihe Germanistische Linguistik 245). Tübingen (Niemeyer): 209-229.
- Fesenmeier, L. (en préparation, in Vorb.). *Zentraldifferenzen im Wortschatz des Französischen, Italienischen und Spanischen* (vorauss.: Frankfurt a.M.: Klostermann; zugleich Habilitationsschrift, Köln, 2009).
- (2008a). Komplementsätze bei span. *conocer* (und *saber*), in: E. Stark, R. Schmidt-Riese & E. Stoll, *Romanische Syntax im Wandel*, Tübingen: Narr, 399-415.
- (2008b). Quasi-sinonimia e ‘differenze centrali’: la coppia *aspettare – attendere*, in: E. Cresti, *Prospettive nello studio del lessico italiano. Atti del IX Congresso SILFI (Firenze, 14-17 giugno 2006)*, Firenze: Firenze University Press, 267-274.
- (2009). Synonymie: Zur Semiotik der Absenz, in: A. Grutschus & P. Krilles, *Figuren der Absenz*, Berlin: Frank & Timme: 219-233.
- (2010). ‘Se souvenir’ en français et en italien: différence(s) de centrage, in: M. Iliescu, H. Siller-Runggaldier & P. Danler, *Actes du XXV^e Congrès International de Linguistique et de Philologie Romanes*, Tübingen: Niemeyer, Vol. 3 : 85-96.
- Fillmore, C., Kay, P. & O’Connor, C. (1988). “Regularity and idiomaticity in grammatical constructions: the case of *let alone*”, *Language* 64: 501-38.
- Firth, J.R. (1957). “A synopsis of linguistic theory 1930-1955”, in: J.R. Firth e.a., *Studies in Linguistic Analysis*, Oxford: Philological Society: 1-32.
- Fischer-Starcke, B. (2010). *Corpus linguistics in literary analysis: Jane Austen and her contemporaries*. London: Continuum.

- Fletcher, W.H. (2007). *kfNgram* <http://www.kwicfinder.com/kfNgram/kfNgramHelp.html>.
- François, J. (1989). *Changement, Causation, Action – Trois catégories sémantiques fondamentales du lexique verbal français et allemand*. Genève: Droz.
- Frow, J. (2006). *Genre : The New Critical Idiom*. London: Taylor and Francis.
- Gallagher, J. D. (2007). Traduction littéraire et études sur corpus. in: M. Ballard & C. Pineira-Tresmontant (Eds.), *Les corpus en linguistique et en traductologie*. Arras: Artois Presses Université: 199-230.
- Goldberg A. (1995). *Constructions. A construction grammar approach to argument constructions*. Chicago: The University of Chicago Press.
- Gonon, L. (2010). “Mythes et démythification dans le roman policier de Fred Vargas”, in: *Recherches et Travaux*, n°77, *Le Devenir-roman des Mythologies de Roland Barthes*, G. Bellon & P. Vachaud (dir.), Grenoble: ELLUG, 119-135.
- (2012), *Le Fait divers criminel dans la presse quotidienne française du XIX^e siècle*, Paris: Presses Sorbonne Nouvelle.
- Granger S. & Lerot, J. & Petch-Tyson, S. (2003). *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*, Rodopi: Amsterdam and Atlanta.
- Greaves, C. (2005). *ConcGram concordancer with ConcGram analysis*, HongKong: HKUST
- Greimas, A. J. (dir.) (1982 [1972]), *Essais de sémiotique poétique*, Paris: Larousse.
- Grossmann, F. & Tutin, A. (2002). “Collocations régulières et irrégulières: esquisse de typologie du phénomène collocatif”, *Revue française de linguistique appliquée* n°VII-1: 7-25.
- Grünbeck, B. (1976). *Moderne deutsch-französische Stilistik auf der Basis des Übersetzungsvergleichs (Sammlung romanischer Elementar- und Handbücher)*, Heidelberg: Winter.
- Grutschus, A. & Fesenmeier, L. (2013). ‘Inter metum, timorem et pavorem interest ...’ – et qu'en est-il des différences entre leurs successeurs romans?, in: E. Casanova Herrero & C. Calvo Rigual, *Actes del 26é Congrès de Lingüística i Filologia Romàniques (València, 6-11 de setembre de 2010)*, Berlin: de Gruyter, Vol. 3, 171-182.
- Guiraud, P. (1954). *Les caractères statistiques du vocabulaire*, Paris: P.U.F.
- Gymnich, M. (2000). *Entwürfe weiblicher Identität im englischen Frauenroman des 20. Jahrhunderts*. Trier: WVT, 2000.
- (2002), „Linguistics and Narratology: The Relevance of Linguistic Criteria to Post-colonial Narratology.“ In: Marion Gymnich, Ansgar Nünning & Vera Nünning (Hg.). *Literature and Linguistics: Approaches, Models and Applications*. Trier: WVT, 61-76.
- (2004a), „Metasprachliche Reflexionen und Sprachkritik im postmodernen amerikanischen Roman.“ In: Reinhard Kacianka & Peter V. Zima (Hg.). *Krise und Kritik der Sprache: Literatur zwischen Spätmoderne und Postmoderne*. Tübingen/Basel: Francke, 233-249.
- (2004b), „Konzepte literarischer Figuren und Figurencharakterisierung.“ In: Vera Nünning & Ansgar Nünning (Hg.). *Erzähltextanalyse und Gender Studies*. Stuttgart/Weimar: Metzler, 122-142.
- (2007), *Metasprachliche Reflexionen und sprachliche Gestaltungsmittel im englischsprachigen postkolonialen und interkulturellen Roman*. Trier: WVT, 2007.
- (2007), „Vorschläge für eine Relationierung verschiedener Aspekte und Dimensionen des Gattungskonzepts: Der Kompaktbegriff Gattung.“ In: Marion Gymnich, Birgit

- Neumann & Ansgar Nünning (Hg.). *Gattungstheorie und Gattungsgeschichte*. Trier: WVT, 31-52. (mit Birgit Neumann)
- (2010), „Gattung und Gattungshistoriographie.“ In: Rüdiger Zymner (Hg.). *Handbuch Gattungstheorie*. Stuttgart/Weimar: Metzler, 131-158.
 - (2011), „„Decolonizing Genre?’ – Das Konzept der literarischen Gattung in englischsprachiger postkolonialer und interkultureller Literatur.“ In: Stephan Conermann & Amr ElHawary (Hg.). *Was sind Genres? Nicht-abendländische Kategorisierungen von Gattungen*. Berlin: EB-Verlag, 299-315.
 - (2013), "Gender and Narratology." In: *Literature Compass* 10.9: 705-715.
- Gymnich, M., Neumann, B. & Nünning, A. (2007). *Gattungstheorie und Gattungsgeschichte*. Trier: WVT.
- Habert, B. (2005). *Instruments et ressources électroniques pour le français*. Gap/Paris : Ophrys, 2005, 170 p.
- Halliday, M.A.K. (2004). *An Introduction to Functional Grammar*, 3th ed. London: Edward Arnold.
- Hausmann, F.J. (1979). “Un dictionnaire des collocations est-il possible?”, *Travaux de littérature et de linguistique de l’université de Strasbourg n°XVII/1*: 187-195.
- (2007). *Collocations, phraséologie, lexicographie. Études 1977-2007 et Bibliographie*. Aachen: Shaker 2007.
- Hengeveld, K. & Mackenzie, J. L. (2010). Functional Discourse Grammar, in: B. Heine & H. Narrog (Eds), *Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press. 367-400. Reviewed by F. J. Newmeyer in *Studies in Language* 36.1 (2012): 215-224.
- Herbst, T. & M. Klotz (2003). *Lexikographie: Eine Einführung*. Paderborn: Schöningh.
- Hoey, M. (2005). *Lexical priming: a new theory of words and language*, London/New York: Routledge.
- Hunston, S. & G. Francis (2000). *Pattern Grammar*. Amsterdam: Benjamins.
- Jakobson, R. (1960). “Closing Statement: Linguistics and Poetics.” In: T.A. Sebeok, *Style in Language*, Cambridge, MS: MIT Press: 350-377.
- Kern, B. & Grutschus, A. (2014a): „Decepción, surprise, colère et furia : exploration d’une méthode statistique en lexicologie“, in: *Zeitschrift für romanische Philologie* 130/3.
- (2014b). „Surprise vs. étonnement : comportement discursif et perspectives contrastives“, in: Blumenthal P., Novakova I., Siepmann D. (Eds.), *Les émotions dans le discours. Emotions in discourse*. Peter Lang, Berlin : 187-198.
- Kraif O. (2008) Comment allier la puissance du TAL et la simplicité d'utilisation ? l'exemple du concordancier bilingue ConcQuest, *JADT 2008 : actes des 9es Journées internationales d'Analyse statistique des Données Textuelles*, Presses universitaires de Lyon, vol. 2, pp. 625-634
- Kraif, O. & Diwersy, S. (2012). “Le Lexicoscope: un outil pour l’étude de profils combinatoires et l’extraction de constructions lexico-syntaxiques”, *Actes de la conférence TALN 2012*, Grenoble: 399-406.
- Kraif, O. & Diwersy, S. (2014). “Exploring Combinatorial Profiles Using a Parsed Corpus: a Case Study in the Lexical Field of Emotions”, in: Blumenthal P., Novakova I., Siepmann D. (Eds), *Les émotions dans le discours. Emotions in discourse*. Peter Lang, Berlin : 381-394.

- Lebart, L. & Salem, A. (1988). *Analyse statistique des données textuelles. Questions ouvertes et lexicométrie*, Paris: Dunod.
- (1994). *Statistique textuelle*. Paris: Dunod.
- Lee, D. (2001) “Genres, registers, text types, domains, and styles: clarifying the concepts and navigating a path through the BNC jungle”, *Language learning and technology* 5/3: 37-72.
- Leech, G. & Short, M. (2007). *Style in Fiction. A Linguistic Introduction to English Fictional Prose*. 2nd ed. London: Pearson.
- Legallois, D. (2006). “Quand le texte signale sa structure la fonction textuelle d'une certaine catégorie nominale”, *Corela* n° spécial corela.edel.univ-poitiers.fr.
- (2012a). “La colligation : autre nom de la collocation grammaticale ou autre logique de la relation mutuelle entre syntaxe et sémantique?”, *Corpus* n°11, URL: <http://corpus.revues.org/2202>.
- (2012b - avec Cellier, P., Quiniou, S. & Charnois, Th.) “Fouille de données pour la stylistique: l'exemple des motifs émergents”, *Journées Internationales d'analyse statistique des données textuelles*, Liège, 13-15 juin.
- Legallois D. & Tutin A. (2013). “Présentation: Vers une extension du domaine de la phraséologie”, *Langages* n°189: 3-25.
- Lits, M. (2011). *Le Roman policier dans tous ses états: d'Arsène Lupin à Navarro*, Limoges: Pulim.
- 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.
- Magri-Mourgues V. (2006). “Stylistique générique et statistique”. *Les Cahiers de la MSH Ledoux* 8 : 655-666.
- Mahlberg, M. (2013). *Corpus stylistics and Dickens's fiction*. Routledge.
- Maingueneau, D. & Philippe, G. (2002). “Les conditions d'exercice du discours littéraire”, in: E. Roulet & M. Burger (dir.), *Les Modèles du discours au défi d'un “dialogue romanesque”: l'incipit du roman de R. Pinget “Le Libera”*, Nancy: PUN : 351-377.
- Malblanc, A. (1961). *Stylistique comparée du français et de l'allemand. Essai de représentation linguistique comparée et étude de traduction*. Paris : Didier.
- Marion, F. (2009). “Le stéréotype dans le roman policier”, *Cahiers de narratologie* n°17: <http://narratologie.revues.org/1095>
- Mel'čuk, I., Clas A., & Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve: Duculot (Champs linguistiques: Manuels).
- Melnikova E., Novakova, I. & Kraif O. (2009). “Quels corpus pour l'analyse contrastive? L'exemple des constructions verbo-nominales de sentiment en français et en russe”, *Actes des 6èmes Journées de la Linguistique de Corpus* (<http://web.univ-ubs.fr/corpus/jlc6.html#publi2009>).
- Monte, M. & Philippe, G. (dir.) (2014). *Genres et textes. Déterminations, évolutions, confrontations*, Lyon: PUL.
- Novakova I. (2015). “Les émotions entre lexique et discours“, in :Rabatel A., Ferrara-Léturgie A, Létrugie A. (eds) *La sémantique et ses interfaces*, Actes du Colloque 2013 de l'ASL, Lambert-Lucas 181-204.

- Novakova I. & E. Melnikova (2013). Vers un modèle fonctionnel pour l'analyse du lexique des émotions dans cinq langues européennes, in: *le Bulletin de la Société de linguistique de Paris*, Vol. CVIII (2013), fasc. 1: 131-160.
- Novakova I. & Sorba J. (2013a). Argumentation et émotion dans les séquences textuelles journalistiques. Le cas de *stupeur* et de *jalousie*, in: T. Muryn, S. Mejri, W. Prazuch & I. Sfar (Eds.), *Phraséologie et littérature*, Bern: Peter Lang, 137-149.
- (2013b). *Stupéfier* et *jalouser* dans les séquences textuelles journalistiques: quel profil discursif pour quelle stratégie argumentative?, *Le discours et la langue* Vol. 4-1, (2012 [2013]): 203-220.
- Novakova I. & Sorba J. (2014a): “L'émotion dans le discours. A la recherche du profil discursif de *stupeur* et de *jalousie*”, in: P. Blumenthal, I. Novakova & D. Siepmann (Eds.), *Les Emotions dans le discours. Emotions in Discourse*, Berlin: Peter Lang : 161-175.
- Novakova I. & Sorba J. (2014b) L'évaluation à travers les émotions : le cas d'*estime* et de *déception*, *Langues française* n°184, 74-89.
- Philippe, G. (2002). *Sujet, verbe, complément. Le moment grammatical de la littérature française 1890-1940*. Paris: Gallimard.
- (2004). *Flaubert savait-il écrire? Une querelle grammaticale (1919-1921)*, Grenoble: Ellug.
- Philippe, G. & Piat J. (2009). *La langue littéraire. Une histoire de la prose en France de Gustave Flaubert à Claude Simon*. Paris: Fayard.
- Piat, J. (2006). “Vers une stylistique des imaginaires langagiers”, *Corpus* n°5: 113-141.
- (2011), *L'Expérimentation syntaxique dans l'écriture du Nouveau Roman (Beckett, Pinget, Simon). Contribution à une histoire de la langue littéraire dans les années 1950*, Paris: Honoré Champion.
- (2013). “Langue littéraire / style : d'un *continuum* et d'une complémentarité”, in: C. Badiou-Monferran (dir.), *La Littérarité des belles-lettres. Un défi pour les sciences du texte?*, Paris: Garnier : 359-369.
- Quiniou, S., Cellier, P., Charnois, T. & Legallois, D. (2012). Fouille de données pour la stylistique: l'exemple des motifs émergents, *Actes des 11es Journées Internationales d'analyse statistique des données textuelles*, Liège, 13-15 juin 2012 : 821-833.
- Rabatel, A. (1998). *La Construction textuelle du point de vue*, Lausanne-Paris, Delachaux & Niestlé, «Sciences des discours».
- (2008). *Homo narrans. Pour une analyse énonciative et interactionnelle du récit*, Vol. II, *Dialogisme et polyphonie dans le récit*, Limoges : Lambert-Lucas.
- Rastier F. (2011). *La mesure et le grain. Sémantique de corpus*. Honoré Champion.
- Renouf, A. & Sinclair, J. (1991). “Collocational frameworks in English”, in: K. Aijmer & B. Altenberg (Eds.), *English Corpus Linguistics. Studies in honour of Jan Svartvik*. London: Longman, 128-144.
- Seretan V., Nerima L. & Wehrli E. (2003). “Extraction of Multi-Word Collocations Using Syntactic Bigram Compositio”, *Proceedings of the Fourth International Conference on Recent Advances in NLP*, (RANLP-2003): 424–431.
- Scholes R. (1982). *Semiotics and interpretation*. New Haven: Yale University Press.

- Siepmann, D. (2002). "Eigenschaften und Formen lexikalischer Kollokationen: Wider ein zu enges Verständnis", *Zeitschrift für französische Sprache und Literatur* n°112: 240-263.
- (2004). "Kollokationen und Fremdsprachenlernen: Imitation und Kreation, Figur und Hintergrund", *Praxis Fremdsprachenunterricht* 2/2004: 107-113.
- (2005a). *Discourse Markers across Languages. A Contrastive Study of Second-Level Discourse Markers in Native and Non-native Text with Implications for General and Pedagogic Lexicography*. Abingdon/New York: Routledge (Routledge Advances in Corpus Linguistics 6).
- (2005b). "Collocation, Colligation and Encoding Dictionaries. Part I: Lexicological Aspects", *International Journal of Lexicography* n°18, 4: 409-444.
- (2011). "Sinclair revisited: beyond idiom and open choice." In: Herbst, Thomas, Faulhaber, Susen & Uhrig, Peter (Eds.), *The phraseological view of language: a tribute to John Sinclair*, Berlin: Mouton de Gruyter 2011, 59-86.
- (2012). *Wissenschaftliche Texte auf Englisch schreiben*. Stuttgart: Klett.
- (2014b). "Extent of collocational difference between English, French and German emotion nouns: a corpus-based study", in: P. Blumenthal, Peter, I. Novakova & D. Siepmann (Eds.), *Les émotions dans le discours/Emotions in discourse*, 39-53.
- (2014c). "Collocations across Languages: Unity in Diversity?" In: Dominguez, M.J., Mollica, F. & M. N. Curcio (Eds.): *Zweisprachige Lexikographie im Spannungsfeld zwischen Translation und Didaktik*. (= *Lexicographica: Series Maior* 145) Berlin: New York: de Gruyter.
- (2015a) Über Vorbilder und Zerrbilder: Literarische Texte als Grundlage von Spracharbeit. W. Hallet, C. Surkamp, U. Krämer (Eds.), *Literaturkompetenzen Englisch: Modellierung - Curriculum – Unterrichtsbeispiele*, 100-113.
- (2015b) Lexicologie et phraséologie du roman: quelques pistes pour le français et l'anglais. Eingereicht bei *Cahiers de lexicologie*.
- (2015c) *The corpus de référence du français contemporain* as the first balanced mega-corpus of French. Eingereicht bei *International Journal of Corpus Linguistics*.
- – (2016). "A corpus-based investigation into key words and key patterns in post-war fiction." Erscheint in *Functions of Language* 2016/2.
- Siepmann, D. & Bürgel, C. (2014). *Le corpus de référence du français contemporain*. Présentation au congrès des francoromanistes, Münster, Allemagne, 29.09.14, [URL : <http://zenodo.org/record/12353#.VORzHi7UJsU>].
- Sinclair, J. M. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- (2004). *Trust the Text: Language, Corpus and Discourse*. London: Routledge.
- Stefanowitsch, A. & Gries S. T. (2003). Collostructions: Investigating the interaction between words and constructions, *International Journal of Corpus Linguistics* 8.2: 209-43.
- Steyer, K. (2013). *Usuelle Wortverbindungen. Zentrale Muster des Sprachgebrauchs aus korpusanalytischer Sicht*. Tübingen: Narr.
- Stubbs, M. (2005). "Conrad in the computer: examples of quantitative stylistic methods. *Language and Literature*." 14 (1): 5-24.
- Stubbs, M. & Barth I. (2003). "Using recurrent phrases as text-type discriminators. A quantitative method and some findings", *Functions of Language* 10 (1): 61-104.
- Talmy, L. (2000). *Towards a cognitive semantics*. Cambridge, MS: MIT Press.

- Todorov, Tz. (1980). *Poétique de la prose: choix, suivi de Nouvelles recherches sur le récit*, Paris: Editions du Seuil.
- Truffaut, L. (1983), *Problèmes linguistiques de traduction allemand-français. Guide de l'étudiant et du praticien*. Munich : Hueber.
- Tutin, A., I. Novakova, F. Grossmann & C. Cavalla (2006). "Esquisse de typologie des noms d'affect à partir de leurs propriétés combinatoires". *Langue française* n°150: 32-49.
- Tutin, A. (2010). *Sens et combinatoire lexicale: de la langue au discours*. Habilitation thesis, University of Grenoble. http://w3.u-grenoble3.fr/lidilem/labo/file/HDR_Tutin.pdf.
- (2013). Les collocations lexicales: une relation essentiellement binaire définie par la relation prédicat-argument. *Langages*, 1(189), 47-63.
- Tutin, A. & Grossmann, F. (2014). *L'écrit scientifique, Du lexique au discours*, Rennes: PUR.
- Vaillant, A. (2003). "Pour une histoire de la communication littéraire", *Revue d'histoire littéraire de la France* 103^e année, n° 3: 549-562.
- Vaudrey-Luigi, S. (2013), *La langue romanesque de Marguerite Duras. Une liberté souvenant*, Paris: Garnier.
- Vinay, J.P. & J. Darbelnet (1958). *Stylistique comparée du français et de l'anglais. Méthode de traduction*. Paris : Didier.
- Zymner, R. (2003). *Gattungstheorie : Probleme und Positionen der Literaturwissenschaft*. Paderborn: Mentis.

4. BEANTRAGTE MODULE/MITTEL

4.1. Mittelbeantragung DFG:

4.1.1. Basismodul

4.1.1.1 Personal

a) Wissenschaftliche Mitarbeiter	
Universität Osnabrück: eine halbe Stelle TV-L 13/Stufe 2:	98100 €
Universität Erlangen: eine halbe Stelle TV-L 13/Stufe 2:	98100 €
Universität Bonn: eine halbe Stelle TV-L 13/Stufe 2	98100 €
b) Wissenschaftliche Hilfskräfte: Universität Osnabrück:	
eine studentische Hilfskraft/10 Std./Monat	9 000 €
Universität Erlangen: eine studentische Hilfskraft/10 Std./Monat	9 000 €
Universität Köln: eine studentische Hilfskraft/10 Std./Monat	9 000 €
Universität Bonn: eine studentische Hilfskraft/10 Std./Monat	9 000 €
c) Universität zu Köln: Werkverträge für Softwareentwicklung	10 000 €
Summe Personalkosten:	331300 €

Die Vergrößerung der deutschen Arbeitsgruppe um M. Gymnich (Bonn) ergibt sich aus der zentralen Kritik der Gutachten zum erstingereichten Antrag, derzufolge ein Mangel an literarischer Expertise auf deutscher Seite bestand; gleichzeitig wird damit eine bessere Komplementarität der deutschen und französischen Arbeitsgruppen im Bereich der literarischen und stilistischen Analyse erreicht. Angesichts der Komplexität des Projekts sind drei Post-

Doc-Stellen in Osnabrück, Bonn und Erlangen notwendig, die Fragen der Phraseologie des Romans, der Schnittstellen zwischen Stilistik und Genretheorie sowie der Kontrastivlinguistik bearbeiten sollen.

Die studentischen Hilfskräfte in Osnabrück, Bonn und Köln sollen bei der Zusammenstellung und Aufbereitung der Korpora (Lemmatisierung, PoS-Tagging, Parsing), beim Training von Software, bei der Literaturrecherche und -beschaffung, bei der Einrichtung der Datenbank und bei der verlegerischen Vorbereitung der Veröffentlichungen helfen. Die studentische Hilfskraft in Erlangen soll bei der Vorbereitung der internationalen Tagung (vgl. 3.1.4) helfen.

4.1.1.2 Sachmittel

4.1.1.2.1 Mittel für Geräte

Universität Osnabrück: Datenserver	2000 €
Universität Osnabrück: Rechner für Mitarbeiter	1000 €
Universität Erlangen: Datenserver	2000 €
Universität Erlangen: Rechner für Mitarbeiter	1000 €
Universität Bonn: Rechner für Mitarbeiter	1000€
Summe:	7000 €

4.1.1.2.2 Mittel für Reisen

Universität Osnabrück 3 x Reisen nach Grenoble (2 Personen) à 800 €/Reise	4.800
2 x Teilnahme an Tagung Ausland (2 Personen) à 1100 €/Reise	4.400
1 x Teilnahme an Tagung Inland (2 Personen) à 500 €/Reise	1.000
Universität Erlangen 3 x Reisen nach Grenoble (2 Personen) à 800 €/Reise	4.800
2 x Reisen nach Osnabrück (2 Personen) à 500 €/Reise	2.000
2 x Teilnahme an Tagung Ausland (2 Personen) à 1100 €/Reise	4.400
1 x Teilnahme an Tagung Inland (2 Personen) à 500 €/Reise	1.000
Universität Bonn 3 x Reisen nach Grenoble (2 Personen) à 800 €/Reise	4.800
2 x Reisen nach Osnabrück (2 Personen) à 500 €/Reise	2.000
2 x Teilnahme an Tagung Ausland (2 Personen) à 1100 €/Reise	4.400
1 x Teilnahme an Tagung Inland (2 Personen) à 500 €/Reise	1.000
Summe	34600 €

Zur Begründung s. 4.2.4.

4.1.1.2.3 Mittel für Gäste

Universität Erlangen: 3 Gastredner bei internationaler Tagung à 800€	2 400
--	-------

4.1.1.2.4 Publikationskosten

Universität Osnabrück/Erlangen, Publikation eines Tagungsbandes	2 250
---	-------

4.1.1.2.5 Sonstige Kosten

Vorbereitung einer internationalen Tagung	15 000
---	--------

4.1.2. Modul Eigene Stelle

-

4.1.3. Modul Vertretung

-

4.1.4. Modul Mercator-Fellow

-

4.1.5. Modul Projektspezifische Workshops

-

4.1.6. Modul Öffentlichkeitsarbeit

-

4.2. Mittelbeantragung ANR:

4.2.1. Personal

- <u>Maschinelle Sprachverarbeitung</u> : 2 ingénieurs de recherches (IR) :	
1 Vertrag über 12 Monate + 1 Vertrag über 13 Monate= 25 Monate, d.h. 2 x 34 000	67 700
- <u>Linguistik</u> : 1 post-doc 12 mois	40 000
- 1 Ingénieur de recherche über 18 Monate	49 000
- <u>Stilistik</u> : ein Ingénieur de recherche über 18 Monate	32 700
Total :	189 400

Kompetenzanforderungen für die 2 Ingénieurs de recherche: Lemmatisierung, Annotation, Kenntnisse von Softwareanwendungen für den Aufbau und die Aufbereitung von Korpora und die Schaffung einer Datenbank (s. Tabelle 1, Aufgaben T1.1. à T1.5. ; T3.1. et T3.3.)

Kompetenzanforderungen für die Post-Doc-Stelle in der Linguistik: Spezialisierung in Korpuslinguistik und Phraseologie ; Aufgaben insbesondere T1, T3, T4, T5, T7.

Kompetenzanforderungen für den Ingénieur de recherche in der Linguistik (Zeitvertrag 18 Monate) : T1, T3, T7.

Kompetenzanforderungen für den Ingénieur de recherche in der Stilistik (Zeitvertrag von 12 Monaten) : T2, T3, T8.

Die Unterschiede in den vorgesehenen Personalausgaben zwischen Deutschland und Frankreich ergeben sich aus den unterschiedlichen Vorschriften der ANR (die Summe für befristete Verträge darf 30% der Summe für festangestelltes Personal nicht überschreiten) und der DFG für diesen Budgetposten.

4.2.2. Sachmittel

Anschaffung von Laptops (10 x 1 000 €)	10 000
1 Internet-Server	2 000
Summe :	12 000

4.2.3. Materialbeschaffung, Kleingeräte usw.

Anschaffung externer Festplatten	2 500
----------------------------------	-------

Anschaffung von Software	4 000
Verbrauchsmaterial, Bürotechnik	4 000
Anschaffung von Büchern	3 500
Photokopien, Digitalisierung von Dokumenten	3 500
Summe	17 500
6 Praktikanten (2 Maschinelle Sprachverarbeitung, 2 Sprachwissenschaft, 2 Stilistik) x 6 Monate	20 000

Die für die Praktikanten vorgesehene Summe (Master 2-Niveau) fällt in den Bereich „Fonctionnement du budget“. Die Praktikanten werden bei verschiedenen Aufgaben als Hilfskräfte tätig werden und dabei gleichzeitig eine wissenschaftliche Ausbildung erfahren.

Summe: 37 500

4.2.4. Reisekosten

Der Anteil der Reisekosten am Gesamtvolumen der Finanzierung beträgt 19%. Während der dreijährigen Laufzeit des Projekts kommen die Mitglieder und Leiter der jeweiligen Arbeitsgruppen zu regelmäßigen Treffen zusammen. Wir planen zwei „Vollversammlungen“ pro Jahr (jeweils eine im März in Frankreich und im Oktober in Deutschland). Hinzu kommen 3 Treffen von Teilen der Gesamtgruppe gemäß Arbeitsplan.

Die Themen der einzelnen Treffen können vorläufig folgendermaßen bestimmt werden :

1. Treffen : Erweiterung der Korpora, Verbesserung der Abfragewerkzeuge, Aufgabenverteilung (Informatik, Linguistik, Stilistik und Genretheorie)
2. Treffen : Abstimmung der Methodik, linguistische und stilistische Pilotstudie zu bestimmten Subkorpora
3. Treffen : Zwischenbilanz zu Datenverarbeitung und -kodierung sowie sprach- und literaturwissenschaftlicher Analyse
4. Treffen : Zwischenbilanz zu Datenverarbeitung und -kodierung sowie sprach- und literaturwissenschaftlicher Analyse
5. Treffen : Zusammenführung der Ergebnisse der Kodierung und der linguistischen und stilistischen Analyse. Erstellung einer Typologie literaturspezifischer Konstruktionen.
6. Treffen: Kontrastierung der literaturspezifischen Konstruktionen in den drei untersuchten Sprachen; Erstellung der Datenbank; Vorbereitung der Konferenz und der Publikation.

Die Kosten werden mit durchschnittlich 800 € pro Person und Reise angesetzt (400 € Hotelkosten, 400 € Reisekosten). Es sind in Abhängigkeit vom Arbeitsplan und dem jeweiligen Stand des Projekts drei weitere Teiltreffen der Koordinatoren und einiger Mitglieder der Arbeitsgruppen vorgesehen.

Um die Projektergebnisse über Deutschland und Frankreich hinaus bekannt zu machen, ist die aktive Teilnahme *aller* Mitglieder der Arbeitsgruppen an internationalen Tagungen notwendig. Nach unseren Erfahrungswerten kostet ein Aufenthalt auf einem internationalen Kongress im Ausland ca. 1100 € pro Person (Reisekosten, Hotelkosten, Teilnahmegebühren). Wir planen die Teilnahme an 2 internationalen Symposien und 1 Tagung in Frankreich pro Person für die Gesamtdauer des Projekts. Für die Tagung in Frankreich werden 700 € pro Person angesetzt (250 € Reisekosten, 300 € Hotelkosten, 150 € Teilnahmegebühren). Zur

erfolgreichen Durchführung des Projekts sind außerdem vier Teilnahmen externer Kooperationspartner an den Treffen der Arbeitsgruppen notwendig (aus Polen, der Schweiz, Deutschland und Frankreich, s. Auflistung). Daraus ergeben sich zusammengefasst folgende Beträge:

3 Gesamttreffen für 8 Personen in Deutschland x 800 €	19200
3 Teiltreffen für 2 Personen x 800 €	4800
2 internationale Konferenzen pro Person 8 x 1100 €	17600
1 nationale Konferenz pro Person 8 x 700 €	5600
Teilnahme an Abschlusskonferenz in Deutschland: 8 x 1100 €	8800
4 Einladungen für externe Kooperationspartner	4000
Dienstreisen von D. Legallois (Caen-Grenoble) 3 x 600€	1800
Summe	61 800

N.B.: Die Reisen und Materialkosten für D. Legallois (Crisco, Caen) werden in das Budget des Lidilem aufgenommen, da sie 15000 € pro Jahr nicht überschreiten.

Die unter Rubrik 4.2.1 bis 4.2.4 angegebenen Summen fassen das Budget der zwei französischen Partner zusammen (Lidilem und Litt&Arts équipe Traverses 19-21).

Die Differenz zwischen den Gesamtsummen im deutschen und französischen Budget ergibt sich aus der größeren Zahl festangestellter Forscher in den französischen Arbeitsgruppen.

4.2.5. Dienstleistungen

Auf Werkvertragsbasis :

Entwicklung und Aktualisierung der Projektwebsite	3000
Übersetzung von Artikeln, Vorschuss für Projektpublikation, sonstiges	9 000
Summe	12 000

5. ÜBERBLICK ÜBER DIE BEANTRAGTEN MITTEL

	FR		D
Dépenses de personnel	189 400	Basismodul: Personalkosten Modul Eigene Stelle Modul Vertretung Modul Mercator-Fellow	331300
Équipements	12 000	Basismodul: Wissenschaftliche Geräte	7 000
Frais de missions	61 800	Basismodul : Reisekosten	34600
Autres dépenses	37 500	Basismodul : Sonstige Kosten	15 000
		Basismodul : Publikationskosten	2 250
Prestations de service	12 000	Basismodul : Mittel für Gäste	2 400
Frais de gestion	12 508		
		Modul Projektspezifische Workshops	
		Modul Öffentlichkeitsar- beit	
Total France	325 208 €	Gesamtsumme Deutschland	392550
Gesamtsumme	717 758€		

6. Weitere Mittel für die Durchführung des Vorhabens

Keine weiteren Mittel außer der Grundausstattung.

7. Weitere Erklärungen

Ein Antrag auf Finanzierung dieses Vorhabens wurde bei keiner anderen Stelle eingereicht. Wenn wir einen solchen Antrag stellen, werden wir die Deutsche Forschungsgemeinschaft und die Agence Nationale de la Recherche unverzüglich benachrichtigen.

Der Vertrauensdozent der DFG an der Universität Osnabrück wurde von der Antragstellung unterrichtet.

Anhang 1²
Mehrsprachige Korpora

1.1. Französisch

Korpus	Zeitraum	Wortzahl	Textsorte	Sprache
Fiction (Emolex)	1950-2010	15.970.000	Literatur	D
Presse (Emolex)	2007-2008	108.740.000	Presse	D
Scientext français	2000-2010	4.880.000	Wissenschaft	D
Summe:		129.590.000		

1.2. Englisch

Korpus	Zeitraum	Wortzahl	Textsorte	Sprache
Fiction	1900-2010	30.000.000	Literatur	E
Presse	2007-2008	105.500.000	Presse	E
Scientifique	1970-2005	30.000.000	Wissenschaft	E
Summe:		165.500.000		

1.3. Deutsch

Korpus	Zeitraum	Wortzahl	Textsorte	Sprache
Fiction (Emolex)	1950-2010	6.370.000	Literatur	D
Presse (Emolex)	2007-2008	93.700.000	Presse	D
Summe:		100.070.000		

² Die deutsch-, französisch- und englischsprachigen Korpora, über die wir in EMOBASE verfügen, werden ergänzt und in ein ausgewogenes Verhältnis gebracht werden.