

STRAND 3: Typology and dynamics of linguistic systems

Head: Katharina Haude, SeDyL; Dmitry Idiatov, LLACAN

Number of Work packages: 14

10 EFL teams involved: CRLAO, HTL, LACITO, LATTICE, LLACAN, LLF, LPP-P3, LPP-P5, MII, SEDYL

Number of participants: 91

Strand 3 investigates the diversity of the world's languages. Most of its members are field linguists with a strong expertise in the description and documentation of lesser-studied languages. The research carried out here focuses on typology (i.e. the study of the universals and differences found in the world's languages independently from their genealogical relation), as well as the variation in space and time found within and across areally or genealogically close languages. It involves, on the one hand, fine-grained cross-linguistic analyses of particular linguistic phenomena, which can only be carried out by experts on the individual languages concerned, and on the other hand, the investigation of the long-term and short-term effects of contact between related and unrelated languages.

Accordingly, Strand 3 is divided into two work packages. The first, "Grammar and Lexicon in Synchrony and Diachrony (GL)", comprises nine projects (work packages) that deal with the analysis of selected grammatical and lexical phenomena, many of which have as yet been insufficiently studied from a typological perspective. GL1 investigates interjections and ideophones, categories that usually occupy a marginal place in grammatical descriptions, but that are interesting for their own sake as they are located between arbitrary and motivated language. Patterns in the semantic organization of the lexicon of diverse languages are the research topic of GL2, which investigates the cross-linguistic structure of several inadequately explored lexical fields and aims at establishing the historical trajectory of semantic change in these domains, including areal contact effects. GL3 focuses on the expression of associated motion and posture in languages from different linguistic areas and families. GL4 aims at a fine-grained typology of nonverbal predicates ("pseudo-verbs"), critically evaluating some of the proposals made in the literature to this respect. GL5 looks at noncanonical inflectional patterns underrepresented in morphological typology. GL6 takes a discourse perspective to investigate language-internal alternations between different grammatical constructions. GL7, a continuation of a project of Strand 6 ("Corpus") of the previous LABEX-EFL, has as its aim to enlarge and refine the corpus of Mande languages. Similarly, GL8 is continuing work on the reconstruction of the Austronesian and Niger-Congo language phyla. Finally, the first work package also hosts a project dealing with the meta-level of grammatical description: GL9, which belonged to Strand 7 of the previous LABEX-EFL, looks at descriptive traditions beyond the Latin-based model. Crucially, all projects in this work package have as their data base a high number of lesser-described and partly endangered languages, and in this way will contribute greatly to a more detailed knowledge of these languages.

Of the five projects in the second work package, "Variation and Contact (VC)", four are continuations of existing LABEX projects. VC1 continues its research on West-Asian languages, now focussing on word order and differential argument marking. VC2 takes a multidisciplinary view on the linguistic variation of the Central Gallo-Romance languages in

Central France, now also including a psycholinguistic dimension. VC3 investigates the striking speaker-dependent syntactic variation that is so frequent in creole languages. VC4, a new project, proposes a fine-grained analysis of contact between closely related languages, mostly Indo-European and Afroasiatic. VC5, finally, investigates code-switching through multifactorial analysis of plurilingual corpora.

Many of the projects in Strand 3 build on results of the first phase of the LABEX-EFL, e.g. by using databases that were created there. This empirical basis will be complemented by the creation of new databases designed for quantitative and qualitative analysis, benefiting from the expertise in more than 100 languages and language families.

The originality of Strand 3 lies in its combination of historical linguistics (macro-families reconstruction, long term typological change), contact linguistics, language documentation and typology, all grounded on both qualitative and quantitative approaches (intensive use of databases and electronic corpora to assess theoretical issues). The increasing interoperability between all these resources and the cross-checking of hypotheses in the different approaches will provide innovative scientific perspectives.

***Work package 1:
Grammar and Lexicon in Synchrony
and Diachrony (GL)***

<p>GL1: Ideophones and interjections in a typological, areal and diachronic perspective</p>
--

PIs	Aimée Lahaussois (HTL) Yvonne Treis (LLACAN)
Participants	HTL: Aimée Lahaussois (CR), Emilie Aussant (CR) LACITO: Sylvain Loiseau (MCF) LLACAN: Yvonne Treis (CR), Marie-Claude Simeone-Senelle (DR ém.), Tatiana Nikitina (CR), Loïc-Michel Perrin (MCF), Paulette Roulon-Doko (DR ém), Nicolas Quint (DR) LLF: Patrick Caudal (CR) CRLAO: Guillaume Jacques (DR) autres: Luca Nobile (MCF) (U Bourgogne)

1. Presentation

1.1. Abstract

We propose in this project to study interjections and ideophones: both are minor, invariable, and at least partly sound-symbolic word classes. Throughout their history, these word classes have been faced with definitional problems of a rather different nature than other word classes, in part because they are situated at the boundary of arbitrary vs 'motivated' language.

Starting with case studies, this project will take a typological, diachronic, areal and multimodal approach to the study of the semantics, morphology and syntax of interjections and ideophones.

1.2. Objectives

Interjections, invented by Latin scholars to fill a gap in the 8-part word classification after the dismissal of articles (which existed in Greek but not in Latin), are on the whole defined, throughout history, in negative terms on account of their phonological marginality, their non-participation in morphological processes and their syntactic autonomy.

Ideophones, the term for which was initially used interchangeably with 'interjection', are also difficult to define notionally, on account of the great variety of ideophonic lexemes in the world's languages, and many definitions, as with interjections, highlight the non-standard phonology found with this word class.

We have decided to consider these two word classes within the same typological project with the hope that setting them up in opposition to each other will make us more efficient in searching for consistent (and contrasting) definitions.

Considering that in many languages, ideophones and interjections do not participate in morphological processes, it is reasonable to question whether they should, or even can be, topics in

grammatical descriptions. These words express emotions or reactions (interjections) and sensory imagery (ideophones), in other words subjective notions with great cultural variability. The difficulty in translating them and accurately capturing their meaning makes describing them difficult (for non-native linguists), marginalizing them even more compared to other word classes.

Another question of interest is that of the type of corpus which is needed to study these word classes: interjections and ideophones are more frequent in corpora of spontaneous speech, most often interactive, in other words types of linguistic production that are significantly different from those used for research into nominal or verbal morphology. Additionally, they will often only be interpretable to the fieldworker after a relatively long period in the field, adding to the complexity of their description.

Studies of interjections are certainly not a recent phenomenon, but more recent work has tended to be situated within the field of pragmatics (Ameka 1992, Cram 2008, Poggi 2009) rather than from the perspective of typology. Ideophones, on the other hand, have been given cross-linguistic consideration, through work on sound symbolism (Hinton et al 1994), and in typology (Voeltz, F.K.E. & C. Kilian-Hatz 2001), and descriptive work (Dingemans 2011, Reiter 2011). This work on ideophones has resulted in the following implicational hierarchy: sound < movement < visual patterns < other sensory perceptions < inner feelings and cognitive states (languages with ideophones covering a semantic field to the right will also have ideophones in semantic domains to the left of that point)

Working from case studies on individual languages, this program will approach the analysis of ideophones and interjections from a **typological, diachronic, areal** and **multimodal perspective** in the consideration of the semantics, morphology, syntax and pragmatics of these word classes.

2. Bibliography

- Ameka, F.K. (ed.) 1992. *Journal of Pragmatics* 18 [Special issue on Interjections].
- Buridant, C. (ed.) 2006. *L'interjection: jeux et enjeux*. [Numéro spécial de *Langages* 161]
- Cram, D. (ed.) 2008. *The Henry Sweet Society Bulletin* 50 [Special issue on Interjections].
- Cuenca, M.J. 2000. Defining the indefinable? Interjections. *Syntaxis* 3: 29-44.
- Dhoorre, C.S. & Tosco, M. 1998. 111 Somali Ideophones. *Journal of African Cultural Studies* 11, 2: 125-56.
- Dingemans, M. 2009. Ideophones in unexpected places. In: Austin, P.K. et al (eds.). *Proceedings of the 2nd Conference on Language Documentation and Linguistic Theory*, pp. 83-97. London: SOAS.
- _____. 2011. *The meaning and use of ideophones in Siwu*. PhD diss., Radboud University, Nijmegen.
- _____. 2012. Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass* 6, 10: 654-72.
- Hinton, L., J. Nichols & J.J. Ohala (eds.) 1994. *Sound Symbolism*. Cambridge: CUP.
- Jendraschek, G. 2001. Semantic and structural properties of Turkish ideophones. In: Johansen, L. (ed.). *Turkic Languages*, pp. 88-103. Wiesbaden: Harrassowitz.
- Lahaussais, A. 2016. Where have all the interjections gone? A look into the place of interjections in contemporary grammars of endangered languages. In: Assunção, C., F. Gonçalo & R. Kemmler (eds.). *Tradition and Innovation in the History of Linguistics: Contributions from the 13th International Conference on the History of the Language Sciences (ICHoLS XIII), Vila Real, 25-29 August 2014*. Münster: Nodus Publikationen.
- Lahaussais, A. 2017. Ideophones in Khaling Rai. *Linguistics of the Tibeto-Burman Area* 40.2, 179-201.
- Meinard, M.E.M. 2015. Distinguishing onomatopoeias from interjections. *Journal of Pragmatics* 76: 150-168.
- Nobile, L. & E.L. Vallauri 2016. *Onomatopea e fonosimbolismo*. Rome: Carocci.
- Norricks, N.R. 2009. Interjections as pragmatic markers. *Journal of Pragmatics* 41, 5: 866-91.

- Nuckolls, J.B. 1999. The case for sound symbolism. *Annual Review of Anthropology* 28: 225-52.
- Poggi, Isabella 2009. The language of interjections. In: Esposito, Anna, Hussain, Amir, Marinaro, Maria & Martone, Raffaele (eds.) *Multimodal signals: Cognitive and Algorithmic Issues* (Lecture Notes in Artificial Intelligence 5398), Berlin / Heidelberg, Springer-Verlag, p. 170-186.
- Reiter, S. 2011. *Ideophones in Awetí*. PhD diss., Christian-Albrechts-Universität zu Kiel.
- Simeone-Senelle, M.-C. 2017. Expression de l'émotion en sudarabique moderne. Interjection et autres formes figée. In: Tersis, N. & P. Boyeldieu (eds.). *Le langage de l'émotion: variations linguistiques et culturelles*, p. 141-16. Louvain/Paris: Peeters.
- Smith, Benjamin 2012. Language and the frontiers of the human: Aymara animal-oriented interjections and the mediation of mind. *American Ethnologist* 39, 2: 313-324.
- Swiatkowska, M. 2006. L'interjection: entre deixis et anaphore. *Langages* 161: 47-56.
- Tosco, M. 2006. The ideophones in Gawwada. In: Uhlig, S. (ed.). *Proceedings of the XVth International Conference of Ethiopian Studies, Hamburg, July 20-25, 2003*, pp. 885-892. Wiesbaden: Harrassowitz.
- van Gijn, R. 2010. Middle voice and ideophones, a diachronic connection: The case of Yurakara. *Studies in Language* 34, 2: 273-297.
- Voeltz, F.K.E. & C. Kilian-Hatz (eds.) 2001. *Ideophones*. Amsterdam, Philadelphia: Benjamins.

GL2: Lexical typology across time and space
--

PIs	Alexandre François, Lameen Souag, Martine Vanhove
Participants	LACITO: Isabelle Bril, Sebastian Fedden, Marc-Antoine Mahieu, Claire Moyse-Faurie, Catherine Taine-Cheikh, Sylvain Loiseau, Antoinette Schapper LLACAN: Tatiana Nikitina, Konstantin Pozdniakov, Guillaume Segerer SeDyL: Stefano Manfredi, Nicole Tersis HTL: Aimée Lahaussais Lattice: Benjamin Fagard

1 Presentation

1.1 Abstract

The project *Lexical typology across time and space* is at the crossroads between synchronic lexico-semantic typology and the diachronic study of semantic change. It aims at examining several semantic fields whose cross-linguistic structure has yet to be mapped in detail in the perspective of language change. This will be achieved through the study of cross-linguistic variations in the lexicon which show recurrent patterns and typological trends in a limited number of semantic domains (motion, impact and manipulation, cognition), using existing lexical databases and building a dedicated one. Understanding the dynamics of lexical typology is essential not only in order to explain the stability of statistical universals, but more broadly in order to shed light on one of the thorniest problems of historical linguistics: semantic reconstruction.

1.2 Objectives

Languages vary not just in their grammatical structure, not just in the form of lexical items, but also in the structure of their lexicon. Anyone speaking more than one language can easily come up with examples of cross-linguistic differences due to the internal organization of the lexicon – the way English lumps *savoir* and *connaître* together as *know*, the absurdity of translating idioms literally, the way a common word in one language has no lexicalized equivalent in another, the mistake in assuming that *hello* is used in the same contexts as *bonjour*... All these examples illustrate *structural* properties of the lexicon – i.e. aspects of a language’s semantic organization that are stored in its lexical component, and that can be compared across languages.

The lexicon is often defined in terms of unpredictability, by contrast with the regularities of grammar. However, unpredictability does not mean randomness. As suggested by the growing body of typological work summarised in Koptjevskaja-Tamm, Rakhilina and Vanhove (2015), cross-linguistic variation in the lexicon does show regularities and exceptions, recurrent patterns and typological trends, all of which are as much worthy of scientific scrutiny as the more often studied domains of grammar. Works such as Koptjevskaja-Tamm (2015), and even a special issue of *Linguistics* (50:3, 2012), illustrate a promising trend of mapping out the cross-linguistic typology of particular semantic domains, while works such as Urban (2011, 2013) show

that synchronic cross-linguistic data on colexification (François 2008) provides diachronic information on the nature and directionality of semantic change. Databases such as CLICS (<http://clics.lingpy.org>) and DatSemShifts (<http://semshifts.iling-ran.ru/>) illustrate the new possibilities opened up in this domain by digital methods. Meanwhile, semantic elicitation methods have advanced considerably, as illustrated by work on topics such as SMELL (Majid & Burenhult 2014).

Despite this significant progress, however, the literature available on the dynamics of lexical typology – a crucial aspect of semantic change – tends to be fragmentary, interested in one language area, or one semantic field, at a time. This scientific landscape contrasts sharply with the more fully developed theories available for phonological change or for grammaticalization. The time is ripe to take our understanding of the dynamics of lexical typology further.

1.3 Aims and methods

Our project will build on the results of the TUL project "Towards a typology of semantic associations" (2002-2009) directed by Martine Vanhove, as well as on recent advances in the domain. We also intend to make use of existing dictionaries – including the online dictionaries currently under development at LACITO – as well as online databases, such as the RefLex database for African languages at LLACAN. Based on these materials and more gathered in the field, we intend to examine several semantic fields whose cross-linguistic structure has yet to be mapped in detail in the perspective of language change. Understanding the dynamics of lexical typology is essential not only in order to explain the stability of statistical universals, but more broadly in order to shed light on one of the thorniest problems of historical linguistics: semantic reconstruction. This project will result in a better understanding of the forms taken and paths followed by semantic change within the lexicon.

Without wishing to restrict ourselves to a single semantic domain, we intend to explore in depth semantic domains of particular interest, including verbs of motion and associated motion (*go, come, bring, carry, hold, send...*), in association with Guillaume Jacques and Patrick Caudal's working group on associated motion; verbs of impact and manipulation (*put, take, shift, twist, raise, remove, throw...*), extending the 19-language sample of Kopecka & Narasimhan (2012) to a more balanced range of languages and a wider range of notions; and verbs of cognition (*know, feel, believe, understand, forget etc.*).

1.4 Expected outcomes

This project will map out the cross-linguistic structure of several inadequately explored lexical fields, and establish as far as possible the historical trajectory of semantic change in these domains, while identifying areal contact effects. The results of the project will be published in the form of a peer-reviewed edited volume and a database.

1.5 Projects and contracts submitted in relation to one of the topics of the operation

Fedden, Sebastian. (demandé). *History and Typology of Central New Guinea Languages*. FNS-ANR.
Lahaussais, Aimée. 2020-2024 (demandé): Membre du programme du Labex EFL *Associated motion*.
Schapper, Antoinette. 2014-2018. *The evolution of the lexicon. Explorations in lexical stability, semantic shift and borrowing in a Papuan language family*. NWO 275-70-040.
Souag, Lameen & Taine-Cheikh, Catherine. 2014-2018: Membres du programme du Labex EFL Axe 3, *Long-term typological changes in language*. (Responsable: Christiane Pilot-Raichoor.)
Vanhove, Martine. 2002-2008: Responsable du programme 9 de la Fédération Typologie et Universaux Linguistiques, *Typologie des rapprochements sémantiques*.

Vanhove, Martine. 2006-2009: Responsable scientifique pour la partie française du projet européen INTAS (Ref. Nr. 05-1000008-7917) *Core vocabulary in a typological perspective: semantic shifts and form / meaning correlations* (TypVoc) (PI Maria Koptjevskaja-Tamm, Univ. Stockholm).

<p>GL3:</p> <p>Associated Motion and Posture</p>
--

PIs	Patrick CAUDAL (LLF), Guillaume JACQUES (CRLAO)
Participants	<p>LACITO: Isabelle Brill (DR), Claire Moysse-Faurie (DR émérite), Paulette Roulon (DR), Guillaume Jacques (DR)</p> <p>LLF: Patrick Caudal (CR), Guillaume Fon Sing (MCF), Berthold Crysmann (LLF, CR),</p> <p>LLACAN: Marie-Claude SIMEONE-SENELLE (DR)</p> <p>CRLAO: Christine Lamarre (PRF Inalco), Tatiana Yakovleva (MCF U. Paris-Diderot)</p> <p>Lattice: Laure Sarda (CR), Benjamin Fagard (CR)</p> <p>HTL: Aimée Lahaussis (CR)</p>

1. Presentation

1.1 Abstract

This research project will be devoted to the typology of associated movement (Koch 1984; Wilkins 1991) and associated posture (Enfield 2002), with an emphasis on areal typological approaches (Austronesian, Australian, Sino-Tibetan languages, Niger-Congo, in particular). Associated posture and associated movement will be understood here in the broadest possible sense, i.e., ranging from dedicated lexical classes to specifically grammatical phenomena - whether related to particles, clitics, or affixes. Relations between associated posture/movement and description of events will be more specifically targeted. In many languages of the world, in fact, these categories are linked e.g. to serial verb systems, TAM auxiliaries/verbal periphrases, etc.

1.2 Objectives

It has been established that associated movement and posture markers frequently develop new functions, thereby forming complex grammatical (or lexico-grammatical) microsystems. Such systems sometimes combine markers (or former markers) of movement and associated posture, as e.g. in Arnhem Land languages (cf. e.g. Mawng (Singer 2006), Iwaidja (Caudal & Mailhammer 2016), Anindhilyakwa (Bednall to appear), Ndjéšana (McKay 2000), or in some Amazonian languages, (Vuillermet 2012; Vuillermet 2015)), or add markers belonging to other notional micro-classes. Such grammatical systems can be endowed with discourse-structural functions (Bourdin 2008), or TAME functions (giving rise to tense / aspect meanings, or forming novel discursive connectors ...) – but also develop towards deixis, or voice categories e.g. the passive.... See (Newman 2002) or (Devos & van der Wal 2014) for some illustrations of such developments. The crosslinguistic consistency of these lexical and grammatical developments demonstrates the deep grammaticalization tendency (cf. (Bourdin 2014: 119)) potential of such categories. This fact seems to take on a particular importance in the light of recent experimental work highlighting the intimate relationships linking postural perception, perception of movement and linguistic learning, cf. (Pfeifer 2006; Benitez & Smith 2012; Morse et al. 2015). In turn, this suggest that the categories of associated motion and associated posture offers vast

and fecund research perspectives, involving fundamental connections between language, cognition/self-perception, and the world.

1.3 Projected activities

The new operation will first focus on collecting relevant data across as many relevant linguistic areas and phyla as possible; experimental and quantitative methods will be regarded as priority, whenever possible. It should be highlighted that the relevant data remains significantly underdescribed across most linguistic areas / phyla. Novel descriptive and (areal) typological results will be hopefully produced. Temporary contracts, as well as an engineer position to be shared with other operations will be considered.

This operation shall begin in 2020, and will capitalize on the work previously conducted within the GD4 operation (Strand 3), as well as through the 'EvEs' (Événements et Espace) Fédération TUL (Typology & Universals Linguistics Federation) project (2018; coordinators: P. Caudal (LLF) & M. Vuillermet (DDL)). The ultimate goal of this project is to help prepare an ANR project for 2021-2022.

GL4:
The typology of pseudo-verbs

PI	Mena B. LAFKIOUI (LLACAN)
Participants	<p>CRLAO: Hilary Chappell (Directeur d'études, EHESS) LACITO: Lameen Souag (CR), Catherine Taine-Cheikh (DR emeritus) LLACAN: Maximilien Guérin (Postdoc), Mena B. Lafkioui (DR), Ronny Meyer (MCF Inalco), Marie-Claude Siméonne-Sennelle (DR emeritus), Mark Van de Velde (CR), Valentin Vydrin (Professeur des universités Inalco) LLF: Reintges Chris (CR) SeDyl: Stefano Manfredi (CR) Several international co-operators will join the project</p>

1 Presentation

1.1 Abstract

The present project proposes to study for the first time the particular category of “pseudoverbs” from a cross-linguistic and typological perspective so as to deepen our understanding of the nature and functioning of the different part-of-speech categories in the world’s languages. In order to do so, the project puts forward to verify certain concepts from Role and Reference Grammar (RRG), which offer compelling alternatives to clarify the grammatical relations involving pseudo-verbs.

1.2 Objectives

Pseudo-verbs form a specific part-of-speech category of all sorts of non-verbal items dedicated to predication and displaying interesting and challenging morphosyntactic features relevant to theoretical and typological linguistics, which relate to the following primary issues:

First, even though none of the pseudo-verbs are “proper” verbs, all of them concern predication or co-predication. While sharing some features with verbs and co-verbs, like for instance having TAM or specialized pronominal paradigms or specific inflections, pseudo-verbs are in essence distinct grammatical forms with specific functions (see, for instance, in Berber: Aikhenvald 1995, Lafkioui 2001, 2011, Satzinger 2005, and in Semitic: Brustad 2000, Huehnergard 1987).

Second, the diachrony of pseudo-verbs shows interesting grammaticalisation phenomena relating to their development mostly out of non-verbal items which took on verbal forms and functions, such as the prepositional predicate in Berber and in Arabic which conveys ‘to have’ (e.g. with-3sg child ‘he has a child’). Other regular pseudo-verb developments concern demonstratives, locatives, adverbs and pronouns. Furthermore, certain languages show how some of these developments are cyclic in nature, since they are evolving in the reverse direction as they are losing some of the verbal features (de-grammaticalisation), causing a kind of cyclic movement (e.g. Berber, Lafkioui 2011).

Third, the pseudo-verbs cast doubt upon the use of traditional grammatical concepts like “subject” and “object” for their morphosyntactic relations, as the “logical subject” of this kind of configurations is marked in a different way in certain languages and contexts (e.g. Berber,

Arabic), often even matching with the “object position” in verb sentences, at the same time as having many of the morphosyntactic characteristics of subjects. The Unaccusative Hypothesis (Perlmutter 1978) might shed some light on certain aspects of this particular grammatical relationship.

The present project proposes to study for the first time these major issues relating to pseudo-verbs from a cross-linguistic and typological perspective so as to deepen our understanding of the nature and functioning of the different part-of-speech categories in the world’s languages. In order to do so, and hence to address the pseudo-verb category from a typological perspective, the project puts forward to verify certain concepts from Role and Reference Grammar (RRG), such as the concept of “macrorole” (Van Valin 2005, 2007), which offers compelling alternatives to clarify the grammatical relations involving pseudo-verbs.

1.3 References

- Aikhenvald, Alexandra Y. 1995. “Split Ergativity in Berber Languages.” *St. Petersburg Journal of African Studies* 4: 39–68.
- Brustad, Kristen. 2000. *The Syntax of Spoken Arabic: A Comparative Study of Moroccan, Egyptian, Syrian, and Kuwaiti Dialects*. Washington, DC: Georgetown University Press.
- Huehnergard, John. 1987. “‘Stative,’ Predicative Form, Pseudo-Verb.” *Journal of Near Eastern Studies* 46 (3): 215–232.
- Lafkioui, Mena B. 2001. “Typologie des assertions affirmatives non-verbales du rifain. Approche de syntaxe intégrée.” *Frankfurter Afrikanistische Blätter*, no. 13: 97–120.
- . 2011. *Etudes de la variation et de la structuration linguistiques et sociolinguistiques en berbère du Rif*. Köln: Rüdiger Köppe Verlag.
- Perlmutter, David M. 1978. “Impersonal Passives and the Unaccusative Hypothesis.” *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 4: 157–89.
- Satzinger, Helmut. 2005. “On the Assumed Ergativity of the Berber Language(s).(Eds.)” In *Proceedings of the 10th Meeting of Hamito-Semitic (Afroasiatic) Linguistics, Florence, 18–20 April 2001.*, edited by Pelio Fronzaroli and Paolo Marrassini, 381–389. Firenze: Università di Firenze - Dipartimento di Linguistica.
- Van Valin, Robert D. 2005. *Exploring the Syntax-Semantics Interface*. Cambridge ; New York: Cambridge University Press.
- . 2007. “The Role and Reference Grammar Analysis of Three-Place Predicates.” *Suvremena Lingvistika* 63 (1): 31–63.

1.4 Work plan

- A regular seminar (three sessions per year) to present and discuss the data and methodological and theoretical issues concerning pseudo-verbs.
- The study will be mainly based on the data and expertise of the participants. A small corpus sample of elicited data obtained from native informants, who will be paid for the task, will also be constructed for comparative purposes.
- 2 international workshops with invited speakers (2nd, 3rd year)
- 1 international conference which will be held in the 4th year.

1.5 Output

- Paper presentations at seminars, workshops and conferences organized by the group as well as by other research groups and institutions.
- Publication of joint articles in highly classified journals
- 1 collective volume dedicated to the topic of the project with contributions of its member

**GL5:
Typology of inflectional systems with non-canonical inflection**

PIs	Enrique L. Palancar (SeDYL), Sebastian Fedden (LACITO)
Participants	<p>CRLAO: Anton Antonov (MC), Guillaume Jacques (DR), Thomas Pellard (CR)</p> <p>HTL: Aimée Lahaussais (CR)</p> <p>LLACAN: Tatiana Nikitina (CR), Yvonne Treis (CR)</p> <p>LACITO: Isabelle Brill (DR), Sebastian Fedden (Prof), Alexis Michaud (CR), Lameen Souag (CR)</p> <p>LLF: Olivier Bonami (Prof), Berthold Crysmann (CR)</p> <p>MII: Agnes Korn (DR)</p> <p>SeDyL: Claudine Chamoreau (DR), Katharina Haude (CR), Enrique L. Palancar (DR)</p> <p>Other participants:</p> <p>LLACAN: Mark Van de Velde (CR)</p> <p>SeDyL: Stefano Manfredi (CR), Sophie Vassilaki (Prof)</p>

1 Presentation

1.1 Abstract

The purpose of the project « Typology of inflectional systems with non-canonical inflection » is to create better typological knowledge of inflectional complexity in the world's languages by studying non-canonical inflectional phenomena on new data coming from less-described languages and by adopting new approaches to better known morphological systems. The goal is to combine typological, formal and quantitative approaches for the system of inflectional systems. Empirical subjects of particular interest are non-concatenative flexion, the continuum between synthesis and periphrasis and the emergence of non-canonical flexion, as well as the emergence of uninflectedness.

1.2 Objectives

Theoretical: Create typological knowledge of inflectional complexity by the study of deviations from « canonical inflection » (Corbett, 2007), following approaches to morphology inspired by theories of morphology characterised as « word-and-paradigm » (Anderson, 1992; Aronoff, 1994; Baerman *et al.*, 2015; Matthews, 1991; Hippisley and Stump, 2016; Stump, 1991, 2016; Stump and Finke, 2013; etc.). We will pay particular interest to the study of phenomena related to four topics (but there could be others), which can show a great deal of interconnection, namely: uninflectedness, non-concatenative inflection, the periphrasis-synthesis continuum and the emergence of non-canonical inflection.

Applied: Create a space to allow both inflectional morphologists in Paris and language experts with an interest in typological linguistics to exchange ideas about inflectional issues aimed at generating a better understanding of such phenomena from a typological point of view so that we can gain better descriptions.

1.3 Work plan

- Monthly seminar (open to MA and PhD students) to discuss the main theoretical issues about non-canonical inflection, discuss new data from different languages or revisit known data from new perspectives.
- Monthly project meeting next to the monthly seminar to discuss management and progress of the project.
- Organization and participation in workshops and publications
- Field trips to collect relevant data and creation of databases of inflectional morphology on poorly described languages.

1.4 Outcomes and dissemination

- Creation of an on-line database with new inflectional data on lesser-described languages.
- Individual and collective publications on specific inflectional systems in different languages with a variety of approaches
- Organisation and participation in workshops

1.5 References cited

- Anderson, S. R. 1992. *A-morphous Morphology*. Cambridge University Press.
- Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes*. Cambridge, MA and London: MIT Press.
- Baerman, Matthew, Dunstan Brown and Greville G. Corbett. 2005. *The Syntax-Morphology Interface: A Study of Syncretism*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2007, Canonical typology, suppletion and possible words. *Language* 83, 8-42.
- Hippisley, Andrew and Gregory Stump. 2016. *The Cambridge Handbook of Morphology*. Cambridge: Cambridge University Press.
- Matthews, P. H. 1991. *Morphology*. Cambridge University Press.
- Stump, Gregory T. 2001. *Inflectional morphology*. Cambridge: Cambridge University Press.
- Stump, Gregory T. 2016. *Inflectional paradigms: Content and form at the syntax-morphology interface*. Cambridge: Cambridge University Press.
- Stump, Gregory T. and Raphael Finkel. 2013. *Morphological Typology: Word and Paradigm*. Cambridge: Cambridge University Press.

<p>GL6: Morphosyntax in discourse</p>

PI	Katharina Haude
Participants	<p>Labex members: LaCiTo: Isabelle Bril (DR1, CNRS) SeDyL: Claudine Chamoreau (DR2, CNRS), Katharina Haude (CR, CNRS) Alexandru Mardale (EC, SeDyL), Enrique Palancar (DR2) HTL: Aimée Lahaussais (CR) LLACAN: Amina Mettouchi (DE), Tanya Nikitina (CR), Stéphane Robert (DR1), Martine Vanhove (DR1)</p> <p>External participants: Klaus von Heusinger and colleagues (CRC1252, U Cologne) Stefan Schnell (Melbourne University) Eva Schultze-Berndt (Manchester University)</p>

1 Presentation

1.1 Abstract

Based on discourse corpora of genealogically and typologically diverse languages, we will investigate the grammatical constructions that speakers actually use when describing events – in particular, events involving two or more participants – in a larger context. This includes the analysis of verbal voice alternations, the ways in which event participants are referred to (or not), and the linear order of clausal constituents. Particular attention will be paid to factors such as animacy, definiteness, and different types of topicality of discourse referents, as well as to the employment of verbal categories (voice, aspect) used to direct the hearer’s attention.

1.2 Objectives

The morphological, syntactic, and semantic properties of the different constructions of an individual language are generally well understood by the participating language experts. This programme’s goal is to get a better idea of the *conditions* under which a particular construction or argument expression is used. Our research basis will be data from naturally occurring speech, also because structured data like those resulting from psycholinguistic experiments are usually not available for the languages under study here. We will use annotated corpora of spontaneous oral discourse which the project members already dispose of. They stem from typologically and genealogically diverse languages, which have different grammatical possibilities to allow a variable presentation of an event and its participants. The project is a partial continuation of the Axe3/GD1 project of the previous Labex EFL, “The Typology and Corpus Annotation of Information Structure and Grammatical Relations”, at which most of its members participated, and relies to a large degree on the corpora annotated there.

As a starting point, we will analyze how previously identified event representation patterns in a particular individual language are distributed in actual discourse. This will necessarily involve the identification of discourse-structuring strategies in the respective

languages. Once the criteria that may influence the choice of a construction are identified, the functions of the morphosyntactic strategies used by one language can be compared with those of others (e.g. voice alternations; overt vs. covert vs. zero argument expressions; elaborate event vs. elaborate participant descriptions), while bearing in mind the overall morphosyntactic system and the grammatical possibilities of each language.

International collaborations are envisaged with the Collaborative Research Center (SFB) 1252 “Prominence in Language” at the University of Cologne (dir. Klaus von Heusinger), with Stefan Schnell at Melbourne University and his colleagues who developed the GRAID corpus annotation strategy, and with Eva Schultze-Berndt of Manchester University who works on the corpus-based analysis of reference-tracking strategies in lesser described languages.

The planned outcome of the project consists in single-authored articles, at least one co-authored article by project members, and a collective volume with contributions by project members and collaborators.

GL7:
Text corpora for Mande languages

	2015 – 2019	2020 – 2024
PI	Valentin Vydrin (LLACAN)	Valentin Vydrin (LLACAN)
Participants	G�rard Dumestre (LLACAN), Jean Jacques M�ric (INALCO), Kirill Maslinsky (Upper School of Economics, St. Petersburg, Russia), Ekaterina Aplonova (Upper School of Economics, Moscow, Russia), Andrij Rovenchak (Lviv University, Ukraine), Aby Sangar� (Universit� de Cocody, C�te d'Ivoire), Kalilou T�ra (Universit� de Cocody, C�te d'Ivoire), Ibrahima Sory Cond� 2 (Universit� Kofi Annan, Conakry), Mahamadou Konta (Acad�mie Malienne de langues, Bamako)	Damien Nouvel (ERTIM), Jean Jacques M�ric, Kirill Maslinsky (Upper School of Economics, St. Petersburg, Russia), Ekaterina Aplonova (Upper School of Economics, Moscow, Russia), Andrij Rovenchak (Lviv University, Ukraine), Aby Sangar� (Universit� de Cocody, C�te d'Ivoire), Kalilou T�ra (Universit� de Cocody, C�te d'Ivoire), Ibrahima Sory Cond� 2 (Universit� Kofi Annan, Conakry), Mahamadou Konta (Acad�mie Malienne de langues, Bamako), Elena Perekhvalskaya (Insitute for linguistic studies, St. Petersburg, Russia), Maria Khachatryan (University of Helsinki)

1. Summary of the first phase

1.1. Activities since 2015

The main efforts were directed to the building of new corpora for Mande languages and to the increasing the Bambara Reference Corpus (which exists on line in open access since 2011), cf. <http://cormand.huma-num.fr/mandeica/>

Bambara Reference Corpus

In January 2015, the Bambara Reference Corpus included 2 165 534 words, of these 339 699 in the manually disambiguated subcorpus. In April 2018, it reached 5 220 107 words, of these 1 075 790 words in the disambiguated subcorpus. A program of partial automatic disambiguation collocation-based program has been developed, it allowed to decrease the ambiguity rate in the main (non-disambiguated manually) subcorpus from about 70 % to about 30%. Some studies have been carried out on automatic statistical POS and tone-disambiguation; the results of these studies are being integrated into the program of morphological analysis of Bambara texts.

In April 2018, a Bambara-French parallel corpus has been published on line. At the moment of the first publication, it counted some 100,000 words in each part (French and Bambara); further texts for more than 100,000 words are ready for integration.

Also in April 2018, a syntactically annotated subcorpus (according to the Universal Dependencies model) has been published. Its starting volume is about 1000 phrases; further texts are being treated.

Maninka Reference Corpus

Maninka Reference Corpus was published in April 2016. It is composed of two subcorpora: in N'ko (an original writing system Manding languages created in 1949), some 3,100,000 words, and in Latin orthography, about 400,000 words. Since that time, the main effort was directed to the improvement of the automatic annotation, in particular to the cleaning and completion of the lexical database (an electronic Maninka-French dictionary, Malidaba).

Eastern Dan corpus

Published in April 2018, this corpus includes texts in a language spoken by some 600,000 people in Côte d'Ivoire. At the beginning, it counted some 20,000 words; it is planned to increase it soon to some 300,000 words.

Before the end of 2018, it is planned to publish on line corpora for 3 or 4 other Mande languages (among the nearest candidates are Mwan, Mano, Kakabe, and Kpelle languages).

On the base of the Bambara and Maninka corpora, orthography spellcheckers for these two languages have been developed (for Open Office).

Electronic libraries for Bambara and Maninka have been created and put online, in cooperation with Malian Academy of Languages (Bambara) and Koffi Annan University & N'ko Academy (Maninka).

1.2. Publications

- Vydrin, Valentin & Perekhval'skaia, Elena. Systèmes de compte dans les langues mandé. Alexandre Zheltov (éd.). *Collection Africaine-2015*. St. Petersburg: Musée d'Anthropologie et ethnographie, 2015, pp. 356-377. halshs-01287426
- Vydrin, Valentin & Rovenchak, Andriy & Maslinsky, Kirill. Maninka Reference Corpus: A Presentation. In: TALAf 2016: Traitement automatique des langues africaines (écrit et parole) Atelier JEP-TALN-RECITAL 2016 - Paris le 4 juillet 2016. Actes. 8 p. http://talaf.imag.fr/2016/Actes/VYDRIN_ET_AL%20-%20Maninka%20Reference%20Corpus:%20A%20Presentation.pdf
- Vydrin, Valentin. Tonal inflection in Mande languages: The cases of Bamana and Dan-Gwæetaa. In: Enrique L. Palancar, Jean Léo Léonard (Eds.). *Tone and Inflection: New facts and new perspectives*. [Trends in linguistics studies and monographs. Vol. 296.] De Gruyter — Mouton, 2016, pp. 83-105. halshs-01287458
- Vydrin, Valentin. Le parfait en maninka de Guinée. *Acta Linguistica Petropolitana*, Vol. XII, Part 2. Éd. par T.A. Maysak, V. A. Plungian, Ks. P. Semionova. St. Petersburg: Nauka, 2016, pp. 709-741. (en russe) halshs-01375791
- Vydrin, Valentin. Toward a Proto-Mande reconstruction and an etymological dictionary. *Faits de Langues* 47, 2016, pp. 109-123. halshs-01375776
- Valentin Vydrin, Mamadi Diané. L'interrogation en maninka de Guinée. *Mandenkan* 56, 2016, pp. 95-117. halshs-01481035
- Vydrin, Valentin. Tonal inflection in Mande languages: The cases of Bamana and Dan-Gwæetaa. In: Enrique L. Palancar, Jean Léo Léonard (Eds.). *Tone and Inflection: New facts and new perspectives*. [Trends in linguistics studies and monographs. Vol. 296.] De Gruyter — Mouton, 2016, pp. 83-105.
- Vydrin, Valentin. Le parfait en maninka de Guinée. *Acta Linguistica Petropolitana*, Vol. XII, Part 2. Éd. par T.A. Maysak, V. A. Plungian, Ks. P. Semionova. St. Petersburg: Nauka, 2016, pp. 709-741. (en russe) halshs-01375791

- Vydrin, Valentin. L'utilisation des formes singulières et plurielles des parties de corps appariées en bambara. In: V.A.Vinogradov, A.I.Koval, M.A.Kosogorova, A.B.Shluinsky (éds.). *Issledovaniya po yazykam Afriki 6*. Moscou: Kloutch, 2016, pp. 22-35. (en russe) halshs-01481085
- Vydrin, Valentin. Langues mandé. In: V.F.Vydrin, Yu.V.Mazurova, A.A.Kibrik, E.B.Markus (éds.). *Langues mandé*. [Langues du monde.]. St. Petersburg: Institut de Linguistique — Nestor-Istoria, 2017, pp. 16-45. (en russe) halshs-01481236
- Vydrin, Valentin. La langue bambara. In: V.F.Vydrin, Yu.V.Mazurova, A.A.Kibrik, E.B.Markus (éds.). *Langues mandé*. [Langues du monde.]. St. Petersburg: Institut de Linguistique — Nestor-Istoria, 2017, pp. 46-143. (en russe) halshs-01481176
- Vydrin, Valentin & Morozova, Maria. La langue loko. In: V.F.Vydrin, Yu.V.Mazurova, A.A.Kibrik, E.B.Markus (éds.). *Langues mandé*. [Langues du monde.]. St. Petersburg: Institut de Linguistique — Nestor-Istoria, 2017, pp. 414-456. (en russe) halshs-01481249
- Vydrin, Valentin & Aplonova, Ekaterian. La langue goo. In: V.F.Vydrin, Yu.V.Mazurova, A.A.Kibrik, E.B.Markus (éds.). *Langues mandé*. [Langues du monde.]. St. Petersburg: Institut de Linguistique — Nestor-Istoria, 2017, pp. 457-469. (en russe) halshs-01481257
- Vydrin, Valentin. La langue dan. In: V.F.Vydrin, Yu.V.Mazurova, A.A.Kibrik, E.B.Markus (éds.). *Langues mandé*. [Langues du monde.]. St. Petersburg: Institut de Linguistique — Nestor-Istoria, 2017, pp. 469-583. (en russe) halshs-01481262
- Vydrin, Valentin & Rovenchak, Andrij. Les écritures des langues mandé. In: V.F.Vydrin, Yu.V.Mazurova, A.A.Kibrik, E.B.Markus (éds.). *Langues mandé*. [Langues du monde.]. St. Petersburg: Institut de Linguistique — Nestor-Istoria, 2017, pp. 1094-1112. (en russe) halshs-01481270
- Vydrin, Valentin & Koryakov, Yury. Cartes des langues et dialectes mandé. In: V.F.Vydrin, Yu.V.Mazurova, A.A.Kibrik, E.B.Markus (éds.). *Langues mandé*. [Langues du monde.]. St. Petersburg: Institut de Linguistique — Nestor-Istoria, 2017, pp. 141-151. (en russe) halshs-01481279
- Vydrin, Valentin & Rovenchak, Andrij & Maslinsky, Kirill. Maninka Reference Corpus: A Presentation. In: TALAf 2016: Traitement automatique des langues africaines (écrit et parole) Atelier JEP-TALN-RECITAL 2016 - Paris le 4 juillet 2016. Actes. 8 p. http://talaf.imag.fr/2016/Actes/VYDRIN_ET_AL%20-%20Maninka%20Reference%20Corpus:%20A%20Presentation.pdf, halshs-01481777
- Sumbatova Nina, Vydrin Valentin. Stratégies de focalisation dans les propositions à question partielle dans quelques langues du Caucase du Nord et Manding. *Langues des pays de l'Extrême-Orient, de l'Asie Sud-Est et de l'Afrique Occidentale: Actes du XIIe colloque international (Moscou, le 16-17 novembre 2016)*. Moscou: Institut des pays d'Asie et d'Afrique, Université d'État de Moscou — Faculté Oriental, Université d'État de St. Petersburg, 2016, pp. 238-246 (en russe). halshs-01481128
- V.F.Vydrin, Yu.V.Mazurova, A.A.Kibrik, E.B.Markus (éds.). *Langues mandé*. [Langues du monde.]. St. Petersburg: Institut de Linguistique — Nestor-Istoria, 2017 (1150 p., en russe) halshs-01481290
- Valentin Vydrin, Ibrahima Condé. Bibliothèque Électronique N'ko. Bibliothèque électronique des publications dans la langue maninka en écriture N'ko, à libre accès. halshs-01481781

2. Presentation of the follow-up project

2.1. Abstract

The operation "Text corpora for Mande languages" is reaching a stage of maturity. Main electronic tools necessary for the functioning of the parallel corpora, syntactically annotated corpora and audiocorpora developed in the first phase will facilitate, on the one hand, the growth of the existing

text corpora and improvement of their quality and, on the other hand, to create quickly text corpora of additional Mande languages and, eventually, of languages from other families.

In the new phase, the NLP aspects of the project will be emphasized, in particular: automatic (statistically-based) disambiguation; automatic clusterization of polysemy (on the basis of vectorial analysis); development of OCR tools for the concerned languages; furthermore, first attempts of speech recognition and automatic translation problematics will be made.

2.2. Objectives

During the second phase of the project, three types of works are planned:

1) Increase of the existing corpora. In particular, it is intended to bring the **Bambara Reference Corpus** to 8-10 million words. This corpus will include most of the existing literature and press available in this language. Automatic disambiguation of the entire Corpus will be carried out. The statistically (manually) annotated subcorpus will reach 10,000 phrases; based on this, automatic syntactic annotations with an acceptable error rate will be carried out for the entire Bambara Reference Corpus. The parallel Bambara-French/French-Bambara corpus will reach 1,000,000 words (in each part). An annotated and synchronized audio corpus of Bambara will be built. The **Maninka Reference Corpus** will reach at least 5 million words. A disambiguated Maninka subcorpus will be launched; by 2024, it will have reached 400,000 words. A parallel Maninka-French/French-Maninka corpus will be built and reach 200,000 words. An audiocorpus of Maninka will be created. The **Eastern Dan corpus** will reach 500,000 words, of which some 15% will constitute the disambiguated subcorpus. A parallel Dan-French/French-Dan corpus will be built (target size of 100,000 words).

2) Compiling new corpora. Among the big Mande languages, Jula of Côte d'Ivoire (spoken by more than 10 million people) is the first candidate. A Ivorian Jula corpus will be built according to the same model as the Bambara Reference Corpus; it will be based on a 12,000-word Jula-French dictionary produced by Ivoirean colleagues, external collaborators of the project. A series of small corpora for Mande languages will be built: apart from Mwan, Guinean Kpelle, Kakabe and Mano, there are good chances to build corpora for Guro, Gban, Jogo, Kla-Dan. For these languages, the size of corpora will vary between 50,000 and 300,000 words, depending on the amount of the texts available. For each language, furthermore, a parallel corpus and an audio corpus will be built. A Soninke corpus may also be built in collaboration with colleagues from Hamburg and Lyon.

Our operation is open for cooperation with other colleagues willing to build corpora for their respective languages with our informatics tools.

3) Application of NLP methods to the corpora will be enhanced: automatic disambiguation of texts (parts of speech and glosses); application of distributive analysis to the sense discrimination; advanced studies in syntax will be carried out; tools for OCR and automatic spellchecking for the concerned languages will be developed; first approaches to automatic translation and speech recognition will be undertaken.

2.3. Evolution with respect to the first phase

The corpora of the big languages (Bambara, Maninka) will be brought to maturity, so that they can be used both in linguistic studies (fine-grained grammars; lexicological studies) and in practical applications. It is intended to simplify the process of building annotated corpora for Mande languages (and, eventually, for languages of other families), which will serve to advance corpus linguistics for African languages.

2.4. Projects and contracts submitted in relation to the project

It is planned to submit a project directly connected with our operation for an ERG grant in 2019.

**GL8:
Reconstruction, genealogy, typology and grammatical description
in the world's two biggest phyla: Niger-Congo and Austronesian**

	2015 – 2019	2020 – 2024
PIs	Isabelle Bril, Alexandre François, Valentin Vydrin	Isabelle Bril, Alexandre François, Valentin Vydrin
Participants	Pascal Boyeldieu (DR Emerite, LLACAN), Isabelle Bril (DR, LACITO), Alexandre François (DR, LACITO), Dmitry Idiatov (CR, LLACAN), Konstantin Pozdniakov (PU, LLACAN-INALCO), Nicolas Quint (DR, LLACAN), Antoinette Schapper (CR, LACITO), Guillaume Segerer (CR, LLACAN), Mark Van de Velde (CR, LLACAN), Valentin Vydrin (PU, LLACAN-INALCO), Valentina Alfarano (doc, LACITO), David Piekut (doc, LACITO), Sékou Coulibaly (doc, LACITO), Aurora Montebran (doc, LLACAN)	LLACAN: Pascal Boyeldieu (DR Emerite), LACITO: Isabelle Bril (DR), Alexandre François (DR), LLACAN: Dmitry Idiatov (CR), Konstantin Pozdniakov (PU), Nicolas Quint (DR), Guillaume Segerer (CR), Mark Van de Velde (CR), Valentin Vydrin (PU), LACITO: Antoinette Schapper (CR), PhD candidates: LACITO: Valentina Alfarano, David Piekut, Sékou Coulibaly, LACITO: Aurora Montebran

1. Summary of the first phase

1.1. Activities since 2015

At the first stage of Labex, a great effort was made to create an international network of researchers for the reconstruction of Niger-Congo. With the support of Labex, members of the group organized two international congresses "Towards Proto-Niger-Congo: Comparison and Reconstruction" in 2012 and in 2016, and launched a series "Niger-Congo Comparative Series" in Language Science Press.

During this period, the work of the Niger-Congo team turned around the lexical database Reflex, <http://reflex.cnrs.fr/database/index.php>, which exceeded the limit of 1,100,000 words in 2017. Apart from the quantitative growth, Reflex has evolved qualitatively too. It represents today a powerful and comfortable tool for comparative studies, in open access.

An important result for the Niger-Congo reconstruction is the publication of a book on reconstruction of Niger-Congo numerals (Pozdniakov 2018, Language Science Press). A Niger-Congo etymological base (K. Pozdniakov, G. Segerer) has been built in Reflex.

Considerable results have been reached in the reconstruction of the proto-languages of the Atlantic and Mande families. For the Mande family, the publication of the volume *Mande Languages* (1150 pages, edited by V. Vydrin) was an important milestone.

In the period 2009-2018, the **Austronesianists** in our team have worked both on the synchronic description of the languages of our expertise, and on what can be reconstructed of their history (see the bibliography, under Bril, François, Sagart).

Isabelle BRIL has been conducting fieldwork on one northern New-Caledonian language, Zuanga-Yuanga, and working on a dictionary and grammar of those two divergent dialects of the same

language. I. Brill is also documenting and analysing northern Amis, a Formosan language of the east coast of Taiwan (Brill 2016). Alexandre FRANÇOIS is compiling a dictionary of Mwotlap (Banks, Vanuatu), of which a current draft is already available online; he expects to complete it by 2020. Another well advanced descriptive project is a dictionary of Teanu (Vanikoro, Solomons); he plans to have it ready for publication in 2019.

Major achievements in our group include the following:

- participation in a special issue of 2017, *Studies in Language*, on *Lexical flexibility in Oceanic languages* (ed. by E. van Lier)
- publication of the *Linguistic Atlas of French Polynesia* (Charpentier & François 2015)
- a collective volume *The Languages of Vanuatu: Unity and Diversity* (François et al. 2015)
- organisation of the international conference *Austronesian and Papuan Languages and Linguistics* (APLL9). Paris, June 2017.
- the creation of two open-access academic series in the Austronesian domain: *Studies in the Languages of Island Melanesia* (2012–) and *Studies in the Languages of Wallacea* (2017–), both within Asia-Pacific Linguistics.
- a contract has been signed with OUP for editing the *Guide to the Malayo-Polynesian languages of Southeast Asia* by Antoinette Schapper and Sander Adelaar.
- a volume *Contact and substrate in the languages of Wallacea*, ed. by A. Schapper.

1.2. Publications

- Brill Isabelle 2016, Tense, Mood and Aspect in Nêlêmwa (New Caledonia). Encoding events, processes and states. In Zlatka Guentchéva (ed.). *Aspectuality and Temporality. Descriptive and theoretical issues*. Amsterdam: Benjamins, (SCLS 172): 63-107.
- 2016, Information Structure in Northern Amis (Formosan): a morphosyntactic analysis. *Oceanic linguistics*, 55, n° 2, 451-481.
- 2017, Lexical and functional categories in Nêlêmwa (New Caledonia) and some other Austronesian languages: fluid vs. rigid categoriality. In V. Vapnarsky & Veneziano, E (eds.), *Lexical Polycategoriality: Cross-linguistic, cross-theoretical and language acquisition approaches*. Amsterdam: Benjamins [SLCS 182], 207 -242.
- 2017. Roots and stems: Lexical and functional flexibility in Amis and Nêlêmwa. *Studies in Language*. 41:2, 358-407
- Brill, Isabelle, Achraf Lassoued, Michel de Rougemont. 2017, A statistical model for morphology inspired by the Amis language. *Second Workshop on Language, Ontology, Terminology and Knowledge Structures* (LOTKS 2017, IWCS 2017).
- Charpentier, Jean-Michel & Alexandre François. 2015. *Atlas Linguistique de Polynésie Française – Linguistic Atlas of French Polynesia*. Berlin–Papeete: Mouton de Gruyter, Université de la Polynésie Française. 2559 pp.
- Dom, Sebastian, Koen Bostoen & Guillaume Segerer (2015). Antipassive/associative polysemy in Cilubà (Bantu, L31a), *Studies in Language* 39-2, pp. 354-385.
- François, Alexandre (2015). The ins and outs of *up* and *down*: Disentangling the nine geocentric space systems of Torres and Banks languages. In *The Languages of Vanuatu: Unity and Diversity*, ed. by A. François, S. Lacrampe, M. Franjeh & S. Schnell. *Studies in the Languages of Island Melanesia*, 5. Canberra: Asia Pacific Linguistics Open Access. 137-195.
- François, Alexandre (2016). The history of personal pronouns in northern Vanuatu. In K. Pozdniakov (ed.), *Comparatisme et reconstruction: tendances actuelles. Faits de Langues*. Bern: Peter Lang. 25–60.
- François, Alexandre (2017). Méthode comparative et chaînages linguistiques: Pour un modèle diffusionniste en généalogie des langues. In J.-L. Léonard (ed.), *Diffusion: implantation, affinités, convergence*. Mémoires de la Société de Linguistique de Paris, xxiv: 43-82. Louvain: Peeters.
- François, Alexandre (2017). The economy of word classes in Hiw, Vanuatu: Grammatically flexible, lexically rigid. In Eva van Lier (ed.), *Lexical Flexibility in Oceanic Languages*. Special issue of *Studies in Language*. 41 (2): 294–357..
- François, Alexandre; Sébastien Lacrampe; Michael Franjeh & Stefan Schnell (eds). 2015. *The Languages of Vanuatu: Unity and Diversity*. *Studies in the Languages of Island Melanesia*, 5. Canberra: Asia–Pacific Linguistics Open Access. 271 pp. (isbn 978-1-92-218523-5)

- Kalyan, Siva & Alexandre François (2018). Freeing the Comparative Method from the tree model: A framework for Historical Glottometry. In *Let's talk about trees: Tackling Problems in Representing Phylogenetic Relationships among Languages*, ed. by R. Kikusawa & L. Reid (Senri Ethnological Studies, 98). Osaka: National Museum of Ethnology. 59–89
- Pozdniakov, Konstantin (2015). Diachronie des classes nominales atlantiques: morphophonologie, morphologie, sémantique. In: (Ed.) Denis Creissels et Konstantin Pozdniakov. *Les classes nominales dans les langues atlantiques*. Köln, Rüdiger Köppe Verlag, p.54-102.
- Pozdniakov, Konstantin (2016). Statistics and reconstruction: «quantitative diachrony». In: (Ed.) Konstantin Pozdniakov. *Reconstruction et classification généalogique: tendances actuelles*. Faits de Langues, Paris, 2016, p.223-250.
- Pozdniakov, Konstantin (2016). Tendances actuelles en comparatisme et reconstruction (Présentation générale). In: (Ed.) Konstantin Pozdniakov. *Reconstruction et classification généalogique: tendances actuelles*. Faits de Langues, Paris, p.3-23.
- Pozdniakov, Konstantin (forthcoming). *The Numeral System of Proto-Niger-Congo: A step-by-step reconstruction*. Berlin: Language Science Press.
- Sagart, L. (2015). East Formosan and the PAN palatals. Paper presented at the 13th International Conf. on Austronesian Linguistics. Taipei, 18-22 juillet.
- Sagart, L. (2016). Liangdao Man and Austronesian origins. *Journal of Chinese Linguistics* 44, 1: 239-242.
- Sagart, L. (2016). The Wider Connections of Austronesian: A Response to Blust (2009). *Diachronica* 33, 2: 255-282
- Sagart Laurent, Tze-fu HSU, Yuan-Ching TSAI and Yue-le HSING, (2017), Austronesian and Chinese words for the millets. *Language Dynamics and Change* 7, 187-209. [DOI: 101163/22105832-00702002.
- Segerer, Guillaume (2016). RefLex: 'la reconstruction sans peine', Faits de langues 47, Comparatisme et reconstruction: tendances actuelles. Berne: Peter Lang.
- Segerer, Guillaume (2016). 'Some hypotheses about possible isolates within the Atlantic branch of the Niger-Congo phylum' in ROLLE, Nicholas, Jeremy STEFFMAN & John Sylak GLASSMAN (éds), Proceedings of the 36th Annual Meeting of the Berkeley Linguistic Society - Feb, 6-7, 2010. Berkeley: BLS.
- Segerer, Guillaume (2015). 'Les classes nominales en keerak' in Creissels, Denis & Konstantin Pozdniakov (éds.), *Les classes nominales dans les langues atlantiques*. Cologne: Rüdiger Köppe Verlag, pp. 107-153
- Segerer, Guillaume (2015). 'Les classes nominales en manjaku' in CREISSELS, Denis & Konstantin POZDNIAKOV (éds.), *Les classes nominales dans les langues atlantiques*. Cologne: Rüdiger Köppe Verlag, pp. 187-212.
- Van de Velde, Mark & Dmitry Idiatov. 2017. Morphological classes and gender in Ḑáná-Yungur. In Shigeki Kaji (ed.), *Proceedings of the 8th World Congress of African Linguistics Kyoto 2015*, 53–65. Tokyo: ILCAA.
- Vydrin, Valentin (2016). Tonal inflection in Mande languages: The cases of Bamana and Dan-Gwɛetaa. In: Enrique L. Palancar, Jean Léo Léonard (Eds.). *Tone and Inflection: New facts and new perspectives*. [Trends in linguistics studies and monographs. Vol. 296.] De Gruyter — Mouton, pp. 83-105.
- Vydrin, Valentin (2016). Toward a Proto-Mande reconstruction and an etymological dictionary. *Faits de Langues* 47, pp. 109-123.
- Vydrin, Valentin. Le parfait en maninka de Guinée. *Acta Linguistica Petropolitana*, Vol. XII, Part 2. Éd. par T.A. Maysak, V. A. Plungian, Ks. P. Semionova. St. Petersburg: Nauka, 2016, pp. 709-741. (en russe) halshs-01375791
- Valentin Vydrin, Mamadi Diané. L'interrogation en maninka de Guinée. *Mandenkan* 56, 2016, pp. 95-117. halshs-01481035
- Valentin Vydrin, Mamadi Diané. L'interrogation en maninka de Guinée. *Mandenkan* 56, 2016, pp. 95-117. halshs-01481035
- Vydrin, Valentin. L'utilisation des formes singulières et plurielles des parties de corps appariées en bambara. In: V.A.Vinogradov, A.I.Koval, M.A.Kosogorova, A.B.Shluinsky (éds.). *Issledovaniya po yazykam Afriki* 6. Moscou: Kloutch, 2016, pp. 22-35. (en russe) halshs-01481085
- Vydrin, Valentin. La langue bambara. In: V.F.Vydrin, Yu.V.Mazurova, A.A.Kibrik, E.B.Markus (éds.). *Langues mandé*. [Langues du monde.]. St. Petersburg: Institut de Linguistique — Nestor-Istoria, 2017, pp. 46-143. (en russe) halshs-01481176
- Vydrin, Valentin & Morozova, Maria. La langue loko. In: V.F.Vydrin, Yu.V.Mazurova, A.A.Kibrik, E.B. Markus (éds.). *Langues mandé*. [Langues du monde.]. St. Petersburg: Institut de Linguistique. Nestor-Istoria, 2017, pp. 414-456. (en russe) halshs-01481249
- Vydrin, Valentin & Aponova, Ekaterian. La langue goo. In: V.F.Vydrin, Yu.V.Mazurova, A.A.Kibrik, E.B. Markus (éds.). *Langues mandé*. [Langues du monde.]. St. Petersburg: Institut de Linguistique. Nestor-Istoria, 2017, pp. 457-469. (en russe) halshs-01481257

- Vydrin, Valentin. La langue dan. In: V.F.Vydrin, Yu.V. Mazurova, A.A.Kibrik, E.B.Markus (éds.). *Langues mandé*. [Langues du monde.]. St. Petersburg: Institut de Linguistique — Nestor-Istoria, 2017, pp. 469-583. (en russe) halshs-01481262
- Vydrin, Valentin & Rovenchak, Andrij. Les écritures des langues mandé. In: V.F.Vydrin, Yu.V.Mazurova, A.A.Kibrik, E.B.Markus (éds.). *Langues mandé*. [Langues du monde.]. St. Petersburg: Institut de Linguistique — Nestor-Istoria, 2017, pp. 1094-1112. (en russe) halshs-01481270
- Vydrin, Valentin & Koryakov, Yury. Cartes des langues et dialectes mandé. In: V.F.Vydrin, Yu.V.Mazurova, A.A.Kibrik, E.B.Markus (éds.). *Langues mandé*. [Langues du monde.]. St. Petersburg: Institut de Linguistique — Nestor-Istoria, 2017, pp. 141-151. (en russe) halshs-01481279
- Vydrin, Valentin & Mazurova, Julia & Kibrik, Andrej & Markus, Elena (eds.). (2017) *Jazyki mira: Jazyki mande* (Языки мира: Языки манде) [Languages of the world: Mande languages]. St. Petersburg: Nesto-Historia, 1152 p.
- Vydrin, Valentin & Perekhvalskaia, Elena. Systèmes de compte dans les langues mandé. Alexandre Zheltov (éd.). *Collection Africaine-2015*. St. Petersburg: Musée d'Anthropologie et ethnographie, 2015, pp. 356-377. halshs-01287426

2. Presentation of the follow-up project

2.1. Abstract

The rising prominence of typological approaches to the study of language universals and of evolutionary approaches to explanation in linguistics entails a renewed interest in the classification of the languages of the world as well as in the reconstruction of their proto-languages. This in turn requires careful lexical and syntactic analysis of a great sample of languages based, as much as possible, on first-hand data collected by specialists.

This project aims at a better understanding of the history and internal genealogy of the two biggest language phyla of the world, viz. Niger-Congo and Austronesian. Bringing together some of the world's leading experts, we will strengthen the empirical basis of comparative work in these phyla by describing and documenting individual languages, and we will propose better classifications and genealogies. In every domain of our research project (description, classification and reconstruction), specialists of Austronesian and Niger-Congo will exchange methods and insight from their own field.

2.2. Objectives

In the next two years, I. BRIL is planning a collaboration with Stavros Skopeteas on the prosody of Amis, in particular in relation to informational hierarchy. Collaboration with Elizabeth Zeitoun (Academia Sinica, Taipei) is expected to start in 2020 on a joint new project on Sakizaya, a neighbouring, nearly extinct Formosan language, with some comparative aims. Various articles, a grammar and a dictionary of northern Amis are planned for publication from 2019 onward. Some additional fieldworks are planned for Amis and Sakizaya.

Alexandre FRANÇOIS plans to finish writing a reference grammar of Teanu (Vanikoro, Solomons) in 2022. The reference grammar of Hiw (Torres, Vanuatu), is planned for publication in 2024.

Antoinette SCHAPPER, who just joined the LACITO team, will be joining the RT1 group in 2018. Her work aims to describe and explain the typology of the languages of Wallacea, at the crossroads of the Southeast Asian and Melanesian worlds. She has worked on documenting Papuan and Austronesian languages in the Timor region and on the Austronesian languages of the Aru Islands, south of New Guinea.

Valentin Vydrin is planning to finish writing reference grammars of Guinean Maninka and Eastern Dan languages, and also to work on dictionaries and annotated text corpora for these languages and Bambara. He is also planning to coordinate work on completion and publication of a set of dictionaries for Southern Mande languages (Mwan, Guro, Mano, Kla-Dan) and Southwestern Mande languages (Guinean Kpelle, Looma).

Guillaume Segerer will continue collecting data for underdescribed Atlantic (and also ex-Atlantic) languages, in particular, Jola varieties and Sua.

Pascal Boyeldieu, Dmitry Idiatov, Mark van de Velde and their PhD students will continue collecting data on Adamawa languages, and Nicolas Quint, on Kordofanian languages (Heibanian and Koalib).

2.3. Evolution with respect to the first phase

The new Labex project will put to work the data accumulated in the RefLex database and other electronic sources developed at the first stage. It is planned to carry out reconstructions of phonology and morphology of protolanguages at various levels (proto-languages of groups, of mid-range families, and, finally, of Proto-Niger-Congo). Electronic annotated corpora of texts will serve a major support to this reconstruction.

Among the tasks of that stage is a verification of the hypothesis on belonging of the Mande family to Niger-Congo (recently put in doubt by some scholars). A very prospective direction is a study of the regular homophony by K. Pozdniakov and G. Segerer on the base of RefLex that opens a path to a reliable semantic reconstruction of proto-languages and to a search of regular phonetic correspondencies. Another major direction will be the verification of the validity of the Adamawa family (whose splitting into two or more mid-range families cannot be excluded) and its link to the Gur family.

Compared to other language families, the Austronesian phylum – including the Oceanic subfamily within it – has been relatively well studied by historical linguists, since the end of the 18th century. While the comparative method has proven its efficiency, a question that remains is the degree to which the internal genealogy of the Austronesian family really does fit the structure of a tree. Most historical scholars assume that the Tree model provides the best approach to reconstructing genealogy; this is especially the case among Austronesianists. Building on a new model called Historical glottometry (François 2014, 2017; Kalyan & François 2018), the new RT1 project (2020-2024) should help us question further which model is most suitable to account for the history of Austronesian family.

The new project will also investigate interferences between Papuan and Austronesian languages. Antoinette Schapper will use the significant documentary materials that exist on the West Papuan languages to develop a comparative database of West Papuan languages. This will permit a preliminary reconstruction of the phonology, the morphosyntax and the lexicon of Proto-West Papuan, giving us a picture of a Papuan language that was probably one of the first encountered by speakers of pre-Proto-Oceanic and earlier. In turn, this will enable the first principled, systematic comparison between a reconstructed Papuan language before Austronesian influence and reconstructed members of the Austronesian language family.

**GL9:
Extended Grammars**

	2015 – 2019	2020 - 2024
Strand	7	3
Work package	EM1c	GL9
Project title	Extended grammars	Extended grammars
PI	Émilie AUSSANT	Émilie AUSSANT
Participants	<p>LabEx members: A. Donabedian, F. Jacquesson, T. Nikitina, N. Quint, P. Samvelian, M. Vanhove, C. Donati, S. Auroux, É. Aussant, B. Colombat, J.-M. Fortis, A. Lahaussais, C. Puech</p> <p>External participants: N. Balbir, T. D'Hubert, A. Griffiths, J. Olszowy-Schlanger, D. Patel, V. Raby, J.-L. Chevillard, J.-M. Fournier, J.-P. Guillaume, C. Laplantine, F. Mazière</p>	<p>LabEx members: SeDyL: A. Donabedian (PU) HTL: M. Farina (CR), S. Auroux (émérite, HTL), É. Aussant (CR), A. Lahaussais (CR), C. Puech (PU émérite) LaCiTO: F. Jacquesson (DR émérite) LLACAN: T. Nikitina (CR), N. Quint (DR), M. Vanhove (DR) MII: P. Samvelian (PU, Paris 3), LLF: C. Donati (PU), B. Colombat (PU émérite)</p> <p>External participants: N. Balbir (PU), T. D'Hubert (Ass. Prof.), A. Griffiths (DE), J. Olszowy-Schlanger (DE), D. Patel (Ass. Prof.), V. Raby (PU), J.-L. Chevillard (CR), J.-M. Fournier (PU), J.-P. Guillaume (émérite), C. Laplantine (CR), F. Mazière (émérite)</p>

1. Summary of the first phase

1.1. Activities since 2015

Research seminar HTL-LabEx EFL: 2015-16: in collaboration with Aimée Lahaussais and Bernard Colombat organisation of the research seminar « Grammaires étendues. Aux sources de la typologie » at the lab HTL. Speakers: S. Auroux, B. Colombat, J.-L. Chevillard, J.-P. Guillaume, M. Quadros Leite, C. Laplantine, M. Haspelmath, G. van Driem, F. Simon & C. de Castelnau-L'Estoile, O. Zwartjes, J.-F. Cottier (see: <http://carnetshtl.hypotheses.org/1317>).

Workshops:

- a) In collaboration with Aimée Lahaussais, organisation of two workshops « Grammaires étendues » on November 3-4, 2016 (November 3rd: « Extended Sanskrit Grammar », speakers: É. Aussant, N. Balbir, J.-L. Chevillard, G. Ciotti, V. D'Avella, A. Griffiths, D. Patel (the proceedings of this workshop were published as n° 39.2 of *HEL*); November 4th: « Descriptive Linguistics », speakers: A. Lahaussais, F. Jacquesson, M. Vanhove, T. Nikitina, R. Lambert-Bretière, N. Quint (the proceedings will be published as a special issue of *Faits de langues*).
- b) Mai 2017: in collaboration with Anaïd Donabédian and Agnès Ouzounian, organisation of a workshop dedicated to the exogenous descriptions of Armenian. Speakers: P. Cowe, A. Orengo, O. S. Vardazaryan, C. de Lamberterie. The proceedings will be published in a special issue of the *Revue des études arméniennes*.
- c) In collaboration with Jean-Luc Chevillard organisation of a workshop « Extended Grammars » At the conference ICHoLS XIV in Paris in 2017 (22 proposals received, 17 presentations, 7 articles as a result of these presentations will be selected and published in a special issue of *Beiträge zur Geschichte der Sprachwissenschaft* in 2019).
- d) November 2017, organisation of a workshop « Grammaire arabe étendue » in collaboration with Jean-Patrick Guillaume. Speakers: G. Bohas, J. Kogel, A. Sidarus, R. Ermers, K. Versteegh, J. Landau. The proceedings will be published in a volume of *HEL*.
- e) organisation of a planned workshop "Grammaire grecque étendue", to take place in November 2018. Provisional list of participants: S. Archaimbault, G. Bohas, M. Farina, C. Conduché (to be completed).

Submitted projects:

- a) In line with the call for projects for the ERC Consolidator Grant 2017, E. Aussant (PI) submitted the project TransTools (« How human languages have been described - A global history of transfers »). The project was not selected.
- b) Aussant submitted a project "Typologie et terminologie" for the period 2019-23 of the Fédération de recherche TUL, which was accepted. As the Fédération will not continue, the project will form part of the project "Grammaires étendues" of the new Axe 3 of LABEX EFL.
- c) In line with the call for proposals for the ERC Synergy Grant 2017, Aussant has submitted, as one of three PIs¹, the project MultiLingICA ("Translation, Adaptation, Amalgamation: Multilingual Interactions in the Indic Cultural Area"), which is still in course of evaluation (in Phase 1 the project has been accepted).
- d) Aussant has submitted a doctoral research project within the campaign « contrat doctoral avec mobilité internationale 2018 de l'INSHS », entitled « Extended Sanskrit Grammar and the description of X [Indian language] ». The project was not selected.

Participations at various conferences

1.2. Publications

2017 (parution en 2018). Numéro thématique "La grammaire sanskrite étendue", n° 39.2 de la revue *Histoire Épistémologie Langage* (une introduction détaillée par É. Aussant + 5 articles: N. Balbir, D. Patel, V. D'Avella, G. Ciotti, J.-L. Chevillard).

¹ PI 1 coordinateur: Eva Wilden (Univ. d'Hambourg), PI 2: Harunaga Isaacson (Univ. d'Hambourg), PI 4: Hugo David (EFEO Pondichéry).

2. Presentation of the new project

2.1. Abstract

The history of language description can be seen as a major, more or less successful, enterprise of adapting linguistic tools that were initially shaped to describe or represent a (very) small amount of languages: of the many thousands of languages known in the world, only a few have produced « reflexive » linguistic tools (i.e. tools which describe the language in which they are composed); this is the case for Arabic, Chinese, Greek and Sanskrit. Spearheaded by S. Auroux, a number of works have been carried out on the transfer of the Latin grammatical model onto European vernaculars and « exotic » languages – a phenomenon called « Extended Latin Grammar » by Auroux. It is now appropriate to 1) document widely the other cases of « extended grammars »; 2) test the commensurability of these different transfer phenomena; 3) investigate and reconstruct the long and complex history of linguistic terminology.

2.2. Planned activities

1) Documentation of the various « extended grammars »: This will take the form of workshops with proceedings. Directly linked to the Extended Sanskrit Grammar is the project MultiLingICA, which has been submitted to the ERC (Synergy Grant 2017) and is still under evaluation, as well as a PhD project (proposed for an INSHS PhD grant with international mobility, but not selected).

2) For a better understanding of the history of language description all around the world, it is necessary to **test the commensurability of the different transfer phenomena**, which has never been done. A re-submission to the ERC of the project TransTools, which precisely aimed at comparing the different transfer phenomena, is planned.

3) The reflection on the transfer of descriptive models and terminologies meets some contemporaneous descriptivists' and typologists' concerns: if one wants to grasp all the issues addressed by a number of questions today in descriptive linguistics and in linguistic typology, and if one wants to contribute meaningfully to these debates, **one has to investigate carefully the linguistic terminology in order to reconstruct its long and complex history**; this inevitably involves a history of terminological transfers. Activities carried out in this direction are those which had been planned within the TUL Federation.

2.3. Changes with respect to the previous period

Two changes: 1) the comparison of the different transfer phenomena, 2) the articulation with some issues currently addressed in descriptive linguistics and linguistic typology. The goals set at the beginning of the program (in 2015) were primarily linked to the documentation of the various extended grammars.

2.4. Submitted and/or obtained projects/contracts linked to the program

Documentation of the various « extended grammars »:

a) Project MultiLingICA (submitted to the ERC, Synergy Grant 2017, still under evaluation).

b) PhD research project "Extended Sanskrit Grammar and the description of X [Indian language]" (submitted to the INSHS in 2018, not selected).

Testing the commensurability of the different transfer phenomena: project TransTools (submitted for an ERC Consolidator Grant 2017, not selected).

Investigation into linguistic terminology: project « Typology and terminology » (submitted—and selected—for the 2019-23 mandate of the TUL Federation).

***Work package 2:
Variation and Contact (VC)***

**VC1:
Languages, Dialects and Isoglosses of West Asia**

	2015 – 2019	2020 - 2024
Strand	3	3
Work package	LC3	VC1
Project title	Languages, Dialects and Isoglosses of the Iran-Caucasus-Anatolia-Near-East Area	Languages, Dialects and Isoglosses of the West Asia
PIs	Anaïd Donabédian, Pollet Samvelian	Anaïd Donabédian, Pollet Samvelian
Participants	<p><u>LABEX members</u> Victoria Khurshudyan (SeDyL) Jonny Cheung (MII) Agnès Korn (MII) Anton Antonov (CRLAO) Fida Bizri (SeDyL) Emmanuel Giraudet (MII) Matteo De Chiara (MII) Pollet Samvelian (MII) Anaïd Donabédian (SeDyL)</p> <p><u>Post-doc with Labex contract:</u> Thomas Jügel</p> <p><u>Doctorants:</u> <u>PhD defended during the project:</u> Hélène Gerardin (membre individuel) Ophélie Gandon (MII) Pegah Faghiri (MII)</p> <p><u>PhD defense expected:</u> Katherine Hodgson (SeDyL) Pinar Karakilcik (SeDyL) Masud Mohammadirad</p> <p>Collaborations extérieures: Hosep Dolatian</p>	<p><u>LABEX members</u> Victoria Khurshudyan (SeDyL) Agnès Korn (MII) Anton Antonov (CRLAO) Fida Bizri (SeDyL) Emmanuel Giraudet (MII) Matteo De Chiara (MII) Pollet Samvelian (MII) Anaïd Donabédian (SeDyL)</p> <p><u>End of Phd thesis, Post-docs</u> Pinar Karakilcik (SeDyL) Hélène Gerardin (membre individuel) Katherine Hodgson (SeDyL) Masud Mohammadirad (MII)</p> <p><u>Expected PhD candidates</u> Arayik Hairiyian (SeDyL)</p> <p><u>External participants:</u> Nina Sumbatova (ex- chaire EFL) Yuri Koriakov (Académie des Sciences de Russie) Georgi Moroz (HSE, Moscou) Ioanna Sitaridou (Cambridge) Christiane Bulut (Nicosie) Pegah Faghiri (U. de Cologne)</p>

1. Summary of the first phase

1.1. Activities since 2015

Three topics have been investigated by the participants of the operation since 2015:

- 1) The diversity of relative constructions in the languages of the area has been one of our main topics of investigation since 2015:
 - We followed up on the series of international workshops initiated in 2013 by organizing 4 new workshops on July 3, 2015, December, 10-11, 2015, December 12, 2016, October 12, 2016.
 - One of the major outcomes of the operation was the PhD thesis of Ophélie Gandon (Doctoral funding, Université Sorbonne Nouvelle), entitled *La relativisation dans une perspective aréale: l'aire Caucase – Anatolie de l'Est – Iran de l'Ouest*, defended on December 12, 2016. The thesis was a thorough overview of relativization strategies in the languages of the area and also proposed a questionnaire including more than 90 parameters and features in order to describe various aspects of relative constructions.
 - Katherine Hodgson (International doctoral funding Inspire – Marie Curie) has almost completed her PhD on Subordination in Armenian dialects: an experimental study from an areal-typological perspective. This is an extensive investigation of relative constructions in spoken and colloquial Armenian, combining field work and experimental methods. The defense is planned for February 2019.
 - A volume on relative constructions in the languages of West Asia is in preparation. It will include a selection of the talks given during our international workshops and also a few chapters by renowned specialists of the languages of the area.
 - About 19 presentations at workshops and conferences.
- 2) Several participants of the operation investigated clitic pronouns:
 - Thomas Jügel was hired as a postdoctoral researcher. He investigated the evolution of clitics from pronouns into agreement markers in West Iranian languages. He co-authored 2 papers with Pollet Samvelian: an article in *Bulletin de la Société de Linguistique de Paris* in 2016 and a book chapter in a volume edited by Richard Larson. There were also several presentations at conferences.
 - Masud Mohammadirad obtained an international doctoral funding, Cofund - Marie Curie, in 2016 to work on clitics in Western Iranian languages.
- 3) Cartographic modeling of language contact. A training session was organized by the operation on January 19, 2018, in order to initiate our participants to the cartographic package of R. We also organized a workshop on January 22, 2018. We are currently collaborating with researchers from Russia and Armenia for the realization of an atlas of Armenian dialects, with new data never investigated in previous studies. The ongoing collaboration with Ioanna Sitaridou (University of Cambridge) aims at the elaboration of georeferenced map of some other languages of the area (e.g. Greek, Aramaic). Data for these languages are contributed by our partners. The purpose is to achieve interoperability of various maps in order to investigate language contact in the area in a more systematic way. The chapter on language contact in Anatolia will be co-authored by A. Donabedian and I. Sitaridou.

1.2. Publications

Journal papers

- Agnes Korn. 2016. Arménien karmir, sogdien krm'yr et hébreu karmīl « rouge ». *Bulletin of the School of Oriental and African Studies* 79:1, pp. 1-22.
- Jügel, Thomas et Pollet Samvelian. 2016. Les « pronoms » clitiques dans les langues iraniennes occidentales, *Bulletin de la Société de Linguistique de Paris*, t. 391-432. [41 pages]

Book chapters and proceedings

- Gandon, O. [in press] "Linguistic divergence with respect to relativization strategies in East Anatolia". *Turkish, Turkic and the languages of Turkey*, UMass University, Amherst, November 21st-22nd.
- Gandon, O. [in press] "Convergence areas in the Caucasus-Western Iran area concerning the relativization strategies". RJC 2015 - Actes des 18èmes Rencontres Jeunes Chercheurs de l'ED 268: *Contact de langues: situations, représentations, réalisations*.
- Gandon, O. 2015. "Stratégies de relativisation atypiques dans l'aire Caucase-Iran-Anatolie". COLDOC 2014 - Actes des 11èmes Rencontres Jeunes Chercheurs: *Diversité des langues, Les universaux linguistiques à l'épreuve des faits de langue*.
- Thomas Jügel: in press. Ditransitive Constructions in three Balochi Dialects from a typological perspective, 19pp. Avec Maryam Nourzaei. In: Proceedings of the workshop Ditransitive constructions in a cross-linguistic perspective, held 2014 in Pavia, ed. by Agnes Korn et al.
- Thomas Jügel: in press. Deixis and Near Future Expressions in Kurdish, 22 p. Avec Cemile Çelebi et Diako Nahid. In: Proceedings of the symposium *The Prospective as a Grammatical Category: Evidence from Turkic, Iranian and beyond*, ed. by Agnes Korn et al. (final proofs 2016)
- Thomas Jügel: in press. The development of the object marker in Middle Persian, 15 p. In: *Proceedings of the Eighth European Conference of Iranian Studies*, St. Petersburg: State Hermitage Publishers, to appear in 2018.
- Thomas Jügel: in press. Enclitic Pronouns in Middle Persian and the Placeholder Construction, 14 pp. In *Professor Mansour Shaki Memorial Volume*, ed. by M. Jaafari-Dehaghi. Tehran: Center for the Great Islamic Encyclopaedia.
- Agnes Korn. 2016. The languages, their histories and genetic classification: Iranian. In: Hans Henrich Hock & Elena Bashir (eds.): *The Languages and Linguistics of South Asia: A Comprehensive Guide [The World of Linguistics 7]*. Berlin: de Gruyter, pp. 51-66
- Samvelian, Pollet and Pegah Faghiri, 2016, Rethinking compositionality in Persian Complex Predicates, Proceedings of *Berkeley Linguistic Society 39th Annual Meeting, February 16-17 2013*, Berkeley: Berkeley Linguistics Society, pp. 212-227.
- Samvelian, Pollet. 2018. Specific Features of Persian Syntax: The *Ezâfe* Construction, Differential Object Marking and Complex Predictaes, in *Oxford Handbook of Persian Linguistics*, A. Sedighi and P. Shabani-Jadidi (Eds.), Oxford: Oxford University Press.
- Donabédian Anaïd. 2018. Western Armenian at the crossroads: A sociolinguistic and typological sketch. In C. Bulut (éd.), *Linguistic Minorities in Turkey and turkic speaking minorities of the peripheries.: Turcologica* 111, Wiesbaden: Harassowitz Verlag, 2018, 89-148.

2. Presentation of the new project

2.1. Abstract

As in the first phase, the main purpose of the operation is the modeling of the notion “linguistic area” and the interaction between dialectology, contact linguistics and linguistic typology. The linguistic configuration in West Asia is extremely promising for investigating these topics. The operation LC3 brings together several specialists of the languages of the area (such as Armenian, Iranian languages, Caucasian languages) and aims to actively contribute to a better understanding of contact phenomena.

2.2. Objectives

In the new phase of the Labex EFL (2020-2024), we will investigate 2 new aspects of the languages of the West-Asian area:

1) OV word order/variation/convergence: A set of typological properties have been shown to converge with SOV basic order (e.g, agglutination, preverbal wh- interrogatives, preverbal focus, pro-drop, lack of morphological agreement in NPs, non-finite relativization, and, maybe less reported, bare objects). In the Anatolia-Caucasus-Iran continuum, several languages share these features, and in addition show a similar pattern of information packaging, characterized by a default focus preverbal position bearing sentence nuclear stress, and a focus marked by prominent stress on any constituent of the clause regardless of its linear position. Furthermore, the linear order in those languages is flexible since it depends on discourse configuration and not on syntactic constraints, as accounted for as discourse configurationality.

Our research will focus on word order variations in the OV languages of the area in order to identify the parameters intervening in word order preferences, in particular information structure, and also the influence of linguistic contact in these variations.

Although many languages of the area are described as SOV (Armenian, Persian, Kurdish, Turkish, Georgian), they display significant differences with respect to the degree of flexibility in word order. In Eastern Armenian, for instance, VO order seems to be quite frequent, without “backgrounding” postverbal constituents, while in Persian and in Turkish, postposed constituents are generally claimed to be backgrounded. Kurdish, in the other hand, seems to display strict OV order, and no postposing is possible.

2) Differential object marking (DOM) is another feature shared by many languages of the area. While it has been extensively investigated in Persian, in other languages, such as Armenian, it has received less attention. Apart from studying the parameters intervening in DOM, we will especially focus on the relation between DOM and word order.

2.3. Outcome

- An interoperable map of the languages and dialects of Anatolia (follow up of the work initiated in the first phase of the project)
- Theoretical modeling of word order variations, information structure and differential object marking

2.4. Evolution with respect to the first phase

The two main topics of investigation are new and have not been investigated during the previous phase.

**VC2:
Central Gallo-Romance: linguistics and ecology of a transitional
zone**

	2015 – 2019	2020 - 2024
Strand	3	3
Work package	LC4	VC2
Project title	The linguistic Crescent : a contact area between Occitan and Oilic varieties	Central Gallo-Romance: linguistics and ecology of a transitional zone
PI	Nicolas Quint (LLACAN)	Nicolas Quint (LLACAN)
Participants	<p>LABEX members:</p> <p>Nicolas Quint (LLACAN), Heather Burnett (LLF), Judit Gervain (LPP), Maximilien Guérin (LLACAN), Barbara Hemforth (LLF)</p> <p>Jean-Léo Léonard (Université Paris 4), Sylvain Loiseau (LACITO, UMR7107), Stefano Manfredi (SeDyL), Marijana Petrovic (LACITO)</p> <p>External participants:</p> <p>Gilles Adda (LIMSI), Philippe Boula de Mareuil (LIMSI), Guylaine Brun-Trigaud (BCL), Jean-Christophe Dourdet (FORELL), Marc-Olivier Hinzelin (Université d’Hambourg), Philippe Maurer (Université de Zurich), Stephan Mietzke (Université de Rennes 1), Jérôme Picard (UPS2259)</p> <p>Rafèu Sichel Bazin – postdoc (CLLE, UMR5263)</p>	<p>LABEX members:</p> <p>LLACAN: Nicolas Quint (DR), Maximilien Guérin (postdoc)</p> <p>LLF: Heather Burnett (CR), Barbara Hemforth (DR)</p> <p>LPP: Judit Gervain (DR)</p> <p>Jean-Léo Léonard (PU Paris 4)</p> <p>LACITO: Sylvain Loiseau (MdC, Paris 13)</p> <p>SeDyL: Stefano Manfredi (CR)</p> <p>Marijana Petrovic (LACITO)</p> <p>External participants:</p> <p>Gilles Adda (IR, LIMSI), Philippe Boula de Mareuil (DR, LIMSI), Guylaine Brun-Trigaud (IA, BCL), Jean-Christophe Dourdet (MdC, FORELL), Louise Esher (CR, CLLE), Marc-Olivier Hinzelin (Université d’Hambourg), Philippe Maurer (Université de Zurich), Stephan Mietzke (MdC, Université de Rennes 1), Jérôme Picard (IR, CLT), Rafèu Sichel Bazin (postdoc, CLLE)</p>

1. Summary of the first phase

The first funding the LC4 project received from EFL came in 2016, but our research group managed to secure an external funding from DGLFLF (see below), which allowed us to begin fieldwork in 2015.

Since the beginning of the project, LC4 members have been active on the field:

- The descriptivists of the team have been collecting data from the Crescent varieties spoken in the following communities: Anzême (23), Azérables (23), Bonnat (23), Bussièrès-Poitevine (87), La Celle-Dunoise (23), Dompierre-les-Églises (87), Fresselines (23), Fursac (23), Lourdoueix-Saint-Michel (36), Saint-Agnant Versillat (La Prugne, 23). For most of these locations, monographs (grammatical analysis + lexicons) are now in progress.
- At least two typological surveys have been initiated: (i) one on phonological data, which were used for specific publications and contributed also to enrich the online data base on regional languages (<https://atlas.limsi.fr/>; the Crescent area appears in grey on the map); the phonological data were collected in places not previously contemplated by descriptivists (eg. Eguzon-Chantôme (36), Naves (63), Pleuville (86)); (ii) a few preliminary psycholinguistic tests, which will be elaborated upon in the proposed new phase of the project.

Each fieldwork resulted in the recording and archiving of several hours of audio/video material which are systematically stored on a hard disk and will be made available on a specific section of HumaNum. This linguistic material will be used as a common resource both for researchers of the project (in particular for comparative analyses) and for the general public.

Regular meetings were held both in Paris and in the Crescent region. The first international scientific conference dealing exclusively with Crescent varieties took place on March 25-27 2017 in Le Dorat.

The high commitment and scientific excellence of the team members allowed us to secure two new fundings, from DGLFLF in 2016 and ANR in 2017 (see details below).

At present, various monographs are being prepared and at least two should be released before the end of 2019. Several papers will also be published in 2018 and in the following years.

Overall, the asset of the team resides in its multidisciplinary, as it regroups both typologists/descriptivists and specialists of various branches of Language Sciences (notably Morphology, Natural Language processing, Phonetics and Phonology, Psycholinguistics, Semantics).

1.1. Publications

Chapters in edited volumes

Boula de Mareuil, Philippe, Rafèu Sichel-Bazin, Nicolas Quint, Gilles Adda (2017) « Norme et variation à l'âge des corpus informatisés pour les langues régionales de France », in C. Feuillard (éd.), *Usage, norme et codification*, EME, Bruxelles.

Brun-Trigaud, Guylaine (2017) « La dialectométrie dans les parlers de la Creuse. » *Mémoires de la Société des sciences naturelles, archéologiques et historiques de la Creuse*, t. 62, 2016/2017, 259-272.

- Brun-Trigaud, Guylaine (2017) « La Marche: frontière et carrefour dans l'espace linguistique du domaine gallo-roman. Marches, limites, frontières en France centrale. » Actes du XXe congrès des Sociétés savantes du centre de la France (Guéret 2017), *Etudes creusoises*, XXIV, 105-110.
- Léonard, Jean Léo (2015/18). « La taxinomisation comme forme de grammaticalisation à échelle systémique: l'émergence de classes flexionnelles dans le diasystème d'oïl », in Actes du colloque international Normes et grammaticalisation: le cas des langues romanes / Normes et transgressions dans les littératures romanes (Sofia, Bulgarie), 20 et 21 novembre 2015: Velinova, Malinka (éd.) 2018. *Normes et grammaticalisation: le cas des langues romanes*, Sofia, CU Romanistica, pp. 101-124.
- Léonard, Jean Léo (2018). "Anchoring Evidential, Epistemic and Beyond in Discourse: alào, vantae and vae in Noirmoutier island (Poitevin-Saintongeais)", in Zlatka Guentcheva (éd.) 2018. *Epistemic Modalities and Evidentiality in Cross-Linguistic Perspective*, Berlin, Mouton de Gruyter, pp. 131-152.

Conferences and seminars

- Brun-Trigaud, Guylaine et Flore Picard (2017) Regards croisés de mesures dialectales sur une tranche de Croissant. *1ères Rencontres sur les parlers du Croissant*. Projet LC4 (Labex EFL - CNRS), Le Dorat, 24-25 mars 2017.
- Guérin, Maximilien (2017) Le marchois: Une langue entre oc et oïl. *Séminaire de l'équipe Linguistique du FoReLL (EA 3816)*. Université de Poitiers, 17 février.
- Guérin, Maximilien (2017) Description du parler marchois de Dompierre-les-Églises (Haute-Vienne): phonologie, conjugaison et lexique. *1ères Rencontres sur les parlers du Croissant*. Projet LC4 (Labex EFL - CNRS), Le Dorat, 24-25 mars.
- Guérin, Maximilien (2017) Les paradigmes de conjugaison en marchois (Croissant limousin): entre oc et oïl. *XIle Congrès de l'Association internationale d'études occitanes*. Albi, 10-15 juillet.
- Guérin, Maximilien (2018) The Crescent Languages: marginalised minority languages. *Promoting or demoting: the transmission of minority languages from past to present*. Université de Poitiers, 6-7 avril 2018.
- Léonard, Jean Léo (2015) Typologie sociographémique de corpus d'Oïl: normand, gallo et bourguignon-morvandiau, *Journée du mercredi 9 décembre 2015 organisée par Fabrice Jecic CNRS-Paris 1-LAMOP et Jean-Léo Léonard (Paris-Sorbonne)*, Campus CNRS de Villejuif.
- Léonard, Jean Léo (2017) « Des 'miniatures' narratives au système: des textes en langues/dialectes d'oïl aux structures (phonologie et morphologie) », *conférence au séminaire de master de Sciences du Langage de l'Université de Dijon*, 28 mars 2017.
- Léonard, Jean Léo (2015). « Typologie et complexité de la gamme des affriquées et des fricatives palatales ou dépalatalisées en poitevin-saintongeais et en bourguignon-morvandiau », *conférence*
- Petrovic, Marijana et Jean Léo Léonard (2015) « Interaction pronominale dans des micro-aires transitionnelles poitevines /saintongeaises & Croissant », *SyMiLa 2015: La microvariation syntaxique dans les langues romanes de France*. Journées d'études internationales des 11-12 juin, 2015, CLLE-ERSS (Université de Toulouse 2-Le Mirail).
- Quint, Nicolas & Maximilien Guérin (2017) Les parlers du Croissant: une aire de contact entre oc et oïl. *XIle Congrès de l'Association internationale d'études occitanes*. Albi, 10-15 juillet.
- Organisation of conferences and summer schools
- Maximilien Guérin & Nicolas Quint. *1ères Rencontres sur les parlers du Croissant*, Le Dorat (87), 24-25 March 2017. <http://llacan.vjf.cnrs.fr/labex/lc4/colloques.html>

2. Presentation of the follow-up project

2.1. Abstract

Our research group intends to develop a multidisciplinary approach to the study of the Gallo-Romance varieties of the linguistic Crescent. These varieties represent fascinating case studies for contact linguistics, as they display simultaneously Oilic and Occitan features. Special emphasis will be put on the following three topics: (i) the linguistic Crescent and its environment (study of the Oilic or Occitan varieties spoken around the Crescent area and of the regional French varieties spoken alongside the Crescent varieties), (ii) checking the non-linguistic (anthropological, historical etc.) correlates of the isoglosses observed in the field, (iii) in cooperation with members of Strand 4, development of psycholinguistic studies devoted to the Crescent. All in all, this new project aims at proposing a global, multidisciplinary and ecological approach to the linguistic Crescent.

2.2. Planned activities and outcomes

Central Gallo-Romance: linguistics and ecology of a transitional zone will continue and enrich the work initiated within the scope of LC4. The following deliverables are planned:

- 10 monographs (grammatical analyses + basic lexicon) devoted to diverse Crescent varieties.
- 20 papers dealing with various features displayed by these varieties in several domains (semantics, morphology, phonetics/phonology, natural language processing, psycholinguistics etc.).
- Organisation of several international conferences within and outside the Crescent area (the next one is scheduled for February 2019 in Montluçon).
- Participation at several international conferences (general and Romance linguistics).
- Development and on-line publication of questionnaires specifically devoted to the Crescent varieties.
- Archiving on Huma-Num of over 50h of audio and video recordings of data from least 30 different Crescent varieties.
- Development of a searchable online data base combining linguistic data and cartographic functions.
- Collective volumes dealing with the Crescent varieties.

2.3. Evolution with respect to the first phase

LC4 'Les parlers du Croissant' was begun and developed during the second half of the first phase of Labex EFL (see above), and the project proposed here will be a direct continuation of this enterprise. The varieties of the linguistic Crescent will be the main focus of our multidisciplinary team. At the same time, we intend to enlarge our perspective by taking into account new, more interdisciplinary approaches regarding the Crescent.

The linguistic Crescent and its environment

The Gallo-Romance linguistic varieties of the Crescent, which display typical Oilic and Occitan features simultaneously and are traditionally used in the central part of Gallo-Romania, represent essentially a transitional area between two diasystems. Today, only a minority of the population of the linguistic Crescent still speak a variety of the Crescent as their mother tongue.

In order to understand the actual practices of these speakers, we must take into account the overall ecology of the transitional zone into which they are inserted. This means that:

- Occitan (Auvergnat, Limousin) and Oïlic (Poitevin-Saintongeais, Berrichon) varieties spoken in the vicinity of the Crescent area will be taken systematically into consideration in order to understand what exactly happens in this traditional zone.

- Local varieties of standard French are spoken on a daily basis by all speakers of Crescent varieties as well as by most of the population who do not speak any Crescent variety. It is therefore crucial to study the main characteristics of these local varieties of standard French in more detail in order to ascertain if any of their features can be traced back to the Crescent varieties; for instance, compare local Crescent standard French 'ça pleut' (*ça* = neutral expletive pronoun distinct from 3rd masculine singular) and Saint-Priest-la-Feuille Crescent variety 'qua pleu' (*qua* = neutral expletive pronoun distinct from 3rd masculine singular) vs. Paris standard French 'il pleut' (expletive pronoun identical to 3rd masculine singular).

The linguistic Crescent: linguistic and non-linguistic aspects of the transitional zone

Apart from linguistic features, other traits seem to characterize the Crescent area as a transitional zone. For instance, the limit between Occitan and Oïlic varieties closely follows the limit between the places where people used to plough with oxen (Southwards) and the places where people used to plough with horses (Northwards). Within the framescope of an ecological approach of the linguistic Crescent, the non-linguistic correlates of the Crescent will be systematically identified and analysed (possibly in cooperation with non-linguists, e.g. historians or anthropologists, thereby enhancing the interdisciplinarity of our project).

Psycholinguistic characterization of bilingualism in the Crescent

Today, psycholinguistics clearly represents a new frontier for language sciences and linguistic research. The Crescent area is of particular interest for psycholinguistic research. As a matter of fact, virtually no study was ever produced about the effects that bilingualism has on the brain of people practicing simultaneously French and a regional language (in particular a Romance variety). Good experimental data about regional languages are still missing in the domain of psycholinguistics (most cases of bilingualism that are studied within this framescope regard people who practice simultaneously two official languages, e.g. Spanish and French, English and German etc.). All of today's speakers of a Crescent variety also speak French on a daily basis and generally with (near)native proficiency. The psycholinguistic study of their productions and abilities is liable to provide us with a theoretical model accounting for the bilingual profile national language + regional language. This is the reason why typologists and psycholinguists involved in LC4 have agreed to further develop their synergy and propose a PhD project 'bidialectalism' (supervisors: Judit Gervain & Nicolas Quint) about the psycholinguistic aspects of bilingualism between a local language and French in the Crescent area.

VC3: Syntactic variation in creole languages

	2015 – 2019	2020 - 2024
Strand	3	3
Work package	GD2	VC3
Project title	The syntax of complex sentences in creole languages	Syntactic variation in creole languages
PIs	Stefano Manfredi (SeDyL), Nicolas Quint (LLACAN)	Stefano Manfredi (SeDyL), Nicolas Quint (LLACAN)
Participants	<p>LABEX members: Alain Kihm (DR émérite, LLF), Stefano Manfredi (CR, SeDyL), Nicolas Quint (DR, LLACAN), Bernard Caron (DR, LLACAN), Guillaume Fon-Sing (MdC, LLF), Isabelle Léglise (DR, SeDyL), Bettina Migge (PU, SeDyL), Jean François Nunez (postdoc, SeDyL), Karina Moreira (PhD candidate, LLACAN), Eliane Vieira (PhD candidate, LLACAN).</p> <p>External members: Noël Bernard Biagui – MdC (CLAD, Dakar), Susanne Michaelis – PU (Max Planck Institute, Jena), Sibylle Kriegel – PU (LPL, UMR7389), Paula Prescod – MdC (Université de Picardie)</p>	<p>Labex members: LLACAN: Nicolas Quint (DR), Bernard Caron (DR), SeDyL: Isabelle Léglise (DR), Stefano Manfredi (CR), Bettina Migge (PU), Jean François Nunez (postdoc) LLF : Guillaume Fon-Sing (MdC), PhD candidates : Karina Moreira (LLACAN), Eliane Vieira (LLACAN). External participants: Noël Bernard Biagui (CLAD, Dakar), Susanne Michaelis (Max Planck Institute, Jena), Minella Duzerol (DDL, Lyon), Sibylle Kriegel (LPL), Paula Prescod (Université de Picardie)</p>

1. Summary of the first phase

The research project GD2 “The syntax of complex sentences in creole languages” only began officially in 2016 (first year when funding were made available), although its members already started to collaborate informally during the second half of 2015. The main aim of the project is to develop a typology of subordinate/coordinate clauses in creole languages on the basis of a searchable database.

Since the beginning of the project,

- we have organised 2 internal meetings per year that allowed us to discuss the main problems of the analysis and comparison of complex sentences in creole languages. During the meetings we had the possibility to collaborate with invited experts in the field (Sonia Cristofaro, Susanne Michaelis – LABEX international chair 2017). Until now,

we have focussed our attention on subordination; we plan to deal with coordination and serial verbs during the last two years (2018 and 2019) of our programme.

- Following the discussions held during the meetings, we have developed a typological database devoted to the syntax of complex sentences in Creole languages and related languages (lexifiers and substrates).
- In 2018, we set up a workshop on Creole syntax for the Syntax of the World's Languages conference held in Paris, September, 3rd-5th (<https://swl8.sciencesconf.org/>)
- Furthermore, we (i) presented our topic of studies to many scientific conferences, (ii) did linguistic fieldwork in order to collect data for our database and check our hypotheses and (iii) published a considerable bulk of material dealing with Creole languages and in particular their syntactic characteristics; (iv) note also that several PhD students are integrated to GD2, three of which at least have syntax as the main or one of the dominant focus of their research.

1.1. Publications

International journals

Fon Sing, Guillaume, forth. *Creoles are not typologically distinct from non-Creoles*. In *Language Ecology*. John Benjamins.

Manfredi, Stefano. 2017, Demonstratives and the emergence of a definite article in Juba Arabic and Ki-Nubi. *Journal of Pidgin and Creole Languages* 32 (2): 205-232.

Migge, Bettina & Léglise, Isabelle. 2015. Assessing the sociolinguistic situation of the Maroon Creoles. *Journal of Pidgin and Creole Languages* 30(1): 63-115.

Migge, Bettina. 2017, Putting Matawai on the Surinamese Linguistic Map. *Journal of Pidgin and Creole languages* 32(2): 233-262

Quint, Nicolas, 2015 [with Denis Creissels & al.] L'impersonnalité dans les langues de la région sénégalaise, in: *Africana Linguistica*, N°XXI, pp. 29-86.

Quint, Nicolas, 2015, Capeverdean words derived from Portuguese non-infinitive verbal forms: a descriptive and comparative study, *Papia* (Revista Brasileira de Estudos Crioulos e Similares), N°25/2, 189-215.

Quint, Nicolas, Eliane Araujo Vieira Semedo & Márluce Coan, forth. Orações completivas em cabo-verdiano: estudo baseado em corpus. In: *Papia* (Revista Brasileira de Estudos do Contato Linguístico), e-ISSN: 2316-2767.

Tavares, Karina, forth. Le créole afro-portugais de l'île de Fogo: présentation historique et linguistique. *Annales de la fondation Martine Aublet*.

Books

Biagui, Noël Bernard. 2018. *Grammaire du créole afro-portugais de Ziguinchor*. Karthala.

Manfredi, Stefano. 2017. *Arabi Juba*. Peeters.

Quint, Nicolas, forth. *Atlas de la langue capverdienne* (Îles Sous-le-Vent), Lambert-Lucas.

Edited volumes

Manfredi, Stefano and Mauro Tosco (eds), 2018. *Arabic in Contact*. John Benjamins.

Chapters in edited volumes

Manfredi, Stefano, forth. The Arab World. In U. Ansaldo, M. Meyerhoff and G. Fon Sing (eds.), *The Routledge Handbook of Pidgin and Creole Languages*. Routledge.

Manfredi, Stefano. 2017. The construction of linguistic borders and the emergence of national identity in South Sudan: some insights into Juba Arabic (árabi júba). In Reem Bassiouney (ed.), *Dialect and Identity Performance*. Edinburgh University Press. 138-163.

- Migge, Bettina. 2015. The Role of Discursive Information in Analyzing Multilingual Practices. In Yakpo, Kofi & Stell, Gerald (eds.), *Code-Switching Between Structural and Sociolinguistic Perspectives*, 185-206. DeGruyter.
- Migge, Bettina. 2017a. From grammar to meaning: Towards a framework for studying synchronic language contact. In Kofi Yakpo & Pieter Muysken (eds.), *Boundaries and Bridges: Language Contact in Multilingual Ecologies*. De Gruyter. 363-382.
- Migge, Bettina. 2017b. Epilogue: of theories, typology and empirical data. In Peter Bakker, Finn Borchsenius, Carsten Levisen & Eeva Sippola (eds.), *Creole Studies & Phylogenetic Approaches*. Benjamins. 389-394.
- Quint, Nicolas & Bart Jacobs, 2016, On the relevance of Classical Portuguese features in four Atlantic creoles. In Armin Schwegler, Liane Stroebel & John McWhorter (eds.) *The Iberian Challenge: Creoles beyond the plantation setting*. Klaus Vervuert/Iberoamericana, 67-83.
- Quint, Nicolas, 2015, *Let's Speak Capeverdean*. Londres/Colombo: Battlebridge
- Quint, Nicolas, Ana Josefa Cardoso & Karina Moreira, forth. Enseignement de la langue maternelle au Cap-Vert: un processus en voie de mise en place. In Raphaël Confiand (ed.), *La place des langues créoles dans l'enseignement*. Caraïb éditions.
- Quint, Nicolas, Bart Jacobs & Armin Schwegler, forth. Spanish-based Creoles. In Francesco Gardani & Michele Loporcaro (eds.), *Oxford Encyclopedia of Romance Linguistics*. Oxford University Press.
- Quint, Nicolas, Noël Bernard Biagui & Jean-François Nunez, forth. Casamance Creole. In Friederike Lüpke (ed.), *Oxford guide to the world's languages: Atlantic*. Oxford University Press.
- Léglise, Isabelle & Migge, Bettina. 2015. Language practices and language ideologies in Suriname: Results from a school survey. In *In and out of Suriname: Language, mobility, identity*, Eithne Carlin, Isabelle Léglise, Bettina Migge & Tjon Sie Fat, Paul (eds), 13-57. Leiden/Boston: Brill.

Databases

- Manfredi, Stefano & Nicolas Quint. 2017. Mise en ligne de la base de données sur la syntaxe de la phrase complexe dans les langues créoles. <http://llacan.vjf.cnrs.fr/labexGD2/form/index.php>

Conferences and seminars

- Fon Sing, Guillaume. 2015. About some innovative features in Mauritian Creole. Workshop GDR SEEPiCLa. CLUL/FLUL, Universidade de Lisboa, Portugal.
- Fon Sing, Guillaume. 2015. Morphological and syntactical innovations in Mauritian Creole: extraction and analysis from regional French data. Tenth Creolistics Workshop: "Innovations" - with special attention to parallels between creole and sign language creation. Aarhus University, Denmark.
- Fon Sing, Guillaume. 2015. Fixed Verbal Expressions in Mauritian Creole. In Workshop GDR SEEPiCLa. University of Amsterdam, Holland.
- Fon Sing, Guillaume. 2016. Description et analyse des expressions verbales figées en créole et en français régional de l'île Maurice. In Colloque de l'AFLS 2016, "French as a shared language / Le français partagé". Belfast, UK: Queen's University.
- Kriegel, Sybille, and Guillaume Fon Sing. 2016. Le factitif dans les créoles de l'Océan Indien. In Colloque international: "Le factitif: perspectives croisées", Université Paris-Sorbonne/Université Paris-Diderot (17-19 novembre 2016), Paris.
- Manfredi, Stefano 2015. Adverbial subordinators in Juba Arabic. Summer Meeting of the Society for Pidgin and Creole Linguistics. Graz University, 7-9 July 2015.
- Manfredi, Stefano 2015. The history of kedé in Sudanese Arabic and Arabic-based pidgins and creoles. From speaker oriented modal marker to general subordinator. 11ème conférence

- de l'Association Internationale de Dialectologie Arabe. Université de Bucarest, 25-29 Mai 2015.
- Manfredi, Stefano 2016. Le créole arabe de Juba entre pratiques et représentations. 3ème Rencontres d'études Africaines, INALCO Paris, 5-7 Juillet 2016.
- Manfredi, Stefano 2016. La syntaxe de la phrase complexe dans le créole arabe de Juba , séance mensuelle de la Société de Linguistique de Paris, 16 Mars 2016.
- Manfredi, Stefano. 2015. Varietà di contatto a base araba. Keynote speaker for the workshop I dialetti arabi: prospettive linguistiche, sociolinguistiche e letterarie, Università di Catania, Catania, 15 Mai 2015.
- Manfredi, Stefano. 2016. Juba Arabic. Invited speaker for the workshop Languages or Dialects? Celebrating the Diversity in Arabic, SOAS, London, 3 June 2016.
- Manfredi, Stefano. 2018a. Determinanti sociostoriche della pidginizzazione e la creolizzazione in arabo. Invited speaker for the workshop L'arabo e la storia. Contatti e confini dell'arabofonia, University of Turin, 22 March 2018.
- Manfredi, Stefano. 2018b. Arabic-based (pidgin)creoles in East Africa. Invited speaker for the workshop Sociolinguistics and Dialectology in the Arabic-speaking Worlds, Freie University Berlin, 2 May 2018.
- Migge, Bettina. 2015. Des structures à la pratique: les recherches sociolinguistiques sur les créoles. Linguistic Conference at Rennes University: France. 10 April.
- Migge, Bettina. 2015. Documentation linguistique dans des contextes linguistiques hétérogènes. DOCULANG, SeDyL. Villejuif: France. 11 June.
- Migge, Bettina. 2015. Linguistic Variation, integration and identity in contexts of migration: An assessment of current approaches. Sociolinguistic Summer School 2015 – Trinity College Dublin. Dublin, Ireland. 4-7 August.
- Migge, Bettina. 2015. Mobility, urbanization and the negotiation of new & old language practices among Maroons. African multilingualism: motivations, modalities, movement and meaning. London: UK. 8-9 September.
- Migge, Bettina. 2015. The role of multilingual practices in negotiating social change. Minor Translingualism. London: Institute of Modern Languages Research, School of Advanced Study, University of London. 28 October.
- Nunez, Jean François 2017. Le créole de Casamance: une langue en zone d'hétérogénéité linguistique . Workshop Interactions en milieu plurilingue. Université d'Orléans, 20 mars 2017.
- Nunez, Jean François 2017. Constructions comparatives en créole de Casamance. Journée d'études Constructions comparatives dans des variétés de langues créoles et non créoles, Université de Picardie Jules Verne. Amiens, 6 Avril 2017.
- Nunez, Joseph Jean François, 2014, Conséquences linguistiques du contact de langues chez les créolophones casamançais: l'exemple de l'expression du temps, Vème Séminaire International du Grupo de estudos de linguas em contato, 5-6 juin, Université de Praia, Cap-Vert.
- Nunez, Joseph Jean François, 2014, L'expression du génitif chez les créolophones casamançais: un cas de contact de langues, 5-8 Août, 20th Biennial Conference of the Society for Caribbean Linguistics (SCL), Aruba.
- Nunez, Joseph Jean François, 2016, Casamancese creole in contact, Invited lecture 25 April, Bremen University.
- Nunez, Joseph Jean François, 2016, The sociolinguistic situation in Senegal, Invited lecture 27 April, Bremen University.
- Nunez, Joseph Jean François, 2016, Repères temporels en contact au Sénégal, 4ème Rencontre des Études Africaines en France, 5-7 July, Paris.

- Quint, Nicolas & Eliane Araújo Vieira Semedo, 2017, Cláusulas completivas em língua cabo-verdiana: estudo baseado em corpus. ACBLPE annual meeting, Stockholm University.
- Quint, Nicolas & Eliane Araújo, Frase complexa em cabo-verdiano: estudo baseado em corpus. ACBLPE annual meeting, Universidade de Cabo Verde, Praia, 2016/06/25.
- Quint, Nicolas & Joseph Jean-François Nunez, 2015, A linguistic account of the translations of 'Le Petit Prince de Saint-Exupéry' in two Portuguese Upper Guinea Creoles, Santiaguense Capeverdean and Casamancese. Meetings of the Society for Pidgin and Creole Linguistics (SPCL), Associação de Crioulos de Base Lexical Portuguesa e Espanhola (ACBLPE) Universität Graz.
- Quint, Nicolas & Joseph Jean-François Nunez, 2015, Contact de langue chez les créolophones casamançais: l'exemple des déterminants. Society for Pidgin and Creole Linguistics (SPCL) and Associação de Crioulos de Base Lexical Portuguesa e Espanhola (ACBLPE) joint conference, organized by Institut für Romanistik de la Karl-Franzens-Universität Graz (Graz University), Graz (Austria).
- Quint, Nicolas & Noël Bernard Biagui, 2016, Un aperçu des structures factitives en portugais dans les créoles à base portugaise de l'Afrique de l'Ouest (Upper Guinea Creoles). Colloque Le factitif: perspectives croisées. Université Paris-Sorbonne & Paris-Diderot, Paris.
- Quint, Nicolas, 2015 (keynote speaker), L'atlas linguistique du capverdien. Séminaire de l'équipe "systèmes linguistiques et dialectologie" du Gipsa-Lab (Grenoble University /CNRS), Grenoble.
- Quint, Nicolas, 2015, Um atlas linguístico das variedades caboverdianas de Sotavento: para uma nova abordagem da variedade diatópica em caboverdiano. Um atlas linguístico das variedades caboverdianas de Sotavento: para uma nova abordagem da variedade diatópica em caboverdiano, Feira de Santana (Etat de Bahia, Brésil). Keynote speaker.
- Quint, Nicolas, 2016. Capeverdean grammar(ian)s: one more account of the difficulties of describing a language for its own sake. Workshop "Grammaires étendues", Paris Diderot University, Paris.
- Quint, Nicolas, 2017. Un aperçu des structures comparatives en capverdien (santiagais). Paper presented at the workshop "études des constructions comparatives dans des variétés de langues créoles et non créoles" Universités des Antilles (CRREF-EA 4538) et de Picardie Jules Vernes (CERCLL-EA 4283), 4 June 2017
- Quint, Nicolas, Eliane Viera Semedo & Stefano Manfredi 2017. Towards a typology of complex sentences in creole languages. Annual meeting of the Association of Portuguese and Spanish-lexified Creoles, Stockholm University. 13-15 June 2017.
- Samedo, Eliane Vieira, 2016. Frase complexa em cabo-verdiano: estudo baseado em corpus. ACBLPE annual meeting, University of Cape verde, 25 June 2016.
- Samedo, Eliane Vieira, 2017. Proposições completivas em cabo-verdiano. 11th ABECS meeting, University of Brasilia, 30 November 2017.
- Samedo, Eliane Vieira et Nicolas Quint 2017. Cláusulas completivas em língua cabo-verdiana: estudo baseado em corpus. Annual meeting of the Association of Portuguese and Spanish-lexified Creoles, Stockholm University. 13-15 June 2017.

1.2. Organisation of conferences and summer schools

- Manfredi, Stefano & Christopher Lucas. 2017. *Workshop Arabic and contact-induced change, 23th International Conference on Historical Linguistics*, University of Texas in San Antonio, 31 July – 4 August 2017.
- Manfredi, Stefano & Nicolas Quint. 2018. *Workshop The Syntax of Complex Sentences in Creole Languages, 8th Syntax of the World's Languages conference*, INALCO 365 September 2018.

Quint, Nicolas & Jean-François Nunez. 2018. *18th ACBLPE annual meeting*. Université de Ziguinchor, 11-14/06/2018.

1.3. Popular science

Biagui Noel Bernard, Nunez, Joseph Jean François & Quint, Nicolas, 2015, *Rey Siñu*, Casamance Creole translation of Antoine de SAINT-EXUPÉRY's *le Petit Prince*, Tintenfass.

Manfredi, Stefano. 2016. *L'arabe de Juba*. Entretien filmé. Réalisateur Maryline Leducq, Service Audiovisuel du CNRS, collection Langues Rares,

http://www.canal-u.tv/producteurs/cnrs_ups2259/documentaires/collection_langues_rares

Quint, Nicolas, 2018, *カーボ・ヴェルデ・クレオール語への誘い* (Kabo Verude Kureorugo E No Izanai, Capeverdean conversational for Japanese speakers), Koyobosho.

2. Presentation of the follow-up project

2.1. Abstract

One of the main problems in typologizing syntactic structures on the basis of (semi-)spontaneous data is represented by the high degree of individual variation affecting these structures. This is especially true for creole languages, whose syntactic structures are affected by different linguistic processes such as substrate/adstrate interference, creole-lexifier attrition, as well as innovative ongoing changes. This new project aims at integrating the analysis of syntactic variation with the typologisation of different kinds of syntactic structures (subordinated and coordinated clauses) in creole languages as well as improving the categorization of the semantic values and origins of creole coordinating and subordinating conjunctions.

2.2. Objectives

The project "Syntactic variation in creole languages" is a prolongation of the project GD2 "The syntax of complex sentences in creole languages", which aimed at producing a database of complex sentences in a number of creole languages with different lexifiers and substrate languages. While compiling our database, it immediately became clear that a standard synchronic comparison of subordinate clauses was particularly difficult because of the concurrence of many constructions for encoding the same meaning. The main aim of this operation is thus to analyze different factors intervening in the syntactic variation of creole languages and to integrate them within a typological approach to syntactic structures. Firstly, we will focus our attention on ongoing internal changes linked to different degrees of grammaticalization and reanalysis. When necessary, we will also take into account the outputs of different kinds of creole-lexifier bilingualism. Lastly, we will analyze instances of matter and pattern transfer from substrate/adstrate languages into creole languages. The analysis of the aforementioned factors of variation will be implemented while expanding our database on the syntax of complex sentences in creole languages (<http://llacan.vjf.cnrs.fr/labexGD2/form/index.php>)

2.3. Evolution with respect to the first phase

- Development and integration of a multifactorial grid for the analysis of syntactic variation within the GD2 database.
- Development of semantic maps for visualizing the grammaticalization of subordination and coordination markers.
- Development of a concordancer for doing complex queries in the database.

VC4: Variation and contact between closely related languages

PIs	Isabelle Léglise (SeDyL) Stefano Manfredi (SeDyL)
Participants	SeDyL: Bettina Migge, Isabelle Léglise, Stefano Manfredi LLACAN: Nicolas Quint LACITO: James Costa LLF: Heather Burnett, Guillaume Fon Sing

1. Presentation

1.1. Abstract

This project aims at developing a fine-grained analysis of the contact dynamics involving closely related languages. Based on heterogeneous (i.e. multistylistic and multidialectal) corpora, its ambition is to typologize different kinds of variation at play in contact settings so as to revise the notion of “language boundaries” and to refine the analysis of the social implications of language contact.

1.2. Objectives

In the attempt to embrace different synchronic aspects of bi/multilingualism, we will work on diverse sociolinguistic settings involving contact between closely related languages. On the one hand, we will investigate contact between creole languages and their lexifiers while adopting a dynamic perspective of analysis which goes beyond the idea of “decreolization”. The theoretical reflection about this particular contact setting will be founded on the annotation and study of heterogeneous corpora including, among others, French, English, Portuguese, Arabic-based creoles, and their respective lexifiers. On the other hand, we intend to analyze the practices characterizing contact between varieties of the same language. For this part of the project, we will work on multidialectal corpora of Spanish, English, Arabic, French, Occitan, and Scottish Gaelic. This will eventually allow us to identify a number of indexical differences between dialects in contact and their role in the selection of linguistic features. By typologizing the previous contact settings, we will also contribute to the debate about the social structuring of linguistic diversity, which goes under the name of “sociolinguistic typology”, as well as to our understanding of the synchronic emergence of new linguistic varieties. On a methodological ground, we will adopt both quantitative and qualitative methods of investigation for working on our corpora of annotated first-hand transcriptions.

**VC5:
Towards a typology of codeswitching through multifactorial
analysis of plurilingual corpora**

	2015 – 2019	2020 – 2024
Strand	3	3
Work package	LC1	VC5
Project title	Multifactorial analysis of language changes	Towards a typology of codeswitching through multifactorial analysis of plurilingual corpora
PI	Isabelle Léglise (SeDyL)	Isabelle Léglise (SeDyL)
Participants	<p>Evangelia ADAMOU (LACITO), Sophie ALBY (SEDYL), Fida BIZRI (SEDYL), Bernard CARON (LLACAN), Claudine Chamoreau (SEDYL), Françoise GUERIN (LACITO), Isabelle LEGLISE (SEDYL), Sylvain LOISEAU (SEDYL), Stefano Manfredi (SEDYL), Bettina MIGGE (SEDYL), , Nicolas QUINT (LLACAN), Claire SAILLARD (LLF), Guillaume SEGERER (LLACAN), Marie-Claude Simeone-Senelle (LLACAN), Duna TROIANI (SEDYL), Martine VANHOVE (LLACAN), Pascal VAILLANT (Paris 13/ SEDYL), Alice VITTRANT (LACITO)</p> <p>xDSDoctorants: Suat ISTANBULLU (SEDYL) Post-docs: Hans-Joerg DOEHLA, Sabine GOROVITZ, Joseph Jean-François Nunez, Santiago SANCHEZ</p>	<p>SeDyL: Sophie Alby, Isabelle Léglise, Stefano Manfredi, Bettina Migge LLACAN: Nicolas Quint LACITO: James Costa LLF: Heather Burnett, Claire Saillard</p>

1. Summary of the first phase

1.1. Activities since 2015

Language contact phenomena have received growing attention in the last 30 years. Different brands of Contact Linguistics (Thomason, 2001, Winford, 2003) proposed typologies of language contact settings based on parameters such as the type of setting (language maintenance or

loss), social factors (length and intensity of contact, degree of bilingualism), language contact mechanisms and results (borrowing, shift, calque, grammaticalization, reorganization etc.; see Croft (2000), Field (2002), Heine and Kuteva (2003 & 2005), and Matras and Sakel (2007)). Most of the studies with a typological perspective focus on the description of the outcomes of contact and its impact on the languages involved – they were realized with a diachronic perspective, and to date little attention has been paid to synchronic phenomena. Whereas most of the studies with a synchronic perspective focus on code-switching and code-alternation from a structural and cognitive perspective in that they try to determine the linguistic patterns and constraints that govern code-switching (cf. Poplack, 1980, Myers-Scotton, 1993, Auer, 1998, Muysken, 2000), research on multilingualism and synchronic contact settings is more heavily concerned with the social constraints and the meaning-making potential of code alternation; this research is mostly conducted by researchers interested in the sociology of language, interactional sociolinguistics or second language acquisition. A few projects proposed in the last years the interaction and cross-fertilisation between these two lines of research. The research project LC1 within the first phase of LABEX EFL is one of them. It proposed to identify multiple causation for both internally motivated changes and contact-induced processes, through a fine-grained approach integrating both social and linguistic factors (see e.g. Chamoreau & Léglise 2012, Migge & Léglise 2013). This involved devising a multifaceted methodology and a multi-model approach to explain contact-induced language changes. It involved furthermore the development of common computer-based tools as an XML editor and a multilingual concordancer that were tested by the various participants in the LC1 research program (see Vaillant & Léglise 2014 and publication lists) available on a platform: <http://lc1-labexefl.vjf.cnrs.fr>.

1.2. Publications

International journals

- Adamou, Evangelia & Granqvist, Kimmo. (2015). Unevenly mixed Romani languages. *International Journal of Bilingualism* 19(5): 525-547.
- Léglise, Isabelle & Sophie Alby. (2016). Plurilingual corpora and polylinguaging, when corpus linguistics meets contact linguistics. *Sociolinguistic Studies* 10 (3): 357-381.
- Migge, Bettina 2017a Putting Matawai on the Surinamese Linguistic Map. *Journal of Pidgin and Creole languages* 32(2): 233-262.
- Ó Murchadha, Noel P. & Migge, Bettina 2017 Support, transmission, education and target varieties in the Celtic languages: an overview. *Language, Culture and Curriculum* 30(1): 1-12.
- Saillard, Claire. Sous presse, Adjectival Modification in Truku Seediq. *Language and Linguistics*: 23 p. 1606-822X
- Sanchez, Santiago, 2017, Enoncés de type OV et positionnements sociaux dans l'espagnol parlé par les Quichuas équatoriens à Cali (Colombie) *Sociolinguistic Studies Equinox Publishing*. Vol. 11, N° 1, 65-105.
- Sanchez, Santiago. 2018. Negociación de posicionamientos sociales (stances) a través del debilitamiento del fonema /s/ en el español hablado por los quichuas ecuatorianos en Cali (Colombia), *Pragmática Sociocultural / Sociocultural Pragmatics*, DOI: 10.1515/soprag-2017-0019.
- Vaillant, P. & Léglise, I., 2014, « A la croisée des langues: Annotation et fouille de corpus plurilingues », Numéro spécial *Fouille de données et humanités numériques*, RNTI *Revue des Nouvelles Technologies de l'Information*, Editions Hermann Paris, 81-100.

Chapters in edited volumes

- Léglise, Isabelle & Sánchez Moreano Santiago. 2017. From varieties in contact to the selection of linguistic resources in multilingual settings. In R. Bassiouney (ed.), *Identity and Dialect Performance*. Routledge, 143-159.
- Manfredi, Stefano and Suat Istanbullu, forth. A multilayered and data-driven approach for exploring Arabic in multilingual context. In M. Klimiuk et W. Arnold (eds.), *Arabic Dialectology: Methodology and Field Research*. Harrassowitz.
- Migge, Bettina 2017. From grammar to meaning: Towards a framework for studying synchronic language contact. In Kofi Yakpo & Pieter Muysken (eds.), *Boundaries and Bridges: Language Contact in Multilingual Ecologies*, 363-382. Berlin: De Gruyter Mouton.
- Migge, Bettina 2018. Research on Maroon Languages and Language Practices among Matawai and Kwinti Maroons. In *Maroon Cosmopolitics: personhood, creativity and incorporation*, Olivia Maria Gomes da Cunha (ed). Amsterdam: Brill.
- Nunez, Joseph Jean François & Léglise Isabelle. 2017. Ce que les pratiques langagières plurilingues au Sénégal disent à la linguistique de contact, in M. Auzanneau, M. Bento, M. Leclère (eds), *Espaces, mobilités et éducation plurilingues: éclairages d'Afrique ou d'ailleurs*, Paris: Editions des archives contemporaines, 99-119.
- Simeone-Senelle, Marie-Claude & Mohamed Hassan Kamil (2018). 'Conséquences linguistiques des contacts sur les rives de la mer Rouge. Remarques sur le lexique halieutique en afar (Djibouti, Erythrée)'. In Kelig Yann Cotto & Lukian Prijac (Eds) *Dans le sillage de Sindbad. Autour des relations maritimes entre l'Arabie et l'Afrique de l'Antiquité à nos jours*. Chateaulin: Locus Solus.
- Simeone-Senelle, Marie-Claude. (2018). Arabic on the Dahlak Islands (Eritrea). In Stefano Manfredi & Mauro Tosco (eds), *Arabic in Contacts. Studies in Arabic Linguistics*. Amsterdam & Philadelphia: John Benjamins.

Books

- Adamou, Evangelia. (2016). *A corpus-driven approach to language contact. Endangered languages in a comparative perspective*. Boston & Berlin: de Gruyter Mouton.

2. Presentation of the follow-up project

2.1. Abstract

This research program investigates codeswitching and contact phenomena through multifactorial analysis of plurilingual corpora. It aims at 1) critically examining the typology of terms proposed in the literature and 2) devising a typology of codeswitching phenomena from corpora from a great range of situations and social contexts.

2.2. Objectives

Typologies of code-switching or code mixing were proposed 20 years ago (see e.g. Auer 1998, Muysken 2000) and are still being discussed. Most studies in contact linguistics and code-switching assumed languages to be discrete and bounded systems in contact; terms like matrix language or code-alternation are good examples of this hypothesis (Léglise 2017). Within the last decade, there has been a shift in focus from the investigation of how relatively bounded communities come into contact and the kind of linguistic outcomes that gives rise to, to that of how social actors produce and negotiate social meanings, social positions and relations through their language practices. Hall & Nilep (2015), for instance, shows very clearly how contemporary

research, much centered on linguistic repertoires, wonderfully illustrates disruption of presumed connections between language, community, identity and spaces. Within the scope of multilingualism, lots of terms were coined in the last 10 years to give voice to the variety of practices that are observable such as crossing (Rampton 2005), trans-idiomatic practices (Jacquemet 2005), translanguaging (Garcia 2009, Garcia & Li 2013, Creese & Blackledge 2010) or (poly)linguaging (Jørgensen et al 2011), heterogeneous language practices (Léglise & Alby 2013). One of the aims of this research program is to critically examine the typology of terms proposed in the literature in the light of the codeswitching phenomena observable in our plurilingual corpora that include data from a range of contexts, such as formal, educational and informal settings and a diverse range of language pairs. A second aim is to devise a typology of codeswitching phenomena from corpora from a great range of situations and social contexts. This typology will crucially aim to also integrate structural-cognitive and socially oriented approaches to code alternation phenomena. Finally, it involves exploring and developing computer-based tools for analysing plurilingual and linguistically heterogeneous language corpora. It will facilitate identifying social and structural factors that condition the appearance of zones of bivalency between languages, types of switches, word order phenomena or linguistic forms belonging to neither of the languages in contact.

2.3. Evolution with respect to the first phase

During the last years, LC1 proposed to identify multiple causation for both internally motivated changes and contact-induced processes, through a fine-grained approach integrating both social factors and linguistic factors. The computer-based tools developed in that project as well as the multifaceted methodology and multi-model approach that have been set up will be further applied to synchronic issues in order to better handle, describe and explain codeswitching phenomena and ongoing contact-induced language changes.